UNIVERSITY OF CALIFORNIA
RIVERSIDE

Computational Analysis of Receptor-Odor Interactions and Prediction of Behavior-Modifying Chemical Space

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics, and Bioinformatics

by

Sean Michael Boyle

March 2012

Dissertation Committee:
Dr. Anandasankar Ray, Chairperson
Dr. Anupama Dahanukar
Dr. Thomas Girke

The Dissertation of Sean Michel Boyle is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgments

Personal Acknowledgements

I would like to acknowledge my parents Laura and Mike Boyle for their constant support and encouragement throughout both my education and life. They have always been a wonderful source of guidance and support while traversing the endeavors of the journey we call life. I would also like to acknowledge the daily guidance, humor, and inspiration provided by my advisor Dr. Anandasankar Ray. He has always been a constant source of wonderful ideas, almost tangible intensity toward science, and at times very needed off topic tangents. I would also like to acknowledge all past and present members of both the Ray and Dahanukar labs for their endless inquisitiveness, humor, and good nature. Without all of you this work would not be so complete and would certainly not have been so enjoyable.

Professional Acknowledgements

- Anandasankar Ray, Anupama Dahanukar, and Thomas Girke for providing research guidance.
- Shane McInally for validating *Drosophila* receptor-odor predictions using electrophysiology.
- Chih-Ying Su, Eliza Kelly-Swift, and John Carlson for validating *Anopheles* receptor-odor predictions using electrophysiology.
- Sana Tharadra for performing *Drosophila* larvae behavior experiments.
- Christine Pham for performing *Drosophila* larvae behavior experiments.
- Dyan MacWilliam for performing CO_2 receptor and repellency validations using electrophysiology.
- Tom Guda for performing Mosquito behavior experiments.
- Will Pitt and Tom Blundell for assisting with structure-based virtual screening.
- Sachin Suardre for validating predicted EthR inhibitors.

ABSTRACT OF THE DISSERTATION

Computational Analysis of Receptor-Odor Interactions and Prediction of Behavior-Modifying Chemical Space

by

Sean Michael Boyle

Doctor of Philosophy, Graduate Program in
Genetics, Genomics, and Bioinformatics
University of California, Riverside, March 2012
Dr. Anandasankar Ray, Chairperson

Coding of information in the peripheral olfactory system and resulting olfactory dependent behavior is thought to depend primarily on two fundamental factors: the interaction of individual odors with different subsets of the odor receptor (Or) repertoire, and the mode of signaling that an individual receptor-odor interaction elicits, activation or inhibition. In order to better understand these processes, we design and implement a structure-based virtual screening approach that identifies common structural features that are highly correlated with odor activity for individual receptors. We then apply these features to rapidly screen for putative ligands *in silico* from a large untested odor space (>240,000 putative volatiles) for the majority of odor receptors in the *Drosophila* antenna, allowing for analysis of odor coding for the majority of receptors for the first time. Functional experiments support a high success rate (~71%) for the screen and we validate numerous new activators and inhibitors for the receptors. Following our initial application in *Drosophila*, we extend our approach to predict activating and inhibiting odors for a large number of important pest and disease vector species including 50 *Anopheles gambiae* Ors (65% validated accuracy), the CO2 receptors of multiple species (48% validated accuracy), 9 newly identified Citrus Psyllid ORNs, and a large number of functionally distinct mammalian receptors. We next extended our in silico

screening approach to identify shared structural features important for a behavioral response to DEET-like repellents for which the molecular target has not yet been identified, identifying ~150 natural compounds as candidate repellents. We select 4 candidates, 3 approved as safe for human food use, and demonstrate that they are strong olfactory and gustatory repellents to both mosquitoes and *Drosophila*. As only a small region of odor space has been explored, there remains potential to uncover previously unidentified patterns of odor coding. Through a combination of *in silico* and electrophysiology screens, we identify odors with ultra-prolonged termination kinetics that are delayed for several minutes, resulting in a memory trace that affects subsequent odor detection. Finally, we successfully perform structure-based virtual screen, identifying potential inhibitors of an important Tuberculosis drug target EthR.

Table of Contents

CHAPTER I: Introduction

- Complexity in the olfactory system p1
- Only a very small portion of the vast odor space has been tested for activity p6
- Or gene identification in insects was largely aided by computational approaches p8
- Applying ligand-based virtual screening to predict receptor-odor interactions p12
- Applying structure-based virtual screening to predict receptor-odor interactions p16
- High-throughput approaches are effective tools for olfactory analysis p18

CHAPTER II: Designing a Ligand-Based Virtual Screening Approach to Decode *Drosophila* Odor Receptor Chemical Space *In Silico*

- Introduction p25
- Results p27
- Discussion p37
- Figures p42
- Tables p60

CHAPTER III: Applying Chemical Informatics to Decode Odor Receptors of Several Important Disease Vector and Pest Insect Species as Well as Mammals

- Introduction p71
- Results p76
- Discussion p95
- Figures p99
- Tables p141

CHAPTER IV: A New Generation of Safe DEET Substitutes that are Strong Olfactory and Gustatory Repellents of Mosquitoes and Flies

- Introduction p201
- Results and Discussion p204
- Figures p214

CHAPTER V: Analyzing Termination Dynamics of Prolonged Activating Odors and their Effects on Receptor-Mediated Olfactory Behavior

- Introduction p227
- Results p230
- Discussion p241
- Figures p245
- Tables p265

CHAPTER VI: Applying Structure-Based Virtual Screening to Identify Potent Inhibitors of the Tuberculosis Target EthR

- Introduction p266
- Results and Discussion p268
- Figures p281
- Tables p301

CHAPTER VII: Methods

- Chemical informatics p302
- Electrophysiology p302
- Natural odor compound library p303
- Pubchem compound library p303
- eMolecules compound library p303
- Calculation of 3D conformations p303
- Calculation of molecular descriptors p304
- Classification of active compounds p304
- Calculation of Accumulative Percentage of Actives (APoA) p305
- Determination of optimized descriptor subsets p305
- Clustering Ors by most common descriptors p306
- Clustering compounds by activity of Or p306
- Clustering Ors by predicted ligand space p307
- Calculation of Or prediction distribution frequencies p307
- Or-ligand interaction map p307
- Computational validation of ligand-based virtual screening (Non-SVM) p308
- Calculation of LogP and vapor pressure values p309
- Repellency behavior testing p309
- Classification of repellent compounds p311
- Support Vecotor Machine (SVM) predictions p311
- Computational validation of ligand-based virtual-screening (SVM) p311
- Vinyl solubility assay p312
- Olfactory avoidance assay trap assay for *Drosophila* p312
- T-maze Assay Methods p313
- Modified hand in-glove olfactory repellency assay for mosquitoes p313
- Humidity and warmth attraction assay p315
- Larval behavior assays p315
- Assembly of EthR screening library p316
- Calculating molecular descriptors for EthR analysis p316

- Applying machine learning with Pipeline Pilot p317
- Performing structure-based virtual screening with Gold p317

REFERENCES p319

List of Figures

CHAPTER I:

1.1: Overview schematic describing neuronal wiring of the insect olfactory system p24

CHAPTER II:

2.1: Comparing efficacy of various structure analysis methods to analyze odor receptor ligands p42
2.2: A receptor-optimized molecular descriptor approach has strong predictive power to find new ligands p44
2.3: Accumulated percentage of actives analysis p46
2.4: Electrophysiology validates that odorant receptor-optimized molecular descriptors can successfully identify new ligands for *Drosophila* p48
2.5: Predicted receptor-odor interactions are highly specific p50
2.6: High-ranking odors are more likely to be active than distantly ranked odors p52
2.7: Analysis of receptor-odor relationships and breadth of tuning p54
2.8: Analysis of receptor-natural odor interactions p56
2.9: Predicted odor space and network view of odor coding p58

CHAPTER III:

3.1: A molecular descriptor optimized approach is able to explain odor activity for individual *Anopheles* Ors p99
3.2: Optimized molecular descriptor sets are able to cluster either active or inhibitory odors p101
3.3: Structural relationships between the 75 most diverse aromatic odors p107
3.4: Newly optimized molecular descriptor sets are able to cluster odors for aromatically tuned Ors p109
3.5: Optimized descriptors for AgOrs 2, 6, and 10 effectively describe training set activity p111
3.6: A Support Vector Machine (SVM) integrated approach is highly effective at explaining odor activity for individual Ors p113
3.7: SVMs trained using optimized descriptors for AgOrs 2, 6, and 10 effectively describe training set activity p115
3.8: A molecular descriptor optimized approach is able to explain odor activity for CO2 receptors p117
3.9: Optimized molecular descriptor sets are able to cluster odors by CO2 receptor response p119
3.10: Active compounds cluster into three distinct structural classes p121
3.11: Chemical structures of validated activators and inhibitors p123
3.12: A SVM integrated molecular descriptor optimized approach is able to explain odor activity for individual Citrus Psyllid ORNs p125
3.13: Optimized molecular descriptor sets are able to cluster ORN activators p127
3.14: SVMs trained using descriptor sets that were optimized for individual ORNs effectively describe training set activity p129
3.15: A molecular descriptor optimized approach is able to explain odor activity

for individual mammalian ORs p131
3.16: Mammalian Odorant receptor-optimized molecular descriptors can successfully cluster known ligands p133
3.17: Optimized descriptors for mammalian ORs effectively describe training set activity p135
3.18: Analyzing relationships between important features for mammalian ORs p137
3.19: Analysis of mammalian OR tuning breadth p139

CHAPTER IV:

4.1: Contribution of olfaction and gustation in DEET avoidance p214
4.2: Mosquito behavioral assay glove setup p216
4.3: A chemical informatics method to predict repellency p218
4.4: Identification of repellents using in-silico screening of a large chemical space p220
4.5: A new class of mosquito repellents with desirable safety profiles p222
4.6: Mosquito escape index p224
4.7: Natural compounds are effective at repelling aedes aegypti in the hand in glove assay p226

CHAPTER V:

5.1: Activity of a behaviorally important neuron, Ab1A, can be described by an odor receptor neuron-optimized chemical inormatics approach p245
5.2: Electrophysiology validates that ORN-optimized molecular descriptors can successfully identify new ligands for ab1A p247
5.3: Functional identification of Ultra-Prolonged activators and analysis of their long-term effects on ab1A using electrophysiology p249
5.4: Ectopic expression confers Or42b UP activation p251
5.5: Behavioral effects of Ultra-Prolonged activators on odor detection in ab1A p253
5.6: Behavioral effects of inhibitory odor on odor detection in ab1A p255
5.7: Functional identification of Ultra-Prolonged activators of additional odor receptors using electrophysiology p257
5.8: Identification of ultra-prolonged activating odors for additional Ors p259
5.9: Long-term effects of Ultra-Prolonged activators on odor detection p261
5.10: Modes of signaling and their behavioral responses p263

CHAPTER VI:

6.1: Schematics of Ethionamide Activation Pathway p281
6.2: Solved EthR structures p283
6.3: The shape of the EthR cavity p285
6.4: Previously identified EthR ligands bind within the proposed cavity p287
6.5: Visualizing previously identified EthR ligands p289
6.6: Single example of successful EthR self-docking p291
6.7: The results of many EthR site 1 self-docking runs p293
6.8: The predicted orientation of our strongest identified inhibitors in EthR p295
6.9: The predicted orientations of 4 modest inhibitors of EthR p297
6.10: The predicted orientations of the poor inhibitors of EthR p299

List of Tables

CHAPTER II:

2.1: Optimized descriptor sets for each *Drosophila* Or p60
2.2: Top 100 predicted compounds for each *Drosophila* Or p68

CHAPTER III:

3.1: Optimized descriptor sets for each *Anopheles* Or p141
3.2: Natural odor library predictions found in the top 500 predictions for each *Anopheles* Or p152
3.3: Predicted odors validated as effective activators for several *Anopheles* Ors p160
3.4: Activity of a large panel of aromatic odors was tested against three aromatically tuned Ors p162
3.5: Optimized descriptor sets for Or2, Or6, and Or10 p164
3.6: Top 75 predicted compounds for each Drosophila Or p166
3.7: Dividing training odors into three distinct sets based upon odor structure and receptor response p168
3.8: Optimized descriptors selected for the Aromatic Activator Screen p170
3.9: Optimized descriptors selected for the Broad Activator p172
3.10: Optimized descriptors selected for the Inhibitor Screen p174
3.11: Top predicted natural library compounds for the aromatic activator, broad activator, and inhibitor screens p176
3.12: Predicted odors validated as activators and inhibitors of the CO2 receptor p178
3.13: Optimized descriptor sets for each Citrus Psyllid ORN p180
3.14: Top predicted natural library compounds for Citrus Psyllid ORNs p184
3.15: Optimized descriptor sets for each mammalian OR p188
3.16: Top 100 predicted compounds for each mammalian OR p195

CHAPTER V:

5.1: Optimized descriptor sets for ab1A (Or42b) p265

CHAPTER VI:

6.1: A breakdown of our structure-based virtual screening accuracy for EthR p301

CHAPTER I:

INTRODUCTION

Complexity in the olfactory system

The olfactory system of insects is a very large and intricate neural network, consisting of a peripheral input layer, a middle processing layer, and high level wiring into the brain (Figure 1.1). Odors are detected at the periphery and contextual information about the odors are aggregated and organized in the middle layer. Here additional stages of processing are performed and the information is then provided to the brain for interpretation and decision-making.

The periphery olfactory system in *Drosophila* is housed in two olfactory organs: the antennae and maxillary palps. Both of these organs are covered with many tiny hairs called sensilla, which house between 1 and 4 distinct Olfactory Receptor Neurons (ORNs), however the predominant number of sensilla house 2 (de Bruyne et al., 2001). Interestingly, ORNs generally express only a single odor selective Odor receptor (Or) along with an obligate co-receptor Orco, which is essential for olfaction (Couto et al., 2005; Larsson et al., 2004; Vosshall and Hansson, 2011). *Drosophila* express 60 Or genes, which are translated into 62 proteins, that are housed in ORNs (Robertson et al., 2003). Both Or gene expression by ORNs and housing of ORNs in the antenna and maxillary palps is highly conserved, allowing for sensilla types to be classified according to what ORNs they house and ORNs to be classified by the Or genes they express (Couto et al., 2005; Dobritsa et al., 2003). While the peripheral olfactory system is highly structured and conserved, it also represents a very complex network of gene expression.

It is the response of the Or to environmental odors that determines whether an ORN sends information, in the form of either neuronal activation or inhibition, to the higher centers in the brain (Dobritsa et al., 2003; Hallem et al., 2004). Or responses to odors have been demonstrated to range from total inhibition of the ORN to strong activation (>250 spikes/sec) and each Or is uniquely tuned to odors (Hallem and Carlson, 2006). While some Ors are narrowly tuned and respond to only 1 or a few known ligands, such as DmOr67d to the pheromone cVa, others are more widely tuned and respond to a broad array of both functional groups and molecule sizes (Hallem and Carlson, 2006; van Naters and Carlson, 2007). Interestingly, an individual odor may activate multiple receptors, with each receptor being tuned to a different range of odor concentration (Hallem and Carlson, 2006).

While originally identified in mammals, Ors of insects are unique. Mammalian ORs were first discovered in 1991, resulting in award of the prestigious Nobel prize (Buck and Axel, 1991). Insect Ors were discovered for the important model organism *Drosophila* nearly a decade later in 1999 (Clyne et al., 1999; Vosshall et al., 1999). Being extremely divergent from mammalian ORs, the signaling pathway of insect Ors is currently a hotly contested topic. While mammalian ORs function as true GPCRs, Insect receptors have several unique characteristics, including an inverse orientation in the neuronal membrane and the previously mentioned obligate co-receptor Orco (Belluscio et al., 1998; Benton et al., 2006). Even the phyolgenetic similarities between insect Ors are far more divergent than for mammalian ORs (Robertson et al., 2003). It was this dissimilarity between insect Ors and mammalian ORs that was responsible for the 8 additional years that were required for the identification of the first insect Ors, as traditional wet lab techniques attempting to identify genes using mammalian OR

sequence knowledge were highly ineffective and computational approaches designed on the hypothesis of extreme divergence were required (Clyne et al., 1999). Recently two groups have proposed two alternative, albeit slightly overlapping, hypotheses on insect Or function. While one group proposes they function as ligand gated ion channels, another contests that they are both ligand gated and G-protein mediated cyclic-nucleotide-activated cation channels (Kain et al., 2008; Sato et al., 2008; Wicher et al., 2008). It is possible their function is an aggregate of both, as there is compelling evidence supporting each hypothesis.

Chemosensation is not limited to Ors. Identification of the Ors in *Drosophila* paved the way for identification of the Gustatory receptor (Gr) family, which was discovered only one year following Or discovery (Clyne et al., 2000). While gustatory receptors are predominately involved in the chemosensation of chemicals in liquid medium, a pair of Gr receptors (Gr21a and G63a) expressed in *Drosophila* is interestingly responsible for detection of CO_2, which is a highly volatile odor (Dahanukar et al., 2001; Kwon et al., 2007). Even more intriguingly, both Gr21a and Gr63a are amongst the rare Grs expressed in the antenna, which almost exclusively houses Or expressing ORNs. As in *Drosophila*, the function of Grs is currently a debated topic including evidence for both GPCR like function and cation channel activity (Ishimoto et al., 2005; Sato et al., 2011; Yao and Carlson, 2010).

Several years after identification of the Or and Gr gene family, a third, more divergent, chemosensory family of Ionotropic Glutamate Receptors (iGluRs) was discovered in *Drosophila* (Benton et al., 2009). While these receptors are related to previously identified known families of glutamate receptors, they appear to be highly

divergent. IRs are expressed in coeloconic sensilla and appear to be tuned to acids, ammonia, and humidity (Abuin et al., 2011; Yao et al., 2005). IRs are believed to function in a heteromeric complex and conduct signals as ion channels (Abuin et al., 2011).

Neuronal wiring of the olfactory system in *Drosophila* is highly complex, containing multiple levels of processing. As discussed above, the peripheral system contains many chemosensory inputs, consisting of three unique classes of receptors, each with their own chemsensory detection spectrum and possibly unique function. The axonal extensions of ORNs project into the antennal lobe, with each ORN class projecting to a unique glomerular location (Fishilevich et al. 2005, Couto et al., 2005). Information processing in the antennal lobe has also been a contested topic, as research groups have debated over the role of local interneurons (LNs). While it was originally claimed that information was faithfully transmitted to higher brain centers without manipulation, it has since been demonstrated that both excitatory and inhibitory local interneuron's play an important role in signal processing prior to transmission (Olsen et al., 2007; Olsen and Wilson, 2008; Root et al., 2007; Shang et al., 2007; Wang et al., 2003; Wilson and Laurent, 2005; Wilson et al., 2004). The main reason for the discrepancy of the two differing claims lies in experimental technique. The Wang research group attempted to tease apart the biological question using calcium imaging, which was not as effective in identifying interneuron communication as Rachel Willson's electrophysiological approach. These different experiments provide a wonderful basis for the importance of teasing apart a biological question from multiple angles to obtain insight that may otherwise be missed. After processing in the antennal lobe, olfactory information is sent to higher brain regions by Projection Neurons (PNs), where it is finally

processed in the mushroom body and lateral horn (Jefferis et al., 2007). It is important to note there are both narrowly and broadly wired neuronal systems. While the majority of ORNs appear to provide information to highly interconnected networks of LNs, a few specialized ORNs form direct channels to PNs (Jefferis et al., 2007). These direct line systems have been termed labeled line and are often observed with pheromone detecting ORNs. The entire 11-cis-vaccenly acetate (cVA) pheromone circuit was beautifully traced from sensory input all of the way to descending output in what is perhaps the best single analysis of olfactory neuronal wiring in *Drosophila* to date, providing an illustrative example for labeled line systems (Ruta et al., 2010). However, these systems appear to be the exception rather than the rule, as processing for the vast majority of ORN input is far to complex for current experimental approaches to have teased apart.

To summarize the system as a whole, the olfactory system is a massive, highly interconnected, and complex neuronal network. It contains inputs from many periphery Ors/ORNs, processes the information in the antennal lobe, and projects processed information into higher brain regions. Luckily for olfactory investigation, the system also has characteristics lending itself to analysis, including the general expression of only a single odor sensitive Or in each ORN, glomerular aggregation of ORN input, and neuronal activity from different ORNs housed in the same sensilla being distinguishable by spike amplitude. While these challenges are considerable, major advances have been made in the field, many of which have been aided by computational approaches.

Only a very small portion of the vast odor space has been tested for activity

It has been estimated that the number of possible carbon based chemical configurations with similar molecular masses to biologically relevant chemicals could exceed 10^{60} unique structures (Bohacek et al., 1996; Dobson, 2004). To put this number into context, it is roughly twice the number of stars than are believed to exist in the universe. Considering that the number of unique chemical structures found in the human body is likely in the hundreds of thousands, this number of potential compounds is staggering (Goto et al., 2002). While volatile chemical space is undoubtedly considerably smaller that 10^{60}, it still likely contains hundreds of thousands of unique compounds, each of which have the potential to activate or inhibit Ors, IRs, or Grs.

Take into consideration that the number of unique Ors expressed by each species has thus far varied from XX on the small end of the scale to over a thousand in mice and each receptor responds independently to a given odor. Clearly, the number of possible Or-odor combinations is staggering. Furthermore, positive classification of a particular odor or blend, such as one produced by a piece of ripe fruit, by the brain requires the integrated response of all expressed Ors to each individual odor in the blend. As one principle goal of olfactory research is to understand how organisms perceive their chemical environment, the aim of understanding the responses of each Or expressed in a particular species to all odors from a large number of behaviorally important source is incredibly challenging, yet still important.

In the 12 years of olfactory research since the discovery of *Drosophila* Ors, 251 unique odors have been experimentally tested for activity in the *Drosophila* olfactory system (de Bruyne et al., 1999; de Bruyne et al., 2001; Dobritsa et al., 2003; Goldman et

al., 2005; Hallem and Carlson, 2006; Hallem et al., 2004; Kreher et al., 2005; Kreher et al., 2008; Kwon et al., 2007; Pelz et al., 2006; Schmuker et al., 2007; Stensmyr et al., 2003; Turner and Ray, 2009; van Naters and Carlson, 2007; Yao et al., 2005). Responses to these screens have recently been assembled and integrated into a single database (Galizia et al., 2010). While some screens have been narrow and focused on testing a small number of odors against a specific Or, others have involved testing a panel of odors across multiple Ors. For example, In 2001 47 odors were tested for activity against 16 ORN classes (de Bruyne et al., 2001). In the largest screen to date 109 odors were tested for activity against 24 Ors expressed in the antenna of *Drosophila* adults using single unit electrophysiology (Hallem and Carlson, 2006). Additionally, 21 Ors expressed in *Drosophila* larvae were tested for activity against 27 odors in 2008 (Kreher et al., 2008). While these three analyses certainly do not represent all of the analyses performed in *Drosophila*, they do represent the largest bodies of work.

Smaller scale olfactory experimentation has also been performed in other insect and mammalian species. While there is great need for an increased understanding of host or food preference in the many important disease vector and pest species, relatively few experiments have tested a sizeable set of odors for activity against multiple receptors. In the largest screen of any insect species outside of *Drosophila*, 110 odors were recently individually tested for activity against 50 *Anopheles* Ors using single unit electrophysiology (Carey et al., 2010). In the largest single mammalian screen a combination of 62 mouse and human ORs were individually exposed to 63 odors and measured for activity using calcium imaging (Saito et al., 2009). The limited number of screened odors tested on a very narrow set of species makes it difficult to tease apart important olfactory cues and pathways that are responsible for attraction, mating, etc.

The odors tested in many screens are meticulously selected, both to cover a broad range of chemical space and for ecological importance. For example, many odors tested in *Drosophila* were selected due to their previous identification in rotting and ripe fruit, an important food source, and a number of odors tested in the *Anopheles* screen performed by Allison Carey were selected for their previously recognized importance as human host volatiles (Carey et al., 2010; Hallem and Carlson, 2006). Odors that were not selected for ecological significance were chosen to cover a broad area of chemical space. These odors cover an expansive range of carbon chain lengths and contain a broad assortment of functional groups including ketones, aldehydes, carboxylic acids, alcohols, and esters. While the majority of screened odors are aliphatic, a number of aromatic odors are usually included as well. These analyses provide the groundwork from which to further expand upon explored odor space, with the ultimate goal of finding effective and species specific behavior modifying chemicals.

Or gene identification in insects was largely aided by computational approaches

Linda Buck correctly made three very important hypotheses in her identification of the OR gene family (Buck and Axel, 1991). Firstly, ORs were likely to belong to a superfamily of proteins involved in GPCR cellular communication. Secondly, the *OR* gene family should be very large and diverse to recognize so many odors. Thirdly, *OR* expression should be restricted to the olfactory epithelium. Guided by these hypothesis, Linda used degenerate primers of known GPCRs to amplify olfactory epithelium cDNA, resulting in the historic identification of band 13 and subsequently the first *OR* gene family (Buck and Axel, 1991).

As previously stated, the Or gene families of Insects are highly divergent from mammalian Ors. While the same approach was applied to identify the *Drosophila* Or gene family, they were unsuccessful. By 1999 bioinformatics approaches were emerging as successful and important tools for biological analysis. Additionally, the first draft of the *Drosophila* genome was underway, providing a ripe time for a computationally guided approach to Or gene family identification (Adams et al., 2000). As conditions were now suitable, two research groups independently identified the *Drosophila* Or gene family in tandem.

Peter Clyne and Coral Warr from John Carlson's lab teamed up with Junhyong Kim and applied bioinformatics approaches to identify 7-transmembrane proteins from the incomplete 1998 *Drosophila* genome build (Clyne et al., 1999; Kim and Carlson, 2002). Recent evidence had shed light that insect Ors would likely share structural features to GPCRs, yet as previous wet lab attempts had been unsuccessful in identification, it was likely these genes were highly divergent from mammalian ORs. Guided by this insight the team created an algorithm to identify potential members of a large 7TM gene family from the roughly 10% completed *Drosophila* genome build. The algorithm began by identifying Open Reading Frames (ORFs) that were > 300 bases. They then identified important sequence descriptors for 7TM GPCRs, testing 70 parameters for the ability to separate 750 training GPCRs from 1000 non-GPCR proteins by sequence analysis alone. The 5 most important parameters were then applied to screen all predicted ORFs of >300 bases, alternative splicing was considered in order to obtain full length 7TM protein sequences, and the top scoring hits that were not previously known GPCRs were selected. RT-PCR primers were designed for the final list of hits and 2 genes were identified that were only expressed in the antennas. The

well known sequence search tool BLAST was next applied, searching for ORFs with sequence similarity to the two potential Or genes, resulting in identification of 16 Or genes shown to be expressed solely in the antennae and maxillary palps (Altschul et al., 1990). A year later the same approach was applied to identify *Drosophila* Gr genes (Clyne et al., 2000; Kim and Carlson, 2002).

In an independent analysis Leslie Vosshall in Richard Axel's lab identified 10 *Drosophila* Or genes using a combination of wet lab and computational approaches (Vosshall et al., 1999). Vosshall applied difference cloning to identify antennal/maxillary palp specific cDNAs, one of which (dor104) was then applied to scan the *Drosophila* genome for transmembrane proteins using the computational programs Dense Alignment Surface (DAS) and TMAP. 10 of the 12 genes with the highest BLAST determined sequence similarity, albeit very low, to dor104 were demonstrated to be expressed solely in either the antennae or maxillary palps.

Initial insect Or gene discovery has led to an explosion of computationally guided Or identification and analysis. All 60 Or and 68 Gr genes in *Drosophila* had been identified within two years of the initial discovery through iterative application of BLAST on the fully completed *Drosophila* genome (Robertson et al., 2003). Since then, the Or gene families of 12 *Drosophila* species have now been identified, allowing for a comparative analysis identifying both gene conservation prior to subgenera division roughly 40 Mya and providing interesting signs of species specific positive selection (Guo and Kim, 2007). In addition to *Drosophila*, computation has aided in Or family discovery of many other species, including both pest and disease vectors. Examples of species in which Or families have been identified include *Anopheles* gambiae,

Acyrthosiphon pisum (Pea Aphid), *Aedes aegypti*, *Nasonia vitripennis* (Jewel Wasp), *Rhagoletis pomonella* (Apple Maggot), *Bombyx mori* (Silk Moth), *Danaus plexippus* (Monarch Butterfly) and several others (Bohbot et al., 2007; Fox et al., 2001; Robertson et al., 2003; Schwarz et al., 2009; Smadja et al., 2009; Zhan et al., 2011). In a slightly different direction, computational approaches have also allowed for analysis of Or genes across distantly related species, such as was performed for the CO2 receptors for 12 *Drosophila*, 3 mosquito, the silk moth, and flour beetle species (Robertson and Kent, 2009), providing insight on gene evolution on a very important behavior-regulating receptor.

Or gene family identification for all of these species was made possible by computational advances in genome sequencing, assembly, and search methods able to identify distantly related genes. Without these approaches, identification of the highly divergent Or and Gr families, would certainly have taken far longer to complete, if at all. As sequencing methods continue to improve, the genomes of additional species will begin pouring in, resulting in an increased importance for bioinformatics in Or and Gr gene identification. In turn, the increased volume of Or gene information will further open the door to cross species comparisons. While many of the same tools will likely be applied for Or and Gr gene identification, many new tools will be required for analysis of sequence-activity relationships. Once motifs and protein regions important for odor selectivity are identified, an entirely new field will be available. Traditionally, 3D structures of receptors are required for ligand docking and binding prediction. However, as the Or gene family is highly divergent while likely still retaining core structural features, key motifs or binding regions on Or sequences may be responsible for odor selectivity. As approaches improve it may be possible to computationally predict

receptor-odor interactions from sequence alone. As Or are the largest gene family in several species, this will provide both large challenges and incredible rewards if successful.

Applying ligand-based virtual screening to predict receptor-odor interactions

The field of chemical informatics has been successfully applied to drug discovery by the pharmaceutical industry for years. With the incredible number of chemicals available and the cost required for purchasing and screening, an approach that can computationally identify chemicals with desired features is incredibly useful for initial compound selection (Maldonado et al., 2006). Many distinct chemical informatics approaches have been designed to explain the relationships shared between molecules. Maximum Common Substructure (MCS) determines the largest substructure that is shared between two molecules (Cao et al., 2008). Atom Pairs (AP) identifies atom types with similar distances that are shared between molecules (Carhart et al., 1985). Another approach Molecular Fingerprints (MFs) calculates bit strings, where each 0 or 1 represents the presence or absence of a particular feature, such as a functional group (Nikolova and Jaworska, 2004). Some commercially available programs are available that calculate thousands of individual molecular descriptors for thousands of query compounds, providing quantitative information on many aspects of a compound, such as molecular weight, carbon atom number, predicted vapor pressure, or complex 3-dimensional relationships between atoms of a molecule. Large virtual chemical libraries, which can contain hundreds of thousands or millions of chemicals, can be screened for desired characteristics by applying these approaches.

Interestingly, the relatively newly developing field of chemical informatics has only been minimally applied to olfaction. There are several reasons why chemical informatics approaches analyzing receptor-odor interactions from the odor (ligand) side should be of high priority to the olfactory community. As a large portion of drugs currently on the market target GPCRs and insect receptors are hypothesized to have GPCR like function and structure, it is suggestive that these techniques should apply well to olfactory ligand prediction. This is especially true for mammalian ORs, as they are GPCRs. Additionally, the structures of Ors have not been solved. As Odor receptors are membrane bound and in the case of insect receptors pair with an obligate co-receptor, obtaining x-ray crystal structures suitable for computational docking will be challenging. Finally, we currently have all of the tools required for chemical informatics analysis. Diverse subsets of odors have been tested for activity against several species, resulting in large number of decoded receptors in both mammals and several important insect species.

While chemical informatics approaches are promising, additional challenges unique to olfaction must be overcome. Odors are far smaller than chemicals that are traditionally investigated by the pharmaceutical industry (Wishart et al., 2008). While the reduction in size may make some processes, such as 3D structure optimization, more straightforward, it may also make structural feature differentiation more challenging. Olfactory researchers will need to identify chemical informatics approaches that lend themselves well to the smaller number structural features separating odors than would be typically found between larger pharmaceutical chemicals. Additionally, Odors are volatile while the majority of pharmaceutical targets are not. Volatility could effect how well odors reach their receptors targets and should to be taken into account. The few

attempts at chemical informatics application to olfactory research have been fairly successfully and broadly accepted.

In 2000 Jan Kaluza and Heinz Breer demonstrated that carbon chain length of aliphatic aldehydes was an important feature for the prediction of binding (Kaluza and Breer, 2000). Compounds with similar lengths were more likely to bind to the same OR and just as importantly changing the length of a chain by a single carbon could entirely abolish activity. While this may now appear obvious to the community, prior to large panel screens this feature of odor selectivity had not been documented. It was the first support for Olfactory receptors that similar chemical shape may result in similar activity.

Seven years later Michael Schmuker attempted the first prediction of receptor-odor activity (Schmuker et al., 2007). Schmuker et al. applied Artificial Neural Networks (ANNs) to predict the responses of 21odors against 7 ORNs. The group calculated ~200 molecular descriptors from MOE (Chemical Computing Group, Montreal) for 47 training odors and 21 test odors. They selected descriptors that were best able to separate the active from inactive molecules and applied them to train the ANN. While the approach was well performed, it unfortunately was only marginally accurate with an accuracy of ~25% (percentage of predicted test odors that activated >50 spikes/sec), which may be due to either molecular descriptor selection, ANN training, or application of ANNs instead of other available approaches.

In 2008 Rafi Haddad from Noam Sobel's laboratory identified molecular descriptors that were correlated among odorants. Haddad et al. calculated 1,664 molecular descriptor values for each odors tested in 7 experimental studies, containing both mammalian and insect receptors (Haddad et al., 2008). The group then selected

which descriptors were the most correlated with activity, resulting in a final set of 32 molecular descriptors that best explained general and non-receptor specific olfactory activity. These 32 descriptors explained activity of the 7 datasets better than carbon atom number or the full set of descriptors. While this approach does not attempt to explain receptor specific selectivity, it was a step in the correct direction.

Most recently in 2010, Sheng Guo and Junhyong Kim applied their own Quantitative Structure Activity Relationship (QSAR) technique to explain the activity of Elissa Hallems analysis where 108 odors were previously tested against 24 *Drosophila* Ors (Guo and Kim, 2010; Hallem and Carlson, 2006). Their model mapped 3D features of each odor and identified which features were most important for receptor activity. Additionally, they performed a sequence analysis for each of the 24 receptors, hypothesizing that the binding pocket of *Drosophila* Ors is on the extracellular halves of it Trans Membrane (TM) domains. The model suggests that the binding pocket is 15 angstroms deep and 6 angstroms wide.

Each of these approaches provides compelling evidence that chemical informatics can be successfully and beneficially applied to olfactory research. However, none of the approaches is successful in explaining receptor specific activity, which should be a major goal of the computational olfaction community. Testing odors using current experimental approaches take a great deal of time and the purchase of a large number of chemicals, many of which may be ineffective. Both of these drawbacks are very expensive. An easy to apply tool that is highly effective at predicting receptor-odor interactions would be immensely beneficial to the olfactory community, allowing for intelligent and quick prioritization of odors for experimental validation. As the number

species for which an initial set of odors has been tested against is currently increasing rapidly, the olfactory field is ripe for such an approach.

Applying structure-based virtual screening to predict receptor-odor interactions

The field of structural-based virtual screening, which involves docking the 3D structure of a chemical into a solved protein structure *in silico*, is regularly used in the field of drug discovery(Kitchen et al., 2004; Nikolova and Jaworska, 2004). Both pharmaceutical companies and academic institutions can have large research groups dedicated to docking libraries of untested chemicals onto the 3D protein structures of potentially drugable targets. The most promising hits are usually selected for experimental validation. The benefits of this system over the ligand-based molecular descriptor prediction techniques are that the resulting dockings can be visually inspected to check that the interaction appears valid, verifying that shape constraints and important polar and non-polar interactions are satisfied. Additionally, manual inspection can reveal important characteristics that can improve structures of current dockings, such as the addition an atom to fully fill a cavity or addition of an oxygen atom to satisfy a structurally important hydrogen bond. Both ligand-based and structure-based and virtual screening can be integrated into a single highly effective pipeline. Molecular descriptors can be applied to select ligands with the most promising characteristics and selected compounds can then be virtually docked into a protein-binding site.

Structure-based virtual screening has been successfully applied for a limited number of mammalian Ors. While structures for insect or mammalian odor receptors have not yet been reported, mammalian receptors are structurally related to rhodopsin,

which has been structurally solved. By applying the structural knowledge from rhodopsin to mammalian Ors research groups have been able to predict a basic structure of mammalian receptors including Rat Or5, Mouse S25, and Rat I7 (Floriano et al., 2000; Singer, 2000; Singer and Shepherd, 1994). The first virtual screening tied to a site directed mutagenesis for functional validation was performed for the mouse mOR-EG in 2005 (Katada et al., 2005). Katada et al. were able to identify the location of an odorant-binding site, as well as how the odor eugenol specifically interacted with it, for the first time, representing a major step forward for the mammalian olfactory community. Since the first successful demonstration of molecular modeling of mammalian ORs, several computational groups have performed similar analyses in additional receptors, including mOR147-9, hOR17-210, and rat OR-I7, without wet lab experimental validation (Khafizov et al., 2007; Kurland et al., 2010; Lai et al., 2008).

In the largest and most complex OR virtual screening attempt to date, 758 compounds from the CAP database were computationally screened for activity against the 5.24 receptor from goldfish, which is activated by all 20 naturally occurring amino acids (Triballeau et al., 2008). Both ligand-based and structure-based virtual screening was applied, reducing the number of potential ligands to 46. Activities of the top 4 predictions were validated using an electro-olfactogram (EOG), each of which produced activation.

Interestingly, structure-based virtual screening has loosely been applied to the human taste receptor TAS2R46 (Brockhoff et al., 2010). Protein chimeras between receptors hTAS2R46 and hTAS2R31 were analyzed to identify important amino acid binding regions, classifying the C-terminus as an important region for ligand selectivity.

Focusing their attention on the extracellular side of trans-membrane spanning alpha helices 6 and 7, they swapped amino acid residues to identify the location of the ligand-binding site. Once a region was identified, a structure-based virtual screen was performed to predict the exact biding site, important interacting residues, and ligand orientation in the membrane bound receptor. This effective blending of wet lab experimental and computational analyses hypothesized the exact orientation of Strychnine on the extracellular side of hTAS2R46.

The field of protein structure determination is rapidly advancing, allowing for determination of structures that were previously untenable. As methodologies continue to advance, our ability to solve membrane bound receptors with heteromeric partners will become possible in the near future, paving the way for structure-based virtual screening of the structurally elusive insect Ors. As the sequences of insect Ors are highly divergent, an additional challenge will be to model the structures of many divergent Ors based upon the few initially solved Insect Ors. As these structures become available, a combined approach involving an initial ligand-based screen of a very large chemical space followed by a focused structure-based virtual screen will likely provide accurate predictions of receptor activations for hundreds of thousands of odors, which would take an unknown number of years to test using wet lab experimentation alone.

High-throughput approaches are effective tools for olfactory analysis

While virtual screening is proving to be an effective tool for identification of receptor-odor interactions, expression analyses are also highly aided by computational approaches. Recent advances in next generation analysis tools, such as microarrays,

RNA-seq, and chip-seq have provided wonderful new ways to understand gene regulation. What would have been impossible just a few years ago is quickly become standard experimental procedure today. Since the initial discovery of the structure and organization of DNA in 1953, a strong charge has been made to understand as much about the genetic makeup of as many species as possible (Watson and Crick, 1953). A major step in this direction was made when Fredrick Sanger and colleagues published the Sanger sequencing method (Sanger and Coulson, 1975). While this method was in no way high throughput by today's standards, it allowed for the sequencing of individual sequences in a straightforward manner that could be performed in a general lab setting. Similar approaches were applied for roughly twenty years until the field-changing introduction of current high throughput techniques. Using today's approaches, such as microarray analysis, RNA-seq, Chip-seq, it is possible to determine the expression of thousands of unique genes of interest or chromatin protein-association in a single experimental protocol. Millions of base pairs can be sequenced. This explosion in sequencing was been greatly aided by computational approaches.

The sheer volume of data produced by a single experiment is astonishing and analysis would simply not be possible without computation. Every aspect from genome assembly to quantification of transcriptional analysis is almost exclusively performed by bioinformatics. For example, transcriptional analysis using Illumina sequencing requires the reading and recording of millions of base pairs. Each read sequence must be checked for quality and aligned to a reference genome. Relative expression profiles must be determined for every gene in the genome prior to promoter sequence, alternative splicing, or SNP analysis is performed. By applying bioinformatics to these approaches a researcher can literally determine SNP placement for every gene in an

individual against a reference genome. As many transcriptional analyses compare gene regulation across experimental conditions, such as time, the entire analysis is performed multiple times and compared, further increasing the complexity. Clearly, computational approaches are advantageous to biological research. High-throughput approaches are increasingly being successfully applied to olfactory research.

Diego Rodriguez-gil from Stuart Firestein's lab recently applied micro-arrays to analyze Or gene expression across mouse development (Rodriguez-Gil et al., 2010). His analysis demonstrated that OR expression is first detectable at embryonic day 9 and that by day 13 expression of nearly all OR genes is detectable. Interestingly, they also noted that gene expression is not constant throughout the course of a mouse life as the "olfactome" expression decreased beyond 3 months of age.

Ewald Grosse-Wilde of Bill Hanson's lab recently identified the Or gene family of *Manduca sexta* using transcriptome analysis by applying 454 sequencing to an antennal cDNA library (Grosse-Wilde et al., 2011). Gene Ontology (GO) annotation using the software Blast2GO was applied to identify olfactory related ontologies. Ors were additionally identified by Blast and HMM profile searches from custom build databases. Predicted amino acid sequences for putative Ors were aligned to previously known *Heliothis* and *Bombyx* sequences, allowing for positive gene identification. Odor Binding Proteins (OBPs) and IR genes were also identified using the same method. In total 54 unique Or gene fragments were identified in a species that contains 73 unique globeruli. If *Manduca* follows the generally applicable one neuron to globerulus map, then roughly 73% of functional Or genes were identified. This approach is considerably significant as it represents application of high throughput transcription analysis in order to identify Or

genes in a species that have not been sequenced by using related species as a reference. As Or identification is traditionally achieved through genome sequencing, this faster and far less expensive transcriptome analysis demonstrates and approach that can be applied to many more species in the future.

R Jason Pitts of Laurence Zwiebel's lab has recently applied transcriptome profiling in *Anopheles gambiae* (Pitts et al., 2011). Both sex specific (male vs female) and olfactory appendage specific (antenna vs whole body) transcriptional regulation was compared using Illumina sequencing (RNA-seq). This approach was able to identify nearly all chemosensory genes from antennal tissue at quantifiable and statistically significant levels. The analysis revealed that expression levels of olfactory genes were present in both sexes, however the levels were much higher in females than males. Additionally, males had a significantly larger number of genes involved in audition. These results were interesting and make sense as it is the female mosquitoes that need to identify hosts for blood feeding and males that need to hear females for mating. This approach demonstrates the sheer volume of information that can be acquired using RNA-seq and its effective application to understand tissue specific regulation in a species and it's broader implications.

Each of the previously highlighted analyses demonstrates how high throughput approaches can be successfully and advantageously applied to answer direct questions or identify targeted genes in the field of olfaction. With the continual reduction in cost as well as increased computational analysis tools available for these methods, more information can be gained at a less expensive price. As computational power has steadily increased, analyses that was once impossible can now be easily performed on a

desktop computer. Large numbers of need-driven computational approaches are being designed to assist in answering biological and chemical questions that can now be asked as the technology that was once considered science-fiction is now simply science.

Figure 1.1: Overview schematic describing neuronal wiring of the insect olfactory system

Odors are detected by Olfactor receptor (Or) expressing neurons in the periphery organs (antennae and maxillary palps). The axons of Olfactory Receptor Neurons (ORNs) expressing the same Or proteins, thus being housed in the same sensilla class, extend into unique glomerular locations within the antennal lobe. Local interneurons (LNs) allow for both excitatory and inhibitory information to be shared between specific sets of glomeruli. Projection Neurons (PNs) then send information from the antennal lobe to higher brain centers in the Mushroom Body (MB) and Lateral Horn (LH), where behavioral decisions occur. The figure was taken from (Carey and Carlson, 2011).

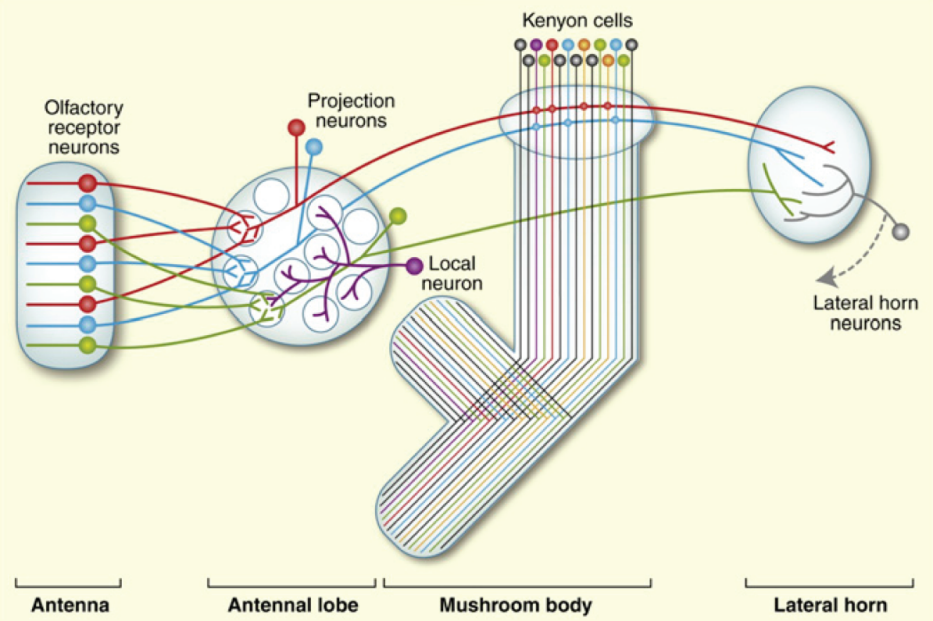

Kenyon cells
Olfactory receptor neurons
Projection neurons
Local neuron
Lateral horn neurons
Antenna
Antennal lobe
Mushroom body
Lateral horn

Figure 1.1

CHAPTER II:

Designing a Ligand-Based Virtual Screening Approach to Decode *Drosophila* Odor Receptor Chemical *Space In Silico*

INTRODUCTION

The peripheral olfactory system is unparalleled in its ability to detect and discriminate amongst an extremely large number of volatile compounds in the environment. To detect this wide variety of volatiles, most organisms have evolved large families of receptor genes that typically encode 7-transmembrane proteins expressed in the olfactory neurons (Buck and Axel, 1991; Clyne et al., 1999; Dahanukar et al., 2005; de Bruyne and Baker, 2008; Vosshall et al., 1999). Each volatile chemical in the environment is thought to interact with a specific subset of odor receptors depending upon odor structure and binding sites on the receptor. This precise detection and coding of odors by the peripheral olfactory neurons are subsequently processed, transformed and integrated in the central nervous system to generate specific behavioral responses that are critical for survival such as finding food, finding mates, avoiding predators etc (van der Goes van Naters and Carlson, 2006).

Currently there are two major rate-limiting steps in analysis of peripheral coding in olfaction: a very small proportion of chemical space can be systematically tested for its activity on odor receptors, and a very small fraction of the numerous odor receptors have been tested for responses (Araneda et al., 2000; Hallem and Carlson, 2006; Hallem et al., 2004; Kreher et al., 2008; Pelz et al., 2006; Saito et al., 2009). The challenges for overcoming the rate-limiting steps are enormous. First, volatile chemical space is immense, more than 2,000 odors in the environment have been catalogued from a small

fraction of plant sources alone (Knudsen et al., 2006). Second, the complete 3-D structures of the 7-transmembrane odor receptor proteins have not yet been determined and modeling of protein-odor interactions and sophisticated virtual screening methods are not yet possible except in rare instances (Triballeau et al., 2008). In the decade since the first systematic study of 47 odorants on the *Drosophila* antenna in 2001 (de Bruyne et al., 2001), additional studies have only identified a total of 251 novel active odors (de Bruyne et al., 1999; de Bruyne et al., 2001; Dobritsa et al., 2003; Goldman et al., 2005; Hallem and Carlson, 2006; Hallem et al., 2004; Kreher et al., 2005; Kreher et al., 2008; Kwon et al., 2007; Pelz et al., 2006; Schmuker et al., 2007; Stensmyr et al., 2003; Turner and Ray, 2009; van Naters and Carlson, 2007; Yao et al., 2005), which have been assembled and compared in an online database (Galizia et al., 2010).

Here we overcome this challenge by designing a chemical-informatics platform that is effective and fast. In order to do so we focused our attention on one of the most comprehensive quantitative data sets available, where measurements of responses of 24 *Drosophila* odor receptors to a panel of 109 odorants are known (Hallem and Carlson, 2006). We devised a method to identify molecular structural properties that are shared amongst the activating odorants for each receptor. We then utilize information about these shared molecular features of actives, that are presumably required for binding to a receptor, to perform *in silico* screens on a chemical space of >240,000 chemicals and predict the top 500 hits as new ligands for each of the odorant receptors (Ors). We then use single-unit electrophysiology to validate a subset of predictions *in vivo* and find that our method met with an overall success rate of ~71% in identifying novel ligands, as compared with a low 10% receptor activation rate while using non predicted odors. This approach allows us to create a peripheral coding map of a large

chemical space, which substantially improves our ability to study peripheral olfactory coding and provides a powerful tool for the discovery of novel ligands for Ors.

RESULTS

Analysis of odorant structure

Since the structure of receptor protein complexes is not known, we analyzed receptor-odor interactions by applying the *similarity property principle*, which reasons that structurally similar molecules (e.g. activating odorants) are more likely to have similar properties (Hendrickson, 1991; Martin et al., 2002). Although this general approach has been useful in the area of pharmaceuticals (Keiser et al., 2009; Martin et al., 2002), receptor-odor analysis presents significant additional challenges. Not only are odorant molecules generally smaller in size than pharmaceuticals (average MW of known odors ~3-fold less than FDA approved pharmaceuticals (Wishart et al., 2008)) and therefore offer fewer structural features for differentiation, they are also detected by the receptors with specificity at extremely low concentrations in the volatile phase (Hallem and Carlson, 2006; Kreher et al., 2008). Additionally, odor receptors are differentially tuned and can sometimes appear not to follow distinct structural rules: odors that look structurally different can strongly activate the same receptor, while odors that appear very similar may have very different levels of activity (Hallem and Carlson, 2006)(Figure 2.1A).

General measures of odorant similarity

Similarity in chemical structure can be described and measured quantitatively using multiple approaches, however a single method may not be ideal for every single application (Maldonado et al., 2006). In order to test whether non-optimized approaches would be able to identify similarities in shape of known activators we compared four different approaches: Cerius2 (Accelrys Software Inc), Dragon (Talete), Maximum-Common-Substructure (MCS) (Cao et al., 2008b), and Atom-Pair (AP) (Cao et al., 2008a; Carhart et al., 1985). Cerius2 and Dragon represent collections of 200 and 3,224 molecular descriptors, respectively, that calculate values for a broad range of chemical properties such as molecular weight, functional group counts, and in the case of Dragon, 3 dimensional relationships within molecules. The AP method compares shortest path distances between all atom pairs in a molecule. Lastly, MCS identifies the largest 2 dimensional substructure that exists between two compounds. Using each of these approaches, we computed distances between 109 odors that had previously been tested against 24 Ors from *Drosophila melanogaster* (Hallem and Carlson, 2006) (Figure 2.1B). These represent most of the *Or* genes expressed in the *Drosophila* antenna (Hallem and Carlson, 2006). Upon comparison, we find that none of the four methods were vastly superior and that each method varied in the ability to group known activating odorants "actives" close together in distance as measured for each Or using a method called Accumulative-Percentage-of-Actives (APoA)(Chen and Reynolds, 2002) (Figure 2.3A, See Methods), and value of the Area-Under-the-Curve (AUC). Ultimately, Dragon and Cerius2, which utilize a large number of diverse molecular descriptor values to describe each odor structure, performed better than AP or MCS, suggesting that a more diverse

set of descriptors is better at explaining Or activity than 2D measures alone (Figures 2.2B, 2.3B). Atom-Pair and MCS were subsequently ignored from further development.

Identification of unique subsets of optimized descriptors for each *Drosophila* Or

Individual Ors respond to distinct subsets of ligands with some degree of overlap (Hallem and Carlson, 2006; Kreher et al., 2008). We reasoned that rather than using entire Dragon or Cerius2 descriptor sets, which likely includes a number of measurements for features irrelevant for ligand-binding to an individual Or, judiciously selecting subsets of molecular descriptors suited to cluster actives for an individual receptor may be more effective at defining an Or-specific chemical space. To test this hypothesis, we used a Sequential-Forward-Selection (SFS) method to incrementally create unique optimized descriptor subsets for each Or from an initial combined set of 3,424 descriptors from Dragon and Cerius2 (Whitney, 1971) (See Methods, Figure 2.2A). This optimization-based analysis was performed on the 19 Ors from the dataset with known activating odors, excluding Or82a, since it has but a single known strong active (Hallem and Carlson, 2006).

Not surprisingly, the composition of the optimized descriptor sets varied greatly between Ors (Table 2.1). Molecular descriptors can be categorized from 0 to 3 dimensions. 0-Dimensional (0-D) descriptors define features that can be viewed as not directly being shape dependent, such as molecular weight or vapor pressure. On the other end of the scale, 3-Dimensional (3-D) descriptors define features of molecules in three-dimensional space, such as the distance between two atoms of an odor molecule. We find an overwhelming preference for 3-D and 2-D descriptors compared to 1-D and 0-D descriptors, suggesting that structural shape features are more important for

receptor-odor interactions. The Or-optimized descriptor sets were far superior at grouping together activating odors from the training set (Figure 2.2B) when compared to the non-optimized methods (Dragon, Cerius2, MCS, AP) and a previously identified collection of descriptors that were identified without receptor-specific optimization (Haddad et al., 2008).

Computational validation of optimized descriptor sets

In order to validate the predictive ability of the *Or*-optimized method, we performed 5 independent trials of 5-fold cross-validations followed by a Receiver-Operating-Characteristic (ROC) analysis, an established computational approach (Hastie et al., 2001; Tan et al., 2006) (see methods). Briefly, this involved withholding 20% of the 109 previously tested odors for a receptor. Descriptors were optimized using the remaining 80% odors for training, and ligand-predictions were subsequently performed on the 20% of odors that were withheld. This operation was repeated 5 times for each receptor, each time selecting a different 20% as withheld from the training set. The entire 5-fold operation was repeated 5 times for each receptor and a mean ROC curve representing the prediction accuracy determined. This analysis was possible for 12 *Ors* which had >6 known ligands that activated >100 spikes/sec. The Area-Under-Curve (AUC) value (0.815) is very promising and suggests that the *Or*-optimized descriptor sets are effective at predicting novel ligands (Figure 2.2C).

In addition to performing the 5-fold cross-validation, we also clustered the 109 training odors independently for each Or using distances calculated from the previously determined receptor specific descriptor sets (Figure 2.2D). As expected, we find that active odors cluster tightly together for each Or (Figure 2.2D). In a few cases, such as

for Or35a and Or98a, not all the highly active compounds are clustered, suggesting the possibility of multiple or flexible binding sites, or imperfect selection of descriptors. Four of the Ors (Or2a, Or23a, Or43a and Or85f) have few known activators, none of which activate the receptors at greater than 150 spikes/sec, however our descriptor optimization approach is still able to cluster each the few weak activators together (Figure 2.2D).

High-throughput *in silico* screening of odor receptors

Since Or-optimized descriptor sets can efficiently group highly active compounds in chemical space, we used them to rank untested compounds according to their distance from known actives for specific Ors. We assembled a natural odor library, which contains 3,197 naturally occurring odors, and a library derived from Pubchem (Bolton et al., 2008), which contains >240,000 compounds with similar molecular weights and atom type compositions to known volatiles (See Methods). We then systematically screened both libraries using the optimized descriptor sets of 19 *D. melanogaster* Ors representing ~5,000,000 receptor-odor interactions *in silico*. We identify the top 500 (0.2%) hits from this vast chemical library for each Or, the top ~100 of which are reported in Table 2.2.

Electrophysiological validation of *in silico* screen and identification of agonists

To validate our *in silico* screen, we obtained several untested odorants (141 total; ~11-23/Or) belonging to the top 500 predicted ligands for 9 different Ors that were available from commercial sources at high purity and reasonable prices. The 9 receptors were selected on the basis of previous functional mapping studies that enable us to

unambiguously identify the antennal olfactory receptor neurons (ORNs) they are housed in (Couto et al., 2005; Hallem et al., 2004). We systematically tested each predicted receptor-odor combination using single-unit electrophysiology to record from the ORNs to which these 9 Ors have been previously mapped (Couto et al., 2005; Hallem et al., 2004). We find that a majority of the predicted ligands evoked responses from the target ORNs; ~71% evoked either activation (>50 spikes/sec above the spontaneous activity) or inhibition (>50% reduction in spontaneous activity) (Fig. 2.6A). These cutoffs were selected based on study from which the training set was obtained and has been used in other studies in the past that use this type of recordings (Hallem and Carlson, 2006; Kreher et al., 2008). Interestingly, the mean vapor pressure of activating odors (11.84 Torr) is 7.5 times higher than of inactive odors (1.58 Torr), raising the possibility that some inactive odors may not be volatilized and delivered at adequate levels to the ORNs. Additionally, we find that ~13% of the predicted compounds we tested showed an inhibitory effect on baseline activity of the respective neuron (Fig. 2.6A). These inhibitors were identified by virtue of structural similarity to known activators suggesting that they may bind to similar sites on the receptor. Thus as an additional benefit our approach may provide a method to identify inhibitors as well. Such inhibitors would not only provide important tools to investigate mechanisms of odor receptor inhibition but could also be used in blocking specific odor-mediated behaviors. Consistent with our observations three of the receptor-odor interactions had been previously identified independently as well, Or22a (Pelz et al., 2006), and Or49b (Hallem et al., 2004). The electrophysiological analysis provides the most important validation of our Or-optimized descriptor-based *in silico* screen.

Odor response spectra of individual Ors

Since we systematically analyzed responses of a large number of new odorants individually, we were able to characterize the odor-response spectra of several antennal ORN classes to these new ligands (Figure 2.4B). New activators are reported for every receptor, and inhibitors are identified for several. Ligand predictions for 2 of the 3 receptors that do not perform as well are Or10a and Or49b that detect aromatic compounds. Their poor performance is explained by the lack of aromatic ligands in the initial training set (13/109) odorants. We find that a >85% of the predicted ligands activate odor receptors Or7a, Or22a, Or59b, Or85a, Or85b, and Or98a (Figure 2.4B).

Specificity of *in silico* predicted ligands

We rigorously examined the rate of false negative predictions for each Or by systematically testing newly identified ligands of each Or against the other non-target receptors using electrophysiology. Of 504 non-target receptor-odor interactions tested, we found that only 10% evoked a response >50 spikes/sec and 3.7% evoked a response >100 spikes/sec (Figure 2.5A). This represents a high degree of specificity, especially considering that the Or-optimized descriptor method did not incorporate any additional computational screening to rule out non-target activators. Additionally, when we plot the percentage of odors that validated as activators when tested using electrophysiology (considering both predicted and non-target receptor-odor interactions), we find that activity is strongly related to predicted odor ranking (Figure 2.5B). Odors which rank closest to known actives for each Or, particularly within the top 500 hits, are far more likely to be activators than odors further away, and there is a drastic drop-off in activating odors present beyond the 1,000 rank. We see the same trend if we plot mean

activity of odors for the same ranking divisions (Figure 2.6). Highly ranked odors have a far higher mean activity than distantly ranked odors.

Relationship between descriptor sets and Or sequence and activity

Since receptor-optimized descriptor sets and the predicted ligand space they define are a function of shared molecular features that a receptor may employ to recognize ligands, we were now in a position to determine how these characteristics correlate with receptor properties such as their known-activity profiles and amino acid sequences. We used hierarchical cluster analysis to create trees that represent the various receptors based on: shared descriptors selected; known activity-based relationships (Hallem and Carlson, 2006); degree of overlap of predicted ligands; and amino acid sequence (Figure 2.7A, See Methods). We found that the maximum overlap in Or relationships is retained between the descriptor and the activity trees, and the descriptor and the cross activity trees with 11 out of 24 Ors present in subgroups that are common in both cases. However, only two subgroups (yellow and grey) are conserved across the 3 trees. The largest shared overlap existing in the descriptor tree suggests that the Or-optimized descriptors link the known and the predicted receptor-odor interactions and that our analysis may expand upon odor receptor activity relationships beyond those previously known from the training data. We also found that the phylogenetic tree has fewer relationships conserved with each of the trees, consistent with previous observations (Hallem et al., 2004) supporting the idea that, while the most conserved amino acid residues in the Ors provide the structure of the tree, they do not correlate strongly with ligand specificity.

Analysis of breadth of predictions for each Or in chemical space

Coding of odors in a large volatile space (>240,000) by a receptor repertoire is virtually impossible to determine experimentally. However, based on the Or-optimized descriptor sets we computationally derived prediction frequency distributions for each of the *Drosophila* Ors in this large chemical space (Figure 2.7B). As expected, we find substantial variation in the distribution frequency of predicted ligands across different receptors. The predicted response profiles support previous observations made with smaller odor panels that the olfactory system can potentially detect thousands of volatile chemicals, many of which the organism may never have encountered in its chemical environment. We computed similar predicted ligand frequency distribution curves to an assembled set of 3,197 known "natural" volatile compounds from plants, humans, and a fragrance collection (Figure 2.8). Plant volatiles constituted a large portion of compounds that are predicted to be ligands for *Drosophila* Ors. To further analyze odor source representation, we classified odors that belong to top 500 prediction lists according to their source, if known, and find that Ors are not specialized for odors from a single source (Figure 2.7C).

Across-receptor activation patterns in *Drosophila*

To study the ensemble activation patterns of odors predicted across all Ors, we analyzed the across-receptor activation patterns of the 3,197 known compounds for 9 receptors (Or7a, 10a, 22a, 47a, 49b, 59b, 85a, 85b, 98a). Surprisingly, we find that only 25% of compounds from the "natural" odor library found in the top 500 predictions for each Or are predicted to activate multiple Ors (Figure 2.7D, lower left panel). If we consider compounds from the Pubchem library in the top 500 predicted actives for each

receptor, we observe further reduction in the proportion of across-receptor activating compounds (Figure 2.7D, upper right). Consistent with this prediction we find that cross-activation by ligands functionally evaluated in this study for 9 receptors is lower than that reported previously using ligands of comparable strength for the same 9 receptors (Hallem and Carlson, 2006) (Figure 2.7D, lower right panel). These data suggest that a number of natural odors may be detected by only a few receptors, particularly at low concentrations.

Producing a systems level view of receptor activity for the *Drosophila* antenna

One of the ultimate challenges of understanding peripheral coding of olfactory information is to be able to map responses of large number of receptors to their ligands. Following the *in silico* analysis, we were able to create such a network-view of peripheral odor coding in the *Drosophila* antenna by mapping all predicted and tested receptor-odor combinations (Figure 2.9A) as has been done previously for mapping drug-target networks (Keiser et al., 2009). Using our chemical informatics pipeline, it becomes possible to infer the network of odor receptors that are activated from complex odor sources without the expensive and time consuming process of purchasing and testing all possible odors.

DISCUSSION

A primary element of the olfactory code is information about odor identity, represented by the characteristic interaction of an odor with the ensemble of olfactory receptors in the nose. Here we report an *in silico* approach to systematically identify ligands from a vast chemical space for a majority of Ors in the antenna of *Drosophila*. We demonstrate that our predictions are accurate using two different validation approaches- computational validations and functional validation using electrophysiology. There is a strong correlation between ranks of predicted ligands to electrophysiological activity.

Obtaining and testing odors using traditional methods is time and cost intensive. Electrophysiology and calcium imaging are consuming processes that require not only a great deal of time to perform, but also the purchase of each odor to be physically tested. Moreover, large plate-based combinatorial chemical libraries, which are commonly implemented in drug discovery in the pharmaceutical industry, are not available for volatile odor libraries for reasonable costs. Since *Drosophila* is a premier model for understanding neurobiology of olfaction, several laboratories over the last 12 years have together screened ~250 odors, activities of which have been and compiled into a valuable database that standardizes across studies (Galizia et al., 2010). In this study we screen >240,000 chemicals and predict >10,000 new ligands which represents a substantial expansion of the known peripheral olfactory code for this important model organism and provides a system-level view of odor detection (Figure 2.9B).

A similar, yet much smaller, analysis applied chemical informatics on *Drosophila* olfactory neuron activities to 47 odorants and screened ligands from 21 untested compounds in *Drosophila* (Schmuker et al., 2007). Although this study had a relatively

modest success rate of ~25% at predicting untested odorants as activators (by applying the same 50 spikes/sec threshold for comparison), it also highlighted that structure-based ligand prediction is a viable method for further development. In another interesting analysis a Quantitative Structure Activity Relationship (QSAR) model was applied to describe odor-activity for *Drosophila* Ors. Important amino acid residues were identified using information from orthologous Or sequences identifying potential odor-binding regions, which was postulated to be 15 angstroms deep and 6 angstroms wide (Guo and Kim, 2010). Our approach is conservative and designed to search for novel odors that share structural features from a previously tested odor panel. Odor molecules are limited in size as well, and may offer a limited scaffold such that novel isofunctional chemotype identification may not be as prevalent as has been seen in other examples of scaffold-hopping (Schneider et al., 2006). However while compounds that share similar values for the optimized descriptors do have structural similarity for selected parts of the molecule, it does not mean that they are not structurally different in other parts of the molecule. In the future, application of machine learning approaches, such as Support Vector Machines (SVMs) to the receptor-optimized molecular descriptor sets, may be useful to further increase the predictive ability. Additionally, we could replace our SFS approach with sequential floating search techniques, which allows for removal, as well as addition, of descriptors in the growing optimized list.

We predict that a number of odorants at low concentrations may be detected by only one or a few receptors. This contrasts a current model of combinatorial coding in which emphasis is placed on the notion that a majority of volatile chemicals, with the exception of pheromones and CO_2, are detected by combinations of several odor receptors. One possible explanation for this disparity could be that our predictions are

fundamentally conservative in nature because we focus only on structurally similar ligands and 7-transmembrane heteromeric receptors may also contain additional unexplored binding sites. Moreover, receptors may respond to compounds ranked beyond the top 500 hits. Another possibility is that previously tested subsets of odors were potentially selected on the basis of strong responses in electroantennograms and behavior assays, which could bias selection of cross-activating odors. In fact it is known that complex fruit odors activate fewer Ors than the number activated by individual odorants at comparable concentrations (Hallem and Carlson, 2006, Semmelhack and Wang, 2009). The architecture of the olfactory code therefore appears to integrate two different models. On the one hand, most odors are detected by a few Ors from the repertoire, which may enhance the specificity of the olfactory system for detection of a large number of odors. On the other hand, 15-20% of odors are predicted to activate several Ors (up to 50%) at the same time, which may serve to aid the olfactory of the system in discriminating between fine concentration changes of important stimuli by having Ors tuned to low and high concentrations such as shown for Or42a and Or42b (Kreher et al., 2008).

By identifying a large number of new ligands for each odor receptor, we can also begin to systematically compare the ligand tuning profiles for each in the endogenous neurons versus the “empty neuron” decoder system. If clear differences were identified, it could enable the identification of underlying reasons such as differences in levels of receptor expression in the neurons, or presence of different Odorant Binding Proteins (OBPs) in the sensillum lymph.

This cheminformatics pipeline can also be applied for system-level analysis of other insects whose receptors and ORNs have been decoded such as mosquitoes

(Carey et al., 2010), and vertebrates such as mice and humans (Saito et al., 2009). The search for novel insect repellents and attractants for species that transmit disease and destroy crops can be greatly assisted by a rational prioritization using such a cheminformatics approach.

Figure 2.1: Comparing efficacy of various structure analysis methods to analyze odor receptor ligands

(A) Comparison of odor structures for known actives and inactive odorants for Or85b and Or98a from(Hallem and Carlson, 2006) (activity shown in parenthesis). **(B)** Overview of the process by which 4 molecular descriptor methods were compared to determine which one best clusters known actives close together in descriptor space.

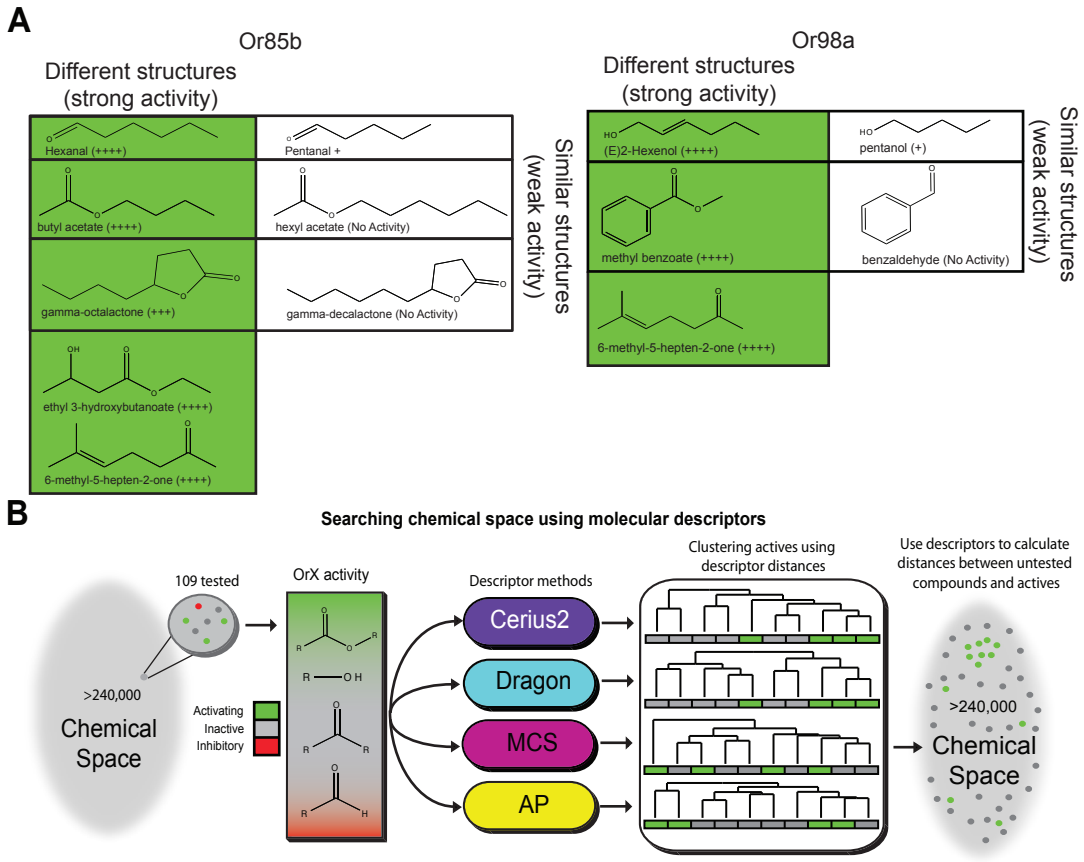

A
Or85b
Different structures
(strong activity)
Hexanal (++++)
Pentanal +
butyl acetate (++++)
hexyl acetate (No Activity)
gamma-octalactone (+++)
gamma-decalactone (No Activity)
ethyl 3-hydroxybutanoate (++++)
6-methyl-5-hepten-2-one (++++)
Similar structures
(weak activity)
Or98a
Different structures
(strong activity)
(E)2-Hexenol (++++)
pentanol (+)
methyl benzoate (++++)
benzaldehyde (No Activity)
6-methyl-5-hepten-2-one (++++)
Similar structures
(weak activity)
B
Searching chemical space using molecular descriptors
109 tested
OrX activity
Descriptor methods
Clustering actives using
descriptor distances
Use descriptors to calculate
distances between untested
compounds and actives
>240,000
Chemical
Space
Activating
Inactive
Inhibitory
Cerius2
Dragon
MCS
AP
>240,000
Chemical
Space

Figure 2.1

Figure 2.2: A receptor-optimized molecular descriptor approach has strong predictive power to find new ligands

(A) Schematic of the cheminfomatics pipeline used to identify novel ligands from a larger chemical space. **(B)** Plot of mean APoA values for 19 *Drosophila* Ors calculated using various methods including a previously identified set (Haddad et al., 2008). (**C)** Receiver-operating-characteristic curve (ROC) representing computational validation of ligand predictive ability of the Or-optimization approach. **(D)** Hierarchical cluster analysis of the 109 odorants of the training set (Hallem and Carlson, 2006) using Or-specific optimized descriptor sets to calculate distances in chemical space for odorant receptors with strong activators (green), and odorant receptors with no strong activators (yellow).

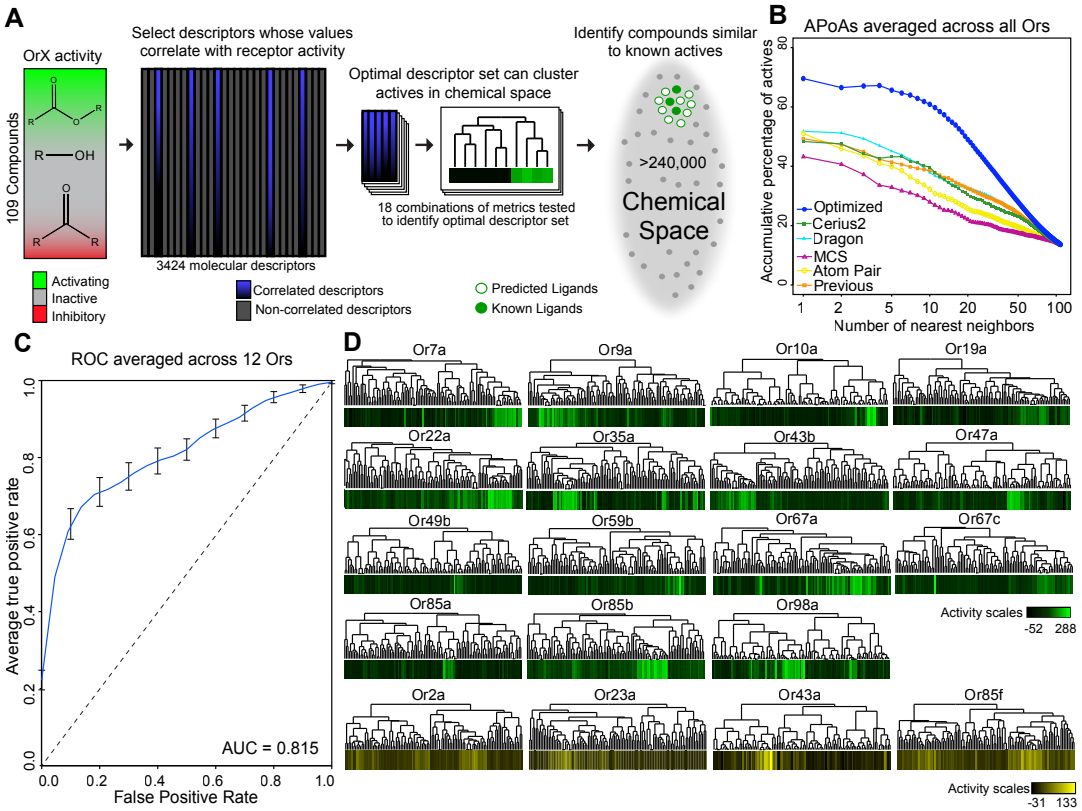
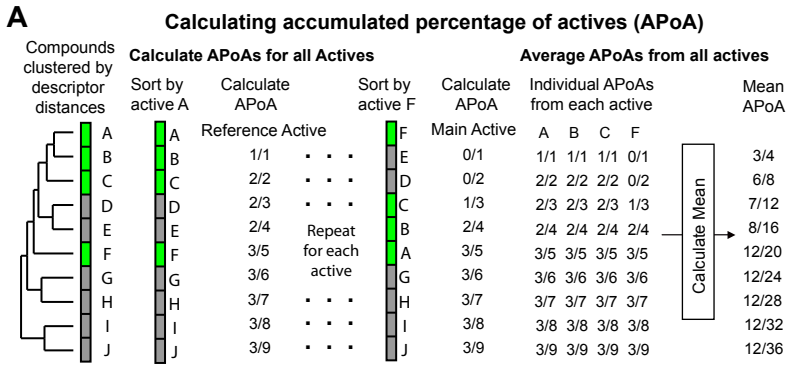

A
OrX activity
109 Compounds
Activating
Inhibitory
Select descriptors whose values correlate with receptor activity
3424 molecular descriptors
Correlated descriptors
Non-correlated descriptors
Optimal descriptor set can cluster actives in chemical space
18 combinations of metrics tested to identify optimal descriptor set
Identify compounds similar to known actives
>240,000
Chemical Space
Predicted Ligands
Known Ligands
B
APoAs averaged across all Ors
Accumulative percentage of actives
Number of nearest neighbors
Optimized
Cerius2
Dragon
MCS
Atom Pair
Previous
C
ROC averaged across 12 Ors
Average true positive rate
False Positive Rate
AUC = 0.815
D
Or7a
Or9a
Or10a
Or19a
Or22a
Or35a
Or43b
Or47a
Or49b
Or59b
Or67a
Or67c
Or85a
Or85b
Or98a
Or2a
Or23a
Or43a
Or85f
Activity scales
-52 288
Activity scales
-31 133

Figure 2.2

Figure 2.3: Accumulated percentage of actives analysis

(A) Representative example for Accumulated Percentage of Actives (APoA) calculation. Green box=active, grey box=inactive. To calculate APoA each active compound was iteratively used as a reference active. Compounds are sorted based upon their increasing descriptor based distance from reference active, and the APoA calculated for each of the other compounds as a ratio of the number of actives over the total number of compounds considered from the reference compound. This process was repeated using each active odorant as a reference active. Reference compound APoAs were averaged to a single mean APoA value. The higher the APoA value while considering a fixed number of nearest neighboring compounds, the greater the proportion of active compounds clustered together. **(B)** Coloured cells mark the method that clusters active ligands best as determined by the highest Area-Under-Curve (AUC) for APoA values.


A
Calculating accumulated percentage of actives (APoA)
Compounds clustered by descriptor distances
Calculate APoAs for all Actives
Sort by active A
Calculate APoA
Reference Active
Repeat for each active
Sort by active F
Calculate APoA
Main Active
Average APoAs from all actives
Individual APoAs from each active
A B C F
Calculate Mean
Mean APoA
A 1/1 1/1 1/1 0/1 3/4
B 2/2 2/2 2/2 0/2 6/8
C 2/3 2/3 2/3 1/3 7/12
D 2/4 2/4 2/4 2/4 8/16
E 3/5 3/5 3/5 3/5 12/20
F 3/6 3/6 3/6 3/6 12/24
G 3/7 3/7 3/7 3/7 12/28
H 3/8 3/8 3/8 3/8 12/32
I 3/9 3/9 3/9 3/9 12/36
J

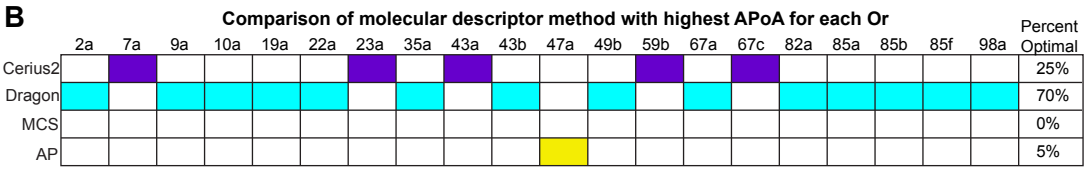

B
Comparison of molecular descriptor method with highest APoA for each Or
2a 7a 9a 10a 19a 22a 23a 35a 43a 43b 47a 49b 59b 67a 67c 82a 85a 85b 85f 98a
Percent Optimal
Cerius2 25%
Dragon 70%
MCS 0%
AP 5%

Figure 2.3

Figure 2.4: Electrophysiology validates that odorant receptor-optimized molecular descriptors can successfully identify new ligands for *Drosophila*

(A) Summary of prediction accuracy percentages obtained by electrophysiology validation. Ligands = Agonists (>=50 spikes/s) + Inhibitors (>50% reduction from baseline activity). **(B)** Mean increase in response of neurons to 0.5-sec stimulus of indicated odors (10^{-2} dilution) predicted for each associated Or. Dashed lines indicate the activator threshold (50 spikes/second). ΔH: Or85b (ab3B) = flies lack expression of Or22a in neighboring neuron, thus all observed neuron activation is unambiguously caused by Or85b. N=3, error bars=s.e.m.

A

Classification	Or7a	Or10a	Or22a	Or47a	Or49b	Or59b	Or85a	Or85b	Or98a	Total
Ligands	88%	31%	86%	39%	27%	91%	92%	87%	100%	71%
Agonists (>50 spikes/sec)	63%	31%	81%	33%	18%	64%	69%	70%	92%	58%
Agonists (>100 spikes/sec)	31%	13%	62%	11%	9%	45%	54%	48%	67%	37%
Inhibitors	25%	0%	5%	6%	9%	27%	23%	17%	8%	13%

B

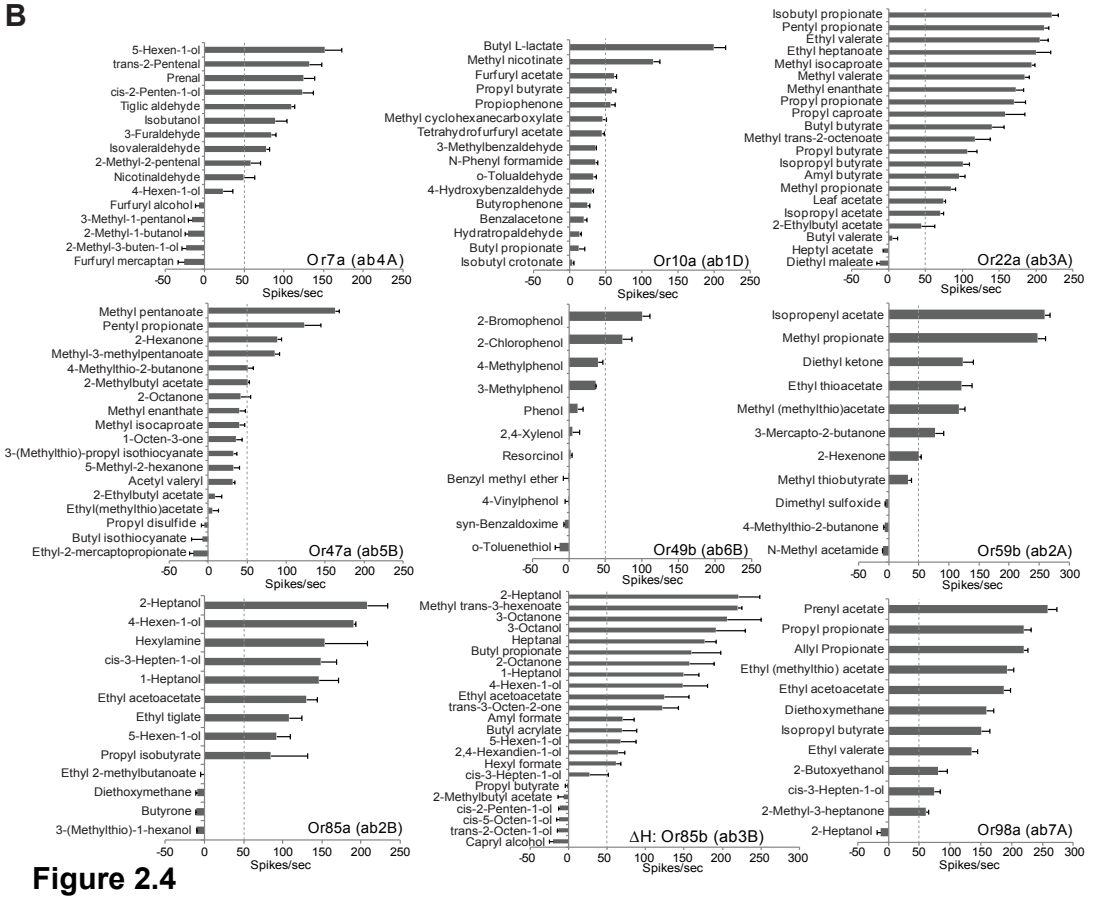


Figure 2.4

Figure 2.5: Predicted receptor-odor interactions are highly specific

(A) Plot of activity (Top) for electrophysiologically tested receptor-odor interactions. (Bottom) Plot indicating locations of predicted receptor-odor combinations (green) and same odorants tested in non-target receptor-odor combinations (gray). **(B)** Plot of percentage of activating odors (>50 spikes/sec) considering all activating or inactive odors (>0 spikes/sec) across ranking bins for all odors tested using electrophysiology.

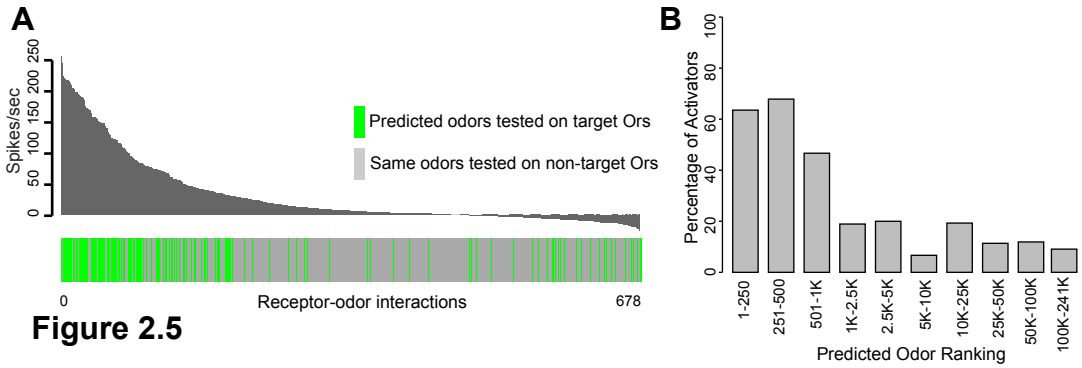

A
Spikes/sec
0
50
100
150
200
250
Predicted odors tested on target Ors
Same odors tested on non-target Ors
0
Receptor-odor interactions
678
B
Percentage of Activators
0
20
40
60
80
100
1-250
251-500
501-1K
1K-2.5K
2.5K-5K
5K-10K
10K-25K
25K-50K
50K-100K
100K-241K
Predicted Odor Ranking

Figure 2.5

Figure 2.6: High-ranking odors are more likely to be active than distantly ranked odors

Bar plot of the mean electrophysiological activity (in spikes/sec) of all activating or inactive odors (>0 spikes/sec), considering both predicted receptor-odor combinations and same odorants tested in non-target receptor-odor combinations, grouped by predicted rankings.

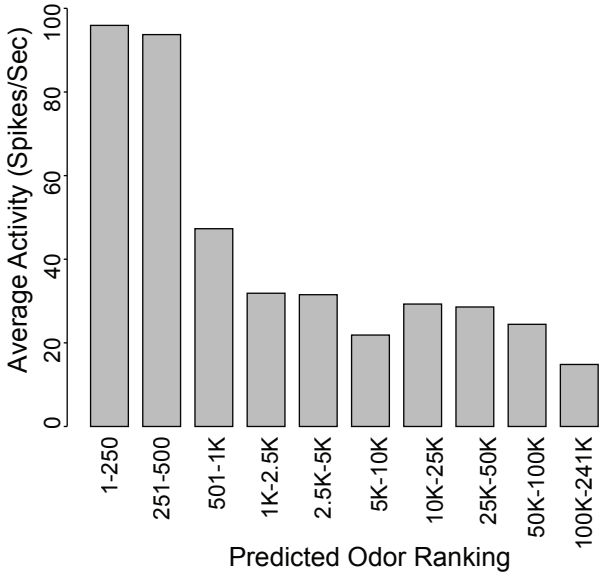

Average Activity (Spikes/Sec)
0
20
40
60
80
100
1-250
251-500
501-1K
1K-2.5K
2.5K-5K
5K-10K
10K-25K
25K-50K
50K-100K
100K-241K
Predicted Odor Ranking

Figure 2.6

Figure 2.7: Analysis of receptor-odor relationships and breadth of tuning

(A) Hierarchical clusters created from Euclidean distance values between *Drosophila* Ors calculated using: (left to right) shared optimized descriptors; known activity to training set odors (Hallem and Carlson, 2006); overlap across top 500 predicted ligands; and Phylogenic tree of receptors (Hallem and Carlson, 2006). Sub clusters shaded with colors or bars. **(B)** Frequency distribution of compounds from the >240K library within the top 15% distance from highest active plotted to generate predicted breadth of tuning curves. Green arrows indicate relative distance of the furthest known activating compound determined by electrophysiology. **(C)** (Left) The numbers of compounds present in the collected volatile library according to source. (Right) The numbers and sources of predicted ligands for the 19 *Drosophila* odor receptors/neurons within the top 500 predicted compounds **(D)** Comparison of plots for percentage of receptors that are: (top left) activated by percentage of known odors from training set (Hallem and Carlson, 2006); (bottom left) predicted to be activated by collected compounds; (top right) predicted to be activated from >240K library; and (bottom right) activated by ligands for 10 shared *Ors* in this study versus activated by comparable actives previously tested (Hallem and Carlson, 2006).

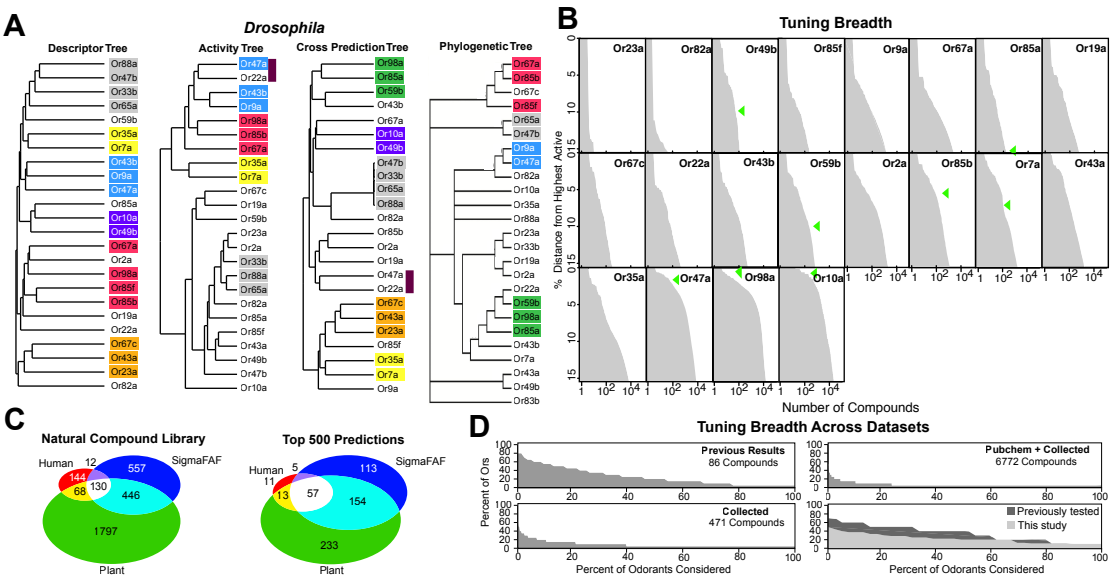

A
Drosophila
Descriptor Tree
Activity Tree
Cross Prediction Tree
Phylogenetic Tree
B
Tuning Breadth
% Distance from Highest Active
Number of Compounds
C
Natural Compound Library
Human
SigmaFAAF
Plant
Top 500 Predictions
D
Tuning Breadth Across Datasets
Previous Results
86 Compounds
Pubchem + Collected
6772 Compounds
Collected
471 Compounds
Previously tested
This study
Percent of Ors
Percent of Odorants Considered

Figure 2.7

Figure 2.8: Analysis of receptor-natural odor interactions

Compounds from the collected compound library that have been cataloged as plant, human and total collected volatiles were ranked according to their relative distance from the compound with highest activity. Frequency distribution of compounds within the top 15% is plotted to generate predicted breadth of tuning curves.

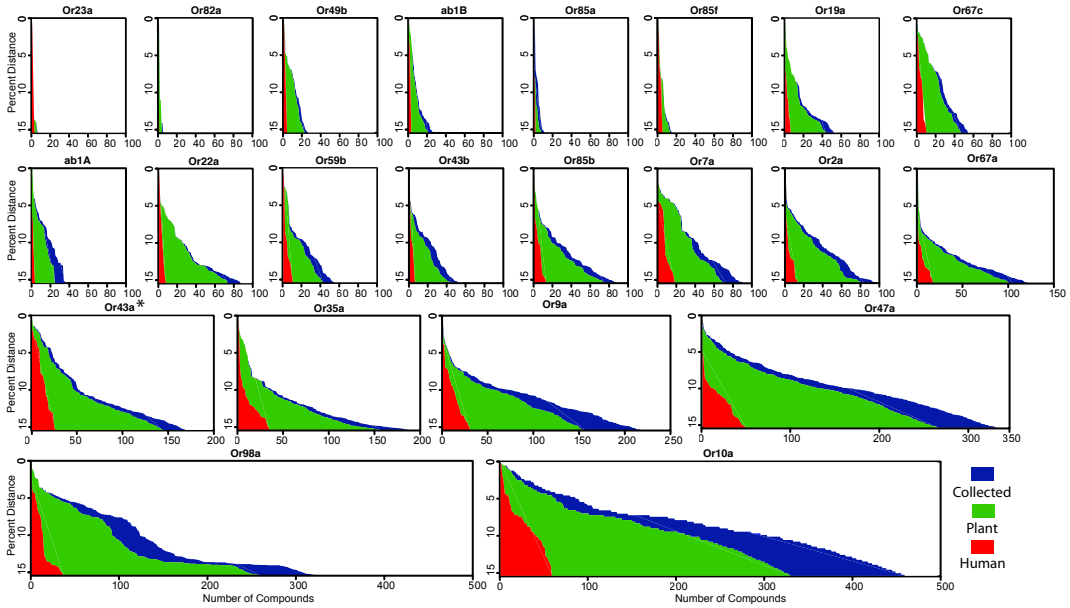

Or23a
Or82a
Or49b
ab1B
Or85a
Or85f
Or19a
Or67c
ab1A
Or22a
Or59b
Or43b
Or85b
Or7a
Or2a
Or67a
Or43a*
Or35a
Or9a
Or47a
Or98a
Or10a
Percent Distance
Number of Compounds
Collected
Plant
Human

Figure 2.8

Figure 2.9: Predicted odor space and network view of odor coding

(A) *Drosophila* receptor-odor network. Each known interaction (>50 spikes/sec) from this and previous studies(Hallem and Carlson, 2006) is linked by a purple edge. Predicted receptor-odor network (top 500 hits) are linked by light-grey edges. All compounds are represented as small black circles and Ors are represented as large colored circles matching the colors used in (Fig. 4A and SI). **(B)** Expansion of the peripheral olfactory code in this study: large increase in numbers of identified activators and inhibitors. The different sized circles represent the approximate ratio of numbers of previously known ligands (top circles), predicted ligands based on a cutoff of the top 500 predicted compounds per receptor and corrected to the validation success rate (lower, diffuse circles).

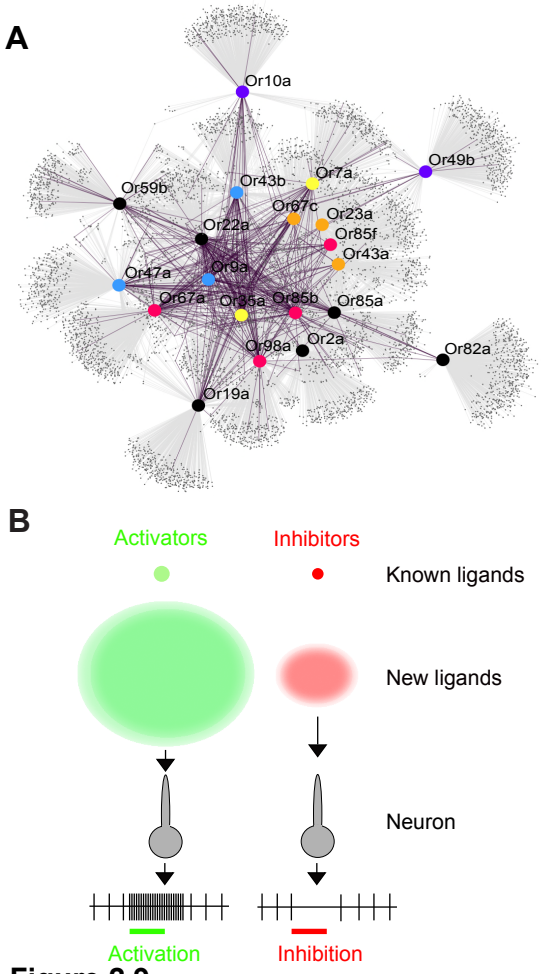

A
Or10a
Or49b
Or7a
Or43b
Or59b
Or67c
Or23a
Or22a
Or85f
Or43a
Or9a
Or47a
Or67a
Or85a
Or85b
Or85a
Or98a
Or2a
Or82a
Or19a
B
Activators
Inhibitors
Known ligands
New ligands
Neuron
Activation
Inhibition

Figure 2.9

Table 2.1: Optimized descriptor sets for each *Drosophila* Or

Optimized descriptors occurrences, symbol, brief description, class, and dimensionality are listed. A summary of the total number of descriptors selected for the receptor repertoire is provided at the beginning. Descriptors are listed in ascending order of when they were selected into the optimized set. Weights indicate the number of times a descriptor was selected in an optimized descriptor set.

Supplementary Table 1

Descriptor #	Weight	Symbol	Description	Class	Dimensionality
Or2a (18 Unique)					
	1	Mor18p	3D-MoRSE - signal 18 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	Mor17e	3D-MoRSE - signal 17 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	Mor28u	3D-MoRSE - signal 28 / unweighted	3D-MoRSE descriptors	3
	1	J3D	3D-Balaban index	geometrical descriptors	3
	2	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	SIC2	structural information content (neighborhood symmetry of 2-order)	information indices	2
	1	EEig10x	Eigenvalue 10 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	MATS5e	Moran autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2
	1	F05[C-O]	frequency of C - O at topological distance 05	2D frequency fingerprints	2
	1	HNar	Narumi harmonic topological index	topological descriptors	2
	1	MATS8m	Moran autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	G3s	3st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	Mor27m	3D-MoRSE - signal 27 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	B04[C-O]	presence/absence of C - O at topological distance 04	2D binary fingerprints	2
	1	H8v	H autocorrelation of lag 8 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	Mor10v	3D-MoRSE - signal 10 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	Mor18v	3D-MoRSE - signal 18 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	2	R8p+	R maximal autocorrelation of lag 8 / weighted by atomic polarizabilities	GETAWAY descriptors	3
Or7a (31 Unique)					
	1	MAXDP	maximal electrotopological positive variation	topological descriptors	2
	1	MAXDN	maximal electrotopological negative variation	topological descriptors	2
	1	B06[C-C]	presence/absence of C - C at topological distance 06	2D binary fingerprints	2
	2	HATS1v	leverage-weighted autocorrelation of lag 1 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	3	Hy	hydrophilic factor	molecular properties	1
	1	S_ssO	S_ssO	atomtypes (Cerius2)	1
	1	JGT	global topological charge index	topological charge indices	2
	2	H-051	H attached to alpha-C	atom-centred fragments	1
	2	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	5	HATS8u	leverage-weighted autocorrelation of lag 8 / unweighted	GETAWAY descriptors	3
	1	G2s	2st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	2	Mor16u	3D-MoRSE - signal 16 / unweighted	3D-MoRSE descriptors	3
	4	B02[O-O]	presence/absence of O - O at topological distance 02	2D binary fingerprints	2
	1	R5p+	R maximal autocorrelation of lag 5 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	EEig08d	Eigenvalue 08 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	DISPp	d COMMA2 value / weighted by atomic polarizabilities	geometrical descriptors	3
	2	C-008	CHR2X	atom-centred fragments	1
	1	R4e+	R maximal autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	nArOH	number of aromatic hydroxyls	functional group counts	1
	1	R2m+	R maximal autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	nRCOOR	number of esters (aliphatic)	functional group counts	1
	1	B02[C-O]	presence/absence of C - O at topological distance 02	2D binary fingerprints	2
	1	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	E2s	2nd component accessibility directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	nRCO	number of ketones (aliphatic)	functional group counts	1
	1	Mor03m	3D-MoRSE - signal 03 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	MATS8m	Moran autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	CIC5	complementary information content (neighborhood symmetry of 5-order)	information indices	2
	1	D/Dr06	distance/detour ring index of order 6	topological descriptors	2
Or9a (29 Unique)					
	1	BEHp8	highest eigenvalue n. 8 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	1	BELv1	lowest eigenvalue n. 1 of Burden matrix / weighted by atomic van der Waals volumes	Burden eigenvalues	2
	1	DISPe	d COMMA2 value / weighted by atomic Sanderson electronegativities	geometrical descriptors	3
	2	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	2	BEHp5	highest eigenvalue n. 5 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	1	E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	Mor25m	3D-MoRSE - signal 25 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	B03[C-C]	presence/absence of C - C at topological distance 03	2D binary fingerprints	2
	3	B07[C-C]	presence/absence of C - C at topological distance 07	2D binary fingerprints	2
	1	B01[C-O]	presence/absence of C - O at topological distance 01	2D binary fingerprints	2
	1	Atype_H_49	Number of Hydrogen Type 49	atomtypes (Cerius2)	1
	1	Infective-80	Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%	molecular properties	1
	3	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	Mor22m	3D-MoRSE - signal 22 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	R1u+	R maximal autocorrelation of lag 1 / unweighted	GETAWAY descriptors	3
	1	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	MATS4v	Moran autocorrelation - lag 4 / weighted by atomic van der Waals volumes	2D autocorrelations	2
	1	R4e+	R maximal autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	G3p	3st component symmetry directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
	1	Hy	hydrophilic factor	molecular properties	1
	1	S_dssC	S_dssC	atomtypes (Cerius2)	1
	1	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	1	B08[C-C]	presence/absence of C - C at topological distance 08	2D binary fingerprints	2
	1	R2m	R autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	HATS5e	leverage-weighted autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	D/Dr06	distance/detour ring index of order 6	topological descriptors	2
	1	RDF030m	Radial Distribution Function - 3.0 / weighted by atomic masses	RDF descriptors	3
	2	Jhetv	Balaban-type index from van der Waals weighted distance matrix	topological descriptors	2
Or10a (11 Unique)					
	3	S_dO	S_dO	atomtypes (Cerius2)	1
	1	BEHm7	highest eigenvalue n. 7 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	E2u	2nd component accessibility directional WHIM index / unweighted	WHIM descriptors	3
	1	HATS8m	leverage-weighted autocorrelation of lag 8 / weighted by atomic masses	GETAWAY descriptors	3
	1	BELe4	lowest eigenvalue n. 4 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	Mor25e	3D-MoRSE - signal 25 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	B08[C-C]	presence/absence of C - C at topological distance 08	2D binary fingerprints	2
	1	JGI3	mean topological charge index of order3	topological charge indices	2
	1	ESpm03u	Spectral moment 03 from edge adj. matrix	edge adjacency indices	2
	1	nR=Ct	number of aliphatic tertiary C(sp2)	functional group counts	1
	2	E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
Or19a (25 Unique)					
	1	Mor31p	3D-MoRSE - signal 31 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	H2m	H autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	L1m	1st component size directional WHIM index / weighted by atomic masses	WHIM descriptors	3
	1	R1m+	R maximal autocorrelation of lag 1 / weighted by atomic masses	GETAWAY descriptors	3
	1	Mor27u	3D-MoRSE - signal 27 / unweighted	3D-MoRSE descriptors	3
	1	HATS6u	leverage-weighted autocorrelation of lag 6 / unweighted	GETAWAY descriptors	3
	3	GGI7	topological charge index of order 7	topological charge indices	2
	1	Gs	G total symmetry index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	1	piPC08	molecular multiple path count of order 08	walk and path counts	2
	2	R7u+	R maximal autocorrelation of lag 7 / unweighted	GETAWAY descriptors	3
	2	G3s	3st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	R4m+	R maximal autocorrelation of lag 4 / weighted by atomic masses	GETAWAY descriptors	3
	1	MATS7p	Moran autocorrelation - lag 7 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	R6u+	R maximal autocorrelation of lag 6 / unweighted	GETAWAY descriptors	3

Table 2.1

	1	Hy	hydrophilic factor	molecular properties	1
	1	ARR	aromatic ratio	constitutional descriptors	0
	1	BEHp7	highest eigenvalue n. 7 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	1	RDF050v	Radial Distribution Function - 5.0 / weighted by atomic van der Waals volumes	RDF descriptors	3
	1	C-005	CH3X	atom-centred fragments	1
	1	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	R5m+	R maximal autocorrelation of lag 5 / weighted by atomic masses	GETAWAY descriptors	3
	2	C-002	CH2R2	atom-centred fragments	1
Or22a (43 Unique)					
	1	Mor29v	3D-MoRSE - signal 29 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	MAXDN	maximal electrotopological negative variation	topological descriptors	2
	1	piPC04	molecular multiple path count of order 04	walk and path counts	2
	1	Mor10e	3D-MoRSE - signal 10 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	Mor27m	3D-MoRSE - signal 27 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	R7p+	R maximal autocorrelation of lag 7 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	S_sCH3	S_sCH3	atomtypes (Cerius2)	1
	2	EEig12r	Eigenvalue 12 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	nRCOOR	number of esters (aliphatic)	functional group counts	1
	4	R6u+	R maximal autocorrelation of lag 6 / unweighted	GETAWAY descriptors	3
	1	Mor32p	3D-MoRSE - signal 32 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	AlogP98	AlogP98 value	structural (Cerius2)	0
	4	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	L3s	3rd component size directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	R1v+	R maximal autocorrelation of lag 1 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	2	nHDon	number of donor atoms for H-bonds (N and O)	functional group counts	1
	2	B10[C-C]	presence/absence of C - C at topological distance 10	2D binary fingerprints	2
	1	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	B04[C-O]	presence/absence of C - O at topological distance 04	2D binary fingerprints	2
	2	Jhetp	Balaban-type index from polarizability weighted distance matrix	topological descriptors	2
	1	STN	spanning tree number (log)	topological descriptors	2
	2	ESpm15u	Spectral moment 15 from edge adj. matrix	edge adjacency indices	2
	1	GATS1v	Geary autocorrelation - lag 1 / weighted by atomic van der Waals volumes	2D autocorrelations	2
	1	F03[O-O]	frequency of O - O at topological distance 03	2D frequency fingerprints	2
	1	GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	2	HATS5e	leverage-weighted autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes	geometrical descriptors	3
	1	R3v+	R maximal autocorrelation of lag 3 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	Mor32u	3D-MoRSE - signal 32 / unweighted	3D-MoRSE descriptors	3
	2	B02[O-O]	presence/absence of O - O at topological distance 02	2D binary fingerprints	2
	1	G3e	3st component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	nCrs	number of ring secondary C(sp3)	functional group counts	1
	2	HOMT	HOMA total	geometrical descriptors	3
	1	B05[C-C]	presence/absence of C - C at topological distance 05	2D binary fingerprints	2
	1	MATS7m	Moran autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	RDF030m	Radial Distribution Function - 3.0 / weighted by atomic masses	RDF descriptors	3
	1	EEig12x	Eigenvalue 12 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	R1m+	R maximal autocorrelation of lag 1 / weighted by atomic masses	GETAWAY descriptors	3
	1	MATS4p	Moran autocorrelation - lag 4 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	B09[C-O]	presence/absence of C - O at topological distance 09	2D binary fingerprints	2
	1	Mor15p	3D-MoRSE - signal 15 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	2	S_sOH	S_sOH	atomtypes (Cerius2)	1
Or23a (37 Unique)					
	1	ATS3p	Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic polarizabilities	2D autocorrelations	2
	2	O-056	alcohol	atom-centred fragments	1
	1	J3D	3D-Balaban index	geometrical descriptors	3
	1	BELm5	lowest eigenvalue n. 5 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	TPSA(Tot)	topological polar surface area using N,O,S,P polar contributions	molecular properties	1
	1	B08[C-O]	presence/absence of C - O at topological distance 08	2D binary fingerprints	2
	2	Mor27v	3D-MoRSE - signal 27 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	2	R6u+	R maximal autocorrelation of lag 6 / unweighted	GETAWAY descriptors	3
	1	DISPe	d COMMA2 value / weighted by atomic Sanderson electronegativities	geometrical descriptors	3
	1	ESpm12d	Spectral moment 12 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	Mor17m	3D-MoRSE - signal 17 / weighted by atomic masses	3D-MoRSE descriptors	3
	2	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	Hy	hydrophilic factor	molecular properties	1
	2	GATS3e	Geary autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2
	1	GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	R4e+	R maximal autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses	3D-MoRSE descriptors	3
	2	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	S_sOH	S_sOH	atomtypes (Cerius2)	1
	1	E3m	3rd component accessibility directional WHIM index / weighted by atomic masses	WHIM descriptors	3
	1	G3s	3st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	2	BELm6	lowest eigenvalue n. 6 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	GATS1m	Geary autocorrelation - lag 1 / weighted by atomic masses	2D autocorrelations	2
	2	EEig08d	Eigenvalue 08 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	F05[C-O]	frequency of C - O at topological distance 05	2D frequency fingerprints	2
	2	nHDon	number of donor atoms for H-bonds (N and O)	functional group counts	1
	1	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	R5p+	R maximal autocorrelation of lag 5 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	BIC	BIC	topological (Cerius2)	2
	2	Infective-80	Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%	molecular properties	1
	1	GATS4p	Geary autocorrelation - lag 4 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	DISPp	d COMMA2 value / weighted by atomic polarizabilities	geometrical descriptors	3
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	Atype_H_49	Number of Hydrogen Type 49	atomtypes (Cerius2)	1
	1	GATS5m	Geary autocorrelation - lag 5 / weighted by atomic masses	2D autocorrelations	2
	1	B02[O-O]	presence/absence of O - O at topological distance 02	2D binary fingerprints	2
	2	JGI5	mean topological charge index of order5	topological charge indices	2
Or33b (32 Unique)					
	6	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	2	EEig08x	Eigenvalue 08 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes	geometrical descriptors	3
	1	TPSA(NO)	topological polar surface area using N,O polar contributions	molecular properties	1
	5	B06[C-C]	presence/absence of C - C at topological distance 06	2D binary fingerprints	2
	4	Atype_H_49	Number of Hydrogen Type 49	atomtypes (Cerius2)	1
	2	R3v+	R maximal autocorrelation of lag 3 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	G1e	1st component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	R2m+	R maximal autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	4	B05[C-O]	presence/absence of C - O at topological distance 05	2D binary fingerprints	2
	1	C-006	CH2RX	atom-centred fragments	1
	2	TPSA(Tot)	topological polar surface area using N,O,S,P polar contributions	molecular properties	1
	1	L/Bw	length-to-breadth ratio by WHIM	geometrical descriptors	3
	1	EEig08d	Eigenvalue 08 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	3	F04[C-O]	frequency of C - O at topological distance 04	2D frequency fingerprints	2
	1	BEHv5	highest eigenvalue n. 5 of Burden matrix / weighted by atomic van der Waals volumes	Burden eigenvalues	2
	1	Mor30p	3D-MoRSE - signal 30 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	nArCO	number of ketones (aromatic)	functional group counts	1
	1	nRCO	number of ketones (aliphatic)	functional group counts	1
	1	R1p+	R maximal autocorrelation of lag 1 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	MATS4p	Moran autocorrelation - lag 4 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	nN	number of Nitrogen atoms	constitutional descriptors	0
	1	B07[C-C]	presence/absence of C - C at topological distance 07	2D binary fingerprints	2

Table 2.1 Continued

	2	JGI4	mean topological charge index of order4	topological charge indices	2
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	nCconj	number of non-aromatic conjugated C(sp2)	functional group counts	1
	1	C-005	CH3X	atom-centred fragments	1
	1	JGI3	mean topological charge index of order3	topological charge indices	2
	1	HATS3p	leverage-weighted autocorrelation of lag 3 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	HATS8u	leverage-weighted autocorrelation of lag 8 / unweighted	GETAWAY descriptors	3
	1	E2u	2nd component accessibility directional WHIM index / unweighted	WHIM descriptors	3
	2	H-051	H attached to alpha-C	atom-centred fragments	1
Or35a (51 Unique)					
	1	ATS4e	Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic Sanderson electroneg	2D autocorrelations	2
	2	TPSA(NO)	topological polar surface area using N,O polar contributions	molecular properties	1
	1	Mor27p	3D-MoRSE - signal 27 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	8	R6p+	R maximal autocorrelation of lag 6 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	6	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	3	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	2	Gs	G total symmetry index / weighted by atomic electrotopological states	WHIM descriptors	3
	9	JGI2	mean topological charge index of order2	topological charge indices	2
	3	EEig12r	Eigenvalue 12 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	7	R4e+	R maximal autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	7	Mor28e	3D-MoRSE - signal 28 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	5	MATS7p	Moran autocorrelation - lag 7 / weighted by atomic polarizabilities	2D autocorrelations	2
	2	L3s	3rd component size directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	6	Mor25v	3D-MoRSE - signal 25 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	4	Mor30e	3D-MoRSE - signal 30 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	5	HATS8u	leverage-weighted autocorrelation of lag 8 / unweighted	GETAWAY descriptors	3
	7	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	3	HATS5m	leverage-weighted autocorrelation of lag 5 / weighted by atomic masses	GETAWAY descriptors	3
	3	Jhetp	Balaban-type index from polarizability weighted distance matrix	topological descriptors	2
	4	JGI8	mean topological charge index of order8	topological charge indices	2
	3	Mor04m	3D-MoRSE - signal 04 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	S_dssC	S_dssC	atomtypes (Cerius2)	1
	2	E1m	1st component accessibility directional WHIM index / weighted by atomic masses	WHIM descriptors	3
	2	nHDon	number of donor atoms for H-bonds (N and O)	functional group counts	1
	2	RDF135u	Radial Distribution Function - 13.5 / unweighted	RDF descriptors	3
	2	D/Dr06	distance/detour ring index of order 6	topological descriptors	2
	3	E2s	2nd component accessibility directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	2	EEig10r	Eigenvalue 10 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	G2s	2st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	3	GATS3p	Geary autocorrelation - lag 3 / weighted by atomic polarizabilities	2D autocorrelations	2
	2	GGI1	topological charge index of order 1	topological charge indices	2
	2	Atype_C_18	Number of Carbon Type 18	atomtypes (Cerius2)	1
	1	nRCO	number of ketones (aliphatic)	functional group counts	1
	1	C-005	CH3X	atom-centred fragments	1
	1	Mor27u	3D-MoRSE - signal 27 / unweighted	3D-MoRSE descriptors	3
	2	F08[C-O]	frequency of C - O at topological distance 08	2D frequency fingerprints	2
	3	G3s	3st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	3	SIC5	structural information content (neighborhood symmetry of 5-order)	information indices	2
	1	G(N..N)	sum of geometrical distances between N..N	geometrical descriptors	3
	2	nR=Ct	number of aliphatic tertiary C(sp2)	functional group counts	1
	2	E3m	3rd component accessibility directional WHIM index / weighted by atomic masses	WHIM descriptors	3
	1	nArCOOR	number of esters (aromatic)	functional group counts	1
	1	HATS6m	leverage-weighted autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
	1	nArCO	number of ketones (aromatic)	functional group counts	1
	1	Jhete	Balaban-type index from electronegativity weighted distance matrix	topological descriptors	2
	1	G(O..O)	sum of geometrical distances between O..O	geometrical descriptors	3
	1	nCt	number of total tertiary C(sp3)	functional group counts	1
	1	H-051	H attached to alpha-C	atom-centred fragments	1
	1	nN	number of Nitrogen atoms	constitutional descriptors	0
	1	P2s	2nd component shape directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	C-025	R--CR--R	atom-centred fragments	1
Or43a (27 Unique)					
	2	O-056	alcohol	atom-centred fragments	1
	1	BELm5	lowest eigenvalue n. 5 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	B07[C-O]	presence/absence of C - O at topological distance 07	2D binary fingerprints	2
	1	R5e	R autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	TPSA(Tot)	topological polar surface area using N,O,S,P polar contributions	molecular properties	1
	1	R6e+	R maximal autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	2	JGI7	mean topological charge index of order7	topological charge indices	2
	3	B04[C-C]	presence/absence of C - C at topological distance 04	2D binary fingerprints	2
	1	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	5	B02[O-O]	presence/absence of O - O at topological distance 02	2D binary fingerprints	2
	3	Mor13m	3D-MoRSE - signal 13 / weighted by atomic masses	3D-MoRSE descriptors	3
	3	nHDon	number of donor atoms for H-bonds (N and O)	functional group counts	1
	1	Mor21m	3D-MoRSE - signal 21 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	JX	JX	topological (Cerius2)	2
	1	R1m+	R maximal autocorrelation of lag 1 / weighted by atomic masses	GETAWAY descriptors	3
	2	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	BELm6	lowest eigenvalue n. 6 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	E3m	3rd component accessibility directional WHIM index / weighted by atomic masses	WHIM descriptors	3
	2	MATS3e	Moran autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2
	1	F04[C-O]	frequency of C - O at topological distance 04	2D frequency fingerprints	2
	1	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	1	Infective-80	Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%	molecular properties	1
	1	EEig09x	Eigenvalue 09 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	GATS1m	Geary autocorrelation - lag 1 / weighted by atomic masses	2D autocorrelations	2
	1	CIC2	complementary information content (neighborhood symmetry of 2-order)	information indices	2
	1	EEig01d	Eigenvalue 01 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	HATS6u	leverage-weighted autocorrelation of lag 6 / unweighted	GETAWAY descriptors	3
Or43b (29 Unique)					
	1	EEig04x	Eigenvalue 04 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	BEHv4	highest eigenvalue n. 4 of Burden matrix / weighted by atomic van der Waals volumes	Burden eigenvalues	2
	1	Mor25e	3D-MoRSE - signal 25 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	2	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	E1p	1st component accessibility directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
	1	BEHe8	highest eigenvalue n. 8 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	R1m+	R maximal autocorrelation of lag 1 / weighted by atomic masses	GETAWAY descriptors	3
	2	B07[C-C]	presence/absence of C - C at topological distance 07	2D binary fingerprints	2
	1	MAXDN	maximal electrotopological negative variation	topological descriptors	2
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	Infective-80	Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%	molecular properties	1
	3	B04[C-C]	presence/absence of C - C at topological distance 04	2D binary fingerprints	2
	1	MATS5e	Moran autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2
	1	Mor24v	3D-MoRSE - signal 24 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	Mor25v	3D-MoRSE - signal 25 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	BEHp4	highest eigenvalue n. 4 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	1	S_sCH3	S_sCH3	atomtypes (Cerius2)	1
	1	HATS3p	leverage-weighted autocorrelation of lag 3 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	H7m	H autocorrelation of lag 7 / weighted by atomic masses	GETAWAY descriptors	3
	1	JGI7	mean topological charge index of order7	topological charge indices	2
	1	STN	spanning tree number (log)	topological descriptors	2
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	MATS6m	Moran autocorrelation - lag 6 / weighted by atomic masses	2D autocorrelations	2
	1	HATS1u	leverage-weighted autocorrelation of lag 1 / unweighted	GETAWAY descriptors	3
	1	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2

Table 2.1 Continued

	1	Atype_H_49	Number of Hydrogen Type 49	atomtypes (Cerius2)	1
	1	EEig08d	Eigenvalue 08 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	nCrs	number of ring secondary C(sp3)	functional group counts	1
	2	H-047	H attached to C1(sp3)/C0(sp2)	atom-centred fragments	1
Or47a (21 Unique)					
	1	piPC04	molecular multiple path count of order 04	walk and path counts	2
	2	DISPm	d COMMA2 value / weighted by atomic masses	geometrical descriptors	3
	1	R7e+	R maximal autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	Mor10p	3D-MoRSE - signal 10 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	Mor20u	3D-MoRSE - signal 20 / unweighted	3D-MoRSE descriptors	3
	1	IC1	information content index (neighborhood symmetry of 1-order)	information indices	2
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	EEig01d	Eigenvalue 01 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	2	Infective-80	Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%	molecular properties	1
	1	MATS4m	Moran autocorrelation - lag 4 / weighted by atomic masses	2D autocorrelations	2
	1	GATS5p	Geary autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	PW4	path/walk 4 - Randic shape index	topological descriptors	2
	1	Mor32p	3D-MoRSE - signal 32 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	Mor09e	3D-MoRSE - signal 09 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	TPSA(NO)	topological polar surface area using N,O polar contributions	molecular properties	1
	1	B04[C-C]	presence/absence of C - C at topological distance 04	2D binary fingerprints	2
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	Atype_H_49	Number of Hydrogen Type 49	atomtypes (Cerius2)	1
	1	ESpm01d	Spectral moment 01 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	PZm	2nd component shape directional WHIM index / weighted by atomic masses	WHIM descriptors	3
	2	Mor06e	3D-MoRSE - signal 06 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
Or47b (14 Unique)					
	3	EEig02d	Eigenvalue 02 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	5	ESpm03d	Spectral moment 03 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	nHBonds	number of intramolecular H-bonds (with N,O,F)	functional group counts	1
	4	X5A	average connectivity index chi-5	connectivity indices	2
	1	EEig08x	Eigenvalue 08 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	C-006	CH2RX	atom-centred fragments	1
	1	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	2	nRCOOR	number of esters (aliphatic)	functional group counts	1
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	EEig08d	Eigenvalue 08 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	X4Av	average valence connectivity index chi-4	connectivity indices	2
	1	GATS6m	Geary autocorrelation - lag 6 / weighted by atomic masses	2D autocorrelations	2
	1	EEig07r	Eigenvalue 07 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	R2m	R autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
Or49b (37 Unique)					
	2	nCb-	number of substituted benzene C(sp2)	functional group counts	1
	1	BEHm6	highest eigenvalue n. 6 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	2	F04[C-O]	frequency of C - O at topological distance 04	2D frequency fingerprints	2
	1	D/Dr06	distance/detour ring index of order 6	topological descriptors	2
	1	BEHp6	highest eigenvalue n. 6 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	3	H-047	H attached to C1(sp3)/C0(sp2)	atom-centred fragments	1
	1	GATS1m	Geary autocorrelation - lag 1 / weighted by atomic masses	2D autocorrelations	2
	3	HATS8p	leverage-weighted autocorrelation of lag 8 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	2	ISH	standardized information content on the leverage equality	GETAWAY descriptors	3
	1	Mor16e	3D-MoRSE - signal 16 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	JGI5	mean topological charge index of order5	topological charge indices	2
	1	R8e+	R maximal autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	Mor25e	3D-MoRSE - signal 25 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	2	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	Mor16p	3D-MoRSE - signal 16 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	JGI4	mean topological charge index of order4	topological charge indices	2
	1	MATS3p	Moran autocorrelation - lag 3 / weighted by atomic polarizabilities	2D autocorrelations	2
	3	CIC	CIC	topological (Cerius2)	2
	1	PZm	2nd component shape directional WHIM index / weighted by atomic masses	WHIM descriptors	3
	1	nHDon	number of donor atoms for H-bonds (N and O)	functional group counts	1
	1	Mor03m	3D-MoRSE - signal 03 / weighted by atomic masses	3D-MoRSE descriptors	3
	2	JGI7	mean topological charge index of order7	topological charge indices	2
	1	Mor23v	3D-MoRSE - signal 23 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	Mor30e	3D-MoRSE - signal 30 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	IC	IC	topological (Cerius2)	2
	1	Mor21m	3D-MoRSE - signal 21 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	Mor13m	3D-MoRSE - signal 13 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	R7v+	R maximal autocorrelation of lag 7 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	piPC07	molecular multiple path count of order 07	walk and path counts	2
	1	nArOH	number of aromatic hydroxyls	functional group counts	1
	1	Mor25v	3D-MoRSE - signal 25 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	Mor08v	3D-MoRSE - signal 08 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	R6e+	R maximal autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	EEig06x	Eigenvalue 06 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	C-001	CH3R / CH4	atom-centred fragments	1
	1	Mor07m	3D-MoRSE - signal 07 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	DISPe	d COMMA2 value / weighted by atomic Sanderson electronegativities	geometrical descriptors	3
	1	nR05	number of 5-membered rings	constitutional descriptors	0
	1	Mor07e	3D-MoRSE - signal 07 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	EEig09x	Eigenvalue 09 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	B05[C-O]	presence/absence of C - O at topological distance 05	2D binary fingerprints	2
	1	X5Av	average valence connectivity index chi-5	connectivity indices	2
	1	HATS3p	leverage-weighted autocorrelation of lag 3 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	R8u+	R maximal autocorrelation of lag 8 / unweighted	GETAWAY descriptors	3
	1	O-060	Al-O-Ar / Ar-O-Ar / R..O..R / R-O-C=X	atom-centred fragments	1
	2	B04[C-O]	presence/absence of C - O at topological distance 04	2D binary fingerprints	2
Or59b (23 Unique)					
	1	piPC06	molecular multiple path count of order 06	walk and path counts	2
	1	R3u	R autocorrelation of lag 3 / unweighted	GETAWAY descriptors	3
	1	S_sCH3	S_sCH3	atomtypes (Cerius2)	1
	4	B06[C-C]	presence/absence of C - C at topological distance 06	2D binary fingerprints	2
	1	R1e+	R maximal autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	ESpm03u	Spectral moment 03 from edge adj. matrix	edge adjacency indices	2
	1	EEig10r	Eigenvalue 10 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	EEig08d	Eigenvalue 08 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	E1u	1st component accessibility directional WHIM index / unweighted	WHIM descriptors	3
	1	nCconj	number of non-aromatic conjugated C(sp2)	functional group counts	1
	1	SP13	shape profile no. 13	Randic molecular profiles	3
	2	S_dO	S_dO	atomtypes (Cerius2)	1
	2	Atype_H_49	Number of Hydrogen Type 49	atomtypes (Cerius2)	1
	1	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	nHDon	number of donor atoms for H-bonds (N and O)	functional group counts	1
	1	R8u+	R maximal autocorrelation of lag 8 / unweighted	GETAWAY descriptors	3
	2	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	Mor10v	3D-MoRSE - signal 10 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	R5m+	R maximal autocorrelation of lag 5 / weighted by atomic masses	GETAWAY descriptors	3
	1	Mor09e	3D-MoRSE - signal 09 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	nOHp	number of primary alcohols	functional group counts	1
	1	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	nCrs	number of ring secondary C(sp3)	functional group counts	1
	1	ESpm01d	Spectral moment 01 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2

Table 2.1 Continued

Or65a (14 Unique)					
	1	F04[O-O]	frequency of O - O at topological distance 04	2D frequency fingerprints	2
	2	Mor30m	3D-MoRSE - signal 30 / weighted by atomic masses	3D-MoRSE descriptors	3
	4	Atype_H_51	Number of Hydrogen Type 51	atomtypes (Cerius2)	1
	1	EEig08d	Eigenvalue 08 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	2	nArOH	number of aromatic hydroxyls	functional group counts	1
	2	JGI7	mean topological charge index of order7	topological charge indices	2
	1	nHBonds	number of intramolecular H-bonds (with N,O,F)	functional group counts	1
	1	Mor13p	3D-MoRSE - signal 13 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	EEig07d	Eigenvalue 07 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	B06[C-O]	presence/absence of C - O at topological distance 06	2D binary fingerprints	2
	1	C-008	CHR2X	atom-centred fragments	1
	1	EEig08r	Eigenvalue 08 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	B01[C-O]	presence/absence of C - O at topological distance 01	2D binary fingerprints	2
	2	Mor32e	3D-MoRSE - signal 32 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
Or67a (37 Unique)					
	2	AlogP98	AlogP98 value	structural (Cerius2)	0
	8	B04[C-O]	presence/absence of C - O at topological distance 04	2D binary fingerprints	2
	6	F08[C-O]	frequency of C - O at topological distance 08	2D frequency fingerprints	2
	1	GGI4	topological charge index of order 4	topological charge indices	2
	3	E2u	2nd component accessibility directional WHIM index / unweighted	WHIM descriptors	3
	2	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	Mor03v	3D-MoRSE - signal 03 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	4	X5A	average connectivity index chi-5	connectivity indices	2
	3	Mor10v	3D-MoRSE - signal 10 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	B03[C-O]	presence/absence of C - O at topological distance 03	2D binary fingerprints	2
	3	X4A	average connectivity index chi-4	connectivity indices	2
	3	nCt	number of total tertiary C(sp3)	functional group counts	1
	1	C-026	R--CX--R	atom-centred fragments	1
	3	RDF075m	Radial Distribution Function - 7.5 / weighted by atomic masses	RDF descriptors	3
	2	C-008	CHR2X	atom-centred fragments	1
	2	B03[C-C]	presence/absence of C - C at topological distance 03	2D binary fingerprints	2
	1	B01[C-O]	presence/absence of C - O at topological distance 01	2D binary fingerprints	2
	1	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	1	Jhetv	Balaban-type index from van der Waals weighted distance matrix	topological descriptors	2
	1	L1s	1st component size directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	Hy	hydrophilic factor	molecular properties	1
	2	C-003	CHR3	atom-centred fragments	1
	1	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	Mor16e	3D-MoRSE - signal 16 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	Mor06u	3D-MoRSE - signal 06 / unweighted	3D-MoRSE descriptors	3
	1	RDF030m	Radial Distribution Function - 3.0 / weighted by atomic masses	RDF descriptors	3
	1	Atype_C_18	Number of Carbon Type 18	atomtypes (Cerius2)	1
	1	F03[O-O]	frequency of O - O at topological distance 03	2D frequency fingerprints	2
	1	nCrs	number of ring secondary C(sp3)	functional group counts	1
	2	nArOH	number of aromatic hydroxyls	functional group counts	1
	1	GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	Jhete	Balaban-type index from electronegativity weighted distance matrix	topological descriptors	2
	1	EEig13x	Eigenvalue 13 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	DISPm	d COMMA2 value / weighted by atomic masses	geometrical descriptors	3
	1	X3A	average connectivity index chi-3	connectivity indices	2
	1	G(N..N)	sum of geometrical distances between N..N	geometrical descriptors	3
	1	Mor32u	3D-MoRSE - signal 32 / unweighted	3D-MoRSE descriptors	3
Or67c (24 Unique)					
	1	BEHe8	highest eigenvalue n. 8 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	O-056	alcohol	atom-centred fragments	1
	1	Mor25m	3D-MoRSE - signal 25 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	BELv4	lowest eigenvalue n. 4 of Burden matrix / weighted by atomic van der Waals volumes	Burden eigenvalues	2
	3	B07[C-C]	presence/absence of C - C at topological distance 07	2D binary fingerprints	2
	1	TPSA(Tot)	topological polar surface area using N,O,S,P polar contributions	molecular properties	1
	1	DISPm	d COMMA2 value / weighted by atomic masses	geometrical descriptors	3
	4	HATS6u	leverage-weighted autocorrelation of lag 6 / unweighted	GETAWAY descriptors	3
	2	EEig08d	Eigenvalue 08 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	2	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	Gs	G total symmetry index / weighted by atomic electrotopological states	WHIM descriptors	3
	3	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	B08[C-C]	presence/absence of C - C at topological distance 08	2D binary fingerprints	2
	1	R1m+	R maximal autocorrelation of lag 1 / weighted by atomic masses	GETAWAY descriptors	3
	1	BELm5	lowest eigenvalue n. 5 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	F03[O-O]	frequency of O - O at topological distance 03	2D frequency fingerprints	2
	1	STN	spanning tree number (log)	topological descriptors	2
	1	Atype_H_49	Number of Hydrogen Type 49	atomtypes (Cerius2)	1
	1	H-051	H attached to alpha-C	atom-centred fragments	1
	1	B01[C-O]	presence/absence of C - O at topological distance 01	2D binary fingerprints	2
	1	Infective-80	Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%	molecular properties	1
	1	Hy	hydrophilic factor	molecular properties	1
	1	Mor22m	3D-MoRSE - signal 22 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	JGI7	mean topological charge index of order7	topological charge indices	2
Or82a (31 Unique)					
	1	GGI9	topological charge index of order 9	topological charge indices	2
	1	Mor02e	3D-MoRSE - signal 02 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	Mor30v	3D-MoRSE - signal 30 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	Mor02v	3D-MoRSE - signal 02 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	Mor30u	3D-MoRSE - signal 30 / unweighted	3D-MoRSE descriptors	3
	2	BLTD48	Verhaar model of Daphnia base-line toxicity from MLOGP (mmol/l)	molecular properties	1
	2	Mor10v	3D-MoRSE - signal 10 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	2	Atype_H_53	Number of Hydrogen Type 53	atomtypes (Cerius2)	1
	1	O-058	=O	atom-centred fragments	1
	1	B02[C-O]	presence/absence of C - O at topological distance 02	2D binary fingerprints	2
	2	R5u+	R maximal autocorrelation of lag 5 / unweighted	GETAWAY descriptors	3
	1	H6e	H autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	MATS7p	Moran autocorrelation - lag 7 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	GATS3p	Geary autocorrelation - lag 3 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	H-051	H attached to alpha-C	atom-centred fragments	1
	2	Mor13p	3D-MoRSE - signal 13 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	SIC2	structural information content (neighborhood symmetry of 2-order)	information indices	2
	1	Mor32u	3D-MoRSE - signal 32 / unweighted	3D-MoRSE descriptors	3
	1	Mor10m	3D-MoRSE - signal 10 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	nR=Cp	number of terminal primary C(sp2)	functional group counts	1
	1	Mor25p	3D-MoRSE - signal 25 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	JGI1	mean topological charge index of order1	topological charge indices	2
	1	E-ADJ-mag	E-ADJ-mag	topological (cerius2)	2
	1	EEig11x	Eigenvalue 11 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	B03[O-O]	presence/absence of O - O at topological distance 03	2D binary fingerprints	2
	1	Mor30e	3D-MoRSE - signal 30 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	Rotlbonds	Number of rotatable bonds	structural (Cerius2)	0
	1	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	2	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
Or85a (15 Unique)					
	1	EEig04r	Eigenvalue 04 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	2	C-006	CH2RX	atom-centred fragments	1

Table 2.1 Continued

	3	ATS6e	Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic Sanderson electroneg	2D autocorrelations	2
	3	JGI5	mean topological charge index of order5	topological charge indices	2
	2	B07[C-C]	presence/absence of C - C at topological distance 07	2D binary fingerprints	2
	1	nCp	number of terminal primary C(sp3)	functional group counts	1
	2	DISPm	d COMMA2 value / weighted by atomic masses	geometrical descriptors	3
	2	GATS4m	Geary autocorrelation - lag 4 / weighted by atomic masses	2D autocorrelations	2
	1	Mor25p	3D-MoRSE - signal 25 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	nHDon	number of donor atoms for H-bonds (N and O)	functional group counts	1
	1	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	R2m+	R maximal autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	JGI4	mean topological charge index of order4	topological charge indices	2
	1	Mor11e	3D-MoRSE - signal 11 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	2	HATS7m	leverage-weighted autocorrelation of lag 7 / weighted by atomic masses	GETAWAY descriptors	3
Or85b (26 Unique)					
	1	piPC05	molecular multiple path count of order 05	walk and path counts	2
	1	BLTF96	Verhaar model of Fish base-line toxicity from MLOGP (mmol/l)	molecular properties	1
	2	GATS4p	Geary autocorrelation - lag 4 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	GGI7	topological charge index of order 7	topological charge indices	2
	3	B05[C-O]	presence/absence of C - O at topological distance 05	2D binary fingerprints	2
	2	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	Mor27v	3D-MoRSE - signal 27 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	HATS4v	leverage-weighted autocorrelation of lag 4 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	Gs	G total symmetry index / weighted by atomic electrotopological states	WHIM descriptors	3
	2	Infective-80	Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%	molecular properties	1
	2	R7u+	R maximal autocorrelation of lag 7 / unweighted	GETAWAY descriptors	3
	2	nCbH	number of unsubstituted benzene C(sp2)	functional group counts	1
	1	B04[C-O]	presence/absence of C - O at topological distance 04	2D binary fingerprints	2
	2	JGI7	mean topological charge index of order7	topological charge indices	2
	2	DISPe	d COMMA2 value / weighted by atomic Sanderson electronegativities	geometrical descriptors	3
	1	R4p+	R maximal autocorrelation of lag 4 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	EEig12x	Eigenvalue 12 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	B06[C-O]	presence/absence of C - O at topological distance 06	2D binary fingerprints	2
	1	MATS5e	Moran autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2
	1	HATS4m	leverage-weighted autocorrelation of lag 4 / weighted by atomic masses	GETAWAY descriptors	3
	1	HATS6u	leverage-weighted autocorrelation of lag 6 / unweighted	GETAWAY descriptors	3
	1	GATS4m	Geary autocorrelation - lag 4 / weighted by atomic masses	2D autocorrelations	2
	1	F03[O-O]	frequency of O - O at topological distance 03	2D frequency fingerprints	2
	1	H8v	H autocorrelation of lag 8 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	2	Mor16e	3D-MoRSE - signal 16 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
Or85f (53 Unique)					
	1	BEHp8	highest eigenvalue n. 8 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	5	F05[C-O]	frequency of C - O at topological distance 05	2D frequency fingerprints	2
	4	BELm4	lowest eigenvalue n. 4 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	HATS8m	leverage-weighted autocorrelation of lag 8 / weighted by atomic masses	GETAWAY descriptors	3
	2	B04[C-O]	presence/absence of C - O at topological distance 04	2D binary fingerprints	2
	6	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	RDF030v	Radial Distribution Function - 3.0 / weighted by atomic van der Waals volumes	RDF descriptors	3
	1	GGI7	topological charge index of order 7	topological charge indices	2
	1	Gs	G total symmetry index / weighted by atomic electrotopological states	WHIM descriptors	3
	4	B07[C-C]	presence/absence of C - C at topological distance 07	2D binary fingerprints	2
	1	E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	MATS2m	Moran autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations	2
	2	Mor28u	3D-MoRSE - signal 28 / unweighted	3D-MoRSE descriptors	3
	3	BEHp5	highest eigenvalue n. 5 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	2	Infective-80	Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%	molecular properties	1
	1	HATS4e	leverage-weighted autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	3	JGI6	mean topological charge index of order6	topological charge indices	2
	6	B05[C-O]	presence/absence of C - O at topological distance 05	2D binary fingerprints	2
	2	JGI7	mean topological charge index of order7	topological charge indices	2
	2	DISPm	d COMMA2 value / weighted by atomic masses	geometrical descriptors	3
	5	RDF030m	Radial Distribution Function - 3.0 / weighted by atomic masses	RDF descriptors	3
	1	R1e+	R maximal autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	HATS8p	leverage-weighted autocorrelation of lag 8 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	Atype_H_49	Number of Hydrogen Type 49	atomtypes (Cerius2)	1
	2	Hy	hydrophilic factor	molecular properties	1
	1	Jhetp	Balaban-type index from polarizability weighted distance matrix	topological descriptors	2
	1	H8v	H autocorrelation of lag 8 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	2	EEig11d	Eigenvalue 11 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	MATS8m	Moran autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	MATS2p	Moran autocorrelation - lag 2 / weighted by atomic polarizabilities	2D autocorrelations	2
	4	B08[C-C]	presence/absence of C - C at topological distance 08	2D binary fingerprints	2
	1	S_sCH3	S_sCH3	atomtypes (Cerius2)	1
	2	HATS1e	leverage-weighted autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	nCconj	number of non-aromatic conjugated C(sp2)	functional group counts	1
	1	B04[C-C]	presence/absence of C - C at topological distance 04	2D binary fingerprints	2
	1	S_aasC	S_aasC	atomtypes (cerius2)	1
	1	R8m+	R maximal autocorrelation of lag 8 / weighted by atomic masses	GETAWAY descriptors	3
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	S_sOH	S_sOH	atomtypes (Cerius2)	1
	1	BELe3	lowest eigenvalue n. 3 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	BEHp4	highest eigenvalue n. 4 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	2	MATS5e	Moran autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2
	1	E3s	3rd component accessibility directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	2	Jhetv	Balaban-type index from van der Waals weighted distance matrix	topological descriptors	2
	1	nR=Ct	number of aliphatic tertiary C(sp2)	functional group counts	1
	1	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	1	HATS8v	leverage-weighted autocorrelation of lag 8 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	Mor28p	3D-MoRSE - signal 28 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	C-003	CHR3	atom-centred fragments	1
	1	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	JGI9	mean topological charge index of order9	topological charge indices	2
	1	B03[C-C]	presence/absence of C - C at topological distance 03	2D binary fingerprints	2
Or88a (19 Unique)					
	3	nHBonds	number of intramolecular H-bonds (with N,O,F)	functional group counts	1
	2	nRCO	number of ketones (aliphatic)	functional group counts	1
	3	GATS6m	Geary autocorrelation - lag 6 / weighted by atomic masses	2D autocorrelations	2
	2	EEig08x	Eigenvalue 08 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	nFuranes	number of Furanes	functional group counts	1
	1	nArCO	number of ketones (aromatic)	functional group counts	1
	1	ESpm15d	Spectral moment 15 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	C-005	CH3X	atom-centred fragments	1
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	L/Bw	length-to-breadth ratio by WHIM	geometrical descriptors	3
	1	nArCOOR	number of esters (aromatic)	functional group counts	1
	1	ESpm15u	Spectral moment 15 from edge adj. matrix	edge adjacency indices	2
	1	E2u	2nd component accessibility directional WHIM index / unweighted	WHIM descriptors	3
	1	EEig08d	Eigenvalue 08 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	H-051	H attached to alpha-C	atom-centred fragments	1
	1	ESpm14d	Spectral moment 14 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	PJI3	3D Petitjean shape index	geometrical descriptors	3
	2	X3A	average connectivity index chi-3	connectivity indices	2

Table 2.1 Continued

Or98a (20 Unique)					
	1	Lop	Lopping centric index	topological descriptors	2
	4	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	2	B04[C-O]	presence/absence of C - O at topological distance 04	2D binary fingerprints	2
	1	GVWAI-80	Ghose-Viswanadhan-Wendoloski drug-like index at 80%	molecular properties	1
	1	HATS7p	leverage-weighted autocorrelation of lag 7 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	HATS5v	leverage-weighted autocorrelation of lag 5 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	MLOGP2	Squared Moriguchi octanol-water partition coeff. (logP^2)	molecular properties	1
	2	GATS5e	Geary autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2
	1	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	1	MATS8m	Moran autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	nCrs	number of ring secondary C(sp3)	functional group counts	1
	3	HATS3p	leverage-weighted autocorrelation of lag 3 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	G1s	1st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	S_aasC	S_aasC	atomtypes (Cerius2)	1
	1	SP18	shape profile no. 18	Randic molecular profiles	3
	1	B05[C-C]	presence/absence of C - C at topological distance 05	2D binary fingerprints	2
	1	JGI2	mean topological charge index of order2	topological charge indices	2
	1	JGI8	mean topological charge index of order8	topological charge indices	2
	1	X4A	average connectivity index chi-4	connectivity indices	2
	1	H5e	H autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3

Table 2.1 Continued

Table 2.2: Top 100 predicted compounds for each *Drosophila* Or

Chemical name or Pubchem compound ID (CIDs), SMILES strings, and distances, of the top ~100 predicted compounds for each Or. All distances represent the minimum distance based on optimized descriptors to the previously known strongest active compound listed in the gray cells for that particular Or.

Or2a		Or7a		Or9a		Or10a		OR19a		Or22a		Or23a		Or35a		Or43a
SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES
CCCCOC(=O)C	0	CCCCC=O	0	CCCC(C)O	0	COC(=O)C1=CC=CC=C1O	0	CCCCCC(=O)C	0	CCCOC(=O)C	0	CCCCCO	0	CSCCCO	0	CCCCCCO
CC(C)CCOC(=O)C	0	CCCCCC=O	0	CCC(C=C)O	0	COC(=O)C1=CC=CC=C1	0	CCCCCCO	0	CCCCOC(=O)C	0	CCCCCCO	0	CCCCCC=O	0	CCCC(C)O
CCOC(=O)CC(C)O	0	CCCC=CC=O	0	CCC=CCCO	0	CCOC(=O)C1=CC=CC=C1	0	CCCCCC(C=C)O	0	CC(C)COC(=O)C	0	CCCC=CCO	0	CCCC=CC=O	0	CCCCCC(C=C)O
CCSC(C)OC(C)O	0.06547077	C1=COC(=C1)C=O	0	CCC=CCCO	0	C1=CC=C(C=C1)C=O	0	CCCCCOC(=O)C	0	CC(C)CCOC(=O)C	0	CCCCC=CCO	0.4489215	CCCCCO	0	CCC=CCCO
CC(C)CCOC(=O)N	0.07017575	C1=CC=C(C=C1)C=O	0	CC(C(C)O)O	0	CC(=O)C1=CC=CC=C1	0	CC(CCC(CC)O)CC	0.1773688	CCCC=CCOC(=O)C	0	C(CCC=CCC)O	0.4645003	CCCCCCO	0	CCCCCC(C#C)O
CC(C)CC=CC(=O)C	0.08148948	CCCO	0	CCCOC(=O)C	0	CC(C)COC(=O)C	0	C=CCCCCC	0.1795821	CCCC(=O)OC	0	C=CCCCCCO	0.4966429	CCCCCCCCO	0	C1CC1CCCCO
C(CCC)OC(=O)C=C	0.08191658	CCCCO	0	CCCCOC(=O)C	0	CC(C)CCOC(=O)C	0	CCCCCC(=N)C	0.1837474	CCCC(=O)OCC	0	CC=CC=CCO	0.5369311	CCCC=CCO	0	CCC#CCC(CC)O
CCCCSCC(C)O	0.08378222	CCCCCO	0	CC(C)COC(=O)C	0	C1=CC=C(C=C1)C=S=O	0	CCCCC(=O)CC	0.1983858	CCC(=O)OCC	0	CC(CCC=C)O	0.5705127	CCCC=CCO	0	CCCCC(CC)O
CCCOC(=O)C(=C)C	0.083826	CC(C)CCO	0	CCOC(=O)CC(C)O	0	CC1=C(C(=CC=C1)C)N=S=O	0	CCCCCC1CO1	0.2055113	CCCCCC(=O)OC	0	C#CCCCCO	0.653074	CCC=CCCO	0	CCCCCC(CC)O
CC(C)CCOC(=O)C=C	0.0868181	CC(=CCO)C	0	CC(CCC=C)O	0.09391671	CC1=CC(=C(C=C1)OC)C=O	0	CCCCOC1CC1	0.2097268	CCCCCC(=O)OCC	0	CCCCCOO	0.6743714	CCC=CCCO	0	CCCCC(C)CO
CC(C)CCOC(=O)CS	0.09010645	CCCC=CCO	0	CC(CCCO)C	0.1227934	CN(C=O)C1=CC=CC=N1	0	CCCCCC1(CO1)C	0.2192309	CCCCCCCC(=O)OC	0	C(CCCCCC)O	0.679884	CCCCCOC(=O)C	0	CCC#CC(CC)O
CCCCOC(=O)CS	0.0962103	CCCC=CCO	0	CC(C)NC(=O)C	0.1291849	CN1C=CC=CC1C=O	0	CCCCCC(=C)C	0.2327693	CC=CC(=O)OCC	0	C#CCCCCCO	0.6854802	CCCCCCOC(=O)C	0	C=CCC(C=C)O
CCC(=O)OCCC(C)C	0.09663025	CCC=CCCO	0	CCC(CCC)O	0.1337314	C1=CC=C(C=C1)C(C=O)C#N	0	CCCC(CCC=C)O	0.233932	CCOC(=O)CCC(=O)OCC	0	COC=CCCCO	0.6878323	CCCC=CCOC(=O)C	0	CCC#CCCO
CCCCOC(=O)NC	0.09927085	CCC=CCCO	0	CC(C)C(C)O	0.1449655	C1=CC=C(C(=C1)NC=O)O	0	CCCCCC(C)(C#C)O	0.2437784	CCCCOC(=O)CCC	0.2657116	C(C=CC=CCC)O	0.730026	CCC#CCCO	0.150652	CC(=C)CCCO
CC(C)CCOC=O	0.1000939	CCC=CC=O	0.06287397	CC=CCOC(=O)C	0.1529503	C1CCC(C1)C(=O)NN	0	CCC(CCCC=C)O	0.2460947	CCC#CCCOC(=O)C	0.2798496	CC=CC=CCCO	0.7327627	CCC=CCCOC(=O)C	0.153849	C=CC(CCCC)O
CCCCOC(=O)N	0.1060414	CC(=CC)CO	0.08745256	CCC(=O)NCC#C	0.165569	C1=CC=C(C=C1)C=C=O	0	CCCCCC(=O)C=C	0.2539876	CCCCC(=O)OC	0.2807561	CC(CCCC=C)O	0.7331066	C=CCCCCO	0.1577711	CCCC#CCO
CCCCOC(=O)CC	0.1064284	CC(=CCCO)C	0.092048	COCCC=C=N	0.1686873	CNC(=O)C1=CC=NC=C1	0	CCCCCC(C)S	0.2540855	CCC(=O)OCC=CC	0.3192234	CCCC(C)CCO	0.7638355	C(CCCCCC=C)O	0.1896431	CCC#CC(C)O
CCCNOC(=O)CC	0.1068854	C1=C(ON=C1)C=O	0.112377	CC(C)NC(=O)C=C	0.1706617	CC(=O)NCN(C)C	0	CCCCC(=O)C(=C)CC	0.2541479	CCCCC(=O)OCC	0.3386281	CCCC#CC#CO	0.7642293	CCC=CC=O	0.1996007	CCCCCC(C)O
CCCC(=O)CCC#CC	0.1072292	CCC(C)(C)C=NO	0.1149527	CC(CCCC)O	0.1771611	CC(C1=CC=CC=C1)N=C=O	0	CC(CC)CCCC	0.2605213	C(CC)OC(=O)CCCC	0.3432405	CCCCOO	0.8340461	CCC#CCCOC(=O)C	0.203439	CC=CC=CCO
CC(=CC(=C)OC(=O)C)C	0.1073104	CC=CC=CC=O	0.1158738	C=CC(CCC)O	0.1824546	CC(=O)C#CC1=CC=CC=C1	0	CCCC(=O)CCC=C	0.268571	CCCC=CC(=O)OCC	0.3470114	CCC#CCCCO	0.8383432	CCCC#CCO	0.2051737	CCC=CCO
CC(=C)C(=O)OCCCN	0.1084215	C1CC(=CC=C1)C=O	0.1165349	CCC#CCCO	0.18493	CC(C)CC(C)C=O	0	CCCCC(=O)OC(=O)C	0.2821946	COC(=O)C=CC	0.3564047	CCCC=CCCO	0.8559539	CC(=O)OCCCCC=C	0.2169564	CC1(CC1)CCCO
CCCOC(=O)C(=CC)C	0.1113651	C1=COC=C1C=O	0.1266509	CCC(CC=C)O	0.1958346	C1CCCN(CCC1)N=O	0	CCCC(=O)CCC	0.2839333	CCCOC(=O)CC	0.3620649	CC(CCC#C)O	0.8633463	CC=CCCCO	0.2327925	CC#CCCO
CCC(C)C(=O)OCC=C	0.1126422	C1=C(OC=N1)C=O	0.1277235	CCC(=O)NCC(=O)C	0.2001053	C1CCC(CC1)C=O	0	CCCCCC(C)(C)O	0.2848814	CCCC(=O)OCCC	0.3642294	CCCCCCOO	0.8935004	C(CCC=CC)OC(=O)C	0.2366404	CC=C=CC(C)O
CCCCC(=O)C=C(C)C	0.1134776	CCC(C)CC=O	0.1310643	CC(CCOC(=O)C)O	0.2011426	C=CC(=O)C1=CC=CC=C1	0	OC(C(=O)C)CCCCC	0.2870452	CCCCOC(=O)CC	0.408598	CCCCCC#CO	0.895056	C(CCC=C)O	0.253342	C=CCCCC(C=C)O
CCOC(C)OCC#C	0.1143103	CCCN=CC=O	0.1384452	CCC(=O)NCC#N	0.2057452	C1CCN(CC1)N=O	0	CCCCC(CC(=C)C)O	0.2945198	CC(=C)CCCC(=O)OC	0.4087118	CC(CCC=C=C)O	0.913551	CC(=O)OCCCC=C	0.2575001	CC(C)CCCCO
CCCCOC(=O)CN	0.1163705	CCC=CCO	0.1388489	CC(CC(C)O)C	0.2065425	C1CCC(CC1)NN=O	0	CCCCCC(C)O	0.3036086	CCCCCC(=O)OCCC	0.4096699	CCC=CCO	0.9216458	C#CCCCCO	0.258166	CCC(CCC(C)C)O
CCC#CC(CC(C)C)O	0.1186069	CCC(=C=CCO)C	0.139234	COC(=O)CC(O)C	0.209219	C=CC(=O)C(=O)C1=CC=CC=C1	0	CCCCCOC=C	0.304017	CCC=CCCOC(=O)C	0.4280228	CC#CCCCCO	0.9630825	C(CCCCCCCC)O	0.262205	CCC(C#CC)O
CCCOC(=O)C=C(C)C	0.1191027	CCC(=C=CC=O)C	0.1407318	CC(CC(=O)C)O	0.2110131	C1=CC=C(C=C1)CN(C=O)O	0	CCCCCC(C=C)S	0.3054288	CCC=CCC(=O)OCC	0.4509044	CCC#CCCO	0.9669537	C(CCO)CCS	0.2659421	CCCCC(C#C)O
CCCCCC(=O)C#C	0.1191398	C=C(CO)CC	0.1424594	CCC(=O)OC(C)(C)O	0.2110216	C1=CC=C(C=C1)C(=O)CS	0	CCCCCC(C)(C=C)O	0.3074979	COC(=O)CCC=CCCC	0.4515538	C=C=CCCCCO	0.9698743	CC=CC=CCOC(=O)C	0.277348	CCCCC(C=CC)O
CC(C)(NCOC)O	0.1211261	CC(=C)CCCO	0.1441704	C=CC(C=C)O	0.2127051	CCN(C=O)C1=CC=CC=N1	0	CC(CC)C(CCCC)O	0.3169439	COC(=O)C=CCC	0.4606538	C=CCCNCCO	0.9822695	C(CC=C=O)CS	0.2857512	CC#CCCCO
COCCCCC#C	0.1212032	CC(=CCCC=O)C	0.1441953	CC(C)C(=O)NC=C	0.2141241	C1=CC=C(C=C1)C(=S=O)C#N	0	CC(C)CCCC(=O)C=C	0.318154	CCC#CC(=O)OC	0.4635536	CCCC=C=CCO	0.9892507	CC(=O)OCCCC=C=C	0.2908757	CCCC(CC)O
CCCCOC(CC)O	0.1213945	C=CC(=C)CCC=O	0.1466388	CC(=O)CNCC(=O)C	0.2158649	COC(=O)C1=CC=CC=C1N=O	0	CC(CCCCC(=C)C)O	0.3204591	CCCCCOC(=O)CC	0.465345	CCCCC#CCO	0.9912028	CCC=CCO	0.2957919	CC(=CCCCO)C
CCCOCCC(C)O	0.1230352	CCCC(=C)CCO	0.1482145	CCC(=O)NCC	0.2169106	CNC(=O)C(=O)C1=CC=CC=C1	0	CCC(C)CC(=O)CC(=O)C	0.3246326	CC(=CCOC(=O)C)C	0.4677529	C=CCCC#CCO	0.9931836	CC(=O)OCCCCC#C	0.3021282	CC(C)C=C=CCO
CCOC(=O)C(C)OC	0.1232314	C1CC1CCO	0.148306	CCC(CC#C)O	0.2179049	CNC(=O)C1=CC=CC=C1C=O	0	CCCCCC(=C)C(=O)C	0.3251662	COC(=O)CC=CC	0.4684388	CCCNCCO	1.000939	C#CCCCCCCO	0.3034524	CCC(CCC)O
CCOCC(=O)C(=C)C	0.1233035	C1CC1CO	0.1521086	CCC(=O)NC(C)C	0.2187382	CN1C=CC=CC1=CN=O	0	CCCCC(=O)CC(=O)C	0.3258552	CCC=CC(=O)OCC	0.4687615	CC(COC=C)O	1.008994	CCCC#CC=O	0.3056329	CC=CCCCO
CCC(=NO)C(=O)OCC	0.1239997	C1=CC=NC(=C1)C=O	0.153194	CCC(C#C)O	0.2188469	CNC(=O)C1CCCC1	0	CCCCCC1C=C1	0.3268252	CC(=C)CCCOC(=O)C	0.4696284	CCCC(CC)CO	1.017518	C=CCCCCCCCO	0.3066109	CC(C)CCC(C)O
CCOC(C)OC=C=C	0.1247327	CC=CCCCO	0.1548523	C(CC#N)CSCO	0.2207365	CC(C)N(CC=C)N=O	0	CCCC(C)CC(CC)O	0.3343905	CC(=COC(=O)C)C	0.4710929	C=CCOCCCO	1.021779	C(CC=CC)O	0.3164314	CCCC(C#C)O
CC(C)CNOC(=O)C	0.1249519	CC=CCO	0.1553325	CC(C)(CO)OC=C	0.2237399	C=CC(=O)N1CCNCC1	0	CCCCCC(C)C(C)(C)O	0.3352282	C(C)OC(=O)CCCCCC	0.4714722	C=CCCCCO	1.022272	CCC=CCCCCO	0.3186713	CC(CCC=C)O
CC(COCCN(C)C)O	0.1254713	CCCC#CCO	0.1555892	CC(=O)OCC(C)(C)O	0.2248765	C1CCC(CC1)SN=O	0	CC(=O)CC(=O)CCC=C	0.3372005	CC=CCOC(=O)C	0.4865902	CCCC(COC)O	1.034345	C=C=CCCO	0.3260218	CC(CCCC=C)O
CCCCCC(=C)C=O	0.1264398	CCC(=CC=O)C	0.1557608	CCC(=O)OCC#C	0.2253047	CNOC(=O)C1=CC=CC=C1	0	CCCC(=O)CCC#C	0.3419066	CCC=CCC(=O)OC	0.4867699	COCCCCCO	1.034478	C=C=CCCCO	0.3268738	CC(C)(CC#C)CO
CCCC(=O)CCC#C	0.1269273	CC(C)CC=O	0.1573001	COCC(CC=C)O	0.2257484	CC(C=O)NC1=NSC=C1	0	CCCCCC(C)(C#N)S	0.3435598	COC(=O)CCCC(=O)OC	0.4925246	C=COCCCCO	1.035535	CC#CCCO	0.3308051	CC=C=CCO
CC(C)CCOCCO	0.1285711	CC(CCCO)C	0.1583737	CCC(COC=C)O	0.226681	CCOC(=O)C(C)CC#N	0	CCCCCSC(C)C	0.3462637	CCOC(=O)C=CC(=O)OCC	0.4932784	CCCC#CCO	1.037642	CCC#CCO	0.3357445	CC(C=CC)O
CC(C)(OC)OCC#C	0.1288759	n1cc(ccc1)C=O	0.1583875	CCON(C)C(=O)C=C	0.2288363	CC(C#C)N1CCCC1=O	0	CCCCC(CC)C#C	0.3503987	CC#CC(=O)OC	0.4977876	CC(C)CCCCO	1.043771	CC=CC=CCO	0.3361849	CC(=C)C#CCO
CC(=CCCCO)C	0.129211	C1=C(C=NO1)C=O	0.160688	CCOC(=O)C(C)C	0.2299445	CCCN1C(=O)C=CC=N1	0	CCCCCC(=C=C)C(C)O	0.3514533	CC#CCOC(=O)C	0.5000781	CCNCCCO	1.057598	CCC#CC=O	0.3414796	CC(C#CC(=C)C)O
CCCOC(=O)CC(=N)C	0.1298602	CC(=C)CCC=O	0.1624599	CCCC(O)SC	0.2308107	CON=C(C=O)C1=CC=CC=C1	0	CCC(CC)CCC(=O)C	0.3538593	CCCCC(=O)OCCCC	0.5017934	CNCCCCCO	1.073321	CCCCCC=CCO	0.3459803	CCCCC(C)(C)O
CCCOS(=O)(=S)CCC	0.1301293	C1=CN=CC=C1C=O	0.1635538	C(CC=CC)O	0.2308405	C1=CC=C(C=C1)CC(C=O)N	0	CC(CCCC(=O)C=C)O	0.354033	CCOC(=O)CCC(=O)OC	0.5026755	C=CCC#CCO	1.09117	CCCC#CO	0.351461	CCCCC(CCC)O
CCCOC(=O)C(C)S	0.1303181	CCC(=C)C=O	0.1636865	C(C)OC(=O)CCC	0.2359009	C#CC(=O)NC1=CC=CC=C1	0	CC(CCC(=O)CC)CC	0.3540939	CC(CCOC(=O)C)C=C	0.5067616	CCC#CCC(C)O	1.100245	CCC#CCOC(=O)C	0.369806	CCC1CC(C1)O
CCCCOC(C)CO	0.1305516	CC1(CC1)CCO	0.1640703	CC(=O)CC(COC)O	0.2359682	COC1=C(C=C(C=C1)N)C=O	0	CCCCC(CC#C)O	0.3577258	CC(=C)CCOC(=O)C	0.5088863	CC#CCCC(C)O	1.104701	C=CCC#CCO	0.3708455	CC#CC(C)CO
CCCCC(=O)C=C	0.1305813	CCC(CC)CC=O	0.1642731	COCC(C=C)O	0.2375351	C1=CC=C(C=C1)CC(N)N=O	0	CCCC(=O)CCC(=C)C	0.3603211	CCOC(=O)CCCC(=C)C	0.5174128	CCNCCO	1.109037	CCCCCC#CCO	0.3716565	C=CC#CCCO
CCCCOCC(C)O	0.1314971	C1CN1CCO	0.1658346	CNN(C)C(=O)C=C	0.2376134	CCC(CC(C)C)C=O	0	CCCCC1CS(=O)(=O)O1	0.3604474	CCC(=O)OCC(C)C	0.5188726	CCCOCCO	1.112199	C=CC#CCCO	0.3768789	CC(CC=C)O
CC(=C)C(=O)OCCCS	0.1325123	CC=C(C)C=O	0.1661592	CC(NC)OC(=O)C=C	0.2379202	C1CCC(C1)N2C(=O)C=CN2	0	CCCCC(CC)C#N	0.3634326	CCOC(=O)CC(=C)C	0.5194836	CCCCCC(C)O	1.112835	CC=CCO	0.3773388	CC(C)(CC=C)O
CCC(C(=O)OCC)O	0.1325821	CC1CC1CO	0.1669874	CC(=O)OCCC=C	0.2380599	CS(=O)C1=CC=CCC1	0	CCC(=O)CCC(CO)C=C	0.3646824	CC(=C)CCC(=O)OC	0.5201922	CC(CC(C)O)C=C	1.115136	CC=C=CCO	0.3773877	CC(CCCC)O
COCC(=O)C(=C)CC	0.1326743	CC=CC=CCO	0.1709591	CC(C(O)OC)O	0.2393481	C=NC(=O)C1=CC=CC=C1C=O	0	CCCCCC(=S=O)C	0.370869	C(C=CC)OC(=O)CCC	0.5210228	CCC(CC)COO	1.117559	CCCCCCCC#CO	0.3803588	CCC(CCCC=C)O
CCC(=O)OCC=C(C)C	0.1326982	CN=NCCO	0.1724168	CCC(=O)OC(C)OC	0.2418581	CC(C1=CC=CC=C1)=O	0	CCCCCOC(C)O	0.3727137	CCC#CCOC(=O)C	0.528372	CC(COCC=C)O	1.123403	C=CCCCC=O	0.3809182	C=CC(CCC)O
CCCCCCC(C#C)O	0.1329399	CC(C)CO	0.1734693	CCC(O)SC	0.2446503	CC(=O)C=CC1=CC=CC=C1	0	CCCCCC(C)C(=O)C	0.3776474	CC(=O)OCCCSC	0.5326966	C=CCCCOO	1.135939	CC#CCCCCCO	0.3838803	CCC#CCO
CCCCC#CC(CC)O	0.134323	CCCN(C)CCO	0.1744309	CC(C)NC(=O)SC	0.2461038	CC(C=O)C1=CC=CC=C1	0	CC(=C)CCCC=C=O	0.3784285	CCC(=O)OCOCC	0.5372453	C=CCCC(C=C)O	1.142012	C#CCC=C=O	0.3956265	C=CCC#CCO
CCCCOC(=O)C(=C)C	0.1347231	C=CCCCCO	0.1755409	CCCCC(=O)NC	0.2463324	C=C(C=O)SC1=CC=CC=C1	0.00444444	CCCC1C(N1)CCC	0.3810508	CCCC(C)OC(=O)C	0.5383471	CNCCO	1.14528	CCC=CCCCCCO	0.4024018	CC(C)CCCC(C)O
CCCCOC=O	0.1358109	C1C=CC=C1C=O	0.1766221	CC(C)COCC#N	0.2466133	CC1=C(N=C(C=C1)OC)C=O	0.00444444	CCCCC(=O)SC(C)C	0.3847285	CC(C)CCC(=O)OC	0.5394631	CNCCCCO	1.157602	CCCSCO	0.4072422	CCC(C=CC)O
CC(C)(CC(=O)OC)O	0.1369096	CC(=C)CO	0.1770184	CC(=O)CC(=N)OC	0.2470889	CC1=CC=CC(=C1)C(=O)NC	0.00444444	CCC(=O)CCC#CC	0.3854414	CCOC(=O)CCSCC	0.5432896	C#CC=CCO	1.160654	C=CCCC=O	0.4187745	C=CCCC(C=C)O
CCCCOCC(=O)C	0.1371675	CCC(=CC)CO	0.17742	CC=CCCCO	0.247501	CC(=O)C=CC1=CC=NC=C1	0.00444444	CCCCC1C(O1)CC	0.3909191	COC(=O)C=CCCC	0.5447552	C=CC#CCCCO	1.168492	CCCCCCC=CCO	0.4207279	CCCC#CO
CCOCC(CC#N)O	0.1377199	C(CC=CC)O	0.1791561	CCC(C=CC)O	0.2496933	C1=CC(=O)N(N=C1)CCC#N	0.00444444	C(C)C(CO)CCCC	0.393805	CC(CCOC(=O)C)CC	0.5477659	CCCN(C)CCO	1.170331	C#CCCCC=O	0.4222186	CC(C)CC#CO
CCCCCC(C#C)O	0.1388412	CCC(C)CCO	0.1795803	CC(COC)OC(=O)C	0.2501114	C=CCC(C=O)C1=CC=CC=C1	0.00444444	CCCCC(CC)C1OO1	0.3938669	CCCOC(=O)C=CC	0.5516671	COCCCCO	1.17205	CCCC#CCCCO	0.4278039	CCCCC(C)OO
CCOC(=O)CC(C)(C)O	0.1396291	C1C=CC(=C1)C=O	0.1804237	CC(C)C(=O)NC	0.2512166	C1CC#CCCCC(=O)C1	0.00444444	CCCCCC(C)(C)S	0.3942348	CCCOCOC(=O)CC	0.5534903	CC(=CC#CCO)C	1.172223	C=CCCO	0.4337125	C1CC(C1)CCO
CCCOC(C)OCC	0.1402788	C=CCO	0.1809406	CCC(COC)O	0.2512212	COC(=O)C1=CC=CC=C1N=C=O	0.00444444	CCC(=O)CC(=O)C1CC1	0.3948309	CCC(=C)CCC(=O)OC	0.5544112	CCOCCCO	1.173143	C#CCCCO	0.4401863	CC=CCCO
CCC(=O)COCC	0.1414631	CC(=CC=O)C	0.1841027	CCSCCCO	0.2515773	C1=CC(=CN=C1)NN=O	0.00444444	CCC(=C)CCCC(C)(C)O	0.3989707	CCCCCCC(=O)OC	0.5646635	CC(CC#CC#C)O	1.176467	CC(=O)OCCCC#C	0.4439934	C(CC=CC)O
CC(C)COCC(C)O	0.1421389	CCC(=C)CC=O	0.1847582	CC(=C)C(=O)OCS	0.2518678	C1=CC=C(C=C1)N=NC(=O)N	0.00444444	CC(C)CCCCC	0.4016131	CCCCCCC(=O)OC	0.5646635	CCCC(CC)OO	1.183238	C#CC=CCO	0.4459709	CC(C1CCC1)O
CC(C)(C(=O)OC=C)O	0.1421985	C(=O)C=CC	0.1847743	CC(=O)NCC=C	0.2532654	CC=CN1C=CC=CC1=O	0.00444444	CCC(=O)CCC(=O)CC	0.4024528	CCCC(=O)OCCN(C)C	0.5674101	CCCCNCCO	1.188692	C(CCS)CO	0.4461035	CCCCCOO
CCC#CCC(C(C)C)O	0.1429746	CC(CO)CC	0.1875941	CC(CC(=O)OC=C)O	0.2542865	C1CCC(C1)NCC(=O)O	0.00444444	CCCCCC(CO)C(C)C	0.4048947	CCCCCOC(=O)CCC	0.5748757	CCC(C)CC(C)O	1.199429	CC#CCCC=O	0.4496623	C1CN1CCCO
CCCCCC#CO	0.1430722	CC(=C)C(=C)CO	0.1882973	CC(C)C(=O)OCCO	0.254431	C1=CC(=CN=C1)C=CC=O	0.00444444	CCCCCNC(=O)C	0.4057246	CCCCCC=CC(=O)OC	0.5755997	C=CCC(CO)C=C	1.206247	CCCCCCSO	0.4525682	C(C)C1C(CC1)O
CCN(C(=O)OCC)N=O	0.14309	CC1(CC1)CO	0.1885827	CC(CCC(=O)C)O	0.2547923	C1=CC=C(C=C1)C(=O)OCC#N	0.00444444	CCC1=CCC(CC1)C(=O)C	0.407917	CC(CCC(=O)OC)S	0.5785319	CCC(CC=C)OO	1.207051	C=CCCCC#CCO	0.4758609	CCC(CC#C)O
CC(C)OC(=O)CC(=O)C	0.143887	CCCC(C)CCO	0.1909857	CNC(=O)CCC#N	0.2557862	C1=CC=NC(=C1)C(=O)NN	0.00444444	CCCCCOC(C)C	0.4095105	CCOC(=O)C=CC=CC	0.5856332	CCCCOCO	1.22489	CCCCC#CCCO	0.4770374	CC1CCC(C1)CO
CC(C)CCCC(=O)C=C	0.1443528	C=CC(=C)CCCO	0.1918209	CC(COC)(C=C)O	0.2578298	C1CCC(C1)NC=O	0.00444444	CCCCCC1C(O1)C	0.4099452	CCC(=C)OC(=O)C	0.5893508	CC(CCCOC)O	1.226391	CCC#CCCCCO	0.4775639	CC#CC(C)O
CCCSCOCC	0.1443588	CCCC(=C)C=O	0.1922224	C(CC=C=O)CS	0.2584825	CCOC(=O)C1CCC=CO1	0.00444444	CCCC(C(=O)CCC)O	0.4100599	CC(C)OCCCC(=O)OC	0.590573	C=CCOCCO	1.242012	C#CCCO	0.4778815	C1CCC(C1)CCO
CCCCCOC(=S)S	0.1448369	C(CCC=C)O	0.1925684	CCCC#CO	0.2588648	C1=CC=C(C=C1)C(=O)C=CO	0.00444444	CCCCCC(=NO)C(=O)C	0.4164583	CCOCCCC(=O)OC	0.5966995	CC(C)CC#CO	1.246353	CC#CCCCO	0.4913373	CCC(C#CC(=C)C)O
CCCCOCC=O	0.1448396	CCC(=CC)C=O	0.1947075	CC(CCCO)C#N	0.259088	CCCSC(=O)NCC	0.00444444	C=CCCC(=O)CC#N	0.4176805	CCC(=C)COC(=O)C	0.6069738	CCCNCCCO	1.2472	CCCCCC=CCCO	0.4944054	CC(CCCO)C
CCOCC(CS)O	0.1452569	CC(CCC=O)C=C	0.1953673	CCCCCOC	0.2592186	CN(C)CCN(C)N=O	0.00444444	CCCCC(CC)S	0.4203377	CCOC(=O)CCC(=C)C	0.6076767	C#CCOCCO	1.247526	C(CC=CCCCC)O	0.4944054	CC1CCCC(C1)O
CC(=CCOC(=O)C)C	0.1463231	CC(=C=C=CC=O)C	0.195764	CCCC(=O)OC(C)O	0.2596483	CC=CC(=O)N1CCCC1	0.00444444	CC(C)CCCC(=O)C(=C)C	0.4206731	CC(CC)OC(=O)C	0.6099463	CCCC(=C)CCO	1.251926	CC=CCOC(=O)C	0.4966221	CCC(C#CC#C)O
CCCOC(C)CO	0.1464299	C(CN=O)CO	0.1965367	CC(CCC#C)O	0.2601054	CN1CCCCCCC1=O	0.00444444	CCCCCC(=C)S(=O)(=O)O	0.4218549	CCCCCCOC(=O)CC	0.6146034	CC(CCO)COC	1.253905	C#CCCC=O	0.5040784	C=C=CCCCO
CCCOC(C)(C)O	0.1467136	CC(C)C#CC=O	0.196942	CCC(=O)OC=C	0.2621356	CC(CC1=CC=CC=C1)N=O	0.00444444	CCCC(=O)CC=C=O	0.4228198	CCN=CC(=O)OC	0.6149937	CC(C)COCCO	1.266434	CC(=O)OCCC=C	0.5056067	C#CCCCC(C#C)O
CCCCC#CC(=O)C	0.1467714	C=C1CC1CO	0.1972782	CCNC(=O)N(C)N=O	0.2628142	CCOC(=O)C=CN(C)C	0.00444444	CCCCC(=O)OC(C)O	0.4236843	CC(C)CCOC=O	0.6166259	COCCCO	1.280553	CCSCCCO	0.5143412	CC#CCC(CC=C)O
CCCOCC(CS)O	0.1468967	C1=CC=C(C=C1)N=O	0.1975616	CCC(=O)OCC=C	0.2642549	CC(=O)OCC1CCCCO1	0.00444444	CCCCCC(C)N	0.4247753	CCC(=O)OC=C(C)C	0.6198678	CCCC(CO)OC	1.281018	C=CC#CCO	0.5166441	CCCCCC(C)(C)O
CCCC#CCC(C)O	0.1471084	C=CC=CC=O	0.1981156	CC(=O)OCCC#C	0.2643481	C1=CN=CC=C1C=CC(=O)O	0.00444444	CCCCC=CCO	0.4273939	CC(=C)CC(=O)OC	0.6225674	CCCC(CCC)O	1.282156	CCC=CC=O	0.5199458	CC(CCC#C)O
CCCCC(=O)C(=C)CC	0.1471127	CC(CC=O)C=C	0.1995606	CCOC(=O)CC=C	0.2651179	C=CC(=O)OC1CCCC1	0.00444444	CCCC(=C=O)CC(C)CC	0.4286117	CCC(=O)OCOC	0.6282884	C(CCC)OCCO	1.282801	CC(=O)OCC#CC=C	0.5288355	CC(CC=C)CO
CC=C(C)C(=O)OCC=C	0.14822	CC1CCC1CO	0.2017347	CN(C)CC#CCO	0.265367	COC(=O)C(=O)C1=CC=CC=C1	0.00444444	CCCCSCCCC	0.4290327	CCCOC(=O)CCCOC	0.630773	CC(CC=C)O	1.293617	C(CS)C=O	0.5322155	CCCCC(C=C=C)O
CCCCCOC=C	0.1482396	CCC=C(C)C=O	0.2020607	COC(=O)CCCS	0.2669651	CC(=O)NC1=CN=CC=C1	0.00444444	CCCCC(=O)NC(=O)C	0.429785	CC(=CCOC=O)C	0.6311287	CCOCCC(C)O	1.294746	CCCCC=C=N	0.5471772	CCC1CC1CCO
C(CCSCCO)CO	0.1482582	CCC#CCCO	0.2021027	CC(CC=C)O	0.2670124	CCN(C=O)C1=CC=CC=C1	0.00444444	CCCCOS(=O)(=O)CCC	0.4309104	CCOC(=O)C=CCOC	0.6346383	COCCCCOO	1.297096	C(CCC=CCCCC)O	0.54918	C=C=CCCO
CCCCOOC(C)C	0.1487555	CCCC(C)CO	0.2036942	CCCC(COC)O	0.2683515	COC(=O)C1CCCC1	0.00444444	CCCCCCN	0.4327508	CCC(=O)OCCOC(C)C	0.6346818	C=C=CCOCCO	1.304615	CCCCC#CCCCO	0.550602	CC(C)(C)CCCCO
CCCCOC(C)(C)O	0.1488735	CC(=CCCCO)C	0.2052078	CC(C)OC(=O)C	0.2684826	CNC(=O)C1=CC=CC=C1O	0.00444444	CCCCCSC=C	0.4333872	C(C)OC(=O)CC(C)C	0.6351463	CCC(CO)NCC	1.316144	CCCCCC#CCCO	0.5541043	CC(C)(CCC=C)O
CC(C)COOC=O	0.1490883	C(=O)C1=CCCCC1	0.2059202	CCC(C)OC(=O)CC	0.2686518	CC=CC(=O)C1=CC=CC=C1	0.00444444	CCCCCC(C)(C(C)C)O	0.4341677	CCC(=O)OCCC(C)C	0.6354924	CCNCCCCO	1.327671	C#CCCCC#CCO	0.5552247	C#CC#CCO
CC(CCCCC=C)O	0.1491159	CC(=C)CCO	0.2070791	CCC(O)SCC	0.2688412	CC1=CC=CC=C1N=C=O	0.00444444	CCCCOC(=O)CC	0.4344767	CC(C)CC(=O)OC	0.6391253	COCC(COC)O	1.333058	CCSCCCC=O	0.5587566	CC(C)C(C=C=C)O
CC(=C)CCC#CCO	0.1491427	CC(C)C=C=CCO	0.2087818	CC1C(CCC1)O	0.2713174	CCC(C)C(C)C=O	0.00444444	CCCC(O)OC(=O)CCC	0.4346575	CCOC(=O)SCC	0.639455	C=CCNCCO	1.335479	C=C=CCO	0.5606924	CCCC(CCC)O
CCOC(=O)C(C)C(=O)C	0.1496697	CCCC(C)CC=O	0.211148	CCCCOC=C	0.2717477	C=CC1=CC(=CC=C1)C=O	0.00444444	CCCCC(=O)N=C=C	0.4349265	CCC(C)COC(=O)C	0.6423194	CCC(CC)O	1.337735	CCCCCCC#CCO	0.5618984	C=CCCCC(C#C)O
CC(C)(C#C)OC(=O)C=C	0.1497625	CC(=C(C)C=O)C	0.2134845	CCCC(=O)NC	0.2743549	CC(=O)N1CCCC=C1	0.00444444	CCCCCSC(C)O	0.4351572	CC=CC(=O)OC(C)C	0.6437383	CC(CC#C)O	1.342632	CC(=O)OCCC#C	0.5627399	CCC(CC=C)O
CCCCC=CC(=O)C	0.1498653	CC(=C1CC1)CO	0.2144683	C(C)(C)OC(=O)CC	0.2750103	CN1CCN(CC1)N=O	0.00444444	CCCCCC1OO1	0.4355869	CCOC(=O)CCCN(C)C	0.6445598	CC(C)C=C=CCO	1.343159	C1CC(CCC=C1)O	0.5761688	CC1CC(C(C1)O)C
CCCOC(=O)CC(C)C	0.1499579	CN(C)CCO	0.2166105	CC1CC=CC1O	0.275073	CCOC(=O)C1=CC=CC=N1	0.00444444	CCCCCS(=O)(=O)OCC	0.4365799	COC(=O)CCCSC	0.6478714	C=COC#CCO	1.345292	CC#CCC=O	0.580464	CC(CC#CC#C)O
CCCCCC(CC#C)O	0.1499741	C=CCCC=O	0.2166117	CCNC(=O)C=C	0.2752052	CCOC(=O)C1=CC=NC=C1	0.00444444	CCCCC1CC1C	0.4365966	CCN(CCOC(=O)C)N=O	0.651088	CC(C)CNCCO	1.347795	C(CC=CCC=CCC)O	0.5830308	C#CC=CCO
CCCCC#CC(C)O	0.1503183	CC(CCCO)C=C	0.2166441	C=CCCCCO	0.2753153	CNC(=O)C1=CC=CC=C1	0.00444444	CC(=O)CCCC(=O)C(C)(C)C	0.4378504	CCCCC=CC(=O)OCC	0.654787	CCCOCC(C)O	1.35553	C1CC=CCC(C1)O	0.591331	C1CCC(CC1)CO

Table 2.2

Or43a	Or43b		Or47a		Or49b		Or59b		Or67a		Or67c		Or82a		Or85a	
Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist	SMILES	Dist
0	CCOC(=O)C	0	CSCCCO	0	C1=CC=C(C=C1)CO	0	CC(=O)C	0	CC(=CCCC(=O)C)C	0	CCCCCO	0	CC(=CCCC(=CCOC(=O)C)C)C	0	CCCCCCO	0
0	CCCOC(=O)C	0	CCCCCC(=O)C	0	C1=CC=C(C=C1)C=O	0	CCC(=O)C	0	C1=CC=C(C=C1)CCO	0	CCCCCCO	0	CCC(OC/C=C(CC/C=C(C)/C)\C)=O	0	CCC=CCCO	0
0	CCCCOC(=O)C	0	CCCOC(=O)C	0	OC1=CC=CC=C1C	0	CC(=O)OC	0	COC(=O)C1=CC=CC=C1	0	CCC(C=C)O	0	O=C(C)CC/C=C(CC/C=C(C)/C)\C	0	CCCC(=O)OCC	0
0	CC(C)COC(=O)C	0	CCCCOC(=O)C	0	C1=CC=C(C=C1)N=O	0.01680465	CCOC(=O)C	0	CCOC(=O)C1=CC=CC=C1	0	CCCC=CCO	0	CC1=CC=C(C=C1)N=C(C)CC(=O)C	0.6201264	CCOC(=O)CC(C)O	0
0.00332052	CCCC(=O)OC	0	CCCCCOC(=O)C	0	CC1=CC(=CC=C1)O	0.02191376	CCCC(=O)OC	0	C1=CC=C(C=C1)C=O	0	CCC=CCCO	0	CCCCCCC#CCCC(=O)C	0.6626511	CCC#CCCO	0.09486577
0.00699056	CCCC(=O)OCC	0	CCCCCCOC(=O)C	0	C1=CC=C(C(=C1)O)S	0.05712351	CCOC(=O)C(C)O	0	CC(=O)C1=CC=CC=C1	0	CC(CC#C)O	0.07067509	CC(=O)OC1=CC=C(C=C1)OCC#C	0.695874	CC=CCCCO	0.1241049
0.01572877	CCOC(=O)CC(C)O	0	CC(C)COC(=O)C	0	Cc1ccc(cc1)O	0.08002216	CCC(=O)OC(C)O	0.08309379	CCOC(=O)CCC(=O)OCC	0	C=CCC(C=C=C)O	0.0775118	CC(=O)CCC(=C)C(=O)CCCC#N	0.7224687	CCCC(=N)OCC	0.1455693
0.01572877	CCC(=O)OCC	0	CCCCCC(=O)OC	0	C1=CC(=CC=C1N)O	0.08291527	CCC(=O)OC	0.08527063	C=CC(=O)C1=CC=CC=N1	0.3008233	CC(C)(CC=C)O	0.0885166	CC(=CCCC1=CC=C(C=C1)CO)C	0.7293173	CCCCNC(=C)C	0.1695939
0.01642782	CC=CC(=O)OCC	0	CCCCCCC(=O)C	0.00616096	C1=CC=C(C=C1)N=C=O	0.08630082	C(C)OC(=O)CC	0.09857435	C1=CC=C(C(=C1)CC#N)C=O	0.3080015	C=CCCCCO	0.09353587	C1CCCN(CCC1)C(=O)CC#N	0.7444323	CCCC#CCO	0.1791638
0.01782593	CCCONC(=O)C	0.0959588	CCC(CC)CCC(=O)C	0.00770119	C1=CC=C(C(=C1)N)O	0.08697793	CCC(=O)COC	0.1112024	CCOC(=N)C1=CC=CC=C1	0.312236	CC(=C)CCCO	0.1018462	CCCCC1=CC=CC=C1OCCO	0.7455565	CCCC(=O)CCO	0.1893542
0.0180007	CCN(C(=O)C)O	0.1130635	CCCSCO	0.00847131	C1=CC(=CC(=C1)O)N	0.08747038	CCCOC(=O)C	0.1141674	COC(=N)C1=CC=CC=C1	0.3311703	CC(CCCC)O	0.1056086	CC(=O)C(=O)OC(=O)C1=CC=CC=C1	0.753441	CC=C=CCCCO	0.1938411
0.01835023	CCOC(=O)SCC	0.1132685	CCCOC#N	0.00924143	C1=CC=C(C(=C1)O)Cl	0.08876304	CCCOC(=O)C	0.1141674	CCC1=COC(=N1)CC	0.3703241	CCC(CC=C)O	0.1068447	CCC=CCCC=CCOC(=O)C	0.7560306	CCN=C(C)CC(=C)O	0.1971383
0.01887452	CCNNC(=O)C	0.1183047	CCOCC#N	0.00924143	c1(ccccc1)NC=O	0.09744237	CCC(O)OC(=O)C	0.1244704	CCC(=O)C1=CC=CC=N1	0.3768891	CCC#CCCCO	0.1081803	CC1=CC=C(C=C1)CCOC(=O)C	0.7645408	CC=CC=CCCO	0.1971396
0.01904928	CC#CC(NC)O	0.1231371	CCCCC(=O)COC	0.01001155	C1=CC(=CC=C1O)S	0.09756548	CC(COC(=O)C)O	0.1292618	C1=CC=C(C(=C1)C=O)C=O	0.3797241	CCC(CC#C)O	0.1259778	C1=CC=C(C=C1)COC(=O)N=C=O	0.7656269	CCOC#CC(C)O	0.2069597
0.02114645	CC=C=CC(C)(C)O	0.1294547	CCCCCCC(=O)C=O	0.01078167	C1=CC=C(C=C1)C#CO	0.1029824	CCC(C(=O)OC)O	0.1352768	CCOC(=O)N1C=CC=C1	0.3857737	CC=C(C)C(C)O	0.1262036	CC1=CC(=C(C(=C1)C)COC(=O)C)C	0.7687964	CCC#CCCCO	0.2311878
0.0223698	CCNCNC(=O)C	0.1317853	CCCCC(=O)CC(=O)C	0.01078167	C1=CC=C(C=C1)C=C=O	0.1066141	CCC(N=C=O)OC	0.1459781	C1=CC(=CC(=C1)C=O)C=O	0.3905579	CC=CCCCO	0.1270505	CCOC(=O)CCCCCC(=O)C#N	0.7747484	CCCC(O)OCC	0.2512179
0.0223698	CCCC(O)OCC	0.1391272	CCCCC(=O)OC	0.01078167	CC1=C(C=CC=C1O)N	0.1069219	COC(=O)CC(O)C	0.1504875	CC1=CC(=CC=C1)CO	0.3917814	CC(CC=C)CO	0.1274556	CC(=CCCC(=CCO)CCC)CCC	0.7747882	CCNC(C)OC=C	0.255045
0.0223698	CCC(=O)NCC	0.1435372	CC(C)C(C)COC(=O)C	0.01155179	CC1=CC(=C(C=C1)C)O	0.1087686	CCCS(=O)OC	0.1531444	CCOCC1=CC=CC=C1	0.399528	CCC(CCC)O	0.1279088	CCC1=CC=C(C=C1)OC(=O)C(=O)C	0.7802926	CCC(=O)NCC(=O)C	0.2887675
0.02341839	CCCC(=O)NCC	0.1476275	CC(C)CCC(=O)OC	0.01155179	CC1=C(C(=CC=C1)O)C	0.1101844	CCC(=O)C(O)OC	0.1567203	COC(=O)N1C=CC=C1	0.4010939	CC(C)(CC#C)O	0.1294606	CC(=O)OC1=NC(=CC=C1)OC(=O)C	0.7820332	CC(CCCO)C=C	0.289174
0.02376791	CCC(O)OCC	0.1478191	CCCCCOC(=O)C=C	0.01232191	C=C(C1=CC=CC=C1)O	0.1105537	CCCC(=O)N(C)O	0.1589646	COC(c1ccccc1)O	0.4035766	CC(CCO)C=C	0.1341464	C1C(COC(O1)CC2=CC=CC=C2)O	0.790118	CC(C)(CCCOC)O	0.2891912
0.02499126	CCNC(=O)OCC	0.1489972	CCCCCOC(=O)CC	0.01232191	CC1=C(C=CC=C1S)O	0.1107999	CC(CC(=O)C)O	0.1612506	C1=CC=C(C=C1)C2=CC=NO2	0.4060794	CC=C=CCCCO	0.1372657	CC(CCC=C(C)C)CCOCC#N	0.799029	CCOC(=O)C(C)OC	0.294917
0.02603985	CC#CC(=O)NC	0.1490445	CC(C)COC#N	0.01309203	C1=CC=C(C=C1)NN=O	0.1138777	CCC(=O)CC	0.1613363	c1(ccccc1)COC	0.4097667	CC(C)C(C#C)O	0.1429075	CCCCC#CCCCC1CO1	0.8023956	CCOC(C)OC(=C)O	0.297801
0.02638937	CC(CN)NC(=O)C	0.1502223	CCCCCCC(=O)OC	0.01309203	CC1=C(C=C(C=C1)N)O	0.1143086	CC(N=C=O)OC	0.1654712	COC1=NN=CC2=CC=CC=C21	0.4106803	CCC#CCC(C)O	0.1430763	CC(OCCC(CC/C=C(C)\C)C)=O	0.8068492	CCCNC(C)C=O	0.3026955
0.0267389	CCNC(=O)C	0.1502272	CCCCCCSSC	0.01386215	C1=CC(=CC(=C1)O)O	0.1145548	OC(C)C(=O)CC	0.165828	COC(=O)C1=CC=CC=C1C#N	0.4139384	CCC(=C)C(C)O	0.1438052	CC(=O)OCC1=CC=C(C=C1)SC	0.8074333	CC(C(C)OCC=C)O	0.3104233
0.02743796	CCC(=O)NC	0.1678103	CCCCC(C)CC(=O)C	0.01386215	C1=CC=C(C=C1)C=CO	0.1148626	CC(O)S(=O)C	0.1659486	CC1=NOCC2=CC=CC=C12	0.4173282	CCC(C)(CC=C)O	0.1477914	COC1=CC=C(C=C1)OCC#C	0.8147536	CC=CC=CCO	0.312425
0.02743796	CCOC(C)ON	0.1752449	CCCCCC(=O)C=C	0.01386215	CC1=C(C(=CC=C1)O)S	0.1148626	COCC(=O)C	0.1665356	COC1=NC=C(C=C1)C=C	0.419944	CCCCC(=C)CO	0.1527974	CC(=O)OC1=CC(=CC=C1)OC(=O)C	0.8162294	OCCCCC(=O)C	0.3180066
0.02761272	CC(C)OC(C)C#N	0.1764415	CCCCCCC(=O)CO	0.01386215	C1=CC=C(C=C1)NO	0.1188637	CCN(C)N=O	0.1718486	CC(=CCOC)C	0.4208605	CCC(=CC)CO	0.1538365	CC(=O)OCC(C)(C)C1=CC=CC=C1	0.8169228	CCCCC(=C)CO	0.3214983
0.02778749	CC(=O)C=CCC	0.1765129	CC(C)OCC#N	0.01386215	C1=CC(=CC=C1O)O	0.1207719	CC(=O)CCOC	0.1721226	CCC=CCC(=O)C	0.4243553	CC(CC=C)O	0.1609561	CC(=O)OC(C)(C)CC1=CC=CC=C1	0.8169847	CCC(=O)NCC(=C)C	0.331915
0.02778749	CC(CNC(=O)C)O	0.1766655	CCCCCC(=O)C=C	0.01386215	CC1=C(C(=CC=C1)C)O	0.1208335	CCC#CC(=O)C	0.1737079	CC1=CC2=CC=CC=C2CO1	0.4298624	CCC(C(C)C)O	0.1618278	CCOC(=O)C1(CCC(=O)CC1)C	0.821562	CCOC(C)C(O)OC	0.3407239
0.02796225	CC(C)C=CC(=O)C	0.1781594	CC(=O)OCC(C)(C)C	0.01463227	CC1=C(C=C(C=C1)S)O	0.1270506	CCS(=O)OC	0.1760367	COC1=NN=C2N1C3=CC=CC=C3O2	0.4389819	CCCC(C)CO	0.1653694	CC(=O)OCC(C)(C1=CC=CC=C1)O	0.8270463	CCCC(C)CN=N	0.3447383
0.02848654	CC=CC(=O)C	0.181488	CCCCSS(=O)C	0.01463227	CC1=C(C=CC(=C1)N)O	0.1271737	CCC(=O)OOC	0.1770918	CC1=CC=C(C=C1)C(=O)OC	0.4404134	CCC(C)CCO	0.1675532	C1C(O1)CCOC2=CC=C(C=C2)C#N	0.8281613	CCCCNC(=O)C	0.346592
0.02848654	CCOC(C)O	0.1828469	CCCCCSSC	0.01540239	c1(c(cccc1)O)O	0.1332061	CS(=O)C	0.1832688	C1=CC(=O)C=CC2=C1C=CO2	0.4413276	CC(CCC=C)O	0.1707291	COC1=CC=CC=C1OCC#CCO	0.8288478	CC#CCCCO	0.3473545
0.02848654	CCOC(C)OC#N	0.1867745	CCCCCC(=O)OC=C	0.01540239	CC1=C(C=CC(=C1)O)N	0.1344988	CC(=O)OCOC	0.18357	CC1=CC=CC=C1CO	0.4426713	CCCCC(C#C)O	0.1736887	CC(=O)OCCC1=CC=C(C=C1)OC	0.8296294	CC(CS)C(=O)NCS	0.348219
0.02848654	CCNC(=O)NCC	0.1871594	CCCCCC#CC#N	0.01540239	C=CC1=CC=C(C=C1)O	0.1346835	CN(N=O)OC	0.1835782	CCOC1=NC(=CC(=N1)C)C	0.4432699	C(CC=CC)O	0.1742751	CC1=CC=C(C=C1)COC(=O)C	0.8318492	CCOC(=O)C(=C=C)C	0.3510932
0.02883607	CCOC(=N)OCC	0.1877558	CCCC(C)COC(=O)S	0.01540239	C1=CC=C(C=C1)SO	0.1365301	CC(CCC(=O)C)O	0.1882776	COC1=NC=CC2=C1C=CO2	0.4440383	C=CC(C#CC#C)O	0.1777934	COC1=CC=C(C=C1)OCCSC	0.8329685	CCC=COCC	0.3549383
0.0291856	CC#CCOC(=O)C	0.1890177	CCCCCNC(=O)OC	0.01617251	CC1=CC=C(C=C1)CO	0.1371457	OC(C(=O)C)CC	0.1900073	C1=CC2=C(C=C1O2)C=O	0.44774	C#CCCCCCO	0.1837031	CC(=O)CC(=O)N1CCCCCC1	0.8330911	CCC(C(=O)OCC)O	0.3581843
0.02970989	CCCNC(=O)C	0.1899005	CCCCCC(=O)COC	0.01617251	CC1=C(C=CC(=C1)O)S	0.1385614	CCCN(C)N=O	0.1946672	COC(=O)C1=CC=CC=C1C=O	0.4490445	CCC(CC)(C#C)O	0.1847531	COC1=CC=C(C=C1)NC(=S)OC	0.8346204	CC(COC(C)C#C)O	0.3585553
0.03005942	CCSC(=O)C(=O)C	0.1899048	CCCCOCC(=O)C	0.01694263	CC1=C(C(=CC=C1)N)O	0.1388077	CN(C)N=O	0.196491	COC1=CC=CC(=N1)CC=C	0.4506459	C=CCCC#CCO	0.1968123	CC(=O)ONCC1=CC=C(C=C1)OC	0.8363373	CCNC(C)N(C)C=O	0.3618774
0.03005942	CCC(C)(C(=C)C)O	0.1922056	CCCCCC(C#N)S	0.01694263	CC1=C(C=CC=C1O)O	0.1390539	COCC(=O)OC	0.1971561	CC1=CN=C(C=C1)OC	0.4523599	CCC#CC(CC)O	0.1990558	C1=CC=C(C=C1)COCOCO	0.8375758	CC(CCCO)C#N	0.3658259
0.03023418	CC(=C)OC(=O)C	0.1940202	CCCCCC#CCS	0.01694263	C1=CC=C(C(=C1)O)Br	0.1426857	CCN(CC)N=O	0.1982886	C1=CC=C2C(=C1)C=CC2=O	0.4541816	CCCC#CCO	0.199963	CC1=CC=C(N1CCCNC(=O)C)C	0.8380146	CCCC(=O)NCC	0.3682986
0.0316323	CCCN(C(C)O)O	0.1952987	CCC(C)CCOC(=O)C	0.01694263	C1=C2C1=CC(=C2)O	0.1458866	OC(C(=O)C)C	0.1999441	CCC1=CC(=O)C=CC1=O	0.4606989	CCCC(C)(C)O	0.2011203	CC1=CC=C(C=C1)C(C)(C)OC(=O)C	0.8389156	CC(=NNCC=C)C	0.3703195
0.0316323	CCC(C(C)O)O	0.1989091	CC(C)(C)SN=C=O	0.01694263	CC1=CC(=CC(=C1)O)N	0.1459481	CC(=N)CC(=O)C	0.2026787	C1=COC2=CC(=NC=C21)C=O	0.4649329	CC(C)CCO	0.2019134	CC(=CCCC(C)C(=O)COC(=O)C)C	0.8435958	CC(=C=CCC=C)CO	0.3764713
0.03198182	CCOCOC(=O)C	0.1994939	CC(CCOC(=O)C)CC	0.01694263	CC1=C(C=CC(=C1)S)O	0.1460097	CC(=C=O)C	0.2037051	C1=CC2=C(C(=C1)C=O)OC=C2	0.467186	C=CCC(C=C)O	0.206483	CC(=O)C1=CC=C(C=C1)OC(=O)C	0.8473312	CC(C)OCC(=N)OC	0.3765162
0.03250612	CCC(COC(=O)C)N	0.2023047	CC(C)C(C)CC(=O)OC	0.01771275	CN(C1=CC=CC=C1)O	0.1469946	CCOC(=O)OC	0.2055868	CC1=CC(=O)C=C(C1=O)C	0.4685768	CC(=C)CCO	0.2086812	C#CCC#CCCOC1CCCCO1	0.8497474	CCCC1C(O1)CO	0.3860526
0.03390423	CCOC(=O)C(C)N	0.2026224	CCCC#CC(=O)C	0.01771275	CC1=C(C=CC(=C1)O)Cl	0.1471792	CCC(=O)N(CC)O	0.2100631	C1=CC(=CC2=C1C=C2)C=O	0.4692471	CCC#CCCO	0.2115267	CCC=C=NC1=CC=C(C=C1)OC	0.8523898	CC(C)C(O)OCC=C	0.3922983
0.03390423	CC(CO)OC(=O)C	0.2029934	CCCCCOC(=O)NC	0.01771275	CC1=C(C(=CC=C1)O)N	0.1492721	CCS(=O)C	0.2101009	CN1CC2=CC=CC=C2CO1	0.4703566	CC=CC=CCO	0.2242905	CC1=CC=C(C=C1)OCCSC#N	0.8547133	CCOC(C)C(=O)OC	0.3930371
0.03425376	CCC(COC(=O)C)O	0.2044477	CCCCOOSC	0.01771275	C1=CC(=CC=C1CS)O	0.1495183	CC#CC(=O)C	0.2128712	COC1=NC2=CC=CC=C2O1	0.4739462	CC=CCO	0.2247846	COC1=CC=C(C=C1)COC(=O)C=O	0.8629236	CCC(COC(=O)C)O	0.3950602
0.03425376	CC(=C)CC(C)(C)O	0.2047883	CCCC(C)CC(=O)OC	0.01848286	C1=CC=C(C=C1)C=S=O	0.1527808	CC(=O)OOOC	0.2131322	COC1=NC=NC2=CC=CC=C21	0.4774684	CC(C=CC)O	0.2253603	CC1=CC=C(C=C1)CC(C)NC(=O)C	0.8636131	CC(O)OC(=O)CCS	0.3964027
0.03477805	CCNC(=O)C(=N)C	0.2060244	CCCCC#CCS	0.01848286	C1=CC=C(C=C1)C(=O)S	0.1532117	CCC(=N)CC(=O)C	0.2184117	CCCC(=N)OCC	0.4790589	C=CCC#CCO	0.2257829	COC1=CC=C(C=C1)OCCC#N	0.8654707	CC(CCCCO)C=C	0.3984752
0.03477805	CC(=C)NC(=O)C	0.2071807	CC(=O)CCCC	0.01848286	CC(C1=CC=CC=C1)O	0.1534579	CCC(=O)C#CC	0.2217868	CC1=CC(=NC(=C1)C)C(=O)C	0.4794396	CCCC(C)(C#C)O	0.2296598	CC(COC(C)C1=CC=CC=C1)O	0.8672233	C(CCCCCC)O	0.3999326
0.03495281	CC(CC(C)O)O	0.2081099	CC(C)(C)COC(=O)C=C	0.01925298	c1(ccccc1)C(C)O	0.1534579	CC(C)S(=O)O	0.222778	CC1=CC=C(C=C1)CO	0.482027	CC(=C)C(CC=C)O	0.231137	CC(=O)C1=CC=C(O1)CC2=CC=CO2	0.8738282	CCCC=CCCO	0.4054452
0.03530234	CC(O)OC(=C)C	0.2091461	CC(C)CCC(=O)C=O	0.01925298	C1=CC(=C(C(=C1)O)S)N	0.1554277	CC(CN(C)N=O)O	0.2300447	C=CC(=O)OC1=CC=CC=N1	0.4822777	CC(C#C)C(C)O	0.2339135	CC(=O)OCCC1=CC=C(C=C1)O	0.8739103	CC(=O)C(CCCS)O	0.4057124
0.03547711	CCCC=CC(=O)C	0.2091766	CC(C)CCOCC#N	0.01925298	C1=CC(=C(C=C1O)N)S	0.155797	CCSS(=O)C	0.2302165	CCC1=CC(=NC=N1)OCC	0.4845314	CC(=CCCO)C	0.2346993	CC1=C(C(=C(C=C1)OCC2CO2)C)C	0.8746223	CCOC(C)NC=O	0.4132463
0.0360014	CCCCNC(=O)C	0.209439	CCCCCCNC(=O)S	0.01925298	CC1=CC(=C(C=C1)N)O	0.1559817	CCCS(=O)C	0.231824	CCOC1=NC=C(C=N1)C	0.4858593	CC(CCCO)C	0.2394392	CC(=O)CCC#CCCCC#N	0.8790764	C=CCCCCO	0.4229389
0.03617616	CCNC(=O)SCC	0.2120554	CC(C)CCCC(=O)OC	0.0200231	c1(ccccc1)COC	0.1562894	CC(C)C(=O)OCO	0.233071	C1=CC2=C(C=C2)C(=C1)C=O	0.4859922	CC#CC(C#C)O	0.2405724	C1=CC(=CC=C1C#N)C(=O)CN=C=O	0.8795918	CC(C)C(CCOC)O	0.4252749
0.03617616	CCOC(=O)CC(C)N	0.2124566	CCCCCCC#N	0.0200231	C1=CC=C(C=C1)CON	0.1570281	CCC(=O)N(C)CO	0.2344225	CC(=O)OCCC#C	0.4874734	CC(C)(C#CC#C)O	0.2431499	COC1=CC=C(C=C1)CCC(=O)S	0.8799519	CCCC#CCCO	0.425384
0.03652569	CC(C)CNC(=O)C	0.2128161	CCCCCC(=O)C(=O)C	0.0200231	C1=CC(=C(C=C1N)O)N	0.161337	CC(C#CC(=O)C)O	0.2347912	C1C2=CC=CC=C2C=NO1	0.4881672	CC(C)C(CC#C)O	0.2442776	CC1=CC=CC=C1OCCOC(=O)C	0.881135	CC(C)C(=O)NCC#C	0.4253995
0.03670045	CCC(=O)NSC	0.215797	CCCCS(=O)SC	0.0200231	CC1=CC(=C(C=C1)O)C	0.1642301	CCN(N=O)OCC	0.2371318	CC1=C(C=CC(=C1)C=O)C=O	0.4887159	CC(C)C(C)(C=C)O	0.2447832	COC1=CC=C(C=C1)OCC#N	0.883729	CCOC(=O)CC=C	0.4304131
0.03704998	CCC(=C)C(=O)C	0.2166584	COC(=O)CCCC(C)C	0.0200231	C1=CC(=C(C=C1N)N)O	0.1653381	CC(C)N=C=O	0.2412284	CCOC(=O)C1=NC(=CC=C1)C	0.489814	CC(=C)C(C)(C#C)O	0.2459762	C1C(O1)CCC2=CC=C(C=C2)C#N	0.88543	CC#CCCCCO	0.4357529
0.03704998	CC(=C=C)C(=O)C	0.2210575	CCC(C)CC(=O)OC	0.02079322	C1=CC=C(C=C1)SN=O	0.1675541	CSS(=O)C	0.244664	C=CC(=O)C1=COC=C1	0.4920158	CCC(C)C(C)(C)O	0.2476242	CC(C=CC=CC#CC#CC#C)O	0.8854347	CC(=CCNC=O)C	0.4373818
0.03722475	CCOC(=O)C(C)S	0.2220287	CCSSSCC	0.02079322	C1=CC(=C(C(=C1)O)N)N	0.1698932	CC(=O)C(CS)NC	0.2448244	C1=CC=C2C(=C1)N=CO2	0.4925286	CC(=C)CC(C=C)O	0.2489664	CC1=CC=C(C=C1)CC(=O)C(=O)OC	0.8894077	CC(CCCC=C)CO	0.4391517
0.03757427	CCC(N)(N)OCC	0.2294489	CCCCCCC#CC#N	0.02079322	CC1=C(C=CC=N1)O	0.1716168	CC(C(=O)OC)O	0.24932	CC(=O)C1=C(N=CC=C1)SC	0.4927137	CCC=CCO	0.249122	CC#CC(C)(C1=CC=C(C=C1)OC)O	0.889933	CC(C)SCC(=O)NC	0.4394719
0.03774904	CCCC(NC)O	0.2310014	COC(=O)CC(C)CC	0.02079322	C=C1C=CC(=O)C=C1	0.1717399	CCC(C)N=C=O	0.2504503	CCC1=C(OC(=N1)CC)C	0.4940735	CCC(C)(C#C)O	0.2491628	CC1=CC=C(C=C1)OCC(=O)C	0.892559	CCC(C)COC=C	0.4396183
0.03809857	CCC(=O)C=CC	0.235215	CCC(C)SCCS	0.02156334	CC1=CC=CC=C1S	0.1721092	CCOS(=O)OC	0.2534911	CN1C(O1)C2=CC=CC=C2	0.4940845	CC(CCC(=C)C)O	0.2493886	CC(=O)CC(C)(C)OC1=CC=CC=C1	0.8925743	CC(COCOC)O	0.4410392
0.03809857	CCNC(=O)C(C)N	0.2359318	CCCSCCS	0.02156334	C1=CC(=CC=C1NN)O	0.1738328	CC(=O)CCCC	0.2554301	CCOC(=O)CC(=C)C	0.4957865	CC=CC=CCCO	0.2502238	C1C(O1)CC2=CC(=CC=C2)CC3CO3	0.8928358	CCOC(=O)C(C)C=C	0.4465477
0.03862286	CCSCNC(=O)C	0.2425486	CC(C)CSSC	0.02156334	C1=CC=C(C=C1)C=NO	0.1743252	CC(C)N=O	0.2618512	C1=CC=C(C(=C1)C=O)C#N	0.4969208	CC(C(C)O)CC	0.2507298	CCC1=CC=C(C=C1)OC(=O)C	0.8931491	CC#CCCNC	0.451089
0.03914715	CC(C)C(=O)C(=C)C	0.2443108	CC(C)OC(=O)CS	0.02156334	C1=CC(=C(C(=C1)S)N)O	0.1769721	CC(C)SN=O	0.263618	CC(=C)CCCC(=O)C	0.4978114	CC(CCC=C=C)O	0.2515521	CC#CC#CC#CC=CC(=O)OC	0.8972107	CC(C)(CCOO)OC	0.4572272
0.03932192	CC(=C=C)C(C)(C)O	0.244501	CC=CCOC(=O)C	0.02156334	CC1=CC=CC=C1N	0.177957	CC(=C=C=O)OC	0.2657267	C=CCC1=CC=CO1	0.4990511	CCC(C#CC)O	0.251604	COC1=CC=C(C=C1)CN2CO2	0.8972258	CC(CCO)OCO	0.4592174
0.03932192	CC(CS)OC(=O)C	0.2462187	CC(C)CCCC#N	0.02156334	C1=C(C=C(C=C1N)O)N	0.1784494	CCC(C)ON=O	0.2662234	C1=COC(=C1)C2=CC=NC=C2	0.4990677	CCC(C)(CC#C)O	0.2525218	COC1=CC=C(C=C1)COCC#C	0.8992848	CC(CCO)(CC=C)O	0.4611569
0.03967144	CCNC(=O)C(=O)C	0.2468909	COCC#CSC	0.02156334	C1=CC=C(C=C1)OS	0.1797421	CCN(C(=O)C)O	0.2684985	CC1=C(C=NO1)C2=CC=CC=C2	0.499121	CC(C)(CCC#C)O	0.2535826	CC1=CC=CC=C1C(C)OCC2CO2	0.8996884	CCCCCCN	0.4650146
0.04019574	CCCC(=N)OCC	0.2473864	CC1C(S1)CC2CS2	0.02156334	CC1=C(C=C(C=C1)O)O	0.182943	OC(C)C(=O)CCC	0.2689118	C1=CN=C2C(=C1C=O)C=CS2	0.499949	CC(CCN(C)C)O	0.2557806	CC(=O)CC1=CC=C(C=C1)C(=O)C	0.9005653	CCOC(CC(=O)C)O	0.4650966
0.04054526	C(C)OC(=O)C(=C)C	0.2489494	CC(C)CCOC(=O)CS	0.02233346	C1=CC(=C(C=C1S)S)O	0.1843587	CCC(=O)CC(C)O	0.2692085	C=CC1=CC(=O)C=CC1=O	0.5020806	CC(C)(CO)C=C=C	0.257962	CC1=C(N=C(C(=N1)C)OOC(=O)C)C	0.901486	CCCC(CCO)O	0.4665137
0.04054526	CCC(OC)ON	0.2500234	CCSCCCS	0.02233346	C(C)c1ccc(cc1)O	0.1845434	CCC(=C=O)C	0.2708551	C=CCCC(=O)CC=C	0.50258	C=CC(CCCC)O	0.2596903	CC(=O)NC1=CC=C(C=C1)CC#N	0.903231	CC(CC(=O)OC)CO	0.4697039
0.04072003	CCCC(=O)SCC	0.2507639	CC(C)CSCSC	0.02233346	CC1=C(C=C(C=C1)O)N	0.1847281	CC(C#N)NC(=O)C	0.2735397	C1=CC2=CC=NN2C(=C1)C=O	0.5027015	CCC(CC)(CC)O	0.2623307	COC1=CC=C(C=C1)CC(=O)C=O	0.9034269	CCCCOCO	0.4731765
0.04072003	CCC(OCC)OO	0.2547335	CCC(=O)CCSC	0.02233346	CC1=C(C=C(C=C1)O)S	0.1865132	CN(COC)N=O	0.274083	C=CCCOCC#C	0.5054086	CCC(=C=CCO)C	0.2654587	CC(=CCCC(=C)CCC(=O)C)C	0.9044843	C=CCC#CCO	0.4738838
0.04124432	CCSC(=O)C	0.2549256	CC(C)(C)CN=C=S	0.02233346	CC1=CC(=CC=C1)S	0.1865747	CCCC(=C=O)C	0.2748403	CC(=C)CCC(=O)C	0.5067956	C=CCC(=C)CO	0.2676966	C1OC2=C(O1)C=C(C=C2)OCC#N	0.9049547	CC(CCCOC)O	0.4757819
0.04124432	CCCSC(=O)C	0.2549442	CCSCCO	0.02233346	C1=CC=C(C=C1)C(=O)N	0.1869441	CN(C)C=C=O	0.2761693	CC1=CC2=C(O1)C(=NC=C2)C	0.5068168	CC(C)C(C=C)O	0.2704222	CC(=O)NCCC1=CC=C(C=C1)OC	0.906019	CC#CC(CCO)O	0.4760775
0.04124432	CCNOC(C)O	0.2561354	CCCCCSCCS	0.02233346	C1=CC(=C(C=C1Cl)O)Cl	0.1876212	CC(=N)CC(=O)OC	0.2778836	C=C(CO)C1=CC=CC=C1	0.5073763	CCC(CC(=C)C)O	0.2704573	CCC1=C(C(=C(C=C1)OC(=O)C)C)C	0.9081612	CC(CC(C)N)COC	0.4766209
0.04176861	CCC(CO)C(=O)C	0.2564131	CCSCCCN=C=O	0.02233346	CC1=C(C=C(C=C1)O)C	0.1879905	CC(OC)S(=O)O	0.2792761	C1=CC2=COC=C2C=C1	0.5089957	CC(C(=C)C=C)O	0.2708031	CCC1(CCC(=O)CC1)C(=O)OCC	0.910161	C=CCCC#CCO	0.4767262
0.0422929	CC=CCCC(=O)C	0.2566022	CSSSCSC	0.02233346	C1=CC(=CN=C1)O	0.1894063	CCCN(CC)N=O	0.282204	CCC1=C(C=NC=C1)C(=O)C	0.509995	CC(C)N(C)CCO	0.2734803	COC(=O)C1=CC=C(C=C1)C(=O)C#N	0.9110655	CC(CCO)(CC#N)O	0.4786804
0.04246767	CCC(=N)OCC	0.2569186	OC(C)C(=O)CCCCC	0.02233346	C1=CC(=C(C=C1O)O)N	0.1905143	CCC(C)SN=O	0.2827767	C1=CC2=NC=CN2C(=C1)C=O	0.5136342	CCN(C)CCO	0.2759709	C#CC(=O)CCCCC(=O)C#C	0.9119213	CC(CCCN)C#N	0.4790261
0.04404055	CCC(C=CC)O	0.2573771	CCCCCSSCC	0.02310358	C1=CC=C(C=C1)N=S=O	0.1909452	CCOC(=O)N(C)O	0.2835829	C=CC1=CN=C2C(=C1)C=CO2	0.514442	CC(C=C)C(C)O	0.2770141	CC(=CCCC(=C=CCC(=O)C)C)C	0.9126302	CC(NCC=O)OC	0.47946
0.04439007	CCOC(=O)C=C(C)N	0.2575125	CCCCOC(=O)NC	0.02310358	C1=CC(=C(C=C1O)N)O	0.1910683	CCS(=O)CC	0.2851874	CCOC1=NN=CC2=CC=CC=C21	0.5157621	CCC(C)(N(C)C)O	0.2834152	CC(=CCCC(=CC(OC)OC)C)C	0.9127961	C=CCCCCCO	0.4882038
0.04439007	C(C)OC(=O)OCC	0.2579565	CCCCCCOC#N	0.02310358	CC1=CC=C(C=C1)N	0.1918685	CC#CCC(=O)C	0.2859649	CCCC(=C)OC	0.5160545	CC(C)C(C=C=C)O	0.2836092	CC(=O)C1=CC=C(C=C1)C(=O)OCO	0.9129944	CC(C)CN=NC=C	0.4906354
0.04439007	CC(CCOC(=O)C)O	0.2589508	CC(CN=C=S)C(=O)C	0.02310358	C1=CC(=C(C(=C1)O)N)O	0.1935305	CC(C)CN=C=O	0.2863445	C1=CN2C=CC1=C2C=O	0.5167967	CC=CC(C#C)O	0.2839289	C1=CC=C(C=C1)C(=O)OCC(=O)OO	0.9130039	CC(=O)C(=C)COCO	0.4915733
0.04456484	CC(=O)NOCOC	0.2591396	CCCC(C)CCCSS	0.02310358	CC1C=CC=CN1O	0.194577	COS(=O)OC	0.2874234	CC1=C(OC=N1)C2=CC=CC=C2	0.5178947	C=C(CO)CC	0.2844362	C1=CC(=CC=C1C#N)OCCC#N	0.9145581	CCCONC(=O)C	0.4945904
0.04631248	CCC(=O)SCC	0.2627113	CCCC(C)OC(=S)NC	0.02310358	c1(ccccc1)C#N	0.1952541	CCOCS(=O)C	0.2924092	C=CC1=NC(=CC=C1)CO	0.517945	CCC(C=CC)O	0.2847231	CC(=O)C(=O)OCCC1=CC=CC=C1	0.9212348	CC(CCCCO)O	0.4963142
0.04631248	CCCC(N)OCC	0.265066	CC(C)C(=O)CCSC	0.02310358	C1=CC=C(C=C1)SC=O	0.1954387	CN(C)CC=C=O	0.2929732	CC#CCCOC	0.5183269	CC(=C)C(=C)CO	0.2860469	C1=CC=C(C=C1)CCC(=O)N=C=O	0.9241146	CCCNC(=O)CC	0.4963989
0.04666201	CC(C)CSC(=O)C	0.2657469	CC(C)(C)CCC(=O)C=O	0.02310358	C1=C(C=C(C=C1O)O)N	0.1963621	COC(=O)OOC	0.2931829	CC(=O)CCCC#C	0.5185916	CC(CCC#C)O	0.2872044	CC(=O)OOC(=O)C1=CC=CC=C1	0.9250149	CCCCC(=O)CN	0.4997457
0.04701153	CC(C)(O)OC=C	0.2672213	CCCCCSCC#C	0.0238737	C1=CC(=C(C=C1S)N)O	0.1982703	CC(=O)SSC	0.2933554	CCC1=CC(=NC=C1)C(=O)C	0.5187441	C(C=CC=CCC)O	0.2873259	CC1=CC=C(C=C1)CN2CC2CC#N	0.9252715	CCCCN(C)C(=N)C	0.5010305
0.0471863	CCC(=C)OC(=O)C	0.2674708	CC(C)CSC(=S)C	0.0238737	CC1=CC(=CC=C1)N	0.1999323	CCCCS(=O)C	0.2936078	CC(=CCCCO)C	0.5192243	CCCC#CCCO	0.2893009	CC1=CC=C(C=C1)CNC(=O)C	0.9274282	CCNC(=O)CC(=O)C	0.501056
0.04771059	CCCC(=O)OSC	0.2675318	CCCCC(C)CN=C=S	0.0238737	CC1=CC=C(C=C1)S	0.1999938	CC(=O)CC(=N)OC	0.295303	CCC(=O)N1C=CC=C1	0.5220931	CCCC(N(C)C)O	0.291846	C1=CC=C(C=C1)C(=O)OOC(=O)CO	0.9299503	CC(COC=C)NC	0.5013229
0.04806012	CC(C#CC(C)O)O	0.2678014	CCCCC(=O)C(=O)C	0.0238737	C1=CC=C(C=C1)CS	0.2007941	COS(=O)C	0.2970129	CCCC1=CC=CCO1	0.5242941	CC(C=C)C(C)(C)O	0.2941807	CC1=C(C=C(C=C1)OOC(=O)C)C	0.9311288	CCC(C)CON=C	0.5052114
0.04823488	CC(C)COC(C)O	0.268703	CCCSSCCC	0.0238737	C1=CC(=C(C(=C1)N)O)N	0.2031332	CCC(=C=O)CC	0.3067286	CC(=O)C=CCC	0.5251579	CCC(C)(C=C)O	0.2964623	C1C(=CC2=CC=CC=C2)COC(=O)O1	0.9313203	CCOC(C)C(=O)C=C	0.505701
0.04875918	CCSS(=O)CC	0.2692035	CCOC(=O)CSC	0.0238737	C1=CC(=C(C(=C1)Cl)O)Cl	0.2041181	CCC(CC)N=C=O	0.3072284	C1=CC=C(C(=C1)CC#N)CO	0.525447	CC(=CC#CCO)C	0.2972936	CC1=CC(=C(C(=C1)C)OCC2CO2)C	0.9320992	CCC(C)C(=C)CN=C	0.5058668
0.0491087	CCN(C(=O)C(C)N)O	0.2699215	CC(C)(C)CSC#N	0.02464382	NC1=CC=CC=C1S	0.2059647	CCC(=C)OC(=O)C	0.3078697	C1=CC2=CSC=C2C=CC1=O	0.5254994	CCC(C)(C(=C)C)O	0.2981777	C1C(=O)COC(O1)C2=CC=CC=C2	0.9351419	CN=CCCCCO	0.5063789
0.0491087	CCC(N)OCC	0.270572	CCC(C)OC(=O)CS	0.02464382	C1=CC(=CC=C1CO)N	0.206211	CC(=C)OC(=O)C	0.3120694	CC(CO)(C#N)C1=CC=CC=C1	0.5256401	CCC(CC#CC)O	0.3007535	C1C(O1)C(=O)OCCC2=CC=CC=C2	0.9353837	COC=CCCCO	0.5072374

Table 2.2 Continued

Or85b		Or85f		Or98a	
SMILES	Dist	SMILES	Dist	SMILES	Dist
CCCCC1CCC(=O)O1	0	CCCCCCO	0	CCCCCCC(=O)C	0
CCCCCC=O	0	CCCC=CCO	0	CC(=CCCC(=O)C)C	0
CCCCCC(=O)C	0	CCC=CCCO	0	CCCCCC(C=C)O	0
CC(=CCCC(=O)C)C	0	CCOC(=O)C(C)O	0	CCCCOC(=O)C	0
CCCCCO	0	CCCC#CCO	0.3215251	CCCCCOC(=O)C	0
CCCCCCO	0	CCCC#CCO	0.3977383	CC(C)CCOC(=O)C	0
CCCCCC(C=C)O	0	CCC#CCCO	0.4721775	CCOC(=O)CC(C)O	0
CCCC=CCO	0	CCCC(=O)OC(C)O	0.5291351	CCCCCC(=O)OC	0
CCC=CCCO	0	CCC(=O)CC(O)C	0.5396708	CC(CCCCO)C=C	0
CCC=CCCO	0	CCCCC(COC)O	0.5401751	CCCC(=O)OCNO	0
CCCOC(=O)C	0	CC(CCCC=C)O	0.574608	CC(CC=C=C(C)C)O	0
CCCCOC(=O)C	0	CC(CC(=O)OC=C)O	0.5760439	CNCC(=O)OCCO	0
CCCCCOC(=O)C	0	CC=CCCCO	0.5830891	CC(C)CCOCC(C)O	0.0006135
CC(C)CCOC(=O)C	0	CCCC(=O)N(C)O	0.5926106	CC(=O)OC=CC=C	0.0006135
CCOC(=O)CC(C)O	0	CN(CCCCO)N=O	0.6121783	C=CCCCC(=C)CO	0.0006135
CCCCCC(=O)OC	0	CC(C)(C1C(O1)C#C)O	0.6193232	C=CCOCC(C=C)O	0.0006135
CC(CCCC=C)CO	0.04010449	C(CC=CC)O	0.6211374	C(CN)C(=O)OCCO	0.0006135
CCCCCCC=O	0.0541304	CC(CCC=C=C)O	0.6297574	C=CC(COCC#C)O	0.0006135
CCC(=O)C=CCCC	0.05661388	CCC(C(=O)OCC)O	0.6342998	CCOCCC(C)O	0.0006135
CCCCC=CC(=O)C	0.05802127	CC(=O)OC(COC)O	0.6391873	CC(CC(=O)OCCO)O	0.0006135
C(CCCCCC=C)O	0.06257155	CCCC1C(O1)CO	0.6497865	CCOC(=O)CC(=O)C	0.0006135
CCCCCC(=O)CC	0.06590403	CCOC(=O)CC(C)O	0.6566422	CCOC#CC(C)O	0.00122699
CC(C(=O)C)CCCC	0.0741716	CC(C#COC=C)O	0.663147	CC(C)C(C#CCN(C)C)O	0.00122699
CN(CCCC=C)O	0.08071376	CN(CCCC=C)O	0.6794308	CCC(=O)COCC	0.00122699
CC(CCC=C=C)C=O	0.08460999	C=CC(CCCC)O	0.6991725	CC(=O)CC(=O)OCCO	0.00122699
CCCCC(=C)C=O	0.08656348	CCC(COC=C)O	0.7054714	CC(C)OC(=O)NCO	0.00122699
CCC(CCC#CC)O	0.0872067	CCCCCN(C)O	0.7137121	CCOC(=O)CS(=O)C	0.00122699
CN(CCCCC=C)O	0.08897917	CC(COCCC#N)O	0.7174705	CC(CO)OCCC=C	0.00122699
CCCCC1CCOC1=O	0.09145651	CC(CC)C=C	0.7219598	CCCNOC(=O)CC	0.00122699
C=CCCCCO	0.09536294	CC(C)(C(=O)OC=C)O	0.7231115	CCCC(=O)OC(=O)C	0.00122699
CCCCC=CCO	0.09564121	CC(=O)CC(COC)O	0.7337541	CC(=C)COCCOC	0.00122699
CCC(CCC=C=C)O	0.0958081	CC(CO)OCC=C	0.7500231	CC(=O)OC(=O)CC=C	0.00122699
CC(C)CCCCOC	0.09698922	CCC=CCO	0.7501492	CCOCOC(=O)C	0.00184049
CCCCCC(CC)O	0.0993378	CCCOOCO	0.7568648	CCC(=O)OOCC	0.00184049
CC(=CC)CCC(=O)C	0.1024115	CCCNCCO	0.7622471	C(COC(=O)NCO)O	0.00245399
CCCCC(C)C(C)O	0.1036823	CCC1C(O1)C(CC)O	0.763967	CCCC(=O)OCCO	0.00245399
CCCCC(C)C(C)O	0.1036823	CCCC(=O)CCO	0.770124	C=CCC(=O)OCCO	0.00245399
CCCCC#CC(C)O	0.1038561	CCC#CCO	0.7729418	CC(O)OOCC=C	0.00245399
CC(C)CC=CC(=O)C	0.1081452	CCC#CC(C#C)O	0.7734482	CN(C)CC#CCO	0.00306749
CCC(C)CCC(C)O	0.1082218	CCC(=O)N(CC=C)O	0.7774074	CCCCC(CC=C)O	0.00306749
CC(CCC=C)N(C)O	0.1085992	CCCOCCO	0.7824314	CC(=O)NOC(=O)NC	0.00306749
CC=CCCCO	0.1090384	CC#CCCCO	0.782923	CCCC(=O)OCN	0.00306749
CCCCC(=C)CO	0.1098731	CC(C(=O)N=C=C)O	0.8017735	CC(=O)COC(=O)NC	0.00306749
CCCCCCCOC	0.1110821	CCCCC(=NO)C	0.8075465	CCCCC(=C)CCO	0.00306749
CC(=O)OCCCC=C	0.1115886	CCCC1C(O1)C(C)O	0.8082863	CC(C)OC(=O)NC(=O)N	0.00306749
C#CCCCCO	0.1129282	CCC(=O)OC(C)O	0.810021	CC(COCC=C)O	0.00368098
CCCCC=CCCO	0.1133289	CCC#CCC(C)O	0.810398	CSCCOC(=O)C=C	0.00368098
CC#CCCCCCO	0.1133403	CCC#CC(C)O	0.8129808	CCOC(=O)NC(=O)C	0.00368098
CCC(C)CCCCO	0.1146856	CCC(=O)NC(C)O	0.8171243	CCCOCC(C)O	0.00368098
CC(C)CCCC(C)O	0.1147666	CC(CC#C)O	0.817815	CCCOC(C)CCO	0.00368098
CCC#CCC(C)O	0.1152258	CC(=O)N(C)CO	0.8226054	CC(=O)OCC(CO)O	0.00368098
CCCCCC(=C)OC	0.1152665	CCC(=O)CC(C)O	0.8308937	CCCC#CC(C#C)O	0.00368098
CC(CCCC(=O)C)C	0.1154874	CN(CCO)CC#C	0.8311732	CNC(=O)ON=C(SC)SC	0.00429448
C=CC(=O)CCCC#C	0.1184745	CCOCC(C=C)O	0.83177	CC(CCOC(=O)C)C=C	0.00429448
CCCCCCOCC	0.1184982	CCOC(=O)C(C)C(C)O	0.8326868	CC(COCOC)O	0.00429448
CCCCCC(C)C=O	0.1189392	CCOC(=O)N(C)O	0.8348134	CCCOC(CO)C=C	0.00429448
CC(C)CCCCO	0.1214928	CC(C)OC(=O)NCO	0.8349976	CCONC(=O)C=C	0.00429448
CCCCC(C)C(C)C=O	0.1218392	CC(CCCC)O	0.8381515	CC(COCC(C)O)N	0.00429448
CC(C)C#CCCO	0.1219294	CCCCCNO	0.845239	CC(=O)CCOCC=C	0.00429448
CC(CO)CCCCC	0.1223597	CCC1CC(C1)O	0.8480502	CC(COCC=C)OC	0.00429448
CC(C)CCCC=O	0.122411	CSCCC(=O)CO	0.8486522	CC(=NOCCOC)C	0.00429448
CC(CCCCCO)C	0.1229771	CC(CO)OC(=O)C	0.851108	CCCOC(C)OC=C	0.00429448
CCCCCC(=C)C=O	0.1232192	CC(C)(COC=C)O	0.85224	COC(=C)COCC=C	0.00490798
C(CCCCCC)O	0.1235551	CCC1=NC(CS1)O	0.8540531	CCOC#CC(C)C)O	0.00490798
C(CCCCCO)C=C	0.1244304	O1C(C1CC)CCO	0.8621313	CC(=O)C#CCOC	0.00490798
C(=O)CC=CCCCC	0.1258339	CC(=O)OC(C=C)O	0.8631899	CCC(COC(=O)C)N	0.00490798
C#CCCCCCCO	0.1260404	CNCCC(=C)O	0.8777239	CCCOCC(C)(C)O	0.00490798
CCC(C)CCOC(C)O	0.1260955	CCOC(=CCO)C	0.8779936	CC(N)NC(=O)C1=CC=CO1	0.00490798
CC(=O)CC(=C)C#C	0.1262294	CC#CCCO	0.8786813	CCOC(=O)NCN=O	0.00490798
CCCCC(C)(C)CO	0.1264312	CCNCC(C)O	0.8810491	COC(=O)C#CCN	0.00490798
CC(=CCCCC=O)C	0.1264406	CC1(C(O1)CO)C#C	0.8818402	C=CCOC(=O)OCCO	0.00490798
CCCCCN(O)OC	0.1272744	CCCOCC(C)CO	0.8849416	C=CCCCC(C#C)O	0.00490798
CC(CC(=O)OC=C)O	0.1279515	CCCCC1(CC1)O	0.886465	CCCON(CC)O	0.00490798
CC(CCCC=C)O	0.1286442	CC(C(=O)OC=C)O	0.8893637	CC(=O)COC(=O)OC	0.00490798
CCCC(C)CO	0.1287862	CC(CN(C)N=O)O	0.8991722	CCCC(O)OCC=C	0.00490798
CCCCCC(C#C)O	0.1289348	CCSC(=O)C(C)O	0.9004022	CC(COC)OC=C=C	0.00490798
CC(C)CCC#CCO	0.1293536	CC(=O)OC(C)(O)OC	0.903124	CCC(=O)OCCN	0.00490798
C#CCCCC(C#C)O	0.129455	CCC(COC(=O)C)O	0.914637	CCCOC(CC)CO	0.00490798
C=CCCCCCO	0.1296563	CCNC(=O)N(C)O	0.9176545	CCCOC(CCN)OC	0.00490798
CCCCCCC(N)O	0.1313166	CC#CCCC(C)O	0.923911	CCCOC(=O)CC	0.00490798
CC(=CCCCO)C	0.1315029	CN(C1=CC=CC=C1)O	0.9244734	CCCOCCOC	0.00552147
CCCC(C)C=O	0.1325794	CCC1COC1CO	0.9259636	CCCCC(CC#C)O	0.00552147
CCCCCC(=O)C=C	0.133011	CCSCCCO	0.9267778	CCOC(=O)CC(C)N	0.00552147
CC=CCC=CC(=O)OC	0.1348199	C(CCC#N)CCO	0.930384	CCOC(=O)C#CCO	0.00552147
CCCCCC(NC)O	0.1354037	CN(CCCNO)N=O	0.9316613	CC(C)CCOC(C)CO	0.00552147
CCCCCOC(C)O	0.1379773	CCCCCC(C)O	0.934426	CCOC(=O)NN=C=O	0.00552147
CCCC1C(C(=O)O1)(C)C	0.1382941	CC(CCOC(=O)C)O	0.937425	CCCC(O)OCCC	0.00552147
C#CCCCCCO	0.1385679	CC=C=CCO	0.9401755	CC(=CC(CCCO)O)C	0.00552147
CC(C)(C)CC#CCO	0.1395223	CN(CCC#N)CCO	0.9412293	CCC(=O)OCC=C	0.00552147
C(CCC=CCC)O	0.140536	CCN(C)CCCO	0.9443744	C=CCCCC(C=C)O	0.00613497
CCCCCCOC=O	0.1409025	CSCCNCCO	0.9449142	CC(=O)C=CC(=O)OC	0.00613497
C1CC1CCCCO	0.1411786	CCOC(=O)N(C)CO	0.9451939	CCNC(=O)OCC=C	0.00613497
CCCCC1CO1	0.1412632	CCN=NC(C)O	0.950467	CCCC(C)(C)OCCCO	0.00613497
CCCCCCOC#N	0.1433369	CC(O)OC(=O)OC=C	0.9587006	CCC(=O)OCC(C)O	0.00613497
C=CCCCCCO	0.1439596	CC(CCCN(C)N=O)O	0.9624226	CC(COC(=O)C=C)O	0.00613497
CCCC#CC(C)(C)O	0.1439988	CC(=O)OC(=O)CO	0.9636813	COCCOC(=O)C=C	0.00613497
CCCCNCN=O	0.1442148	CC(C)(CO)OC=C	0.9655692	CCOC(=O)NC(C)C	0.00613497

Table 2.2 Continued

CHAPTER III:

Applying Chemical Informatics to Decode Odor Receptors of Several Important Disease Vector and Pest Insect Species as Well as Mammals

INTRODUCTION

Olfaction is exceedingly important for a broad array of species living in all ecological regions of this planet (van Naters and Carlson, 2006). While a major effort has been made by the scientific community to decode and analyze Olfactory receptors (Ors) in *Drosophila* over the last decade, there remains a vast number of important pest and disease vector species for which very little is known. Very recently, odor receptors of several high priority pest and disease vector species have been the focus of investigation (Carey et al., 2010; Ditzen et al., 2008; Ghaninia et al., 2007; Hill et al., 2009; Stanczyk et al., 2010; Syed and Leal, 2008; Turner et al., 2011). Olfactory receptors (Ors) or Odor Receptor Neurons (ORNs) in these species are being tested against relatively small panels of odors in an attempt to decode their response profiles. While these panels are limited in size and scope, they do generally consist of multiple functional groups and carbon chain lengths and often a number of odors in the panels were selected for their ecological significance. Humanity could greatly benefit from the identification of highly effective, safe, and receptor-specific odors for these species, ideally a number of which will prove to be behavior modifiers. We demonstrate that our chemical informatics pipeline can be applied to predict active odors for 5 important species.

Anopheles gambiae is the principle disease vector for malaria, which is a devastating disease in many parts of the world, including South America, Africa, India,

and Asia (World Health Organization., 2011). Malaria symptoms include vomiting, retinal damage, anemia, joint pain, convulsions and in some cases coma and death. The World Health Organization (WHO) estimated 216 million episodes of malaria in 2010, resulting in over 655,000 deaths (World Health Organization., 2011). While these numbers do represent a reduction in the incidence of malaria from previous years, growing resistances to current treatments and repellents have been identified. It is estimated that over 2 billion dollars were spent on financing malaria control in 2011, with higher predictions for future years if global control targets are to be achieved. To further complicate matters, current preventative medications are too expensive for widespread use in developing countries. Identification of safer odors with better organoleptic properties that activate or inhibit behaviorally important Ors, such as those used for detection of human hosts, could greatly aid in the global fight against vector-borne diseases, such as malaria. Recently, 110 odors were tested against 50 *Anopheles gambiae* Ors using the "empty neuron" expression system, resulting in newly identified activators and inhibitors (Carey et al., 2010). The 110 compound set consisted of an array of functional groups and sizes, providing a wonderful training set for our computational pipeline.

In addition to *Anopheles* Or responses, we are also interested in predicting odors for the CO_2 receptors of several important species. CO_2 is a primary long-range detection cue that mosquitoes use to identify and navigate towards their blood meal (Gillies, 1980; Grant and Oconnell, 1996). By navigating upstream in CO_2 plumes, mosquitoes are able to identify hosts at long ranges with high effectiveness (Carde and Gibson, 2010; Zwiebel and Takken, 2004). It is believed that other important attractive odors play additional roles in more specific host preference once the mosquito is in close

range (Gillies, 1980; Takken, 1996). As the CO_2 receptor is behaviorally very important, it is not surprising that the genes are conserved across mosquito species including *Anopheles gambiae*, *Aedes aegypti*, and *Culex pipiens*, all of which express a CO_2 receptor in the cpA neuron on the antenna. Interestingly, CO_2 detection is even important in *Drosophila*, which are able to detect CO_2 through gustatory receptors in their ab1C ORN (Turner and Ray, 2009). While no large unified panel of odors has been tested on the CO_2 neuron of the previously mentioned 3 species of mosquito and *Drosophila*, a number of smaller panels have been tested across each and a small panel was tested across all 4 (Turner et al., 2011; Turner and Ray, 2009). As the response profiles for the receptor across species appears to be at least moderately conserved, we integrate the responses of all species into a single unified odor response set, which we use to train our prediction pipeline.

The Asian Citrus Psyllid (ACP), *Diaphorina citri* Kuwayama, is posing a major challenge to the global citrus industry. While originally present only in Asia, this insect has recently invaded North America, including both Florida and Southern California (Halbert and Manjunath, 2004)(http://www.californiacitrusthreat.com/). While the insect itself does not cause major damage to citrus trees, the ACP is a vector for the bacterial disease Huanglongbing (HLB), also known as the Citrus Greening Disease, caused by *Candidatus* Liberobacter (Bove, 2006; Dagraca, 1991). If infected, the resulting HLB causes a chronic decline in fruit production. The limited fruit that is produced is often small, misshapen and bitter (Baldwin et al., 2010; Bove, 2006). Unfortunately, there is currently no cure for a tree once it is infected. Several attempts have been made to repel the insect, with limited success, or to super nutrient infected orange trees, which is ultimately not cost effective as a long-term solution (Onagbola et al., 2011). Recently,

our lab has decoded ORNs in the ACP antenna against a panel of 62 odors, several of which are known citrus vegetative volatiles (unpublished data). Each of these odors in the rhinarial plates on each segment of the antenna, identifying ORNs on rhinarial plates 2, 4, 6, and 7. Each rhinarial plate appears to house 3 ORNs, which have been labeled A, B, and C in decreasing order of spike amplitude. While very few ligands for the A neurons have been identified, a number of activators for the B and C neurons in each plate have been classified. If we can identify inexpensive, safe, and effective activators and inhibitors of these ORNs, a number of which will hopefully be attractants and repellents, the citrus industry can apply them as lures for surveillance and repellents to curtail the progressing ACP population, thus stopping the spread of HLB. We use these odors as our training set to predict additional ACP Or activating odors.

Insect and mammalian Ors are functionally distinct from one another. Insect Or proteins are considered to be 7-transmembrane proteins that have a non-traditional inside-out membrane orientation. They are believed to function as ligand-gated ion channels and form a heteromer with an obligate partner Orco, which is required for function (Benton et al., 2006; Sato et al., 2008). Mammalian odor receptors on the other hand are G-protein coupled receptors with a traditional outside-in 7-transmembrane orientation (Zhang and Firestein, 2002). Additionally, Mammals have far larger families of odor receptors (~1000 in mice, ~350 in humans) and thus pose an even greater challenge to examine odor coding (Saito et al., 2009; Saito et al., 2004). These distinctions provide an additional challenge for our ligand discovery pipeline. While our method is highly effective for insect Ors, we now have the opportunity to predict odors for functionally distinct mammalian receptors. 52 mouse and 10 human ORs were recently decoded using panel of 63 odorants. Unlike the previous analyses, these

receptors were decoded using calcium imaging and are reported as concentration required for 50 percent effectiveness (EC50). If successful, predictions for these receptors both demonstrate the breadth of our pipline and provide a tool that would allow the fragrance industry to identify natural and pre-approved odors that activate specific human odor receptors, providing the possibility of truly tailor made fragrances.

In this chapter we apply our cheminformatics platform to predict activators and inhibitors for *Anopheles* Ors, the mosquito CO_2 receptor, Asian Citrus Psyllid ORNs, Mouse ORs, and Human ORs. Computational validations are performed in all instances, and electrophysiological validations are performed for both the *Anopheles* Ors and CO_2 receptor predictions, demonstrating that our chemical informatics approach can be successfully applied in each of the 5 species. A comparative analysis of predicted receptor-odor relationships between *Drosophila* and mammals is also performed, where interesting properties are uncovered. Most importantly, thousands of new odors are identified for over a hundred odor receptors.

RESULTS

Predictions for *Anopheles gambiae* Odor Receptors

Or tuning of previously identified *Anpheles* receptor-odor interactions

The tuning of the 50 previously tested *Anopheles gambiae* Ors is distinctly different from what was observed in *Drosophila* (Allison 2011, Hallem 2006). Activating odors (>50 spikes/sec) for 21 out of 24 *Drosophila* Ors expressed in the antenna were identified. By contrast, activators for only 35 out of 50 were identified for *Anopheles*. For the 15 Ors that remained, an average of 64% of tested odors were inhibitors. This differs sharply from the other Ors of both *Anopheles* and *Drosophila*, where on average only 30% and 28% of tested odors reduced the activity of Ors below spontaneous respectively. Based on the limited training set there are two hypotheses that can explain this distinction. While it is possibly that the true odors for these receptors have not yet been identified, an alternative explanation could be that these receptors are specifically tuned for inhibition. As a result, we have divided the *Anopheles* Ors into 3 different classes, each of which we individually apply the chemical informatics platform to, predicting receptor-odor interactions for the following three classes: Activating (Or57, Or12, Or50 ,Or38 ,Or56 ,Or21 ,Or20 ,Or18 ,Or4 ,Or75 ,Or11 ,Or15 ,Or39 ,Or48, Or46, Or30, Or9, Or13, Or1, Or10, Or6, Or2, and Or8), no strong activators (Or32, Or26, Or53, Or16, Or64, Or31, Or27, Or65, Or59, Or63, and Or61), and Inhibitory receptors (Or35, Or25, Or3, Or76, Or44, Or66, Or43, Or54, OR67, Or42, Or41, Or33, and Or45).

Chemical Informatics can successfully explain *Anopheles* receptor-odor activity

It is the molecular structures and properties of odors that determine their activity against an Or. As a result, we have designed a chemical informatics pipeline to identify which structural features are important for the activity of odors for individual Ors and have applied these features to predict the activity of the Or to a large set of untested chemicals, as we have previously done for *Drosophila* (Figure 3.1). A single 3D structure was calculated for each of the odors using the Omega software package (See Methods)(Hawkins et al., 2010). We calculated molecular descriptors, which are mathematical values that describe the structure and features of a molecule, from the 3D conformations to quantitatively explain the odor structures. Dragon (Talete) and Cerius2 (Accelrys), which are commercially available software suites, was applied to calculate 3,424 molecular descriptors for each compound tested against an *Anopheles* Or. We then applied a Sequential Forward Selection (SFS) approach to identify which of the molecular descriptors were most highly correlated with activity, resulting in a single subset of descriptors that describe the activity of the training odors against a single Or (See Methods) (Whitney, 1971). While descriptors for Ors classified as activating or fishing “no strong activators” were selected for their ability to describe activation, descriptors for Ors classified as inhibitory were selected for their ability to describe inhibition.

As observed in *Drosophila*, our approach was able to successfully identify molecular descriptors for each *Anopheles* Or that were highly correlated with activity (Table 3.1). We then applied each Ors optimized descriptors to cluster the training odors in order to visualize how well these descriptors grouped the activating or inhibitory odors (Figure 3.2). We find that our approach is able to successfully cluster together odors

tested against Ors for all three classes. While we had observed similar results in our *Drosophila* predictions, this was the first application of our approach to structurally classify inhibition. Inhibitory odors for the majority of Ors were successfully clustered together using optimized structural features. As inhibition can happen both allostericly and competitively, these results pose two interesting hypothesies. Firstly, the very tight clustering for all inhibitory odors observed for several of the Ors could signify competitive or allosteric inhibition in a single pocket. Secondly, the broader clustering of inhibitory odors observed for several of the Ors may signify two active sites or simply an imperfect selection of descriptors.

Predicting receptor-odor interactions

We next applied our approach to predict the activity of the Ors to a large untested odor panel. Our panel consisted of two odor libraries (See Methods). One panel was a natural odor library containing 3,197 odors emitted by plants, flowers, insects, and mammals. The second was large library of Pubchem odors containing over 240,000 compounds with similar molecular weight (<200) and atom type compositions to those found in nature. We ranked the activity of each compound from both libraries using each of the receptor-optimized descriptor sets, resulting in individual rankings for each Or (See Methods). We list odors from the natural odor library that were ranked in the top 500 predictions for each Or (Table 3.2)

Validating predicted receptor-odor interactions using single unit electrophysiology

Several inexpensive and easily obtainable odors were selected from the natural odor prediction lists of 12 *Anopheles* Ors for experimental validation. Members of Dr. John Carlson's lab at Yale performed the experimental validations using single unit electrophysiology. Each *Anopheles* Or gene to be validated along with the obligate co-receptor was heterologously expressed in the empty neuron system in *Drosophila* (Carey et al., 2010; Dobritsa et al., 2003). As this was the same group that created the initial training set using the same experimental procedure, the validations are expected to be highly accurate. A total of 129 receptor-odor interactions were tested across 12 Ors. On average 65 percent of predicted odors activated the Ors (>50 spikes/sec) (Table 3.3). While this is lower than the 71% we had observed for our *Drosophila* predictions, it is still far better than the average activation rate of 16% that was observed across the Ors from our training set. In *Drosophila*, we had observed that our computational pipeline was more accurate at predicting odors for Ors that respond primarily to aliphatic rather than aromatic compounds. We again see this trend for *Anopheles*, with primarily aliphatic responsive Ors having a high accuracy of 84%, whereas aromatic responsive Ors have a relatively low accuracy of 38%.

Increasing the number of aromatic odors in the training set

In *Drosophila* we had hypothesized that our low aromatic prediction accuracy could be due to a lack of aromatic odors in the training set (Hallem and Carlson, 2006). In order to test this hypothesis in *Anopheles*, we increased the size of our aromatic training set by testing a broad panel of aromatic odors. Ideally, we wanted to test as

broad a set of aromatic compounds as possible, providing the largest amount of information on what aromatic structures are or are not active to our computational pipeline for training. As testing odors for activity is a time intensive and expensive process, we applied chemical informatics to select the most structurally diverse set of aromatic odors for testing. Beginning with all 306 aromatic compounds available in the laboratory of Dr. John Carlson, we applied all 3,224 Dragon descriptors to cluster the 306 odors, producing a cluster derived from non-optimized and broad features. We used the dendrogram to select the 75 most diverse aromatic odors (Figure 3.3). Interestingly, the tree is divided into two main branches with small single ring aromatics being in the left subcluster and large multi ring aromatics residing in the right. The Carlson lab then tested 37 of these diverse odors on the primarily aromatic responsive Or2, Or6, and Or10 using electrophysiology. As expected, only a small percentage (9%) of this structurally broad array of 37 aromatic odors activated the Ors (Table 3.4). Interestingly, this number is similar to the random activation chance observed both for *Drosophila* (10%) and in *Anopheles* (16%).

Predicting receptor-odor interactions using our expanded aromatic training set

We applied our ligand prediction pipeline for Or2, Or6, and Or10, using the original 109 odors, odors validated during the initial round of testing, and the newly tested 37 aromatic odors for training, and optimizing descriptor sets for each of the three receptors. (Table 3.5). Since only 29 of the 109 odors in the original training set were aromatics, the inclusion of the additional 37 odors more than doubled the number of aromatics. This large expansion in size, combined with being a very broad array of structural features, increased the challenge of descriptor selection for our pipeline, as

descriptors that once completely separated active from inactive aromatics now resulted in imperfect classification. Additionally, only 22% of the original molecular descriptors are present in the newly optimized sets and the mean number of descriptors per Or has increased from 21 to 30 for these three receptors, further indicating inclusion of the new aromatics into the training set has increased the complexity of the system.

When we cluster the training odors using the optimized descriptors we find that the descriptor sets are successful in clustering training odors (Figure 3.4). Nearly all activating odors are clustered tightly together for Or2, Or6, and Or10. In order to further validate the predictive ability for these three receptors we performed a 5-fold cross validation as was performed for *Drosophila* (See Methods) (Hastie et al., 2001; Tan et al., 2006). The training sets for each Or were divided into 5 equal partitions. One partition was withheld as a test set while the remaining 4 were used to optimize descriptors. The predictive ability of the system was assessed by performing a Receiver Operating Characteristic (ROC) analysis on the withheld test set. This process was repeated 5 times where each of the 5 partitions was used as the test set once. The whole process was repeated 5 times to improve accuracy. A single mean ROC curve was plotted for each Or and an Area Under the Curve (AUC) was calculated (Figure Figure 3.5). We find that the average AUC value across the three receptors was 0.832, which is very close to the value of 0.815 that we observed in *Drosophila*, suggesting that our approach should show similar effectiveness to what we saw in that species.

We applied the newly determined optimized descriptor sets to rank both the Pubchem and Natural odor libraries (Table 3.6). The three prediction lists are quite different from those previously calculated. The next stage will be to validate these predictions using single-unit electrophysiology.

Predicting receptor-odor interactions using machine learning

In order to improve predictive ability we integrated a Support Vector Machine (SVM), which is a well-known and widely applied machine learning approach (Cortes and Vapnik, 1995), into our prediction pipeline (Figure 3.6). We train the SVM using previously determined optimized descriptor set values for Or2, Or6, and Or10 (See Methods Chapter). The trained SVM is then applied to predict the activity of odors from both the Pubchem and Natural Odor libraries. While we have not yet performed electrophysiological validation of our predictions, we have performed a 5-fold cross validation with the SVM, observing far higher AUC values (Figure 3.7). The mean AUC value of 0.988 is far higher than that of the previous value of 0.832 when SVMs were not applied, suggesting that the approach may be highly successful in predicting activating aromatic compounds for these three receptors.

Identifying safer alternative activators with better organoleptic profiles

One target of our approach is to predict odors from a large library of natural odors, many of which are already approved for use in food and/or fragrances. We compare the organoleptic properties of our newly identified activating odors vs those previously identified and discover that odors from our natural odor library have more favorable profiles for several receptors. For example, we have identified new activators for AgOr8, which is believed to be important for host identification through activation by 1-octen-3-ol, that smell more pleasant to the human nose than those previously known. Previously known activators included 3-octanone, 1-octen-3-ol, and 2-heptanone, the first two of which have been described as smelling either moldy or fungal at high

concentrations (www.thegoodscentscompany.com). 2-heptanone is described as fruity, spicy, or woody, however it is also listed as having a high odor strength resulting in a suggested smelling concentration of 10% or less. Our identified compound 2-Heptanol activates the receptor equally as strongly, is described as smelling like fresh lemon grass, and only has a medium odor strength allowing for a pleasant odor experience even at concentrations as high as 100%. Using our approach it becomes possible to easily explore the organoleptic properties of predicted odors, providing a way to select lures and repellent odors based upon their safety profiles and pleasant aroma.

Predictions for CO_2 receptors in a broad range of species

Integrating response profiles to CO_2 from multiple insect species

The olfactory receptors of insects are highly divergent and very little information can be gained from comparing the full-length sequences of Ors between species. As a result of this divergence, we are very rarely able to compare the response profiles of Ors across species. However, the CO_2 receptor is extremely important in many insect species for host detection or predator avoidance, providing a rare case where receptor response profiles are moderately retained across species. The CO_2 receptor of 3 mosquito species as well as *Drosophila* have recently been identified and exposed to panels of odors (Jones et al., 2007; Kwon et al., 2007; Lu et al., 2007; Turner et al., 2011; Turner and Ray, 2009). While several odors were unique to testing in a single species, a panel of 29 odors was tested across all four species. The response profiles of overlapping odors appear in general similar, suggesting the possibility for pooling the responses of the 4 species into a single metric of CO_2 receptor activity.

We wanted our metric to consider the responses of multiple species for odors where more than one species had been tested, however since responses to some odors were significantly lower in one species than the others, we decided upon selection based on customized criteria. We averaged the two highest activity values for each odor that had been tested on more than 1 species, resulting in a metric that was influenced by multiple species, yet not unfavorably reduced by low activity outliers. The resulting activity values were used as our training set for the remainder of our CO_2 analysis.

The resulting training set contained a large variety of odor structures including both aromatic and aliphatic activators. Visual inspection revealed there were two distinct structural classes of known activating odors (aromatic and aliphatic) and one for inhibiting odors. As we have both recognized the challenges posed in describing aromatic compounds in our analyses of *Drosophila* and *Anopheles* Ors and considered that these two structurally distinct classes of odors may bind in distinct pockets on the receptor, we decided to separate our training odors into three groups with different predictive aims: aromatic ligands, structurally diverse ligands, and aliphatic ligands (Table 3.7). Our aromatic set focused on screening for highly active and aromatic structures and did not consider inhibitors. Our Broad activator screen focused on structurally diverse ligands and tested all previously tested odors including both activators and inhibitors. Finally, we performed an inhibition screen to identify CO_2 receptor inhibitors, where we only considered inhibitory or inactive odors and did not include activators.

Chemical Informatics can successfully explain CO_2 receptor-odor activity

We once again applied our chemical informatics pipeline to predict active odors, in this instance predicting activators and inhibitors for the CO_2 receptor (Figure 3.8). We again calculated a single 3D structure using Omega software (Hawkins et al., 2010) and calculated 3,224 molecular descriptors using Dragon (Talete) for each odor from all three training sets. We then implement our computational pipeline to identify a subset of the 3,224 molecular descriptors that are most important for activity in each of the training sets through application of a Sequential Forward Selection (SFS) method (Whitney, 1971). This method identifies molecular descriptors that are the most highly correlated with odor activity by incrementally building a best descriptor set, which involves adding a single molecular descriptor at a time until the correlation ceases to increase (See Methods Chapter VII). We applied this process independently for each of the training sets, resulting in three uniquely optimized molecular descriptor sets (Tables 3.8, 3.9, 3.10).

While there were many important 2D and 3D descriptors selected for all three of the training sets, several more intuitive 0D and 1D descriptors were selected as well. Chiefly, the number of non-terminal nitrogen atoms and number of aldehydes were important for the aromatic activator training set, the number of aldehydes, presence of a carbon-oxygen bond 6 atoms apart and number of donor atoms for H-bonds were important for the broad activator training set, and the presence of a carbon and oxygen atoms 3 atoms apart, number of aliphatic secondary carbons, presence of carbon and carbon atoms 6 atoms apart, and absence of ring secondary carbons were important for the inhibitor training set. The aromatic activator set was the smallest, containing 11 descriptors, and the broad activator set was the largest with 27.

We clustered each of the training odors by the values of their previously determined optimized descriptor sets in order to visualize how well each descriptor set groups activating or inhibiting odors (Figure 3.9). We observed odors of interest are grouped tightly together for all three training sets, supporting that the optimized descriptors do indeed explain odor activity.

Predicting the response of the CO_2 receptor to a broad panel of untested odors

We next applied the optimized descriptor sets to computationally rank two large libraries of untested odors. We assembled both eMolecules and natural odor libraries containing over 440,000 and 3,197 odors, respectively (See Methods Chater VII). The eMolecules library contains all chemicals from the eMolecules library of similar molecular weight (<350 MW) and atom type compositions (C, O, N, H, I, Cl, S, F) of known volatile odors. The Natural Odor library contained 3,197 odors of identified floral, vegetative, insect, or mammalian origins. We ranked each odor from both libraries based on their similarity to known activators using the previously determined optimized descriptor sets, resulting in three independent rankings, one for each training set (See Methods Chater VII). We provide the top 60 predictions for each of the training sets (Table 3.11).

Many predictions validated as either activators or inhibitors of the CO_2 receptor

Electrophysiological tests were conducted by collaborator Dr. Dyan McWilliams for 139 predicted ligands originating from the three prediction lists, that were considered reasonably safe for human use. Each predicted odor was individually tested for activity using single unit electrophysiology on the *Aedes aegypti* CpA neuron, resulting in identification of many new actives for the CO_2 responsive neuron. We observed a broad

range of activity for our validated odors, including both activators and inhibitors (Table 3.12).

In our initial observation we noticed that far fewer (18%) of our predicted CO_2 odors activated the receptor at the threshold (>50 spikes/sec) previously applied for both *Drosophila* (58%) and *Anopheles* (65%) (Table 3.21). As previously described the CO_2 receptor is actually a gustatory receptor, which are believed to function uniquely from Ors. Additionally, the strongest identified activator (143 spikes/sec) for this receptor from either our large screen or any previous screen, which encompasses the responses of over 200 odors, activates significantly lower (~44%) than those observed in either *Drosophila* or *Anopheles* Ors (~250 spikes/sec), leading us to hypothesize that these Gr structured CO_2 receptors may be tuned uniquely from general Ors. As a result we have lowered our activity threshold from 50 spikes/sec to 30 spikes/sec, which loosely relates to the 44% reduction observed from the strongest known activators of other insect species to those observed for the CO_2 receptor. Even with this threshold reduction we observe that only 30% of compounds activated the CO_2 receptor at greater than or equal to 30 spikes/sec.

Interestingly, CO_2 responsive neurons in mosquitoes do not appear to have a spontaneous activity, responding only to the quantity of CO_2 present in the environment, which vary slightly depending on how many people have been in the electrophysiology room that day. As a result, we tested for inhibition by overlaying odors on a 0.15% CO_2 pulse, which on average activates the receptor at 100 spikes/sec, providing a standardized metric by which to compare inhibitory odors. We observed a similar number of tested odors inhibited the CO_2 neruon (18%) as we had observed in *Drosophila*.

In addition to CO_2 receptor structure/function differences, we also must take account of the noise introduced by the multi-species integrated training odor set. As there were no large unified odor panels tested on a single species, we integrated the responses of similarly responding CO_2 receptors for 4 insect species. While this challenge did pose a very interesting intellectual question of whether ligand structural similarity can be compared across species for similarly tuned receptors, it undoubtedly will reduce the accuracy compared to when training is performed on responses of only a single species. Considering these additive challenges, we feel our computational approach was successful in predicting ligands.

Active predicted odors appear to target up to three distinct binding pockets

We next analyzed the structures of our predicted odors that validated as actives. Interestingly, from visual inspection of their structures it appeared the odors could be structurally diverse enough to be acting on distinct binding pockets. In order to identify whether ligands fall into diverse structural classes that could potentially bind to distinct active sites on the receptor, we clustered all activating (>30 spikes/sec) and strongly inhibiting (<-10 spikes/sec) odors. We concatenated all molecular descriptors from each of the three distinct optimized descriptor sets, which we had individually optimized to predict ligands with unique characteristics, creating a single set of optimized descriptors explaining both activation and inhibition of the CO_2 receptor. We then applied hierarchical clustering to organize odor relationships based on the Euclidean distances between odors based on each odors optimized descriptor values (Figures 3.10, 3.11).

The resulting tree had three distinct branches or clusters, each of which contains structurally distinct odor classes. Cluster 1 contains only aliphatic odors that have a

variety of functional groups including aldehydes, ketones, alcohols, and esters. With the exception of the few alcohols, which are H-bond donors, nearly all compounds in this cluster contain an H-bond acceptor group. Cluster 2 contains pyrazine based heterocycles with a variety of small branched side chains. Lastly, cluster 3 contains both small 5 and 6 member non-aromatic cyclic compounds and aromatics.

We hypothesize that each of these three odor classes bind to distinct regions in the CO_2 receptors. As the structures of the CO_2 receptors have not been solved, we can only hypothesize on whether or not they bind to distinct regions, however it is very curious as to whether these distinct structures bind in entirely different protein regions, or to distinct regions in the same binding pocket, or whether the binding pockets of Ors are highly variable, accepting a very wide variety of shapes and sizes.

Interestingly, inhibitory and activating odors are interspersed across sub clusters. While the largest grouping of inhibitors occurs for cluster 2, there are strongly inhibiting odors distinctly residing in both clusters 1 and 3. In the same fashion activating odors exist in all three clusters. If three distinct binding regions do indeed exist within the CO_2 receptors, this would mean that we have identified site selective activators and inhibitors for all three.

It will be interesting to test the proposed three distinct binding region hypothesis by creating odor blends containing odors from each of the clusters. In theory, if these odors do function through distinct binding sites, exposure of a CO_2 receptor to odors from multiple clusters should provide a stronger response than similarly strong odors from within a single cluster, as binding in distinct regions at the same time should further stabilize the active conformation. This can be tested using single unit electrophysiology.

We also plan to test the strongest activating and inhibiting odors from behavioral response. Since we predicted ligands from a large collection of naturally occurring odors, many have safe activity profiles and a few are even approved for use as flavorings, fragrances, or cosmetics. If we can identify safe odors that are highly effective at altering the activity of the CO_2 receptor neuron, we will have made a major step in the crusade for global health.

Citrus Psyllid Odor Receptors

Chemical Informatics can be applied to explain Psyllid receptor-odor activity

Through work performed in our lab, activators for 11 ORNs found on 4 rhinerial plates of Citrus Psyllids have been identified. One activator has been found for PR4_A, two have been found for RP6_A, and one has been found for RP7_A, with none of the actives increasing firing rates more than 70 spikes/sec. As our computational approach requires at least two activating odors in order for optimization, we will focus our efforts on 9 AsCP neurons (RP2_B, RP2_C, RP4_A, RP4_C, RP6_A, RP6_B, RP6_C, RP7_B, RP7_C).

We begin our computational pipeline as performed for *Drosopihla*, calculating 3D structures with Omega (Hawkins et al., 2010) and 3,224 molecular descriptors using Dragon (Talete) for each of the 61 odors tested against the 9 ORNs (Figure 3.12). Molecular descriptors, which are mathematical values attempting to describe the structure of each compound, are then optimized individually for each ORN using a Sequential Forward Selection (SFS) approach, resulting in a single optimized set of descriptors for each ORN (See Methods Chapter VII) (Whitney, 1971)(Table 3.13). When we cluster the training odors based upon their optimized descriptor values, we find

that each optimized descriptor set is successful in grouping together highly active odors (Figure 3.13).

Predicting the response of the Psyllid ORNs to a broad panel of untested odors

We next used each ORN optimized descriptor set to train a Support Vector Machine (SVM), which is a highly applied machine learning approach (Chang and Lin, 2001; Karatzoglou et al., 2006). The SVM was trained to optimally predict the activity of each odor in the training set using only information provided by each ORNs optimized descriptor set, resulting in one trained SVM for each ORN. We next performed a 5-fold cross validation followed by a Receiver Operating Characteristic (ROC) analysis to validate the predictive ability for each ORN (See Methods Chapter VII). The Area Under the Curve (AUC) was calculated and the ROC analysis was plotted for each ORN (Figure 3.14). The extremely high AUC values, with a mean value of 0.99 across all ORNs, indicates that our odor prediction pipeline was extremely efficient at describing the activity of the training sets for all 9 ORNs.

We next assembled both natural odor and eMolecules libraries, representing large collections of previously untested odors (See Methods Chapter VII). We next applied each trained SVM to predict the responses of all odors from both training sets for each ORN. We provide the top 100 predicted natural odors for each ORN (Table 3.14). Predicted ligands can be validated, identifying strong activators and inhibitors for each ACP ORN. Strong actives can then be tested for their ability to modify ACP behavior. Ideally, we will be able to identify strong attractants for trap lures and strong repellents odors to protect citrus trees.

Mammalian Odor Receptors

Identification of unique subsets of optimal descriptors for mammalian receptors

Insect Or proteins are believed to consist of 7-transmembrane regions that are inversely oriented in the membrane. It is believed that they function as ligand-gated ion channels with a possible cyclic-nucleotide-activated cation channel function along with an obligate co-receptor Orco, forming a heteromer (Benton et al., 2006; Sato et al., 2008; Wicher et al., 2008). Alternatively, mammalian Ors function as G-protein coupled receptors with a traditional outside-in 7-transmembrane orientation and do not require an obligate co-receptor such as Orco (Zhang and Firestein, 2002). In order to test whether the chemical informatics platform would be successful at predicting ligands for the functionally distinct mammalian Ors, we performed a similar analysis on 33 ORs from mouse and 4 ORs from humans. The responses of each of these mammalian receptors to a panel of 62 odorants have been determined by functional expression in heterologous cells and >2 actives have been identified for each (Saito et al., 2009).

We applied our chemical informatics pipeline to optimize descriptors for each of the 37 mammalian ORs (Figure 3.15). 3D conformations for each of the training odors were calculated by Omega (See Methods Chapter VII). Dragon and Cerius2, which are commercially available programs for calculating molecular descriptors, were used to calculate 3,224 and 200 molecular descriptors for each of the training odors (See Methods Chapter VII).

We implemented a Sequential Forward Selection (SFS) approach to select a subset of molecular descriptors that best correlated with the activity of the training set odors (Whitney, 1971) (See Methods Chapter VII) (Figure 3.15). The SFS approach iteratively assembles an optimized subset of descriptors, beginning with a single best

descriptor and then incrementally adding additional descriptors one at a time until the correlation between descriptor values and OR activity fails to increase. The SFS approach is applied for each OR, resulting in 37 unique optimized descriptor sets (Table 3.15).

We applied hierarchical clustering on the Euclidean distances calculated between training odors using optmized descriptor set values for each of the 37 ORs in order to visualize how well our approach brought together activating odors (Figure 3.16). Upon visual inspection, clustering of the 62 training odors by each OR's optimized descriptor set indicated that the receptor-optimized descriptor sets were indeed able to effectively cluster activating odors together.

Computational validation of mammalian OR-optimized descriptor sets and *in-silico* prediction of ligands

Since functional testing of predictions for mammalian receptors are beyond the scope of this study, we performed the well established computational 5-fold cross-validation to determine the predictive ability of the *in silico* approach. ORs with >15 known ligands were selected and for each OR 20% of the compounds (12/60) were excluded as a test set, while the remaining were used as a training set to generate the optimized descriptors. As before we performed a ROC analysis for each of the withheld test-set odors by classifying activity based on distance in the training set derived Or-optimized chemical space. We repeated this operation five times for each receptor, each trial performed by excluding a different subset of odors.

We generated mean ROC curves for each OR and calculated AUC values (Figure 3.17). Using this method we demonstrate that the OR-optimized-descriptor sets

generated using the training odors could accurately identify actives from the test odorants. Both the ROC and ApoA values were comparable, if not better (data not shown), than for the *Drosophila* Ors suggesting that the descriptors are able to efficiently explain the activity of the mammalian Ors.

Predicting mammalian receptor-odor interactions for a large untested odor space

We then applied the OR-optimized descriptors to systematically screen both the Natrual Odor and Pubchem libraries *in silico*, comprising over ~8,880,000 receptor-odor interactions and representing 33 mouse ORs, 4 human ORs (See Methods Chapter VII). We identify the top 500 (0.2%) hits from this vast chemical library for each Or/ORN, the top ~100 for each are reported in Table 3.16.

Relationship between descriptor sets and Or sequence and activity

We next analyzed the relationship between ORs based upon their shared molecular descriptor sets, training set activity, predicted odor sets, and phylogenetics, as we had previously performed for *Drosophila* (Figure 3.18). In contrast to what was observed in *Drosophila*, the analysis for the mammalian dataset reveals a greater degree of common relationships across the known-activity, predicted cross-activity and descriptor trees (~77% ORs present in common subgroups). Similarly, the *Drosophila* Or-phylogenetic tree has sparser subgroup relationships conserved with each of the other trees (<45%), as opposed to the mammalian ORs where the majority of subgroups in the phylogenetic tree (>56%) are conserved across the different trees. This difference may reflect the much greater amino-acid similarity across the mammalian receptors (47%) as compared to the divergent *Drosophila* receptors (23%).

Analysis of breadth of predictions for each mammalian OR in chemical space

Determining the breadth of coding in a large volatile chemical space across many receptors is virtually impossible to determine experimentally. However, if we apply the OR-optimized descriptor sets, it becomes possible to computationally predict frequency distributions for each of the mammalian ORs for the odor space of both libraries (>240,000 odors) (Figure 3.19). As expected, we find a great deal of variation across the ORs. While some ORs are predicted to be very broadly tuned, others are decisively narrower. Analyses such as these allow inferences of OR properties that would be very difficult to assess through wet lab experimentation alone.

DISCUSSION

We have applied our chemical informatics pipeline to predict activating and inhibiting odors for a large number of extremely important species, including several pest and disease vector insects. For *Anopheles* Ors we applied our pipeline to predict activators and inhibitors for a large number of receptors. We validated predictions experimentally for 12 Ors, demonstrating an effectiveness similar to that observed for *Drosophila*. For the CO_2 receptor, we considered the responses of multiple species integrating them into a single metric for molecular descriptor optimization. While validations for this receptor were not as successful as seen in either *Drosophila* or *Anopheles* Ors, it still signifies a major progression in the fight for behavior modification on disease vector species. We predicted activating odors for the very first time in ACP, creating a vast expansion upon the previously tested odor space. Finally, we predicted odors for 33 mouse and 4 human Ors.

We have noted challenges involved in predicting aromatic odors in both *Drosophila* and *Anopheles*. In an attempt to improve our accuracy for aromatic responsive Ors we have performed a thorough expansion of aromatic odor space. This is the first time a focused effort has been made to expand upon and explore responses of highly diverse aromatic odors. We applied chemical informatics to select and test 27 aromatic odors that were intelligently selected for their extremely broad composition of molecular features. Further, we integrated SVMs into our prediction pipeline to increase the abilities of our computational approach. Future experimental validation will expose how successful our attempts were.

No computational analyses in olfaction had been previously performed to integrate the responses of multiple species into a single metric for odor prediction. We successfully integrated the responses of 4 divergent species to predict odors for the very important CO_2 receptor. Additionally, while the three mosquito species are attracted to CO_2, *Drosophila* are repelled, further creating an interesting intellectual challenge by integrating odors causing mixed behavioral responses. All of these differences introduce noise into the training set. As a result, the observed lower accuracy (48%) compared to that previously observed in both *Drosophila* (71%) and *Anopheles* (65%) Or predictions was not unexpected. However, due to these multiple challenges we still consider this to be a significant achievement.

It will be interesting to perform experimental validations for both the ACP and mammalian predictions in the future. There are several ways to further improve our computational approach in the future. Sequential Floating Forward Selection (SFFS) allows for the intermittent removal of previously selected descriptors instead of a steady building of an optimized list. Additionally, other predictive approaches beyond SVMs

and other alternative approaches can be considered. Since machine learning has been demonstrated to be highly effective for our pipeline, we plan on applying it for all future analyses.

Figure 3.1: A molecular descriptor optimized approach is able to explain odor activity for individual *Anopheles* Ors

Schematic of our chemical informatics pipeline. Molecular descriptors that are most correlated with activity are selected, resulting in a metric that is able to cluster together highly active odors using important structural features. The optimized descriptor sets can then be applied to predict Or activity against a large panel of odors. This pipeline is applied independently, optimizing descriptors for either activators (Top) or inhibitors (Bottom).

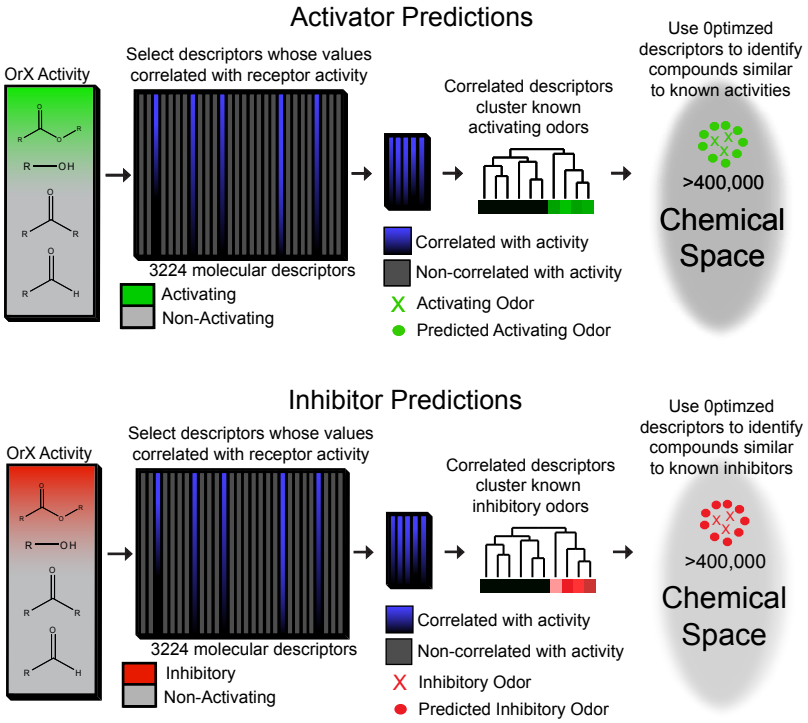

Activator Predictions
OrX Activity
Select descriptors whose values correlated with receptor activity
3224 molecular descriptors
Activating
Non-Activating
Correlated descriptors cluster known activating odors
Correlated with activity
Non-correlated with activity
Activating Odor
Predicted Activating Odor
Use 0ptimzed descriptors to identify compounds similar to known activities
>400,000
Chemical Space
Inhibitor Predictions
OrX Activity
Select descriptors whose values correlated with receptor activity
3224 molecular descriptors
Inhibitory
Non-Activating
Correlated descriptors cluster known inhibitory odors
Correlated with activity
Non-correlated with activity
Inhibitory Odor
Predicted Inhibitory Odor
Use 0ptimzed descriptors to identify compounds similar to known inhibitors
>400,000
Chemical Space

Figure 3.1

Figure 3.2: Optimized molecular descriptor sets are able to cluster either active or inhibitory odors

Optimized molecular descriptors values were applied to cluster training set odors individually for each Or. Clusters are divided into Activators, Fishing Activators, and Inhibitors. Colors for Activators and Fishing activators range from purple to red for lowest activity to highest activity and are individually applied for each Or. Colors for Inhibitors range from blue to bright for highest activity to most inhibitory.

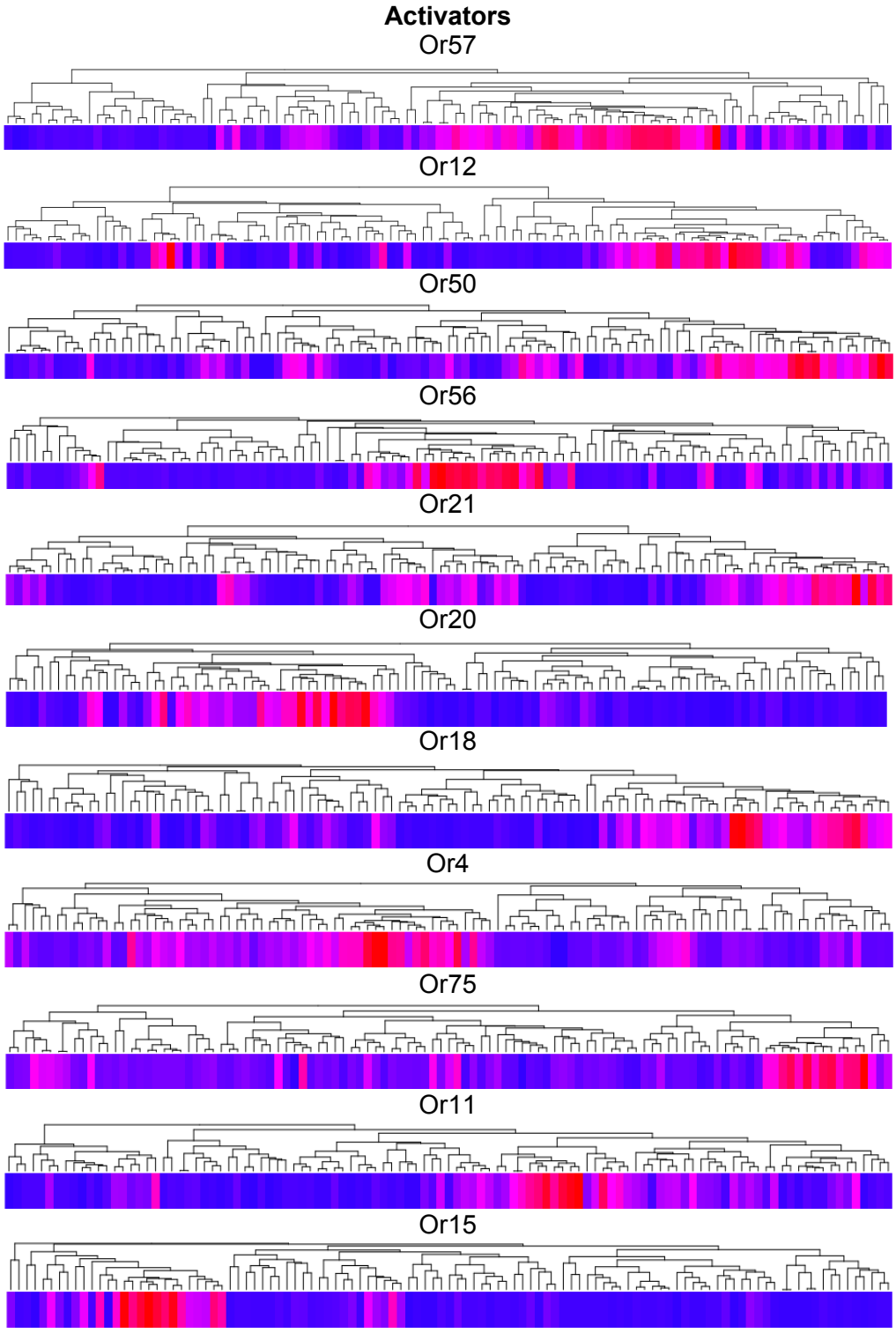

Activators
Or57
Or12
Or50
Or56
Or21
Or20
Or18
Or4
Or75
Or11
Or15

Figure 3.2

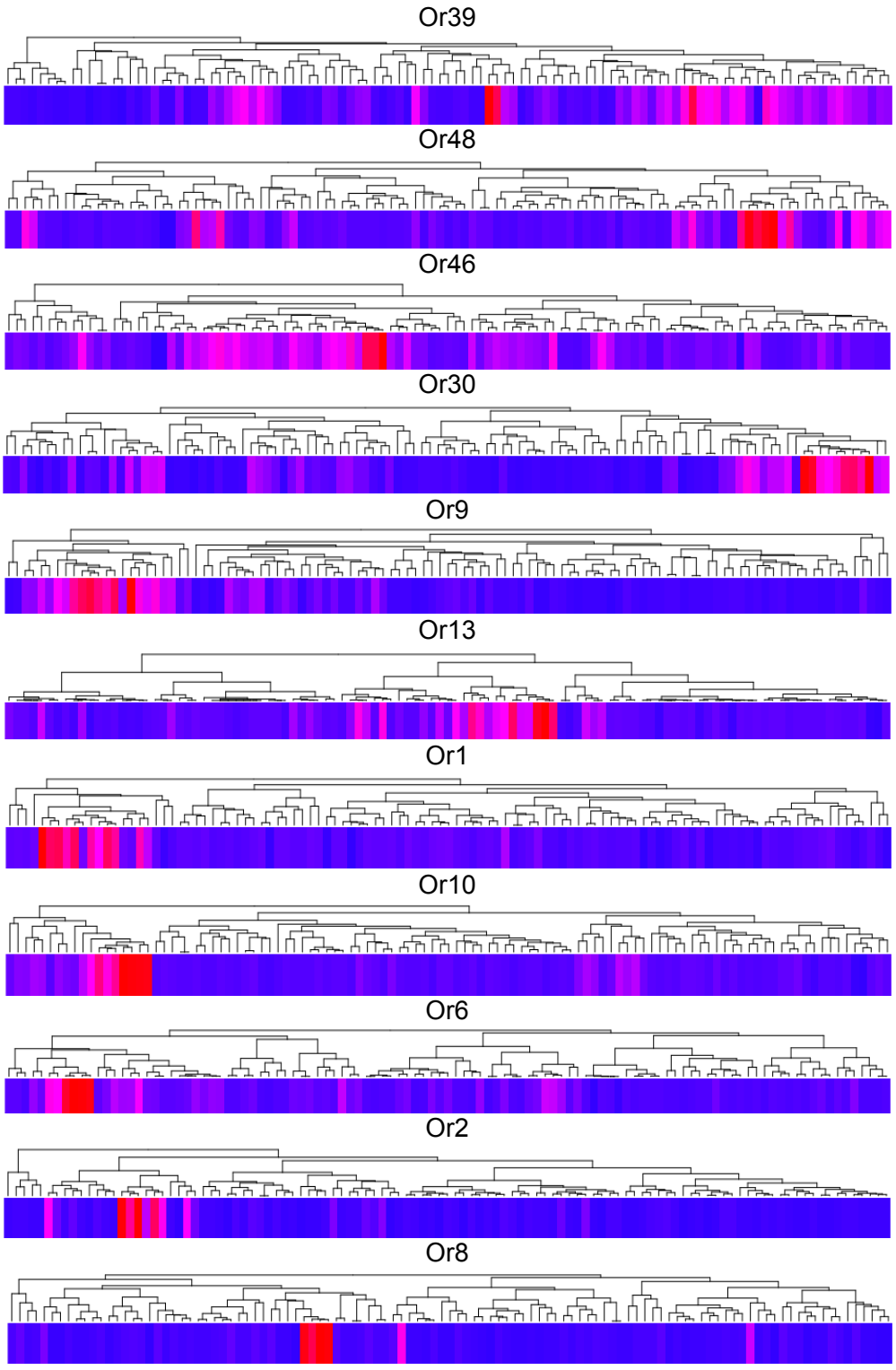


Figure 3.2 Continued

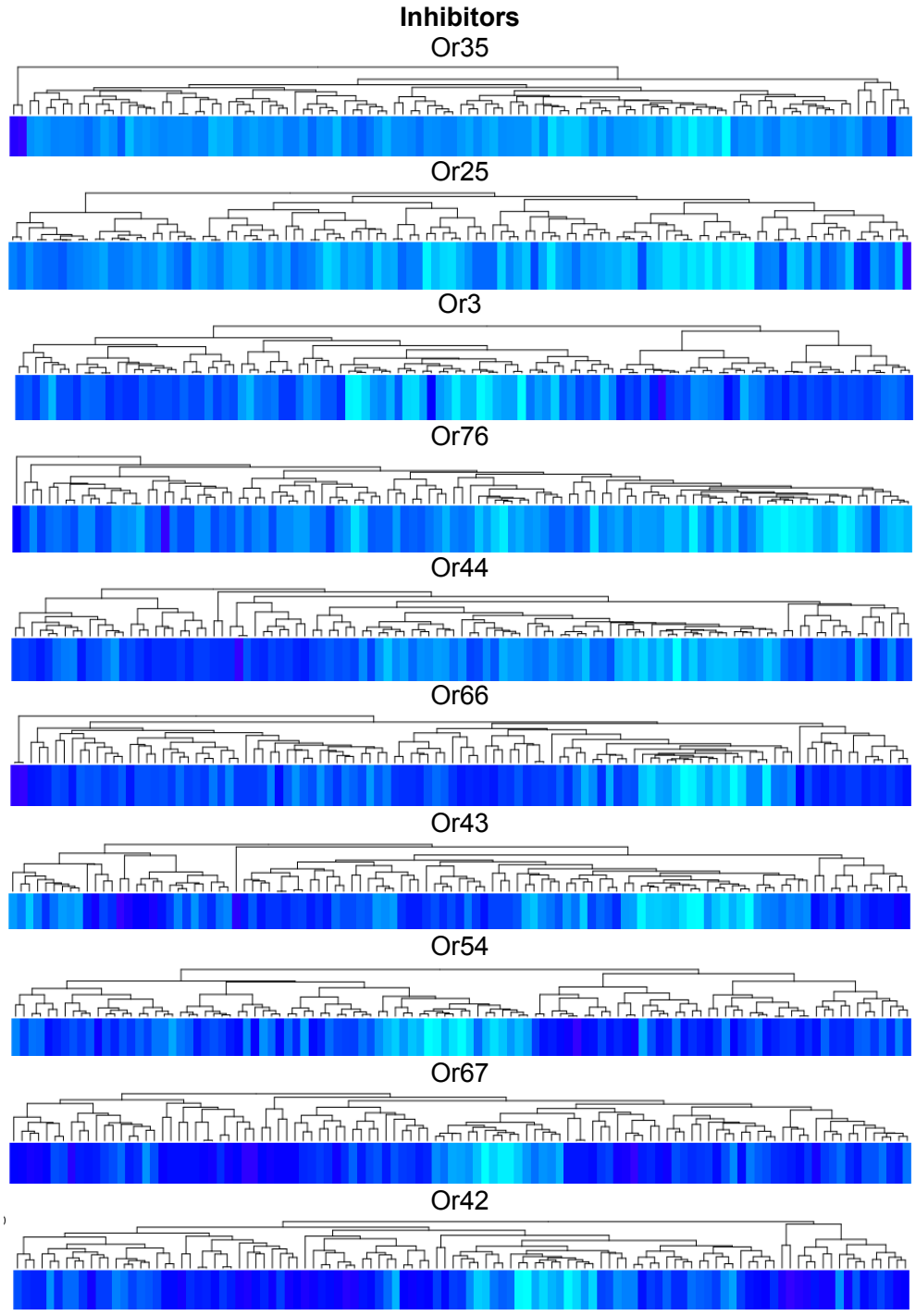


Figure 3.2 Continued

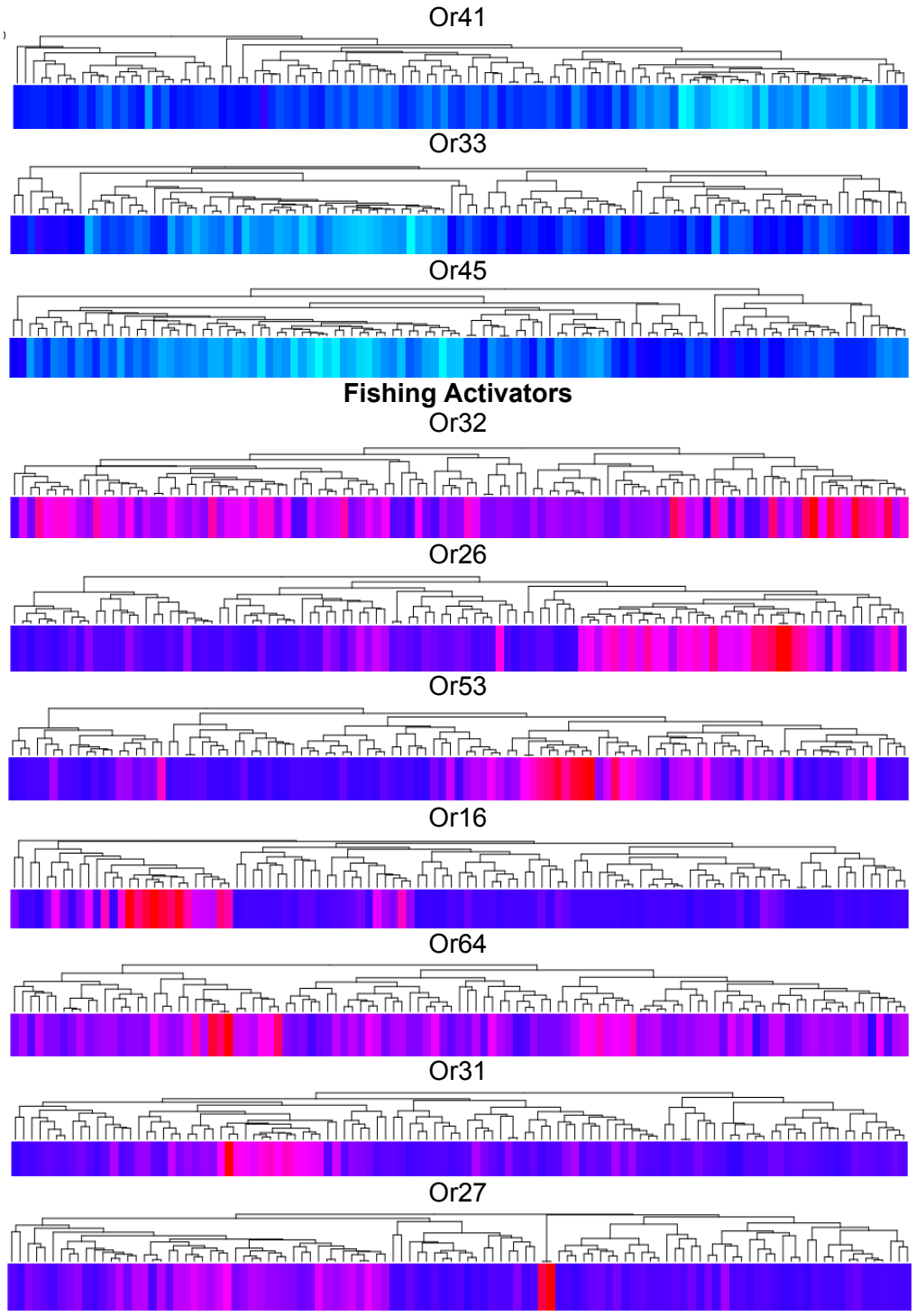


Figure 3.2 Continued

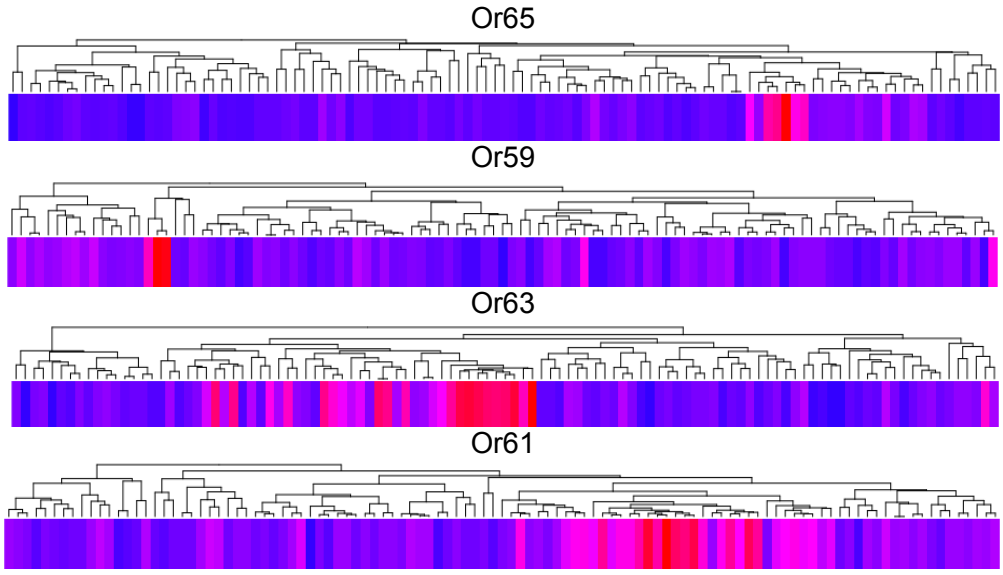


Figure 3.2 Continued

Figure 3.3: Structural relationships between the 75 most diverse aromatic odors

Hierarchical clustering for 75 odors using Euclidean distances between odors calculated using all 3,224 molecular descriptors calculated by Dragon. Compound names are provided for each odor.

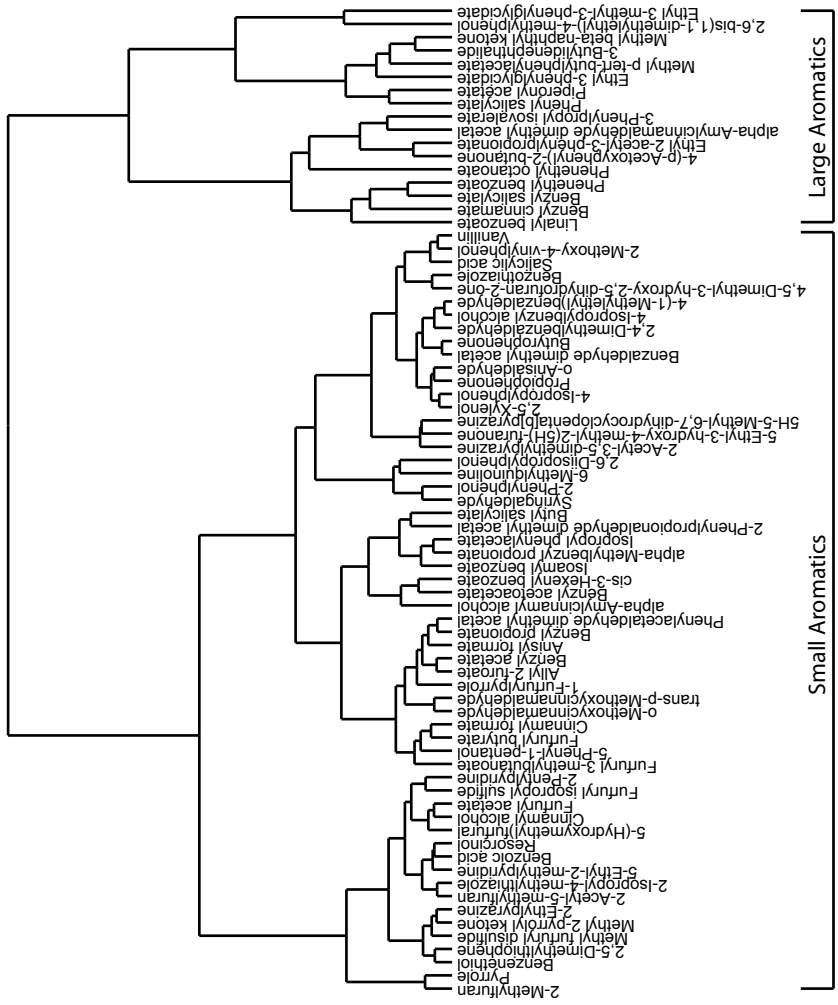

2-Methylfuran
Pyrrole
Benzenethiol
2,5-Dimethylthiophene
Methyl furfuryl disulfide
Methyl 2-pyrrolyl ketone
2-Ethylpyrazine
2-Acetyl-5-methylfuran
2-Isopropyl-4-methylthiazole
5-Ethyl-2-methylpyridine
Benzoic acid
Resorcinol
5-(Hydroxymethyl)furfural
Cinnamyl alcohol
Furfuryl acetate
Furfuryl isopropyl sulfide
2-Pentylpyridine
Furfuryl 3-methylbutanoate
5-Phenyl-1-pentanol
Furfuryl butyrate
Cinnamyl formate
o-Methoxycinnamaldehyde
trans-p-Methoxycinnamaldehyde
1-Furfurylpyrrole
Allyl 2-furoate
Benzyl acetate
Anisyl formate
Benzyl propionate
Phenylacetaldehyde dimethyl acetal
alpha-Amylcinnamyl alcohol
Benzyl acetoacetate
cis-3-Hexenyl benzoate
Isoamyl benzoate
alpha-Methylbenzyl propionate
Isopropyl phenylacetate
2-Phenylpropionaldehyde dimethyl acetal
Butyl salicylate
Syringaldehyde
2-Phenylphenol
6-Methylquinoline
2,6-Diisopropylphenol
2-Acetyl-3,5-dimethylpyrazine
5-Ethyl-3-hydroxy-4-methyl-2(5H)-furanone
5H-5-Methyl-6,7-dihydrocyclopentapyrazine
2,5-Xylenol
4-Isopropylphenol
Propiophenone
o-Anisaldehyde
Benzaldehyde dimethyl acetal
Butyrophenone
2,4-Dimethylbenzaldehyde
4-Isopropylbenzyl alcohol
4-(1-Methylethyl)benzaldehyde
4,5-Dimethyl-3-hydroxy-2,5-dihydrofuran-2-one
Benzothiazole
Salicylic acid
2-Methoxy-4-vinylphenol
Vanillin
Linalyl benzoate
Benzyl cinnamate
Benzyl salicylate
Phenethyl benzoate
Phenethyl octanoate
4-(p-Acetoxyphenyl)-2-butanone
Ethyl 2-acetyl-3-phenylpropionate
alpha-Amylcinnamaldehyde dimethyl acetal
3-Phenylpropyl isovalerate
Phenyl salicylate
Piperonyl acetate
Ethyl 3-phenylglycidate
Methyl p-tert-butylphenylacetate
3-Butylidenephthalide
Methyl beta-naphthyl ketone
2,6-bis(1,1-dimethylethyl)-4-methylphenol
Ethyl 3-methyl-3-phenylglycidate
Small Aromatics
Large Aromatics

Figure 3.3

Figure 3.4: Newly optimized molecular descriptor sets are able to cluster odors for aromatically tuned Ors

Optimized molecular descriptors values were applied to cluster training set odors individually for each Or. Colors range from gray to black, representing the lowest to highest activity.

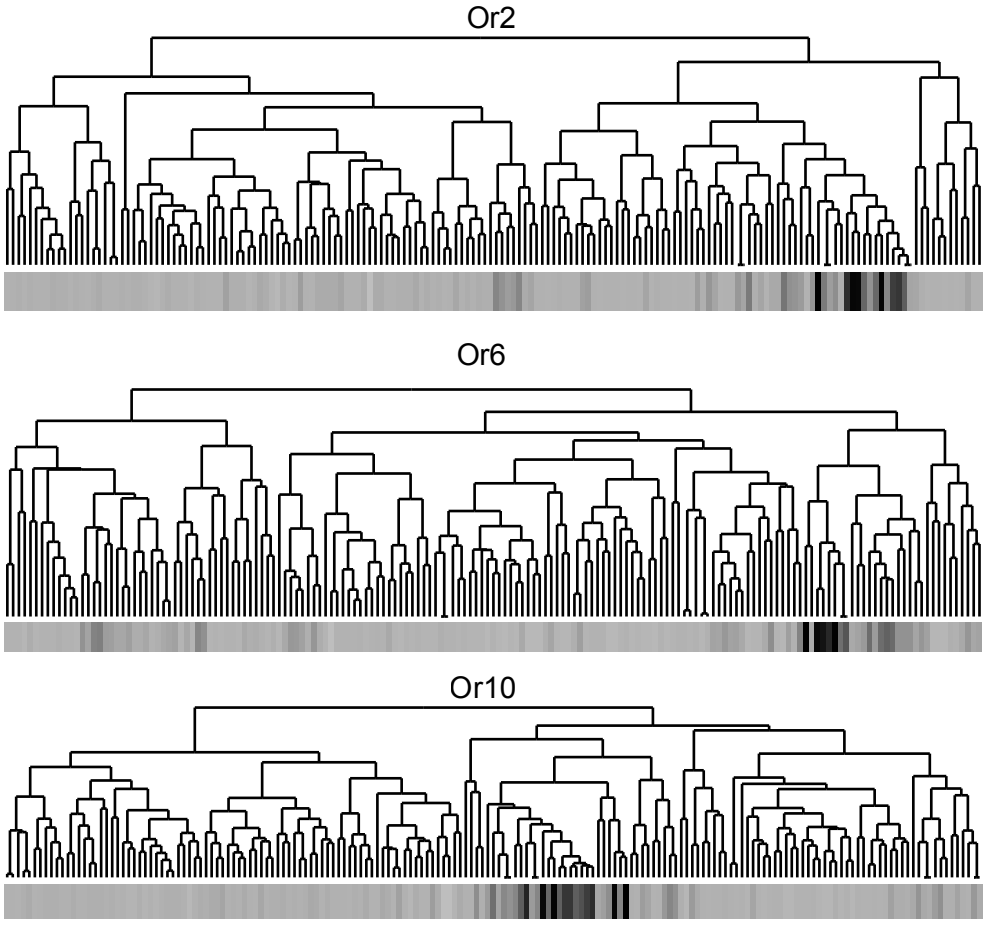


Figure 3.4

Figure 3.5: Optimized descriptors for AgOrs 2, 6, and 10 effectively describe training set activity.

A Receiver-Operating-Characteristic (ROC) curve is plotted for each Or, depicting the computational validation of ligand predictive ability of the Or-optimization approach.

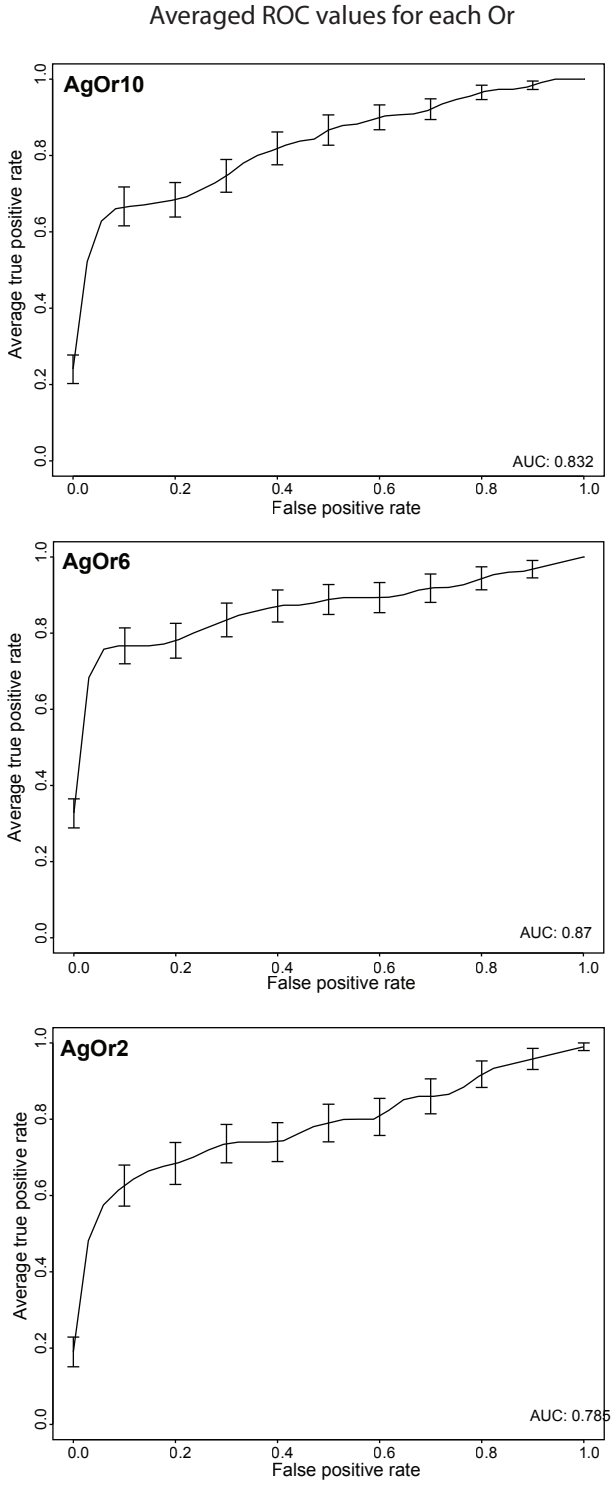

Averaged ROC values for each Or
AgOr10
Average true positive rate
False positive rate
AUC: 0.832
AgOr6
Average true positive rate
False positive rate
AUC: 0.87
AgOr2
Average true positive rate
False positive rate
AUC: 0.785

Figure 3.5

Figure 3.6: A Support Vector Machine (SVM) integrated approach is highly effective at explaining odor activity for individual Ors

Schematic of our SVM integrated chemical informatics pipeline. Molecular descriptors that are most correlated with activity are selected, resulting in a metric that is able to cluster together highly active odors using important structural features. The optimized descriptor sets are then be applied to train a SVM to predict Or activity, which is then applied to predict the activity of a large untested odor space.

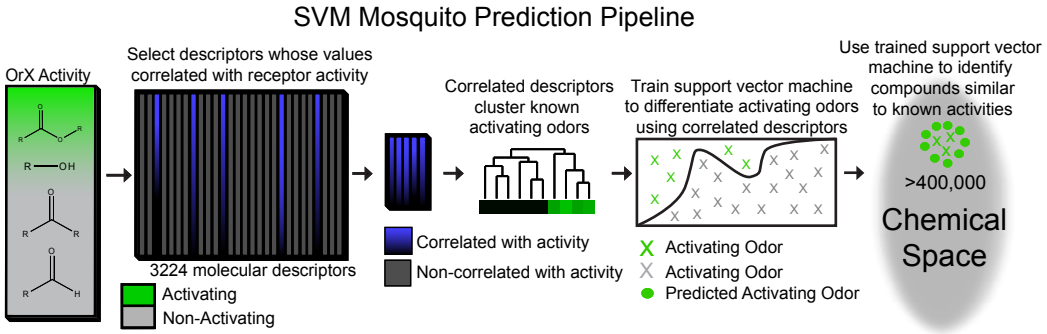

SVM Mosquito Prediction Pipeline
OrX Activity
Select descriptors whose values correlated with receptor activity
3224 molecular descriptors
Activating
Non-Activating
Correlated with activity
Non-correlated with activity
Correlated descriptors cluster known activating odors
Train support vector machine to differentiate activating odors using correlated descriptors
Activating Odor
Activating Odor
Predicted Activating Odor
Use trained support vector machine to identify compounds similar to known activities
>400,000
Chemical Space

Figure 3.6

Figure 3.7: SVMs trained using optimized descriptors for AgOrs 2, 6, and 10 effectively describe training set activity

A Receiver-Operating-Characteristic (ROC) curve determined from 100 independent iterations of a SVM applied 5-fold cross validation is plotted for each Or, depicting the predictive ability of the Or-optimization approach.

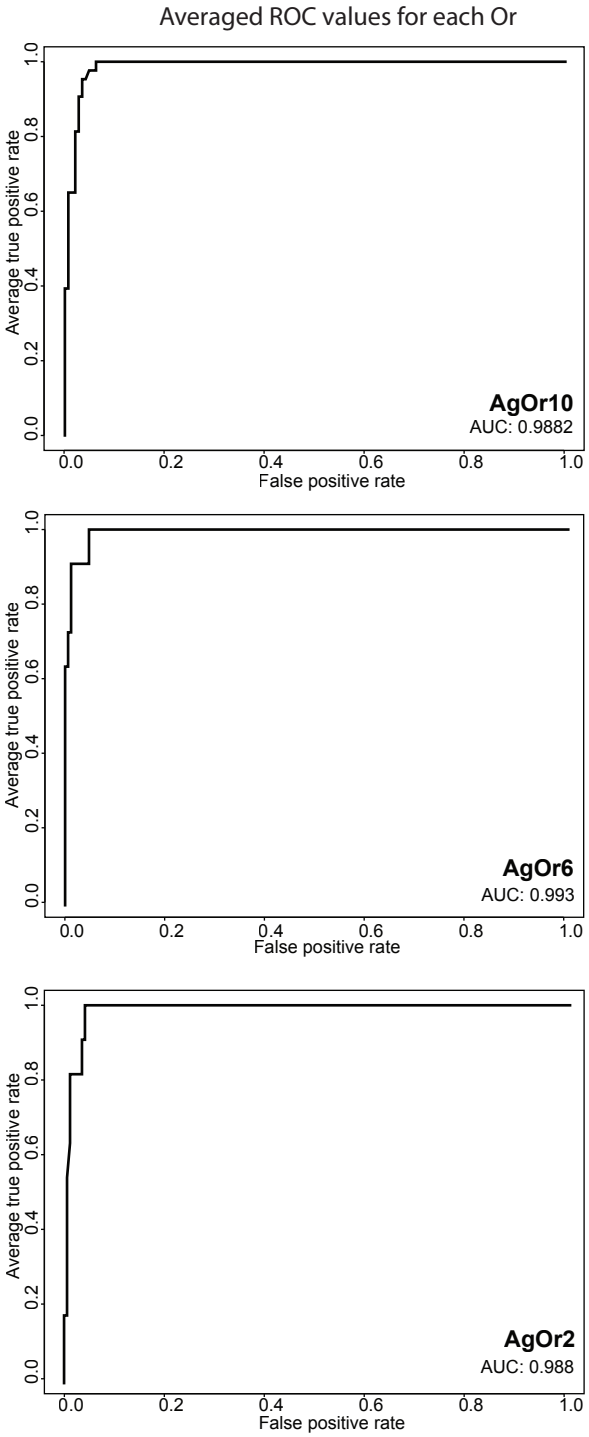

Averaged ROC values for each Or
Average true positive rate
False positive rate
AgOr10
AUC: 0.9882
AgOr6
AUC: 0.993
AgOr2
AUC: 0.988

Figure 3.7

Figure 3.8: A molecular descriptor optimized approach is able to explain odor activity for CO_2 receptors

Schematic of our chemical informatics pipeline. Molecular descriptors that are most correlated with activity are selected, resulting in a metric that is able to cluster together highly active odors using important structural features. The optimized descriptor sets can then be applied to predict CO_2 receptor activity against a large panel of odors. This pipeline is applied independently, optimizing descriptors for either activators (Top) or inhibitors (Bottom).

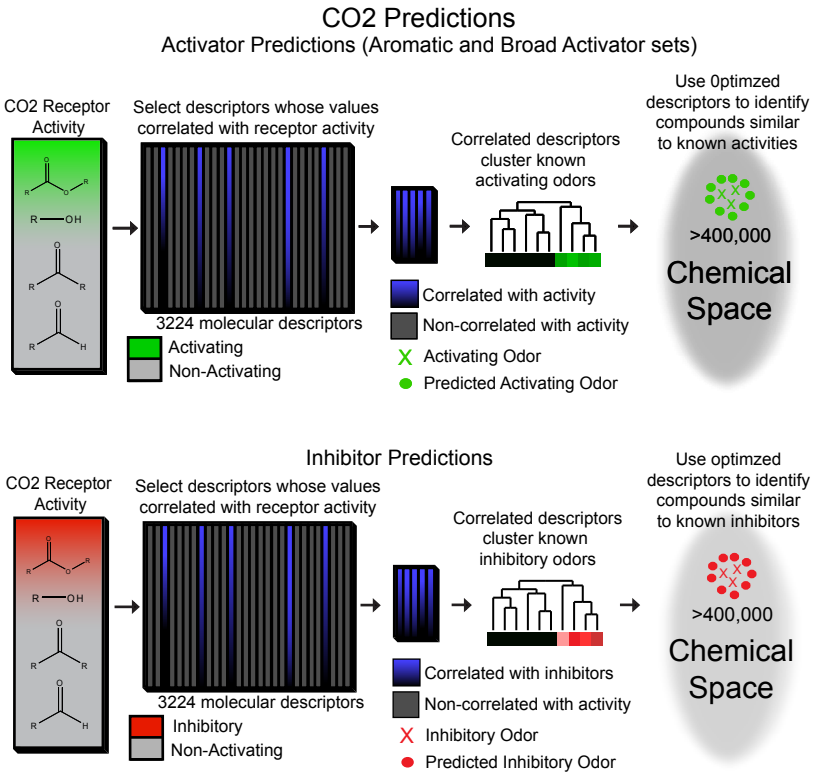

CO2 Predictions
Activator Predictions (Aromatic and Broad Activator sets)
CO2 Receptor Activity
Select descriptors whose values correlated with receptor activity
3224 molecular descriptors
Activating
Non-Activating
Correlated descriptors cluster known activating odors
Correlated with activity
Non-correlated with activity
Activating Odor
Predicted Activating Odor
Use 0ptimzed descriptors to identify compounds similar to known activities
>400,000
Chemical Space
Inhibitor Predictions
CO2 Receptor Activity
Select descriptors whose values correlated with receptor activity
3224 molecular descriptors
Inhibitory
Non-Activating
Correlated descriptors cluster known inhibitory odors
Correlated with inhibitors
Non-correlated with activity
Inhibitory Odor
Predicted Inhibitory Odor
Use optimzed descriptors to identify compounds similar to known inhibitors
>400,000
Chemical Space

Figure 3.8

Figure 3.9: Optimized molecular descriptor sets are able to cluster odors by CO_2 receptor response

Optimized molecular descriptors values were applied to cluster training set odors individually for each screening set. Odor activity is represented individually for each screening set, ranging from gray to black in increasing activity.

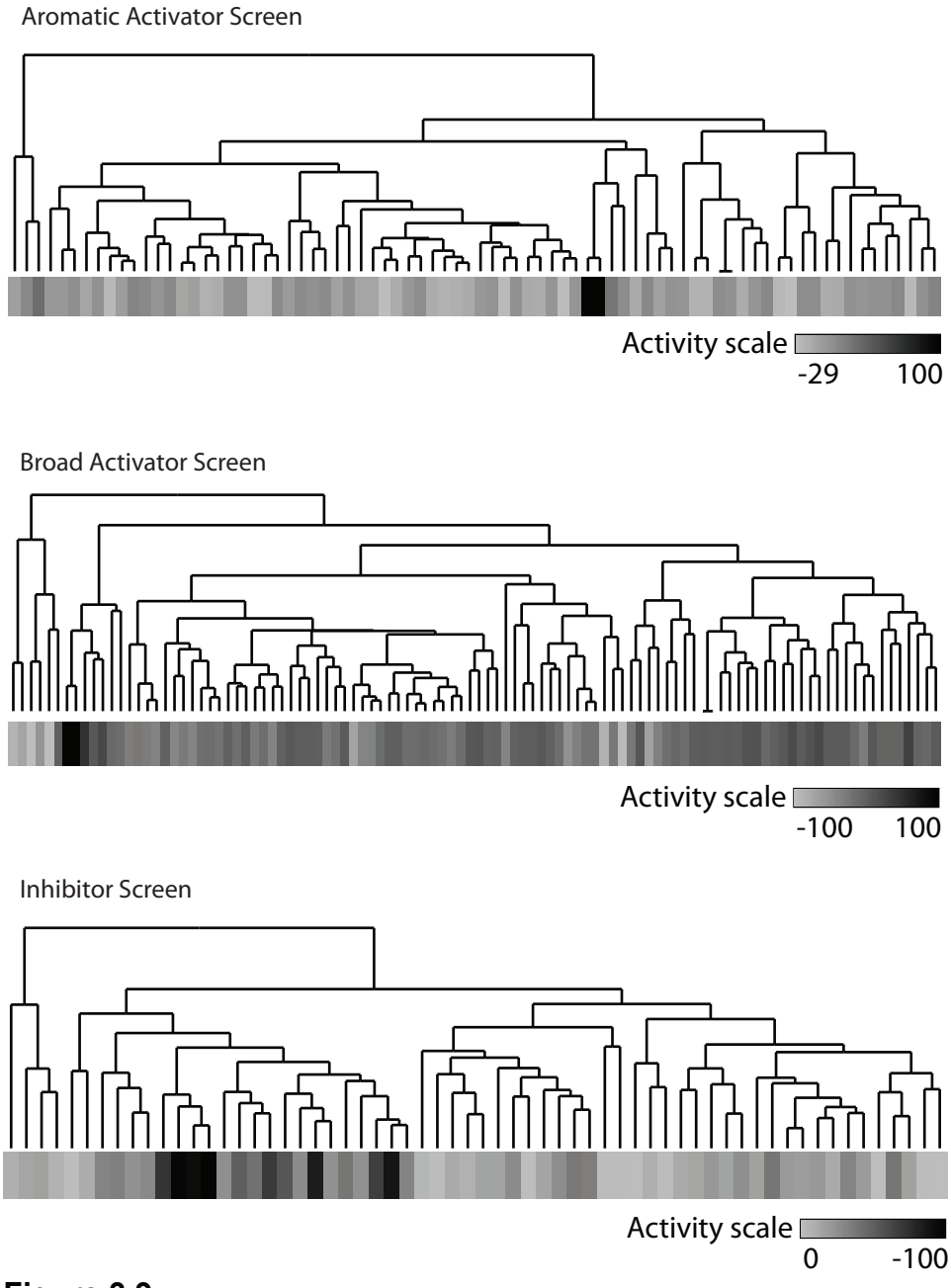

Aromatic Activator Screen
Activity scale
-29
100
Broad Activator Screen
Activity scale
-100
100
Inhibitor Screen
Activity scale
0
-100

Figure 3.9

Figure 3.10: Active compounds cluster into three distinct structural classes

Active compounds were clustered by the combined 3 sets of optimized descriptors. Hierarchical clustering was performed on compound optimized molecular descriptor values, dividing the chemical structures into three distinct classes that have been outlined by both label and color. CO_2 receptor responses (spikes/sec) have been provided for each compound and green and red boxes label activators or inhibitors, respectively. Representative chemical structures for each class are provided.

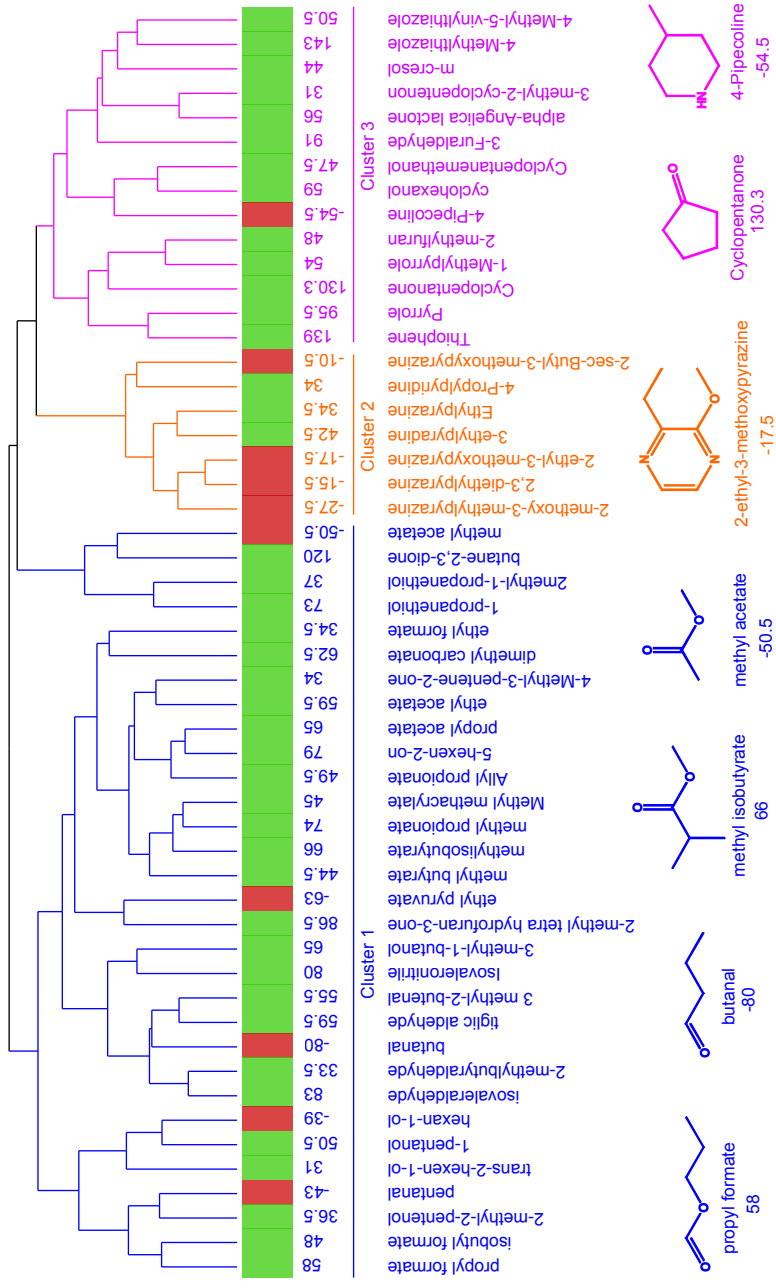

propyl formate
isobutyl formate
2-methyl-2-pentenol
pentanal
trans-2-hexen-1-ol
1-pentanol
hexan-1-ol
isovaleraldehyde
2-methylbutyraldehyde
butanal
tiglic aldehyde
3 methyl-2-butenal
Isovaleronitrile
3-methyl-1-butanol
2-methyl tetra hydrofuran-3-one
ethyl pyruvate
methyl butyrate
methylisobutyrate
methyl propionate
Methyl methacrylate
Allyl propionate
5-hexen-2-on
propyl acetate
ethyl acetate
4-Methyl-3-pentene-2-one
dimethyl carbonate
ethyl formate
1-propanethiol
2methyl-1-propanethiol
butane-2,3-dione
methyl acetate
2-methoxy-3-methylpyrazine
2,3-diethylpyrazine
2-ethyl-3-methoxypyrazine
3-ethylpyradine
Ethylpyrazine
4-Propylpyridine
2-sec-Butyl-3-methoxypyrazine
Thiophene
Pyrrole
Cyclopentanone
1-Methylpyrrole
2-methylfuran
4-Pipecoline
cyclohexanol
Cyclopentanemethanol
3-Furaldehyde
alpha-Angelica lactone
3-methyl-2-cyclopentenon
m-cresol
4-Methylthiazole
4-Methyl-5-vinylthiazole
Cluster 1
Cluster 2
Cluster 3
58
48
36.5
-43
31
50.5
-39
83
33.5
-80
59.5
55.5
80
65
86.5
-63
44.5
66
74
45
49.5
79
65
59.5
34
62.5
34.5
73
37
120
-50.5
-27.5
-15.5
-17.5
42.5
34.5
34
-10.5
139
95.5
130.3
54
48
-54.5
59
47.5
91
56
31
44
143
50.5
propyl formate
58
butanal
-80
methyl isobutyrate
66
methyl acetate
-50.5
2-ethyl-3-methoxypyrazine
-17.5
Cyclopentanone
130.3
4-Pipecoline
-54.5

Figure 3.10

Figure 3.11: Chemical structures of validated activators and inhibitors

The structures for each of the compounds that validated as either activators (>30 spikes/sec) or inhibitors (<0) spikes/sec are provided. Compounds are divided their cluster as described in figure 3.22.

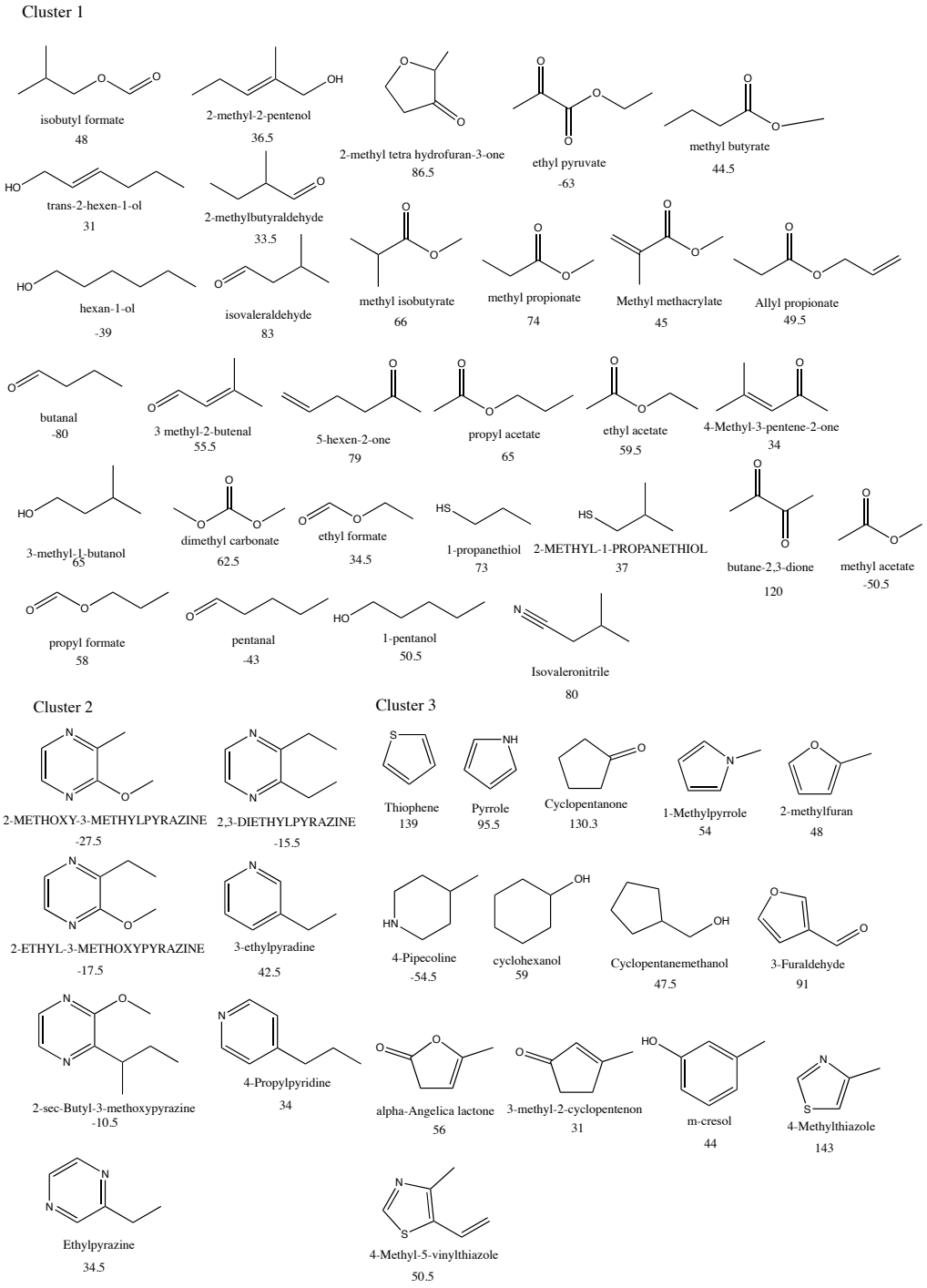

Cluster 1
isobutyl formate
48
2-methyl-2-pentenol
36.5
2-methyl tetra hydrofuran-3-one
86.5
ethyl pyruvate
-63
methyl butyrate
44.5
trans-2-hexen-1-ol
31
2-methylbutyraldehyde
33.5
hexan-1-ol
-39
isovaleraldehyde
83
methyl isobutyrate
66
methyl propionate
74
Methyl methacrylate
45
Allyl propionate
49.5
butanal
-80
3 methyl-2-butenal
55.5
5-hexen-2-one
79
propyl acetate
65
ethyl acetate
59.5
4-Methyl-3-pentene-2-one
34
3-methyl-1-butanol
65
dimethyl carbonate
62.5
ethyl formate
34.5
1-propanethiol
73
2-METHYL-1-PROPANETHIOL
37
butane-2,3-dione
120
methyl acetate
-50.5
propyl formate
58
pentanal
-43
1-pentanol
50.5
Isovaleronitrile
80
Cluster 2
2-METHOXY-3-METHYLPYRAZINE
-27.5
2,3-DIETHYLPYRAZINE
-15.5
2-ETHYL-3-METHOXYPYRAZINE
-17.5
3-ethylpyradine
42.5
2-sec-Butyl-3-methoxypyrazine
-10.5
4-Propylpyridine
34
Ethylpyrazine
34.5
Cluster 3
Thiophene
139
Pyrrole
95.5
Cyclopentanone
130.3
1-Methylpyrrole
54
2-methylfuran
48
4-Pipecoline
-54.5
cyclohexanol
59
Cyclopentanemethanol
47.5
3-Furaldehyde
91
alpha-Angelica lactone
56
3-methyl-2-cyclopentenon
31
m-cresol
44
4-Methylthiazole
143
4-Methyl-5-vinylthiazole
50.5

Figure 3.11

Figure 3.12: A SVM integrated molecular descriptor optimized approach is able to explain odor activity for individual Citrus Psyllid ORNs

Schematic of our SVM integrated chemical informatics pipeline. Molecular descriptors that are most correlated with activity are selected, resulting in a metric that is able to cluster together highly active odors using important structural features. The optimized descriptor sets are then be applied to train a SVM to predict Or activity, which is then applied to predict the activity of a large untested odor space.

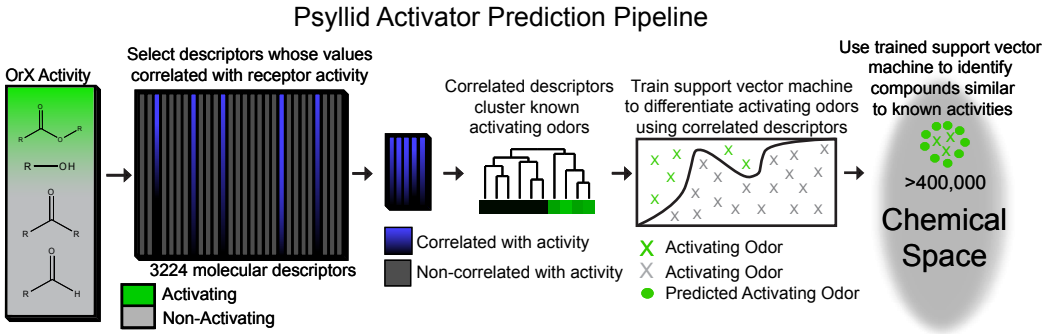

Psyllid Activator Prediction Pipeline
OrX Activity
Select descriptors whose values correlated with receptor activity
3224 molecular descriptors
Activating
Non-Activating
Correlated with activity
Non-correlated with activity
Correlated descriptors cluster known activating odors
Train support vector machine to differentiate activating odors using correlated descriptors
Activating Odor
Activating Odor
Predicted Activating Odor
Use trained support vector machine to identify compounds similar to known activities
>400,000
Chemical Space

Figure 3.12

Figure 3.13: Optimized molecular descriptor sets are able to cluster ORN activators

Optimized molecular descriptor values were applied to cluster training set odors individually for each ORN. Odor activity is represented individually for each screening set, ranging from gray to black in increasing activity.

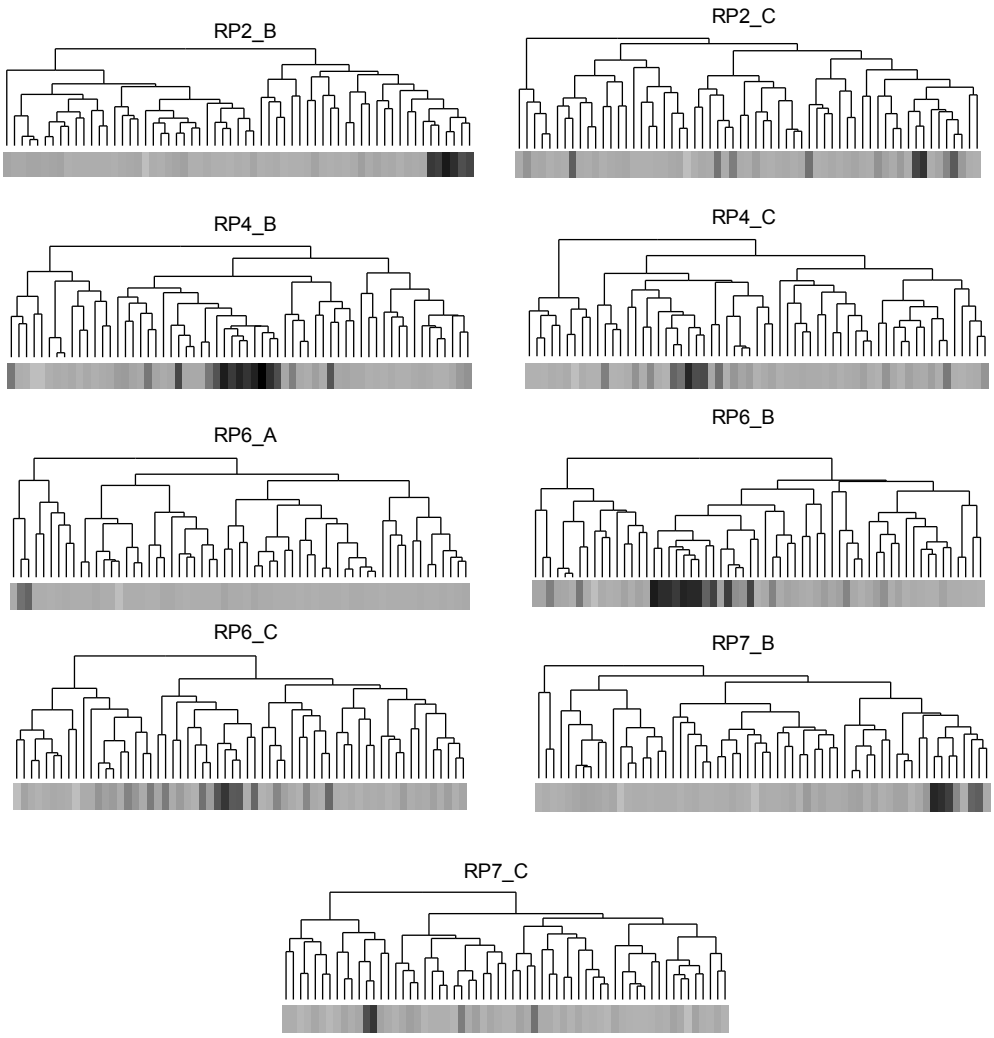


Figure 3.13

Figure 3.14: SVMs trained using descriptor sets that were optimized for individual ORNs effectively describe training set activity

A Receiver-Operating-Characteristic (ROC) curve determined from a SVM applied 5-fold cross validation is plotted for each ORN, depicting the predictive ability of the ORN-optimization approach.

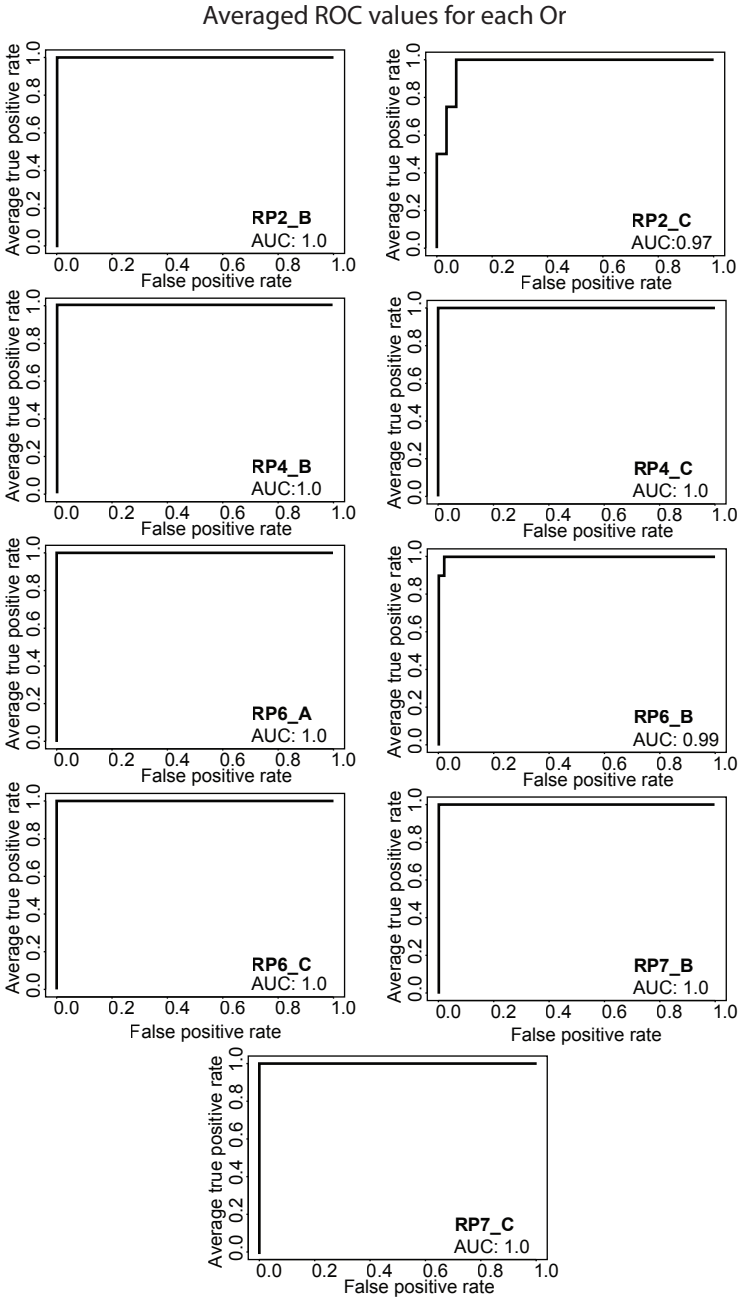

Averaged ROC values for each Or
Average true positive rate
False positive rate
RP2_B
AUC: 1.0
RP2_C
AUC:0.97
RP4_B
AUC:1.0
RP4_C
AUC: 1.0
RP6_A
AUC: 1.0
RP6_B
AUC: 0.99
RP6_C
AUC: 1.0
RP7_B
AUC: 1.0
RP7_C
AUC: 1.0

Figure 3.14

Figure 3.15: A molecular descriptor optimized approach is able to explain odor activity for individual mammalian ORs

Schematic of our chemical informatics pipeline. Molecular descriptors that are most correlated with activity are selected, resulting in a metric that is able to cluster together highly active odors using important structural features. The optimized descriptor sets can then be applied to predict Or activity against a large panel of odors.

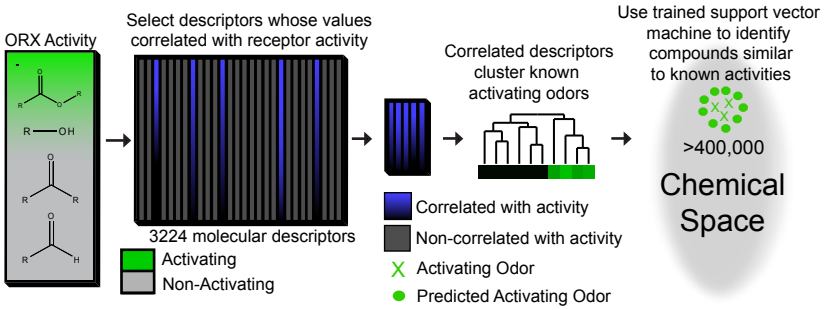

ORX Activity
Select descriptors whose values correlated with receptor activity
3224 molecular descriptors
Activating
Non-Activating
Correlated descriptors cluster known activating odors
Correlated with activity
Non-correlated with activity
Activating Odor
Predicted Activating Odor
Use trained support vector machine to identify compounds similar to known activities
>400,000
Chemical Space

Figure 3.15

Figure 3.16: Mammalian Odorant receptor-optimized molecular descriptors can successfully cluster known ligands

Hierarchical cluster for 33 mouse receptors and 4 human receptors using receptor-optimized descriptor sets using data from (Saito et al., 2009). Known odorant activity scale is indicated using independent color gradient scales.

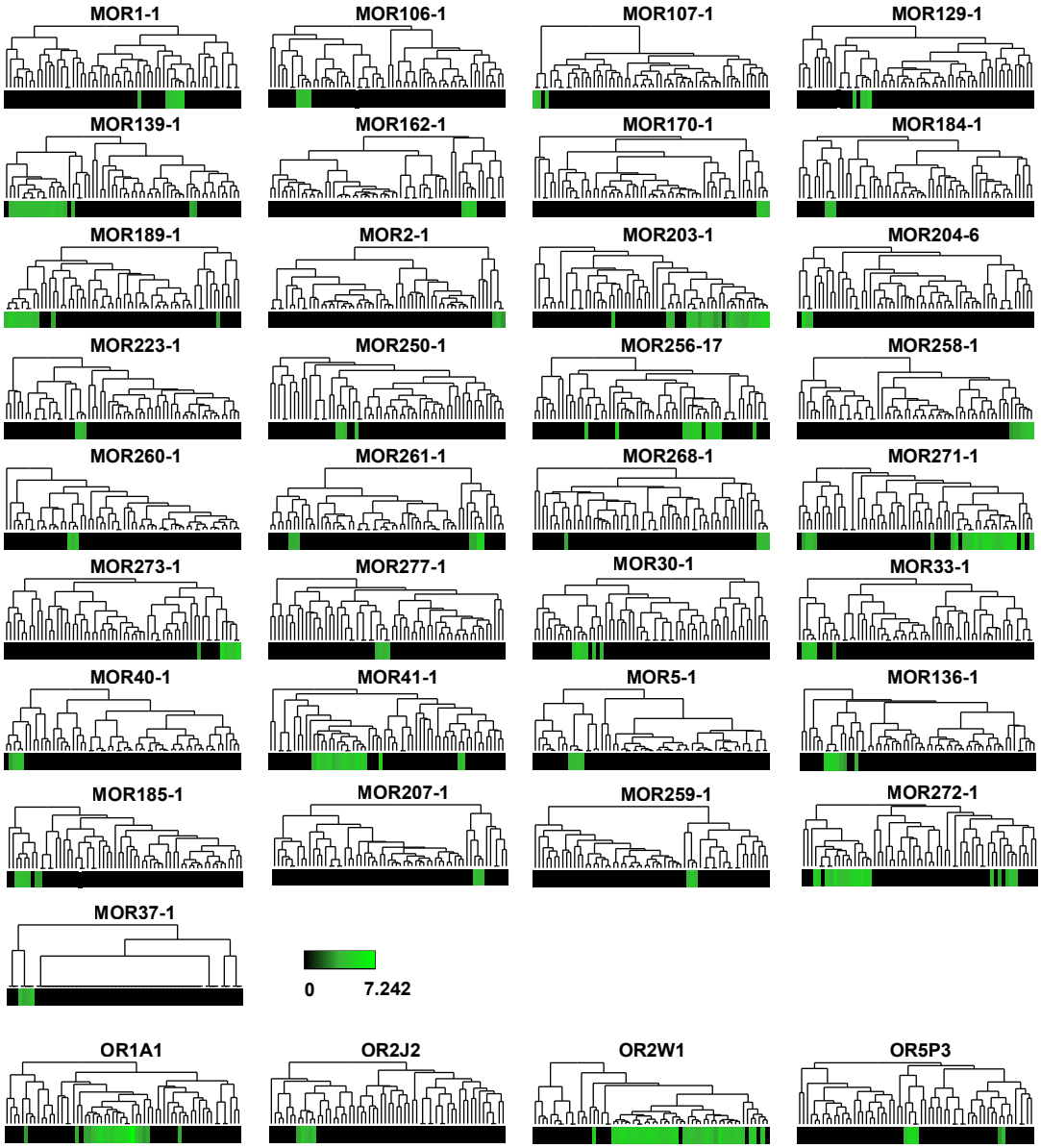

MOR1-1
MOR106-1
MOR107-1
MOR129-1
MOR139-1
MOR162-1
MOR170-1
MOR184-1
MOR189-1
MOR2-1
MOR203-1
MOR204-6
MOR223-1
MOR250-1
MOR256-17
MOR258-1
MOR260-1
MOR261-1
MOR268-1
MOR271-1
MOR273-1
MOR277-1
MOR30-1
MOR33-1
MOR40-1
MOR41-1
MOR5-1
MOR136-1
MOR185-1
MOR207-1
MOR259-1
MOR272-1
MOR37-1
0
7.242
OR1A1
OR2J2
OR2W1
OR5P3

Figure 3.16

Figure 3.17: Optimized descriptors for mammalian ORs effectively describe training set activity

Receiver-operating-characteristic (ROC) curve representing computational validation of ligand predictive ability of the Or-optimization approach. The mean true-positive value from 5 independent 5–fold cross validation runs for 5 mouse and 2 human receptors are plotted.

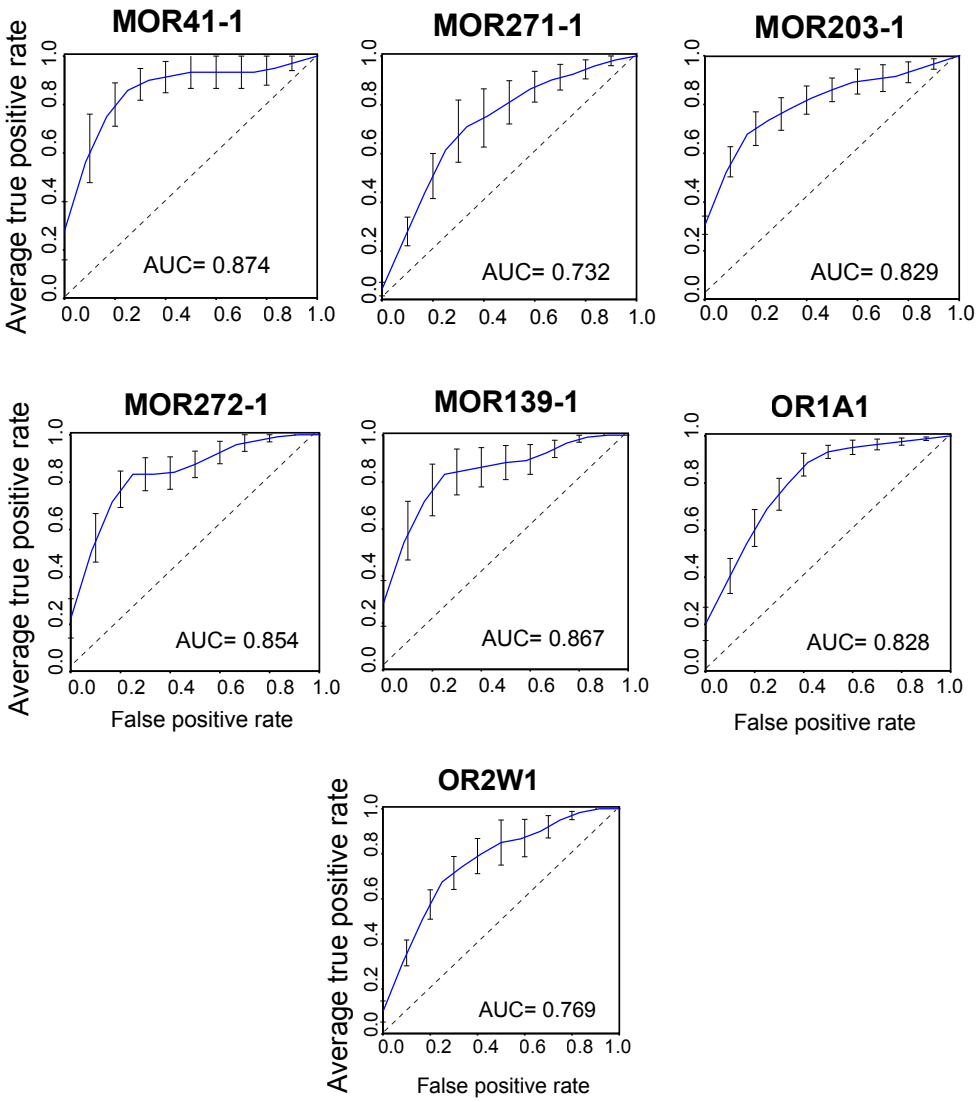

MOR41-1
AUC= 0.874
MOR271-1
AUC= 0.732
MOR203-1
AUC= 0.829
MOR272-1
AUC= 0.854
MOR139-1
AUC= 0.867
OR1A1
AUC= 0.828
OR2W1
AUC= 0.769
Average true positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

Figure 3.17

Figure 3.18: Analyzing relationships between important features for mammalian ORs

Hierarchical clusters created from Euclidean distance values between mammalian ORs calculated from the following sources: (left to right) shared optimized descriptors; activity to training odor set; similarity across top 500 predicted ligands; and Phylogenic tree of receptors. Phylogenetic tree of Mammalian ORs calculated from protein sequences. Subclusters are shaded with different colors or bars to enhance ease of comparison.

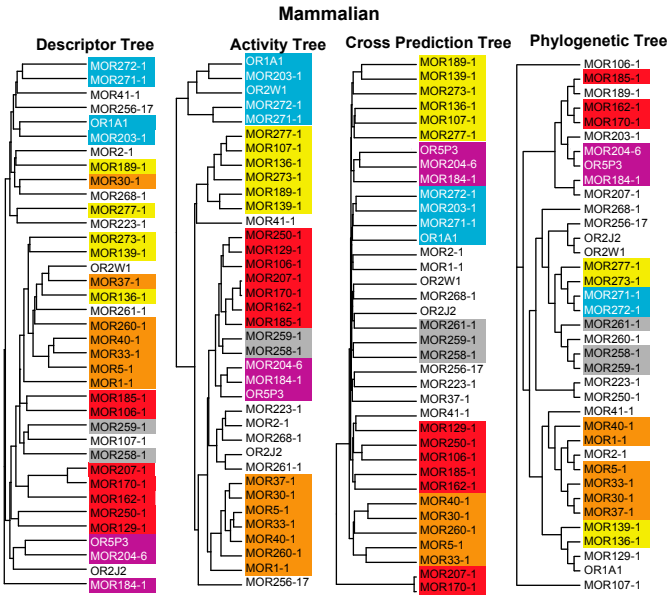

Mammalian
Descriptor Tree
MOR272-1
MOR271-1
MOR41-1
MOR256-17
OR1A1
MOR203-1
MOR2-1
MOR189-1
MOR30-1
MOR268-1
MOR277-1
MOR223-1
MOR273-1
MOR139-1
OR2W1
MOR37-1
MOR136-1
MOR261-1
MOR260-1
MOR40-1
MOR33-1
MOR5-1
MOR1-1
MOR185-1
MOR106-1
MOR259-1
MOR107-1
MOR258-1
MOR207-1
MOR170-1
MOR162-1
MOR250-1
MOR129-1
OR5P3
MOR204-6
OR2J2
MOR184-1
Activity Tree
OR1A1
MOR203-1
OR2W1
MOR272-1
MOR271-1
MOR277-1
MOR107-1
MOR136-1
MOR273-1
MOR189-1
MOR139-1
MOR41-1
MOR250-1
MOR129-1
MOR106-1
MOR207-1
MOR170-1
MOR162-1
MOR185-1
MOR259-1
MOR258-1
MOR204-6
MOR184-1
OR5P3
MOR223-1
MOR2-1
MOR268-1
OR2J2
MOR261-1
MOR37-1
MOR30-1
MOR5-1
MOR33-1
MOR40-1
MOR260-1
MOR1-1
MOR256-17
Cross Prediction Tree
MOR189-1
MOR139-1
MOR273-1
MOR136-1
MOR107-1
MOR277-1
OR5P3
MOR204-6
MOR184-1
MOR272-1
MOR203-1
MOR271-1
OR1A1
MOR2-1
MOR1-1
OR2W1
MOR268-1
OR2J2
MOR261-1
MOR259-1
MOR258-1
MOR256-17
MOR223-1
MOR37-1
MOR41-1
MOR129-1
MOR250-1
MOR106-1
MOR185-1
MOR162-1
MOR40-1
MOR30-1
MOR260-1
MOR5-1
MOR33-1
MOR207-1
MOR170-1
Phylogenetic Tree
MOR106-1
MOR185-1
MOR189-1
MOR162-1
MOR170-1
MOR203-1
MOR204-6
OR5P3
MOR184-1
MOR207-1
MOR268-1
MOR256-17
OR2J2
OR2W1
MOR277-1
MOR273-1
MOR271-1
MOR272-1
MOR261-1
MOR260-1
MOR258-1
MOR259-1
MOR223-1
MOR250-1
MOR41-1
MOR40-1
MOR1-1
MOR2-1
MOR5-1
MOR33-1
MOR30-1
MOR37-1
MOR139-1
MOR136-1
MOR129-1
OR1A1
MOR107-1

Figure 3.18

Figure 3.19: Analysis of mammalian OR tuning breadth

Frequency distribution of compounds from the >240K library within the top 15% distance from highest active plotted to generate predicted breadth of tuning curves.

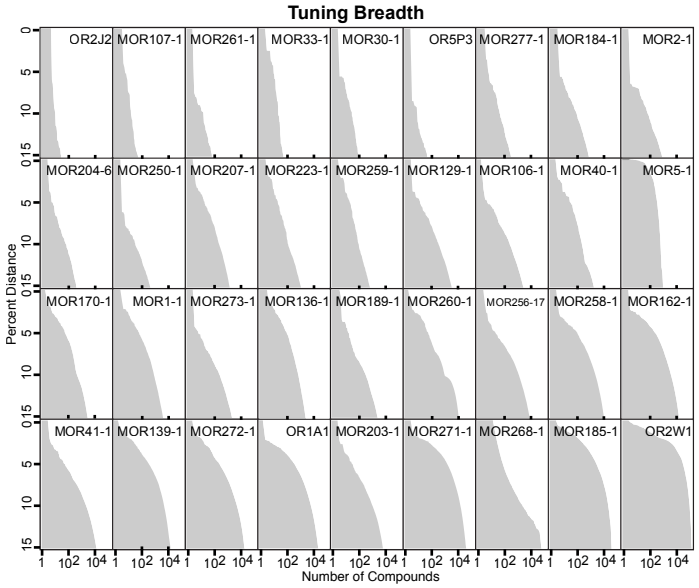

Tuning Breadth
OR2J2
MOR107-1
MOR261-1
MOR33-1
MOR30-1
OR5P3
MOR277-1
MOR184-1
MOR2-1
MOR204-6
MOR250-1
MOR207-1
MOR223-1
MOR259-1
MOR129-1
MOR106-1
MOR40-1
MOR5-1
MOR170-1
MOR1-1
MOR273-1
MOR136-1
MOR189-1
MOR260-1
MOR256-17
MOR258-1
MOR162-1
MOR41-1
MOR139-1
MOR272-1
OR1A1
MOR203-1
MOR271-1
MOR268-1
MOR185-1
OR2W1
Percent Distance
Number of Compounds

Figure 3.19

Table 3.1: Optimized descriptor sets for each *Anopheles* Or

Optimized descriptor symbols are listed for each Or. Descriptors are listed in the order in which they were included into the list. The same descriptor can be present multiple times for an Or representing the importance of the descriptor for a particular Or. The final correlation between the optimized descriptor set and odor activity is provided.

Or	Desc. Symbol		Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol		
AgOr57	AlogP98	C.043	AgOr12	B06.C.C.	AgOr50	SP05	AgOr38	AlogP98	nRCHO	nR.Cp.3
Correlation	O.057	Mor13p	**Correlation**	nRCOOH	**Correlation**	O.057	**Correlation**	DELS	Atype_C_40.1	H.051.2
0.933835	EEig10d	O.057.8	0.8635844	BELe2	0.9031393	piPC08	0.8536688	B06.C.C.	R5m.	nRCOOH.7
	nRCOOH	EEig10d.3		EEig10d		H1p		PW4	C.026	nRCOOR.4
	B07.C.C.	T.N..N.		nRCO		R8m.		nRCOOH	nRCO.2	MATS7 v.2
	Hy	nRCOOH.9		nRCHO		nCconj		EEig10d	MATS2e.1	MATS7 v.3
	Ds	B07.C.C..6		C.040		E2u		GATS2e	C.008	
	EEig02d	BIC		Hy		Gs		HATS8e	B07.C.C..4	
	Mor27p	GV WAI.80.2		B06.C.C..1		JGI6		O.057	Atype_C_40.2	
	B07.C.O.	O.057.9		BEHp4		L2u		G1u	RDF035m.1	
	Ms	B06.C.C.		nRCOOH.1		O.057.1		C.040	MATS7 v.1	
	O.057.1	nRCOOH.10		EEig 11d		B07.C.C.		B06.C.C..1	MATS4m.1	
	HATS8v	nCbH.1		HATS8u		piPC05		JGI2	J.4	
	MATS2e	nRCHO.3		O.057		De		RDF035m	Hy.2	
	EEig10d.1	nRCO.2		C.001		JGI7		B07.C.C.	H.051.1	
	nRCOOH.1	MATS3m.1		JGI6		H0e		Ms	O.057.7	
	nRCHO	B02.C.C..1		Ms		MATS6v		nR.Ct	nR.Ct.1	
	GV WAI.80	nR.Ct.1		Ms.1		C.040		MATS2e	Vindex	
	O.057.2	O.057.10				X5 Av		O.057.1	B06.C.C..8	
	F03.C.N.	nR09.1				nR.Ct		F03.C.N.	Atype_C_40.3	
	BELe2	nRCOOH. 11				RTu.		B06.C.C..2	MEcc.1	
	S_dssC	B07.C.C..7				P2u		JGI3	B07.C.C..5	
	O.057.3	C.026.2				piPC08.1		nRCOOR	O.057.8	
	MATS3v	nRCOOH.12				O.057.2		nRCO	MATS3m.1	
	B07.C.C..1	B07.C.C..8				BEHp4		MATS7v	B06.C.C..9	
	O.057.4	nRCOOH.13				B07.C.C..1		O.057.2	E2s	
	nRCO	EEig10d.4				O.060		GATS2e.1	G2m	
	R3u.	JGI1				Mor25e		Hy	BEHm3	
	EEig09r	JGI1.1				X5 Av.1		B07.C.C..1	B06.C.C..10	
	nRCOOH.2					C.026		nR.Cp	G.N..O..1	
	MATS7v					Hy		O.057.3	O.057.9	
	Hy.1					E2e		MATS4m	nRCOOR.1	
	C.026					Ds		J	nR.Cp.2	
	Infective.80					F04.C.O.		EEig10d.1	Mor10u	
	nRCOOH.3					B07.C.C..2		R4e.	nArCOOH	
	B07.C.C..2					HNar		G.N..O.	nRCOOH.2	
	MATS2e.1					nCconj.1		B07.C.C..2	nRNH2	
	BELe2.1					R7m.		MEcc	B07.C.C..6	
	nRCOOH.4					F03.C.N.		C.040.1	Yindex	
	nR09					X4A		J.1	S_aaaC.1	
	nRCHO.1					B08.C.C.		B06.C.C..3	B06.C.C.. 11	
	nR.Ct					B03.C.C.		O.057.4	nRCO.3	
	O.057.5					RBF		H.051	B02.C.C.	
	DISPe					C.026.1		X4A	B06.C.C..12	
	GV WAI.80.1					MATS2e		B06.C.C..4	O.057.10	
	nCbH					R1e		nCb.	GATS8m.1	
	nRCOOH.5					nR.Ct.1		C.040.2	JGI3.2	
	Hy.2					C.040.1		GATS8m	nRCOOR.2	
	S_dCH2					B07.C.C..3		G1u.1	nRCOOH.3	
	MATS4v					B04.C.C.		MATS3m	C.026.1	
	B07.C.C..3					piPC08.2		J.2	E2e	
	Ms.1					E2e.1		Atype_C_18	MATS3m.2	
	O.057.6					O.057.3		B06.C.C..5	nRCO.4	
	EEig10d.2					O.057.4		nRCOOH.1	B06.C.C..13	
	G2m							R1u.	nRCOOH.4	
	nRCOOH.6							EEig10d.2	JGI3.3	
	B07.C.C..4							nRCO.1	B06.C.C..14	
	F03.C.N..1							O.057.5	nRCOOR.3	
	C.026.1							nR.Cp.1	G1u.2	
	Atype_C_18							Hy.1	nR.Ct.2	
	B02.C.C.							Atype_C_40	nRCOOH.5	
	HATS8e							B07.C.C..3	GATS2e.3	
	nRCOOH.7							S_aaaC	B06.C.C..15	
	MATS3m							JGI1	nRCOOH.6	
	nRCHO.2							B06.C.C..6	MEcc.2	
	nRCOOH.8							O.057.6	C.026.2	
	EEig12x							GATS2e.2	B07.C.C..7	
	O.057.7							B06.C.C..7	Yindex.1	
	nRCO.1							JGI3.1	J.5	
	B07.C.C..5							J.3	MATS4v	

Table 3.1

Or	Desc. Symbol			Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or
AgOr56	Ss	C.024.2	TIC4	AgOr21	SP02	AgOr20	O.058	AgOr18	Mor20m	AgOr4
Correlation	nHAcc	nRCOOH.3	nRCOOH.8	**Correlation**	O.057	**Correlation**	EEig06d	**Correlation**	BELe2	**Correlation**
0.9312913	X4Av	nRCO	S_sNH2.1	0.9052854	piPC08	0.8829924	L2s	0.9201605	O.057	0.896526
	B07.C.O.	Atype_C_18	EEig09d.3		TPSA.NO.		C.040		EEig10d	
	O.057	nR09.1	O.058.5		B06.C.C.		GATS7m		Gu	
	Hy	nRCOOH.4	MATS2e.7		H1p		X3Av		DISPe	
	EEig09d	B07.C.C..5	O.057.16		Yindex		De		nRCHO	
	MATS2m	C.026.2	GVWAI.80.4		ASP		C.040.1		EEig12r	
	S_dssC	MATS3p	nRCHO.6		O.057.1		EEig10r		MATS4p	
	B07.C.C.	BELe2.2	O.057.17		R8m.		BEHm4		J	
	Ms	Atype_C_40	S_ssO.5		nDB		O.058.1		O.057.1	
	nBnz	nRCHO.3	Mor04m.1		H1e		Hy		JGI7	
	nRCHO	nArCHO	MATS2e.8		TPSA.NO..1		JGI7		X3A	
	C.040	GVWAI.80.2	nRCOOH.9		EEig10d		E2e		Mor23m	
	BELe2	nRCOOH.5	EEig09d.4		nRCHO		nCconj		B06.C.C.	
	O.057.1	nR.Ct.3	EEig09d.5		BEHp4		HNar		nHDon	
	EEig05d	nRNH2			O.057.2		JGI6		Mor27p	
	MATS2e	O.057.7			B06.C.C..1		Ku		nRCOOH	
	B04.C.N.	CIC5			B04.C.O.		JGI5		GATS3m	
	O.058	S_ssO.1			C.040		B03.C.C.		Mor28e	
	O.057.2	EEig09d.1			X4Av		C.040.2		C.026	
	GVWAI.80	RDF150u			O.057.3		nRCHO		X4A	
	MATS8v	nRCOOH.6			R7m.		MATS4v		S_dssC	
	IC2	nRCO.1			Hy		C.040.3		E2e	
	O.057.3	O.057.8			E2u		nRCO		MATS2v	
	nR.Ct	B07.C.C..6			O.057.4		R8m.		nRCHO.1	
	C.040.1	nPyrroles			Neoplastic.80		IC1		nRCOOH.1	
	C.024	MATS3p.1			EEig10d.1		O.057		Infective.80	
	MATS3v	S_ssO.2			MATS2p		O.060		nR09	
	Xt	O.057.9			O.057.5		MATS6v		HATS3u	
	B07.C.C..1	nRCHO.4			B06.C.C..2		R1e		R8e.	
	nRCOOH	X4A.1			BEHp4.1		L2e		nRCOOH.2	
	F03.C.N.	O.057.10			F03.C.N.		piPC09		JGI1	
	O.058.1	B07.C.C..7			Ms		nRCOOH		C.026.1	
	C.026	nR.Ct.4			HATS6m		nR.Ct		RARS	
	MATS2e.1	O.057.11			B07.C.C.		De.1		MATS4v	
	nRCHO.1	C.043.1			O.057.6		PW4		B06.C.C..1	
	C.040.2	S_ssO.3			nRCHO.1		O.060.1		nRCO	
	X4A	MATS7v.1			Yindex.1		E2e.1		nRCOOH.3	
	B07.C.C..2	MATS2e.4			C.040.1		G2m		Gu.1	
	Hy.1	O.057.12			C.026		nRCOOH.1		nRCHO.2	
	O.057.4	Mor04m			B06.C.C..3		nRCHO.1		EEig12x	
	GVWAI.80.1	CIC5.1			ESpm05d		S_aaCH		nRCOOH.4	
	MATS7v	B07.C.C..8			BEHp4.2		C.040.4		X4A.1	
	C.043	O.057.13			nRCOOH		BEHe3		EEig10d.1	
	C.040.3	S_sNH2			EEig10d.2		nRCO.1		MATS5p	
	nR.Ct.1	EEig09d.2			H1e.1		JGI6.1		PJI2	
	nR09	O.058.3			H1e.2		Ku.1		PJI2.1	
	Ms.1	Mor24e.1					GATS6v			
	C.026.1	C.026.3					X5Av			
	MATS2e.2	MATS3p.2					C.040.5			
	O.058.2	nR.Ct.5					MATS5v			
	Mor24e	Atype_C_40.1					B07.C.O.			
	G1v	Ms.2					R6e.			
	nRCOOH.1	B07.C.C..9					R6e..1			
	B07.C.C..3	nRCHO.5								
	C.024.1	nRCOOH.7								
	C.040.4	GVWAI.80.3								
	BELe2.1	MATS2e.5								
	EEig10d	S_aaNH								
	nRCOOH.2	RDF130m								
	MATS1v	O.057.14								
	nRCHO.2	S_ssO.4								
	nR.Ct.2	X4A.2								
	O.057.5	O.057.15								
	S_ssO	O.058.4								
	O.057.6	C.026.4								
	Hy.2	MATS2e.6								
	B07.C.C..4	nR.Ct.6								
	MATS2e.3	B07.C.C..10								

Table 3.1 (Continued)

Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol
BLTF96	AgOr75	ATS5p	AgOr32	E1s	AgOr26	Mor31v	AgOr11	ATS6e	AgOr15	ARR
F05.C.O.	**Correlation**	S_sCH3	**Correlation**	B06.C.O.	**Correlation**	O.057	**Correlation**	HNar	**Correlation**	B06.C.O.
B08.C.O.	0.8265765	BELe5	0.6823186	Mor10p	0.9274835	B08.C.O.	0.889735	R8e.	0.9092773	BELe3
PW4		O.057		Atype_H_47		B03.C.O.		O.057		E2e
nRCOOH		ESpm15u		H.049		Mor27v		E2u		nAB
B05.C.O.		HATS7e		nRCOOH		S_dssC		X5A		Mor10u
Infective.80		H2m		HNar		X4A		HATS7p		HOMA
DISPm		G3s		GATS1v		B07.C.C.		Mor13m		R7m.
BEHp1		DISPe		nCconj		R6e.		EEig10d		SPH
nRCOOH.1		Mor07u		R7v.		Mor25e		Mor25e		E2e.1
HATS5v		ESpm15u.1		GATS7m		Mor27v.1		O.057.1		F04.C.N.
B08.C.O..1		P2s		CIC		nRCOOH		S_ssO		HOMA.1
B05.C.C.		nRCOOH		B06.C.O..1		nDB		E2u.1		nOHs
EEig09d		Infective.80		Mor10u		MATS2v		nDB		GATS6m
DISPm.1		BELe4		E1s.1		HATS8m		X5A.1		GATS3e
nR.Ct		GATS5v		RDF020m		Gs		B06.C.O.		B06.C.C.
Gs		R6u.		R8m.		O.057.1		BELv4		nRCOOH
MATS3e		B09.C.O.		B06.C.O..2		De		BELv4.1		X4A
nRCOOH.2		G3s.1		nRCHO		B07.C.O.				ESpm01d
Mor10u		C.026		DISPm		Mor27v.2				B06.C.C..1
R5e.		ESpm15u.2		R7e		BEHm1				HATS4u
B08.C.O..2		nRCOOH.1		MATS2m		nDB.1				MATS4v
G3s		HATS7p		Mor10p.1		F04.C.O.				B06.C.O..1
EEig10d		C.016		P2s		B08.C.O..1				HATS3u
GATS4p		C.016.1		P2s.1		O.057.2				B06.C.C..2
nRCOOH.3						Mv				BIC
EEig12r						E2u				nRCO
F03.C.N.						O.060				E2e.2
RARS						RDF040u				R2v.
nRCOOH.4						nR.Ct				JGI7
R5e..1						MATS3v				nAB.1
R8m.						O.057.3				GATS6v
Mor32v						F02.C.N.				X4A.1
HATS4u						B08.C.O..2				Infective.80
HATS4u.1						piPC05				nRCOOH.1
						C.026				B04.C.N.
						Mor25u				nOHs.1
						R5e.				B06.C.C..3
						B07.C.C..1				X4A.2
						nRCOOH.1				JGT
						Mor27v.3				HOMA.2
						MATS6v				C.026
						R5u.				Mor09u
						B08.C.O..3				G2s
						piPC06				nRCOOH.2
						nCconj				nRCHO
						Mor27v.4				Infective.80.1
						B08.C.O..4				MATS1v
						MATS2v.1				nRCOOH.3
						R1u.				nOHp
						R1u..1				E1p
										R2v
										R6m.
										BIC.1
										BIC.2

Table 3.1 (Continued)

Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or
AgOr39	ALOGP	AgOr53	Ss	AgOr16	BELe3	AgOr48	F06.C.O.	AgOr46	Mor25e	AgOr30
Correlation	RDF035m	**Correlation**	O.057	**Correlation**	nHAcc	**Correlation**	BELv7	**Correlation**	BEHp3	**Correlation**
0.8405153	nRCOOH	0.8962904	EEig09d	0.853112	EEig10r	0.8627068	piPC04	0.8647854	C.006	0.8920303
	R5e.		nHAcc		ARR		DISPe		E2u	
	RBF		BELe2		R5v.		R8u.		O.057	
	ALOGP.1		B04.C.N.		Mor10e		ESpm03u		EEig10d	
	HATS6e		MATS2v		nRCOOH		F06.C.O..1		HATS7m	
	B09.C.O.		HATS6e		EEig10d		O.057		B04.C.O.	
	MATS8m		O.057.1		R2v.		Mor18m		ESpm15u	
	B05.C.O.		Infective.80		X4A		GATS1v		O.057.1	
	nCconj		O.057.2		nRCHO		R8u		S_dssC	
	Mor10e		R7m.		S_dssC		JGT		S_dssC.1	
	F02.C.O.		R1u.		MATS2v		Atype_C_40			
	nRCOOH.1		Mor25p		B04.C.N.		G3s			
	R6u.		BELe2.1		BIC		DISPm			
	Mor27u		MATS3e		JGI7		piPC04.1			
	B06.C.O.		O.057.3		B04.C.C.		ESpm05u			
	Ke		nRCHO		EEig11x		S_dssC			
	nHDon		C.040		IC2		AROM			
	B09.C.O..1		EEig08d		nRCOOH.1		nRCOOH			
	R5e..1		Infective.80.1		E1p		B06.C.O.			
	MATS8m.1		F04.C.N.		EEig10d.1		EEig12x			
	MATS3v		E2s		X4A.1		E1s			
	RDF035m.1		Mor32p		Mor10e.1		E1s.1			
	nRCOOH.2		X3Av		TPSA.NO.					
	RBF.1		HATS6u		S_dCH2					
	R6u..1		O.057.4		EEig09d					
	nR.Ct		B06.C.C.		L2u					
	L3s		HATS4e		F04.C.O.					
	Mor23m		O.057.5		E1s					
	nRCHO		Mor10e		E2e					
	S_sOH		E2u		BIC.1					
	HATS6u		RTu.		Mor10e.2					
	EEig09x		EEig09d.1		EEig10d.2					
	R5e..2		MATS5v		C.027					
	C.008		SRW09		nRCOOH.2					
	SPH		C.040.1		Mor32m					
	DISPe		nR.Ct		B04.C.C..1					
	nDB		O.057.6		JGI7.1					
	X4A		EEig08d.1		Mor08p					
	B09.C.C.		MATS2e		MATS3m					
	R6u..2		E1p		EEig10d.3					
	RDF035m.2		B04.C.N..1		G1p					
	MATS7m		O.057.7		nRCHO.1					
	MATS7m.1		nArCO		nRCOOH.3					
			nRCHO.1		RDF035m					
			C.040.2		BIC2					
			C.026		JGI6					
			BELe2.2		B05.C.O.					
			MATS2v.1		J					
			Infective.80.2		X4A.2					
			nRCOOH		Mor16e					
			G2p		Mor16e.1					
			C.043							
			X4Av							
			Mor27e							
			C.040.3							
			C.026.1							
			BELe2.3							
			MATS6v							
			E2s.1							
			B06.C.C..1							
			Mor10p							
			Mor10p.1							

Table 3.1 (Continued)

Desc. Symbol		Or	Desc. Symbol	Or	Desc. Symbol		Or	Desc. Symbol	Or	Desc. Symbol
Kappa.3.AM	Ks.1	AgOr64	Mor14e	AgOr9	nAB	nAB.7	AgOr13	ARR	AgOr1	ESpm01d
O.057	nRCOOH.8	Correlation	R3v.	Correlation	ESpm01d	MATS3m.2	Correlation	B05.C.C.	Correlation	S_sOH
EEig01x	EEig11x	0.7699497	Mor25e	0.946032	Mv	R7m..2	0.8117099	EEig03r	0.7939964	B06.C.C.
H2m	EEig11x.1		G3s		B07.C.O.	nArCO.1		EEig03r.1		GATS2p
B06.C.O.			HATS8e		Mor18p	nAB.8				MAXDN
B09.C.O.			MAXDN		MATS3v	E1m				Mor16e
O.058			AROM		B06.C.O.	nAB.9				S_sOH.1
O.057.1			Mor13v		nAB.1	GATS6v.1				C.040
R8p.			Mor13m		H0e	F04.C.N.				IC
ESpm15u			MATS2p		nAB.2	BIC.1				E1s
EEig11d			JGI5		EEig11r	Infective.80				B06.C.C..1
G3s			R4e.		Mv.1	X4A				B03.C.N.
nRCOOH			R7m.		R7m.	EEig10r				O.058
Mor18m			G3s.1		HATS3v	nAB.10				GATS2e
MATS3e			Mor13v.1		EEig04x	BELv4.2				E1p
PW5			O.057		nAB.3	MATS2e.1				E1p.1
EEig12r			nRCHO		R3m	nAB.11				
O.057.2			nHDon		F03.C.N.	R7m..3				
G3s.1			Atype_H_51		B06.C.O..1	S_aaaC.1				
ESpm15u.1			C.040		MATS2e	RCI.4				
S_sOH			R2e		nAB.4	nRCOOH.3				
nCconj			RTm.		nCconj	X5A.3				
B01.C.N.			B06.C.C.		Mv.2	AMW				
B09.C.O..1			nCb.		JGI7	G2m				
C.026			nRCOOR		HATS3p	nR.Ct				
X5A			B03.C.C.		Gu	MATS3m.3				
nRCOOH.1			JGI5.1		nRCOOH	Mor18m				
Ks			R3v..1		nBM	B06.C.O..2				
EEig12r.1			Mor13v.2		nArCO	X5A.4				
B06.C.O..1			Mor13v.3		E3p	nRCOOH.4				
ESpm15u.2					nAB.5	E2e.1				
nRCOOH.2					MATS3m	O.056.1				
G3s.2					R8p.	R3m.1				
Mor10u					HATSe	R8p..2				
GATS5e					RCI	BIC.2				
nRCOOH.3					nAB.6	BIC.3				
C.026.1					EEig10d					
ESpm15u.3					GATS6v					
Hypnotic.80					X5A					
C.027					BELv4					
nRCOOH.4					X5A.1					
B06.C.O..2					S_dCH2					
ESpm15u.4					nRCOOH.1					
nR.Ct					R7m..1					
PW5.1					MATS4p					
nRCOOH.5					nBM.1					
nCb.					BELv4.1					
B09.C.O..2					nRCHO					
B06.C.O..3					SPH					
S_dCH2					S_aaaC					
ESpm15u.5					S_dO					
HATS6u					RCI.1					
nRCOOH.6					nArCHO					
nArOR					RCI.2					
B03.C.O.					H3m					
L2s					MATS3m.1					
O.057.3					R8p..1					
PW5.2					G1u					
ESpm15u.6					Mp					
R8p..1					E2e					
EEig12x					BIC					
O.057.4					JGI7.1					
G3s.3					T.O..O.					
MATS5p					RCI.3					
B09.C.O..3					Mor32u					
PW5.3					EEig12d					
nRCOOH.7					X5A.2					
G3s.4					O.056					
nCb..1					EEig09d					
ESpm15u.7					nRCOOH.2					

Table 3.1 (Continued)

Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or
AgOr31	TI2	AgOr10	nCar	AgOr27	Mor31m	AgOr6	nArCO	AgOr2	C.024	AgOr65
Correlation	O.057	**Correlation**	R8e.	**Correlation**	MEcc	**Correlation**	RCI	**Correlation**	HATS6v	**Correlation**
0.8778699	HATS8v	0.9310224	C.025	0.8758972	O.057	0.9545822	BELm5	0.8565036	HATS3v	0.863439
	Mor32e		ZM1V		R8m		O.057		E2e	
	G3s		H.047		MATS2m		ESpm03u		Gs	
	Mor10v		X5A		nBnz		R7m.		Gs.1	
	S_sOH		RTm		Ms		EEig03x			
	B06.C.O.		RCI		MATS8v		nDB			
	ESpm15u		MATS2p		C.006		Mor23m			
	R6u.		R8u.		B09.C.O.		E2e			
	nRCOOH		nBnz		R2e.		Kp			
	E3p		HATS5m		O.057.1		nCconj			
	nHDon		Kp		nCb.		S_sOH			
	nCrs		JGI1		B08.C.C.		ESpm01d			
	R5u.		O.060		X3A		nAB			
	GATS6v		nBnz.1		S_aasC		F04.C.O.			
	nRCOOH.1		GATS6m		C.040		B06.C.O.			
	B06.C.O..1		JGI6		nCq		GATS3e			
	Mor23m		B04.C.C.		MATS8v.1		GATS3e.1			
	Mor32e.1		Mor25p		MATS5p					
	HATS6u		MATS3v		C.024					
	IVDE		O.056		nRCOOH					
	O.057.1		nRCOOR		Neoplastic.80					
	nOHp		Mor25v		C.040.1					
	D.Dr05		H.051		X3A.1					
	B06.C.O..2		E3m		nCb..1					
	nRCHO		P2v		nRCHO					
	O.057.2		nOHs		MATS3e					
	B06.C.O..3		C.006		B07.C.C.					
	C.008		EEig04r		R6u.					
	nRCOOH.2		R4v.		C.040.2					
	R6u..1		Mor11p		MATS6v					
	C.026		GATS6v		C.024.1					
	E1s		B04.C.C..1		REIG					
	MATS5v		JGI6.1		R8m.					
	Mor32e.2		nBnz.2		MEcc.1					
	G3s.1		MATS3e		O.057.2					
	nCconj		HOMA		C.034					
	nOHp.1		HATS5m.1		O.057.3					
	EEig12x		RBF		MATS1v					
	JGI5		Mor07u		nOHp					
	nRCOOH.3		C.027		C.040.3					
	G3s.2		JGI7		HATS8v					
	JGI6		AROM		O.057.4					
	Mor10p		JGI1.1		EEig08d					
	R5e.		R8u..1		EEig08d.1					
	B09.C.O.		RCI.1							
	HATS6u.1		RCI.2							
	GATS6v.1									
	nRCOOH.4									
	GATS5v									
	nArCO									
	B05.C.O.									
	RDF125u									
	C.026.1									
	C.026.2									

Table 3.1 (Continued)

Desc. Symbol		Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol
nArOH	R6u.	AgOr8	R6m.	AgOr59	S_dCH2	AgOr63	nH	AgOr61	nHAcc	AgOr35	Hypnotic.80
TPSA.NO.	Ke.1	**Correlation**	GATS2p	**Correlation**	nOHs	**Correlation**	Dv	**Correlation**	EEig09d	**Correlation**	GGI9
REIG	E3s	0.7551107	F01.C.O.	0.85381	R8u.	0.8400413	O.057	0.8538082	Mor17e	0.8460269	R1u
HATS4m	nAB.2		J3D		nROH		B04.C.O.		O.057		Ui
B07.C.O.	R1m.		B07.C.O.		JGI4		H2m		MATS2e		S_dsCH
E1u	MATS4p.1		B06.C.C.		Hbond.acceptor		H.049		EEig10d		R8v.
SPAM	nRCHO.2		nDB		JGI7		R5e.		Gs		S_sOH
MATS4p	nRCOOH.3		JGI1		JGI5		O.057.1		PW4		GGI9.1
S_dO	nArCO.2		Atype_C_40		R1v.		Mor13p		B08.C.O.		nDB
Gu	nAB.3		ESpm05u		Mor17m		G3s		O.057.1		EEig09d
Mor16v	nArOR		EEig09d		ESpm01d		B08.C.O.		O.060		GATS4e
GATS7m	G2u.1		GATS4e		JGI5.1		GATS5v		PW4.1		JGI9
Mor32e	nAB.4		GATS4e.1		nR.Cs		Vindex		nHDon		DISPv
C.040	JGI5				nR.Cs.1		nRCOOH		O.057.2		S_dssC
B04.C.C.	Mor04m.1						Ui		Infective.80		JGI7
nRCHO	nAB.5						B09.C.O.		MATS2v		O.056
EEig11r	R8m.						B06.C.C.		Mor23m		Mor16p
AlogP	Mor16p.3						MATS7v		Mor23m.1		E1u
nArOH.1	R6u..1						MATS5e				DISPm
G2u	Atype_C_40						G3s.1				GGI9.2
nRCOOH	F03.C.N.						nCconj				Hypnotic.80.1
Mor10e	MATS4p.2						B08.C.O..1				JGI3
nArCO	nCconj.1						PW5				S_dsCH.1
MATS2v	HATS1p						B06.C.C..1				Mor09u
R8u.	E1u.3						EEig10d				R8v..1
nAB	E1u.4						B09.C.O..1				JGI4
F04.C.O.							G3s.2				JGI4.1
H0e							MATS4e				
Ke							O.057.2				
E1u.1							R8m.				
JGI7							GATS6v				
Mor16p							HATS6u				
Hy							Mor17m				
SPAM.1							Mor17m.1				
nRCOOH.1											
R4m.											
nAB.1											
H.049											
MATS4v											
nR.Ct											
C.040.1											
nR09											
HATS3p											
MATS7v											
E1u.2											
E2e											
Mor16p.1											
nCconj											
nOHs											
HATS4p											
nHDon											
MATS2e											
nArCO.1											
C.040.2											
B04.C.C..1											
nRCHO.1											
EEig11d											
JGI7.1											
B04.C.C..2											
Mor16p.2											
nArCOOR											
J3D											
HATSp											
nBM											
HATSe											
nRCOOH.2											
C.006											
Mor04m											
MATS2e.1											
Infective.80											

Table 3.1 (Continued)

Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol		Or	Desc. Symbol	Or	Desc. Symbol		Or
AgOr25	ESpm11u	AgOr3	BLTF96	AgOr76	EEig10d	X4A.3	AgOr44	ATS8m	AgOr66	EEig10x	MATS4p.1	AgOr43
Correlation	nRCOOH	**Correlation**	nRCOOH	**Correlation**	nRCOOH	Neoplastic.80	**Correlation**	EEig12x	**Correlation**	TI2	BEHp5.1	**Correlation**
0.7611209	nOHp	0.8205165	Infective.80	0.7581273	R1v	nRCOOH.6	0.8356094	nRCOOH	0.866496	DISPe	EEig12d.1	0.8854268
	EEig12x		B04.C.O.		MATS8m	EEig10d.13		EEig10r		SPH	JGI4.3	
	HATS3p		JGI7		nOHt	EEig09d.1		B03.C.O.		B05.C.O.	nRCOOH.3	
	B09.C.O.		nCconj		EEig10d.1	G2m.1		B07.C.O.		B08.C.O.	EEig10d.6	
	nN		G2v		X4A	MATS7p.1		Mor08m		Mor08p	B05.C.O..6	
	nCconj		nRCOOH.1		JGI7	nRCOOH.7		nRCOOH.1		TPSA.NO.	B09.C.O..4	
	nOHp.1		F02.C.N.		MATS4v	nR.Ct		EEig12r		EEig10d	SPH.2	
	nRCOOH.1		B09.C.O.		R1e.	nOHt.4		JGI7		JGI4	C.026.1	
	nThiophenes		PW5		B07.C.O.	B02.C.C..2		nRCHO		F04.C.N.	E1u	
	MATS5v		ATS1p		EEig10d.2	EEig10d.14		MATS3m		EEig12d	E1u.1	
	B09.C.O..1		E2s		MATS6v	MATS8m.5		SPH		HATS4u		
	ESpm14d		Infective.80.1		S_dssC	nRCOOH.8		DISPe		EEig09d		
	E1m		nRCOOH.2		nRCOOH.1	EEig10d.15		Gs		DISPe.1		
	HATS8m		EEig04x		EEig10d.3	MATS4v.1		O.057		GATS4v		
	GATS5e		Hy		B02.C.C.	X4A.4		Infective.80		B09.C.O.		
	nOHp.2		EEig10d		G2m	B06.C.O.		Mor16e		MEcc		
	O.057		nRCOOH.3		MATS8m.1	EEig10d.16		Mor16e.1		PW5		
	F03.C.N.		B02.C.C.		E2s	E2s.2				EEig10d.1		
	nCrs		B08.C.C.		EEig10d.4	E2s.3				nRCOOH		
	E1s		nArCO		MATS7p					B05.C.O..1		
	nRCOOH.2		B04.C.O..1		nOHt.1					Mor10u		
	nOHp.3		nRCHO		EEig10d.5					G2s		
	EEig12d		H.052		MATS4p					MATS8m		
	ESpm15u		nRCOOH.4		RDF085e					Mor13p		
	E1u		PW5.1		EEig09d					DISPe.2		
	EEig09d		Infective.80.2		X4A.1					B05.C.O..2		
	S_dsCH		RDF140m		nRCOOH.2					B09.C.O..1		
	X5A		B04.C.O..2		EEig10d.6					Infective.80		
	DISPe		nRCOOH.5		MATS8m.2					nHBonds		
	Ke		JGI7.1		nRCO					C.026		
	Ke.1		PW5.2		GVWAI.80					MATS5v		
			F02.C.N..1		S_aaaC					SPH.1		
			E2s.1		nRCOOH.3					B09.C.O..2		
			Mor10e		CIC4					B05.C.O..3		
			EEig10d.1		nOHp					EEig10d.2		
			nRCOOH.6		MATS8m.3					JGI4.1		
			G2v.1		MATS4e					nRCOOH.1		
			nCconj.1		EEig10d.7					MEcc.1		
			X5A		nOHt.2					EEig09d.1		
			Infective.80.3		nRCOOH.4					F04.C.N..1		
			nRCOOH.7		JGI7.1					Mor32e		
			C.043		EEig10d.8					C.008		
			B08.C.C..1		nCconj					BEHp5		
			PW5.3		B04.C.N.					EEig10d.3		
			Hy.1		R7m.					MATS4p		
			EEig09d		MATS2e					nPyridines		
			nRCOOH.8		X4A.2					B09.C.O..3		
			CIC		E2e					B05.C.O..4		
			MATS8m		EEig13r					MATS8m.1		
			GATS5e		EEig10d.9					HATS4p		
			MATS2m		PJI3					nRCOOH.2		
			EEig04x.1		MATS8v					Infective.80.1		
			Infective.80.4		E2s.1					JGI4.2		
			X5A.1		EEig10d.10					EEig10d.4		
			X5A.2		MATS4p.1					B05.C.O..5		
					B02.C.C..1					Mor10e		
					nRCHO					DISPm		
					nRCOOH.5					B05.C.C.		
					EEig10d.11					B10.C.O.		
					nOHt.3					EEig10d.5		
					JGI7.2					nThiophenes		
					nRCO.1					MEcc.2		
					JGI6					G2s.1		
					E1u					O.057		
					MATS6v.1					F04.C.N..2		
					EEig10d.12					Mor32p		
					MATS4p.2					Mor13p.1		
					MATS8m.4					GATS8v		

Table 3.1 (Continued)

Desc. Symbol	Or	Desc. Symbol		Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol	Or	Desc. Symbol
SNar	AgOr54	TI1	HATS4p	AgOr67	VRv2	AgOr42	BLTF96	AgOr41	Mor29v	AgOr33	EEig06x
nRCOOH	**Correlation**	EEig10d	MATS5m	**Correlation**	EEig12d	**Correlation**	DISPe	**Correlation**	O.057	**Correlation**	nRCOOH
EEig10d	0.8997646	nRCOOH	nOHp	0.8871336	S_sOH	0.84541	PW5	0.8698281	EEig10d	0.890219	EEig10d
Hy		BAC	EEig11x.5		Ui		Infective.80		RBF		Hy
GNar		BLTD48	nRCOOH.6		EEig09d		TPSA.NO.		DISPe		Mor32v
O.057		JGI4	EEig09d.4		nDB		HNar		H2m		Infective.80
Infective.80		B08.C.O.	JGI4.5		H2m		DISPm		Hy		O.057
PHI		B03.C.O.	MATS8v.1		MATS5e		E2s		Dv		Xindex
MATS3p		EEig11x	G3s		G3v		B08.C.O.		O.057.1		EEig09d
MATS8v		R2m.	MATS2p		S_sOH.1		Mor07u		Infective.80		MATS4v
BEHp1		EEig09d	H2m		JGI7		B05.C.O.		EEig09d		B09.C.O.
HATS3p		B05.C.C.	H2m.1		BEHm1		B09.C.O.		JGI4		nRCOOH.1
nRCOOH.1		RDF140e			E2u		Hy		JGI10		Mor10u
EEig09d		GATS8v			B09.C.O.		nRCOOH		E1e		R8m.
G2u		MATS5e			O.057		EEig10d		MATS4e		RARS
nHBonds		HATS5m			ESpm01d		JGI4		JGI7		EEig10d.1
B09.C.O.		Mor30m			EEig12d.1		L3m		O.057.2		F03.O.O.
S_sNH2		nRCOOH.1			IC1		X5Av		BEHm1		Mor30m
BEHp1.1		Infective.80			JGI4		HATS8m		DISPm		MATS5e
EEig12d		GATS4e			EEig09d.1		E2s.1		HATS5e		RDF150e
E1u		B09.C.O.			ESpm15d		JGI10		B09.C.O.		Infective.80.1
EEig10d.1		R4e			E1u		MATS3e		C.026		nRCOOH.2
E1m		Ui			HATS8m		Mor10e		SPAM		G1u
nRCOOH.2		nRCOOH.2			G3v.1		DISPm.1		nRCOOH		EEig09d.1
nRCHO		EEig10d.1			EEig11d		PW5.1		EEig10x		EEig09d.2
PW5		JGI4.1			nROH		EEig09d		nN		
BEHp1.2		nN			R5p.		MATS6p		PW3		
Mor32v		MATS8v			DISPm		nRCOOH.1		S_dssC		
HATS3p.1		R3u.			Infective.80		Infective.80.1		O.057.3		
Infective.80.1		MATS3p			O.057.1		GATS5e		nCrs		
O.057.1		B09.C.O..1			E2e		GATS5e.1		DISPm.1		
Jhetp		EEig11x.1			nN				DISPm.2		
EEig09d.1		Mor28u			B08.C.O.						
E1u.1		nHBonds			Jhetp						
E1u.2		JGI4.2			EEig09d.2						
		EEig10d.2			O.056						
		E1u			RDF125m						
		Ts			nR05						
		EEig09d.1			C.026						
		B05.C.C..1			G3v.2						
		nArOR			nRCOOH						
		B09.C.O..2			nDB.1						
		MATS8m			HATS8m.1						
		B05.C.C..2			HATS8m.2						
		EEig11x.2									
		nRCOOH.3									
		MATS6v									
		R5e.									
		Mor10e									
		JGI4.3									
		EEig09d.2									
		B02.C.C.									
		RDF150v									
		EEig11x.3									
		GATS4e.1									
		nRCOOH.4									
		C.034									
		MATS5p									
		GATS8v.1									
		EEig11x.4									
		B09.C.O..3									
		R5m.									
		JGI4.4									
		EEig09d.3									
		R2m..1									
		EEig04r									
		nRCOOH.5									
		EEig10d.3									
		GATS4e.2									
		B09.C.O..4									

Table 3.1 (Continued)

Or	Desc. Symbol
AgOr45	nRCOOH
Correlation	ATS1m
0.8828472	R2m.
	Hy
	B09.C.O.
	nRCOOH.1
	Mor13m
	nRCOOH.2
	nDB
	Mor10p
	B03.C.C.
	GATS5e
	EEig10d
	nRCOOH.3
	nN
	EEig12x
	MATS4p
	nOHp
	nRCOOH.4
	GGI8
	R2m..1
	nRCOOH.5
	B02.C.C.
	RDF140e
	MATS6p
	Infective.80
	R2m..2
	F03.C.N.
	nRCOOH.6
	JGI4
	nRCOOH.7
	JGI8
	MATS3m
	C.006
	R2m..3
	S_dssC
	nRCOOH.8
	nCrs
	B03.C.C..1
	RDF150p
	Hy.1
	HATS4p
	nRCOOH.9
	EEig09d
	GATS5p
	nHBonds
	B09.C.O..1
	nOHp.1
	nRCOOH.10
	R2m..4
	R2m..5

Table 3.1 Continued

Table 3.2: Natural odor library predictions found in the top 500 predictions for each *Anopheles* Or

The prediction type (Activators, Fishing Activators, or Inhibitors), chemical structure, and predicted distance from the known training activators is listed for each odor. Compounds marked in grey boxes represent odors from the training library.

Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators	
Or57		**Or12**		**Or50**		**Or38**		**Or56**		**Or21**		**Or20**	
Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist
Cc1ncsc1	0	C(=O)(C)c1sccc1	0	O=C1[C@@](C2)(C)CC[C@@]2([H])C1(C)C	0	CCC/C=C/C=O	0	s1cncc1	0	C(=O)(C)c1ccccc1	0	CC1(CCC(O1)C(C)(C)O)C=C	0
C(CCCC)O	0	Cc1ncsc1	0	CC1(CCC(O1)C(C)(C)O)C=C	0	c1(ccccc1)C=O	0	Cc1ncsc1	0	C=CC(CCC)O	0	Cc1sc(nc1C)C	0
CC(CCO)C	0	C(=O)(C)c1sccn1	0	Cc1sc(nc1C)C	0	s1cncc1	0	C(CCCCC)O	0	CC/C=C\CCO	0	C=CC(CCC)O	0
CCC/C=C\CO	0	C(C)Oc1sccn1	0	CC1(C(=O)C2CCC1C2)C	0.1952	C(C)Oc1sccn1	0	CC(CCO)C	0	CC1=CC(CCC1)O	0	C=CC(CCCC)O	0
CC/C=C\CCO	0	CC/C=C\CCO	0	CC(C)C12CCC(=O)C1C2	0.27902	CC(C)CC1=NC=CS1	0	CC1CCC(CC1)O	0	CCC=CCCO	0.13836	C=CC(CCCCC)O	0
CC1CCC(CC1)O	0	CC1CCC(CC1)O	0	CC1C2CC2(CC1O)C(C)C	0.28733	C(CCCC)O	0	C(CC=CC)O	0.32202	CC(CCO)C=C	0.18559	CC1CCC(CC1)O	0
C(CC)OC(=O)C	0	CC1=CC(CCC1)O	0	C(=O)(C)C1OC(CC1)(C)C=C	0.29512	C(CCCCC)O	0	CC(CO)CC	0.32466	CC(=CCO)C	0.23354	CCCC(CCC=C)O	0.28607
CCC=CCCO	0.04578	CCC=CCCO	0.00439546	CC1(C2CCC(C2)(C1O)C)C	0.29931	CCC/C=C\CO	0	CCC=CCO	0.32615	CC1CCC(=O)CC1	0.23747		
CCCC=CCO	0.0467	CC1(CCCC=C1)O	0.02200348	OC(C)(C)C1OC(CC1)C	0.33768	CC/C=C\CCO	0	CC(CCCO)C	0.32777	CC(C=CC)O	0.24233		
C1(CC=CCC1)CO	0.29354	O1C(C1CC)CCO	0.04337664	CC(C)C12CCC(C1C2)(C)O	0.37874	CCC=CCCO	0.131918	CCC(C)CCO	0.32925	CC(=C)CCO	0.2591		
CC(C(C)O)CC	0.39139	CCCC=CCO	0.04354943	CC1(C2CCC(C2)C1(C)O)C	0.38252	CCCC=CCO	0.141631	CCC=CCCO	0.33557	CC1CCCC(=O)C1	0.2718		
CCCC(C)O	0.39271	CC1=CC=CC=C1S	0.05512895	CC1=NC(=CC=C1)C	0.38646	CC=CCCCO	0.43901	CC(CCCC)O	0.34169	CC(CCC(=O)C)C	0.29317		
CC(C=CC)O	0.39862	CC1=CC=C(S1)C=O	0.05544335	CSC1=NC=CN=C1	0.38908	CC=CC=CCO	0.448354	CCC(CCC)O	0.36484	CCCC=CCO	0.31729		
CCC(C)CCO	0.41955	CC(=O)C1=CC=CO1	0.07779456	C[C@@H]1[C@H]2C[C@]2(CC1=O)C(C)C	0.39218	CC=CC=CC=O	0.467956	CC1(CCCC=C1)O	0.40337	CC(=C)CC(C)O	0.32104		
CC(CC(C)O)C	0.42129	CC1=C(SC=N1)C=C	0.07845724	CC1CCCC(C1=O)(C)C	0.40798	CSCCCCO	0.469796	CC(C(C)O)CC	0.40662	CC1CCCCC1=O	0.34758		
CC1(CCCC1)O	0.44045	C(C)C1C(CC1)O	0.08242568	C12CCCC(O1)(CO2)C	0.40825	CCC=CCO	0.501183	CC(CC(C)O)C	0.41673	CC1C(CCC1)C(=O)C	0.35081		
CC(CC)OC(=O)C	0.4433	CC1=C(C=CO1)SC	0.0903326	CC(C)C12CCC(O1)(CC2)C	0.41836	CC(CCCO)C	0.509375	CSCCCO	0.42004	C(CC=CC)O	0.35445		
CC1C(CCC1)O	0.44514	CC(=O)C1=NC=CN=C1	0.09356382	CCC1=NC(=CS1)C	0.41886	CCC=CC=O	0.535788	CSCCCCO	0.43056	CC1=CC=C(C=C1)C=O	0.37139		
C1(CCCCC1)O	0.46992	CCC1=NC=CN=C1	0.09366896	CCC(C)C1=NC=CS1	0.42959	C(CC=CC)O	0.575718	CC1CC(CC1)O	0.43518	CC1=CC=CC(=C1)C=O	0.37682		
CC=CC=CCO	0.47747	CC(=O)C1=CC=NC=C1	0.0980415	CC(C)(C=CC(C)(C)O)C=C	0.43827	CC1=CC=C(C=C1)C=O	0.603019	CCCC=CCO	0.43888	CCC=CCO	0.38434		
C(CC=CC)O	0.47871	Cc1cocc1	0.0993155			C1=CC=C(C=C1)CCO	0.611034	C(CCC=C)O	0.45538	CC=CCCCO	0.38522		
CC1=CC=CO1	0.53738	CC=CC=CCO	0.1063697			CC1=CC=CC(=C1)C=O	0.637007	CC=CCCCO	0.45797	CC1(CCCC=C1)O	0.39915		
CC(CCCO)C	0.54014	CC1=CC=CO1	0.1119358			CC1=CC=CC=C1C=O	0.65316	C1(CC=CCC1)CO	0.46236	C1CC=CC(C1)O	0.41159		
CC1CC(CC1)O	0.55011	CC1=CC=CC(=C1)C=O	0.113021			C1(CC=CCC1)CO	0.666086	CC(C=CC)O	0.47274	C1(CC=CCC1)CO	0.4336		
Cc1cocc1	0.55444	CCC1=CC=CO1	0.1134702			C(CC)c1occc1	0.681512	C=CCCCCO	0.47479	CCC1CCCCC1=O	0.45129		
CC=CCCCO	0.55595	CC(=O)C1=CN=CC=C1	0.1138653			CCSC	0.739457	CCC(CC(C)C)O	0.47641	C=CCCCCO	0.45964		
C1CC=NC1	0.60273					CC(CCCO)CC	0.759834	C1CC=CC(C1)O	0.47712	CCC(=O)CC(C)C	0.48283		
CC(CCCC)O	0.6081					c1(ccccc1)CO	0.762276	CCCC(C)O	0.49317	CC1=CC=C(S1)C=O	0.48388		
C1=COC(=C1)CS	0.64382					CN1C=CC=C1	0.764383	CCCC(CCC)O	0.49792	CC1CC(CC1)O	0.48809		
C1CC=CC(C1)O	0.6465					C1=CNC=C1	0.783687	CC(C)C(C)O	0.50499	CCC(C)CC(=O)C	0.48973		
CC1C(O1)C	0.64984					C1CCC(CC1)CCO	0.786011	CC(CO)C=C	0.5052	CCC(CCC)O	0.49915		
CC(CO)CC	0.65665					CCC1=CC=CO1	0.791385	C(C)C1C(CC1)O	0.50967	CC(=CC)CO	0.50695		
CCC(CCC)O	0.65837					O1COCC1	0.797689	CC1C(CCC1)O	0.52487	CC1=CC(=O)CCC1	0.50946		
CCC1CCC(=O)O1	0.66507					CC(CCCC)O	0.800245	CCC(CC)O	0.53731	CC(C)C=CC(=O)C	0.51043		
CN1C=CC=C1	0.66895					c1(ccccc1)C(C)O	0.806178	CC(CCO)C=C	0.57069	C=C(CO)CC	0.51199		
CCC(C)C(CC)O	0.67588					C=NCC	0.82486	CC=CC=CCO	0.57521	CC1=CCC(=O)CC1	0.51471		
O1CC=CC1	0.68315					CC1NCCS1	0.828945	CCC(C)(CC)O	0.58139	CC1=CCCC(=O)C1	0.52976		
CC1CCCC(=O)O1	0.69201							C1CC=NC1	0.60201	CC1C(=CCC1)C(=O)C	0.54279		
CCC=CCO	0.69244							CC(CCCO)CC	0.61211				
CCC(CC)O	0.70248							CCCCC(CC)O	0.61346				
CC1(CCCC=C1)O	0.7118							CCCC(C)C(C)O	0.61447				
CCCC(C)C(C)O	0.71214							CCC(C=C)O	0.6246				
CC1=CSC(=N1)C	0.71248							C1(CCCCC1)O	0.63862				
COC(=O)C=CC	0.72877							CCC(C)C(CC)O	0.64035				
								CCCCCC(C)O	0.64674				
								CCCCCCS	0.65727				
								CC(C)CCS	0.71423				
								CC1(CCCC1)O	0.7338				
								CC(C)(C=C)O	0.77596				
								Cc1cocc1	0.79495				

Table 3.2

Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators	
Or18		**Or4**		**Or75**		**Or11**		**Or15**		**Or39**		**Or48**		**Or46**	
Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist
C(=O)(C)c1ccccc1	0	CC(=O)CCCCC	0	CC(=O)CCCCCCC	0	c1(ccccc1)C=O	0	c1(ccccc1)C=O	0	CC(=CCCC(=O)C)C	0	CC(=O)CCCCC	0	CC(CCO)C	0
Cc1ncsc1	0	C(CCCCC)O	0	CC(=CCCC(=O)C)C	0	C(C)c1c(ccccc1)C	0	C(=O)(C)c1ccccc1	0	c1(ccccc1)COC(=O)C	0	CCC(=O)CCCCC	0	C(CC)OC(=O)C	0
Cc1scnc1C	0	C=CC(CCCCC)O	0	C(C)C(CO)CCCC	0	C(=O)(C)c1ccccc1	0	C(=O)(C)c1sccc1	0	c1(ccccc1)CCOC(=O)C	0	CC(=O)CCCCCCC	0	C(C(C)C)OC(=O)C	0
Cc1sc(nc1C)C	0	C=CC(CCCCC)O	0	C(C)C(CO)CCCC	0	CC1CCC(CC1)O	0	Cc1scnc1C	0	CC(=CC)CCC(=O)C	0.2872	C(C)C(CO)CCCC	0	CC(C)CCO	0
C(=O)(C)c1ncccc1	0	C(CC(C)C)OC(=O)C	0	C(CC(C)C)OC(=O)C	0	CC(=O)C1=CC=NC=C1	0.03501	C(=O)(C)c1ncccc1	0	Cc1c(ccccc1)COC(=O)C	0.3135	C=CC(CCCCC)O	0	CCCOC(=O)C	0
Cc1cocc1	0.19322	CCCCCC(=O)C=C	0.13557	C(C)C(C=O)CCCC	0.1037	c1(ccccc1)C(C)O	0.042619	CC(=O)C1=CC=CN1	0.28335	CCC(C)COC(=O)C	0.4063	C=CC(CCCCC)O	0	CC(C)COC(=O)C	0
CCC1=NC(=CS1)C	0.25548	CC=CCCC(=O)C	0.15001	CCCCCCC(=O)C	0.1592	C(C)(C)c1ccccc1	0.056389	c1(ccccc1)C(=O)O	0.30392	c1(ccccc1)CCC(=O)C	0.4279	CC(=O)CCCCCCC	0	CC(CCCO)C	0.04093
CC1=CC=C(O1)C(=O)C	0.25945	CCCCCCC(=O)C	0.17174	CCCCCCCC(C)O	0.1661	CC(=O)C1=CN=CC=C1	0.057519	CC1=CSC(=N1)C	0.36739	CCC(=O)OCC1=CC=CC=C1	0.4754	C(C)C(CO)CCCC	0	CC(C)CCOC=O	0.08521
CCC1=NC=CN=C1C	0.26219	CCCC=CCO	0.2067	CC(=CCCC(C)O)C	0.2484	CC1CCCC(N1)C	0.060244	C1C(O1)C2=CC=CC=C2	0.40256	CCCCOC(=O)C	0.4764	CCCCCC(=O)CC	0	CC(C)COC=O	0.08773
CSCC1=CC=CO1	0.27256	CCCCC=CC(=O)C	0.2164	CCCCCCOC(=O)C	0.2561	CC1=CN=C(C(=N1)C)C	0.063861	CC1=CC=C(S1)C	0.43591	CC(=O)OCCC1CCCCC1	0.5423	CCCCCC(=O)C	0	COCC(C)O	0.09717
CC1=NC(=CC=C1)C	0.27501	CC(CC)OC(=O)C	0.21911	CCCCCCC(C)O	0.2751	CNc1ccccc1	0.068495	CC1=NC(=CC=C1)C	0.46329	CC=CCCC(=O)C	0.5564	CCCCCCC(=O)C	0.1093	CCC(=O)OCC=C	0.09928
CC1=CSC(=N1)C(C)C	0.27851	C(CCCCCC)O	0.22923	CC(CCCCCO)C	0.2765	CC1CCCCC1	0.078844	CC1=CC=CC=C1C(=O)C	0.46589	CC(CCCC(=O)C)C	0.5567	CCCCCCCCC(=O)C	0.1569	COC(=O)CSC	0.09955
CSC(=O)C1=CC=CO1	0.28789	CC(CCCC(=O)C)C	0.24361	CC(C)CCNC(=O)C	0.2949	c1(ccccc1)C(C)N	0.07896	CC1=CC=CC=C1C(=O)O	0.47334	COC(=O)CCC1=CC=CC=C1	0.5654	CCCCC(C)CCCO	0.1989	C(C)OC(=O)CC	0.14788
CC1=NC=CN=C1SC	0.30513	CCCC(C)OC(=O)C	0.24656	CC(CCCC(=O)C)C	0.3059	CN1CCCCC1	0.07912	CC(=O)C1=CC=CO1	0.48581	CC1=CC=C(C=C1)OC(=O)C	0.5688	CCCCCCCC(CC)O	0.2019	C(C)OC(=O)CC	0.14937
CCC1=NC(=CN=C1)C	0.30672	CC(=O)C=CCC	0.24856	CCC(C)COC(=O)C	0.3399	CCC1=CC=CC=C1	0.081879	c1(ccccc1)C(C)(C)O	0.4875	CC(CCOC(=O)C)CC	0.578	C(C)C(C=O)CCCC	0.2357	COCC(=O)C	0.15981
CC(=O)C1=NC=CN=C1	0.31157	CC(=O)CCCC	0.25291	COC(COCC(C)O)C	0.3433	Cc1c(ccccc1)C	0.092333	C1=CC=C(C=C1)CS	0.4927	CCC=CCC(=O)C	0.6055	CC(CCCCCO)C	0.2423	CCCCOC(=O)C	0.16227
CCC(C)C1=NC=CS1	0.32113	CCC=CCC=O	0.25456	CCC(=O)OCCC(C)C	0.3438	CC1=CC=C(C)C=C1	0.096179	CC1=CN=CC(=N1)C	0.5322	CC(C(=O)C)CCCC	0.6161	CCCCCC(=O)C=C	0.2538	COC(=O)CCCOC	0.16919
CC1=CN=C(C(=N1)C)C	0.33036	CC=CCC=CC(=O)C	0.2639	CC(CO)CCCCC	0.3461			C1=CC(=CC=C1C=O)O	0.55684	CCCCOC(=O)CC	0.6319	CC(CCCCCCCC)O	0.2596	CCOC(=O)C=CC	0.17785
CSC1=NC=CN=C1	0.34012	CCCCC(=O)C=C	0.29107					CC1=CN=C(C(=N1)C)C	0.57083	CCC(C)CCC(=O)C	0.6371	CC(CCCO)CC	0.26	CCCOC(=O)CC	0.18052
CC1=CC=CC(=C1)C=O	0.34182	CC=CC=CCO	0.30323					n1cc(ccc1)C=O	0.58094	CCC(=O)OCC(C)C	0.664	CCCCCCCC(C=C)O	0.2639	CC(CO)CC	0.20258
CC1=CN=CC(=N1)C	0.34455	C(C=CC=CCC)O	0.32891					C1=CC(=CC=C1C(=O)O)O	0.59189	CCC(=O)OCCC1=CC=CC=C1	0.6745	CCCCCCC(C)O	0.2684	CC(CO)C=C	0.20742
CC1=CC=C(C=C1)C=O	0.34495	CC(C)CC(C)CC=O	0.33532					CC1=CC(=C(O1)C)C(=O)C	0.60931	CCCCCCC(=O)C	0.6779	CCCCCCCC(C)O	0.2761	C(CC(C)C)OC(=O)C	0.20993
C1=CSC(=C1)S	0.34594	CCCCOC(=O)CC	0.33844					CC(=O)C1=CC=NC=C1	0.61367	COC1=CC=CC(=C1)C=O	0.6824	CCCCCCC(C=C)O	0.2841	CC(C)CCOC(=O)C	0.23274
CC1=NC=CN=C1C	0.34797	CC(=O)C=CC=CCC	0.35905							CCC=CCC(=O)OC	0.6867	CC(=O)CCCCCC=C	0.2916	C(C)OC(=O)C	0.2435
c1(ccccc1)CC#N	0.35014	CC(C)C=CC(=O)C	0.40407							CCCOC(=O)CC	0.6889	CC(CCCCCCCCC)O	0.3226	CCOC(=O)C	0.2446
CCC1=CN=CC=C1	0.35849	CCCC=CC(=O)C	0.40857							C0C(=O)CC(C)CC	0.6903	CCCC(=O)CCCCCC	0.3546	CCC(=O)OCC(C)C	0.24929
		C(C)C(C=O)CCCC	0.41382							CC(C)C(=O)OCC1=CC=CC=C1	0.7006	CCCC=CCCC(=O)C	0.3599	CC(CCCCCO)C	0.2626
		CC(CCC(=O)C)C	0.41763							CC(C)CCC(=O)OC	0.7058	CCCC(CCO)SC	0.378	CCCC(=O)OCC=C	0.26289
		CCCC(=O)OCC=C	0.41827							CC(=O)OC1=CC=CC=C1	0.7106	CCCCCCCC(=O)CCC	0.3863	C(C)OC(=O)CCC	0.26564
		CC(C)CCNC(=O)C	0.42243							CCCOC(=O)C(C)C	0.715	CC(CO)CCCCC	0.4119	C(C)OC(=O)CCC	
		CCC=CCC(=O)OC	0.44296							CCC(C)CC(=O)C	0.7196	CC=CCCC(=O)C	0.4155	CC=CC(=O)OCC(C)C	
		C0C(=O)C=CCC	0.45416							CCCCCCOC(=O)C	0.7225	CCCCC=CC(=O)C	0.4273	C(CCCC)OC(=O)C	
		CCCC(C)C(=O)OC	0.45464							C(CCC=CC)OC(=O)C	0.7272	CCC=CCC(=O)C	0.4314	CCCCCOC(=O)C	
		CCC(C)COC(=O)C	0.45911							CCCCCC(OC)=O	0.7315	CCCC=CC(=O)C	0.4385	C(C=CC)OC(=O)CCC	
		CCC(C)CC(=O)C	0.46324							CCCC(C)C(=O)C	0.7377	CCCCCCCC(=O)CC	0.4412	CCCCOC(=O)CC	
		C0C(=O)CC(C)CC	0.46736									CC(CCCCCC(=O)C)C	0.4429	CCC(C)COC(=O)C	
		CCCCCC(OC)=O	0.46915									CC(=O)CCCC=CCC	0.4512	CC(CCOC(=O)C)CC	
		CCCC0C(=O)C	0.47139									CC(CC(C)O)CCC	0.4572	CC(CCO)C=C	
		CCC(=O)C=CCCC	0.5002									CCCCCC(CC)O	0.4608	CCC(C)CCO	
		C0CCCCCC	0.50079									C1CCC(CC1)CCO	0.4614		
		CC(CCCCCO)C	0.52123												
		CCC(C)CCC(=O)C	0.52257												
		CCCCC(=O)OC	0.52557												
		CCC=CCC(=O)C	0.52863												
		CCC=CCCC=O	0.5318												
		CCCC(C)C(C)O	0.53766												
		CC(N(C)CCCC)=O	0.54853												
		OC(C)C(=O)CCCC	0.55514												
		CC(C(O)CC(=O)C)C	0.55791												
		CC(C)C/C=N/OC	0.57019												
		CCOC(=O)CCSC	0.57586												
		OC(C)C(=O)CCC	0.58362												
		CCCC(CCO)SC	0.58721												

Table 3.2 (Continued)

Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators		Predicted activators	
Or30		**Or9**		**Or13**		**Or1**		**Or10**		**Or6**		**Or2**	
Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist
CC(CCC=C(C)C)CC=O	0	Cc1c(cccc1)O	0	c1(ccccc1)C=O	0	c1(ccccc1)O	0	c1(ccccc1)C=O	0	C(=O)(C)c1ccccc1	0	c1(ccccc1)O	0
CCC(=O)CCCCC	0	Cc1cc(ccc1)O	0	Cc1ccc(cc1)O	0	Cc1cc(ccc1)O	0	C(C)c1c(cccc1)O	0	C(=O)(C)c1sccc1	0	c1(ccccc1)C=O	0
CC(=CCCC(=O)C)C	0	Cc1ccc(cc1)O	0	C(=O)(C)c1sccc1	0	Cc1ccc(cc1)O	0	C12=C(NC=C2)C=CC=C1	0	C(=O)(C)c1sccn1	0	Cc1c(ccc1)O	0
C(CCCC)OC(=O)C	0	C(C)c1c(cccc1)O	0	Cc1scnc1C	0	C(CCCC)O	0	CC1=CNC2=C1C=CC=C2	0	C(=O)(C)c1ncccc1	0	C12=C(NC=C2)C=CC=C1	0
COC(=O)CCCCCCC	0	C(C)Oc1sccn1	0	c1(ccccc1)C#N	0	CC1CCC(CC1)O	0	CCC1=CC(=CC=C1)O	0.1015	CC(=O)C1=CC=NC=C1	0.01054	C1=CC(=CC=C1C=O)O	0.010196
CCCCCCCC(=O)CC	0.12523	Cc1scnc1C	0	c1(cccccc1)CO	0	C1(CCCCC1)O	0.12772	c1(ccccc1)C#N	0.1847	CC(=O)C1=CC=CO1	0.01229	c1(ccccc1)C(=O)O	0.0128
CCCC(=O)CCCCCC	0.18977	Cc1sc(nc1C)C	0	Clc1ccc(cc1)Cl	0	CC1=C(C=C(C=C1)O)C	0.18522	CC(C)C1=CC=CC=C1O	0.3233	CC(=O)C1=CC=CN1	0.02809	C1=CC(=CC=C1C(=O)O)O	0.017154
C(C)OC(=O)CCCCCC	0.2053	CC1=CC=CC=C1OC	0.27619	CC(=O)C1=CC=CS1	0	CC1=CC(=C(C=C1)O)C	0.18676	CC1=CC=CC(=C1)C=O	0.5822	CC(=O)C1=NC=CN=C1	0.03161	C1=CC(=CC(=C1)O)C=O	0.017917
COC(=O)CCC=CCCC	0.20875	CC1=CSC(=N1)C	0.41637	C1=CC=C(C=C1)CS	0	CC(C)C1=CC=C(C=C1)O	0.18882	CCC1=CC=CC=C1S	0.6009	CC(=O)C1=CN=CC=C1	0.04039	C1=CC=C(C=C1)S	0.019241
CC(CCCC(=O)C)C	0.21498	CCC1=CC=CC=C1S	0.47183	C1=CC(=CC=C1O)O	0	C=CC1=CC=C(C=C1)O	0.20264	CCC1=CC(=CC(=C1)C)O	0.6019	CC1=CC=C(S1)C=O	0.0597	O=C1SC2=CC=CC=C2N1	0.020891
CCCCCCOC(=O)CC	0.23016	CC1=CC=CC=C1S	0.53176	n1cc(ccc1)C=O	0.02253	CC1=CC(=C(C=C1)C)O	0.20755	CC1=CC=CC=C1C=O	0.6048	o1c(ccc1)C=O	0.06497	CC1=CC=C(C=C1)C=O	0.021228
CCCCCOC(=O)CCC	0.24319	COC1=CC=CC=C1O	0.56118	CC1=CC=C(S1)C=O	0.02492	CC1=C(C(=CC=C1)O)C	0.21059	c1(ccccc1)C(C)(C)O	0.6131	CC1=CC=C(O1)C=O	0.06673	s1cnc2c1cccc2	0.022034
CCCCCCC(=O)C	0.24475	COC1=CC=CC=C1	0.63001	CC(=O)C1=CC=CN1	0.0256	c1(ccccc1)C(C)O	0.21682	c1(ccccc1)C(C)N	0.654	CC1=C(SC=N1)C=C	0.0878	c1(ccccc1)NC=O	0.023269
CCCCCC(OC)=O	0.2877	CC1=CC=C(S1)C	0.70917			CCCC=CCO	0.22901	C1CC2=CC=CC=C2NC1	0.6613	CCC(=O)C1=CC=CC=C1	0.09306	CC1=CC=CC=C1S	0.028825
CCCCCC(=O)OCCC	0.31006	CC1=NC(=CC=C1)C	0.73839			Oc1ccc(cc1)OC	0.23076	Nc1c(cccc1)C=O	0.6737	C1=COC=C1C=O	0.09658	C1=CC=C(C=C1)C=NO	0.032971
CCCCC(=O)OCCCC	0.32045	CC1=C(C=CO1)S	0.74378			CC1=CC(=CC(=C1)O)C	0.23902	CC1=C(C(=CC=C1)O)C	0.6781	CC1=CC=CC(=C1)C=O	0.09833	C1=CC=C(C(=C1)O)Cl	0.03301
CCC=CCC(=O)OC	0.36075	C1=CC=C(C=C1)CS	0.78212			o1c(ccc1)CO	0.24295	C(C)c1c(cccc1)C=O	0.6902	n1cc(ccc1)C=O	0.09833	Cc1ccc(cc1)C(=O)CC	0.033088
CCCCOC(=O)CC	0.39692	C1=CC=C(C(=C1)N)O	0.89065			CCC=CCCO	0.24619	CC1=C(C=CC(=C1)C)O	0.6948	COC(=O)C1=CN=CC=C1	0.1036	COC1=CC=C(C=C1)C=O	0.033708
CC(=CCOC(=O)C)C	0.39767	CSCC1=CC=CO1	0.89947			CC1=C(C(=CC=C1)C)O	0.24803	CC1=C(C=CC(=C1)O)O	0.7107	CC1=CC=C(O1)C(=O)C	0.12116	C1=CC=C(C(=C1)[N+](=O)[O-])O	0.034061
CCCCCCC(=O)CC	0.43843	CSC1=CC=CC=C1	0.90741			CC1=C(C=CC(=C1)O)Cl	0.28643	CC1=CC(=C(C=C1)O)C	0.712			CC1=CC=CC(=C1)C=O	0.034772
CCCCCCC(=O)OC	0.4502	CC1=NC=CN=C1OC	0.90862			CC(C)CO	0.28787	CC1=CC(=C(C=C1)C(C)C)O	0.7147				
CCCCOC(=O)C	0.46677	C1=CC=C(C(=C1)CO)O	0.93744			CCC=CCO	0.2914	CC1=C(C=C(C=C1)O)C	0.7201				
CCCCCOC(=O)CC	0.49066	Cc1cocc1	0.98265			CC(=CCO)C	0.29263						
CCC(C)COC(=O)C	0.50843	CC1=CC(=CC(=C1)O)C	0.99873			CC(CO)CC	0.2991						
C(CCC=CC)OC(=O)C	0.51122	COC1=NC=CN=C1	1.00099			CCC(CC)O	0.32257						
CCCC=CCOC(=O)C	0.51488	CNc1ccccc1	1.00204			CC(CC(C)O)C	0.32611						
CCC=CCC(=O)OCC	0.52598	CCC1=CN=CC=C1	1.00636			CCCC(C)O	0.33756						
CC(C)CCOC=O	0.54372	CC1=CC=CO1	1.00813			c1(ccccc1)CO	0.34707						
CC(=O)CCCC=CCC	0.54985	c1(ccccc1)CO	1.00997			CC1(CCCC1)O	0.35191						
CC(C)CCNC(=O)C	0.55352	CC1=CC=C(O1)C	1.01527			C1CCC2CC(CCC2C1)O	0.35659						
CC=CCCC(=O)C	0.55731	COc1c(cccc1)N	1.02951			CC1C(CCC1)O	0.35719						
CCCC(=O)OCCC(C)C	0.56128	C1=CC=C(C=C1)N	1.04535			CC(C(C)O)CC	0.35725						
CCCCCCOC(=O)C	0.56304	C(CCC)c1occc1	1.05691			CCC(C)O	0.3574						
CCCCOC(=O)CCC	0.57665	CC1=CN=CC(=N1)C	1.1009			C1CC(OC1)CO	0.38537						
CCCC=CCCC(=O)C	0.57963	NC1=CC=CC=C1S	1.10182			C=C(CO)CC	0.38622						
COC(=O)CCCC(C)C	0.59023	CC1=CC=C(C=C1)OC	1.11908			CCC(C)C(CC)O	0.39381						
CCC=CCC(=O)C	0.60008					COCC(C)O	0.39605						
C(C=CC)OC(=O)CCC	0.60267					CC1CC(CC1)O	0.40156						
CC(CCOC(=O)C)CC	0.61102					CCC(CC(C)C)O	0.41058						
CCC(C)CCC(=O)C	0.61765					CC(C=CC)O	0.41171						
CCC=CCCOC(=O)CC	0.61767												
CCC=CCCOC(=O)C	0.62953												
CC(CCC=C(C)C)C=O	0.64155												
CCC(=O)OCCC(C)C	0.64353												
CCCCCC(=O)OC(C)C	0.64726												

Table 3.2 (Continued)

Predicted activators		Predicted Fishing Activators		Predicted Fishing Activators		Predicted Fishing Activators		Predicted Fishing Activators		Predicted Fishing Activators	
Or8		**Or32**		**Or26**		**Or53**		**Or16**		**Or64**	
Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	
CC(=O)CCCCC	0	CC1(CCC(O1)C(C)(C)O)C=C	0	O=C1[C@](C2)(C)CC[C@]2([H])C1(C)C	0	C1(=O)CCCCC1	0	c1(ccccc1)C=O	0	O=C1[C@](C2)(C)CC[C@]2([H])C1(C)C	
CCC(=O)CCCCC	0	CC/C=C\CCO	0	O=C1[C@@](C2)(C)CC[C@@]2([H])C1(C)C	0	s1cncc1	0	C(=O)(C)c1ccccc1	0	O=C1[C@@](C2)(C)CC[C@@]2([H])C1(C)C	
C=CC(CCCC)O	0	C(C)OC(=O)CCCCC	0	CC1(CCC(O1)C(C)(C)O)C=C	0	Cc1scnc1C	0	Cc1c(cccc1)O	0	CC1CCC(CC1)O	
C=CC(CCCCC)O	0	CCC=CCCO	0.23627	CC1(C2CCC(C2)(C1=O)C)C	0.00036	C(CCC)O	0	c1(ccccc1)CO	0.05272	C1(CCCCC1)O	
CCCCC(=O)CC	0.10224	CC(C)CCOC(=O)C(C)C	0.3159	CC(C)C1CCC(C=C1)(C)O	0.54132	CC(CCO)C	0	CC1=CC=CC=C1S	0.05644	CC1(C2CCC1(C(C2)O)C)C	
CC(=CCCC(C)O)C	0.22171	CCCCOC(=O)CCC	0.32239	CC(C)C12CCC(C=C1)(OO2)C	0.65187	CC(CO)CC	0.16223	CC1=NC(=CC=C1)C	0.0826	CC1(C2CCC(C2)(C1O)C)C	
CC(CCCC(=O)C)C	0.22206	CCCCCCOC(=O)C(C)CC	0.32365	CC1(C)C(CC1C2C)CC2=O	0.66225	C1(=O)C=CCCC1	0.19305	c1(ccccc1)C(C)O	0.09479	CCCC(C)C(C)O	
CC(C)C/C=N/OC	0.27028	CCCCCCOC(=O)CC(C)C	0.32449	C12(C(=O)CC(C1(C)C)CC2)C	0.66656	C1(CCCCC1)O	0.2259	C1=CC=C(C=C1)CS	0.09511	CC(=C)C1CCC1(C)CCO	
CC(CCC(=O)CC)C	0.27038	CCCC(=O)CC(=O)OCC	0.35142	CC1OC2(OCCC2)CCC1	0.69256	CC1CCCNC1	0.28112	COC1=CC=CC=C1	0.10214	C[C@]12C[C@@H]3C1C(C)(C)O[C@@H](O3)C2	
CC(=C)CCC(CC)O	0.27425	O1C(C1CC)CCO	0.37265	CC(C1OC(C=C)(C)CC1)CO	0.6968	CC(CO)C=C	0.29902	CC1=CSC(=N1)C	0.10415	OC1C2C(C)=CCC(C2)C1C	
CCC(C)CCC(=O)C	0.34766	CCCCCCOC(=O)C(C)C	0.37356	CC1(C2C(=O)C(CC1C2)C)C	0.701	CC(C)C(C)O	0.30959	CC1=CC=C(C=C1)C=O	0.1136	CC1(C2CC=C(C1C2)CO)C	
CCCC(CCC=C)O	0.3502	CC1(O)CCC(C(C)=C)C=C1	0.4074			CC1C(CCO1)S	0.32482	CSC1=CC=CC=C1	0.11853	CC1=CC(C2CC1C2(C)C)O	
CCCCC(=O)C(C)C	0.35434	CC(C)CC(=O)OCCC(=C)C	0.41182			CC1CCCCN1	0.32927	CC1=CC(=C(C=C1)C)O	0.12943	CC(C(C)O)CC	
CC(C(=O)C)CCCC	0.36722	c1(ccccc1)C(O)C(=O)C	0.44233			CCC(C)(CC)O	0.33909	CC1=C(C(=CC=C1)C)O	0.13249	C1(CC=CCC1)CO	
CCCC(=O)CCC	0.41715	CC(=CC)COC(=O)CCCCC	0.44856			CC(C(C)O)CC	0.34135	C1=CC(=CC=C1C=O)O	0.13379	CC1(C2C(CCC1C2)O)C	
CC(C(=O)CC)CCC	0.42254	CCC(=O)OC1=CC=CC=C1	0.45853			CC(C(C)S)O	0.34424	n1cc(ccc1)C=O	0.13702	CCC(C)C(CC)O	
CCCCCC(=O)C=C	0.43272	CC1(C(CC1C(=C)C)CO)C	0.46148			C1CSCC1=O	0.36189	CC1=C(C(=CC=C1)C)S	0.13936		
CCCCC(CC)O	0.43707	O=C1C(C)(C)OC(C)(C=C)CC1	0.46867			CCC(C)O	0.36354	CC1=C(C(=CC=C1)O)C	0.14		
CCCCCC(C)O	0.44125					C(C)C1C(CC1)O	0.36563	CC1=C(C=CC(=C1)O)Cl	0.14146		
CCCCCC(CC)O	0.44524					CC1=CC=CO1	0.36679	C1=CC=C(C=C1)CCO	0.14307		
CCCCC(=O)C=C	0.47115					CC(C)(CCO)S	0.38597	CC1=CC=C(O1)C	0.14687		
CCCC=C(CC)CO	0.47199					CCCC(C)O	0.38814	CNc1ccccc1	0.14727		
CCC=CCC(=O)C	0.47255					CCC(C=C)O	0.38925	C=CC1=CC=C(C=C1)O	0.14751		
CCC(C)CC(=O)CC	0.47386					CN1CCCCC1	0.41875	CCC1=CC=CO1	0.1488		
CCC(=O)C=CCCC	0.47679					C1CCNCC1	0.42189	c1(ccccc1)COC	0.15768		
CC=CCCC(=O)C	0.48135					[nH]1cncc1	0.42597	CC1=C(C=C(C=C1)O)C	0.1614		
C(C)C(C=O)CCCC	0.48664					C1=COC(=C1)CS	0.42691	c1(ccccc1)C#N	0.16576		
CCCCCCC(=O)C	0.49194					[nH]1nccc1	0.42694	C1=CC=C(C=C1)COCC2=CC=CC=C2	0.16988		
CCCC=CC(=O)C	0.49218					COC(C)CC	0.42754	CC1=CC(=C(C=C1)O)C	0.17198		
CC(=O)CCCC	0.50571					CSCCO	0.43189	Cc1cc(ccc1)C	0.17222		
CC(=CC)CCC(=O)C	0.51221					CC(C)(C=C)O	0.4372	C1=CC=C(C=C1)S	0.17682		
OC(C(=O)C)CCCC	0.51228					CSCCCO	0.43769	C1CCC(CC1)CCO	0.17892		
OC(C)C(=O)CCCC	0.51233					CC(C(=O)C)C	0.44017	CC1=CC=C(S1)C	0.18078		
CCCC(=O)CC	0.52138					CCC(CC)O	0.44354	CC1=CC=C(C=C1)OC	0.18086		
CCCCCCCCCCCCCCC(=O)CC	0.52827					CCCC(=O)C	0.44785	c1(ccccc1)OCC	0.18828		
CC(CCC(=O)C)C	0.52901					CC1NCCS1	0.44857	Cc1c(cccc1)C	0.18909		
CCCCCCCCCCCCC(=O)C	0.53754					n1ccccc1	0.45548	C1=CC=C(C=C1)C=NO	0.18998		
CCCC(=O)C(C)CC	0.54055					CCC1=CC=CO1	0.45744	C1=CC(=CC=C1CO)O	0.19022		
CCCCCCCCCCCCCCC(=O)C	0.54295					CC(C(=O)C)S	0.46011	NC1=CC=CC=C1S	0.19127		
CCCCCCCCCCCCC(=O)C	0.5459					CC(C=CC)O	0.4686	CCCCCCOC=O	0.19216		
CCC(CCC(C)C)O	0.54873					CC(COC=C)C	0.46903	C1=CC=C(C=C1)N	0.1945		
CCCCCCCCC(=O)CCCCCCCCC	0.55375					s1cccc1	0.4784				
CCCCCCCCCCCC(=O)CC	0.55541					Cc1cocc1	0.4798				
CC(=O)CCCCCCCCCCCCCC	0.56119					CC1C(=O)CCS1	0.48568				
CCCC(C)C=O	0.56799										
CCCCCC(=O)CCCCC	0.56822										

Table 3.2 (Continued)

Predicted Fishing Activators			Predicted Fishing Activators		Predicted Fishing Activators		Predicted Fishing Activators		Predicted Fishing Activators	
Or31			Or27		Or65		Or59		Or63	
Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist
0	CC(=CCC/C(=C/COC(=O)C)/C)C	0	O=C1[C@](C2)(C)CC[C@]2([H])C1(C)C	0	Cc1c(cccc1)O	0	C=CC(CCCC)O	0	CCC(=O)CCCCC	0
0	c1(ccccc1)COC(=O)C	0	O=C1[C@@](C2)(C)CC[C@@]2([H])C1(C)C	0	Cc1ccc(cc1)O	0	C=CC(CCCCC)O	0	C=CC(CCCC)O	0
0	COC(=O)CCCCCCC	0	CC1(C2CCC(C2)(C1O)C)C	0.0046	C(C)c1c(cccc1)O	0	CCCCCCC(C=C)O	0.02263	C=CC(CCCCC)O	0
0.07571	C(C)OC(=O)CCCCCC	0.14358	CC1(C(=O)C2CCC1C2)C	0.0062	CCC1=CC(=CC=C1)O	0.10759	CCCCCCCC(C=C)O	0.04185	C(CCCC)OC(=O)C	0
0.09705	CC(CCCC(=C)C)CCOC(=O)C	0.18468	CC1(C2CCC(=O)C1C2)C	0.0114	CC1=C(C(=CC=C1)O)C	0.25083	CCC(C=C)O	0.0791	C(CC(C)C)OC(=O)C	0
0.11173	CC(=CCCC(=CCOC(=O)C)C)C	0.20187	CC1(C2C(CCC1C2)O)C	0.0114	CC1=CC(=C(C=C1)C)O	0.25956	CCCC(CCC=C)O	0.08094	CCC(CC)COC(=O)C	0.18488
0.15284	CCCCCCOC(=O)CC	0.20838	CC1(C2C(=O)C(CC1C2)C)C	0.0129	CC1=C(C=C(C=C1)O)C	0.26811	CC1=CCC(CC1O)C(=C)C	0.47381	CCCCC(=O)CC	0.19518
0.18487	CC(CCC=C(C)C)CCOC(=O)C	0.2304	CC1(C2CCC(C2)C1(C)O)C	0.0172	CCCC1=CC(=CC=C1)O	0.35369	C=CCCCCO	0.49185	CCC(=O)OCC(C)C	0.20736
0.19363	CC(=CCCC(=CCOC=O)C)C	0.23787	CC(C)C12C(C=C(C)C1C2)=O	0.018	CC1=C(C=CC(=C1)O)Cl	0.37792	C(CCC=C)O	0.4922	CCCCOC(=O)C(C)O	0.23524
0.19597	CCCC=CCCC(=O)OCC	0.23893	CC1(C2CCC1(C(C2)O)C)C	0.0192	CC1=CC(=C(C=C1)O)C	0.41393	CC(CCO)C=C	0.49409	CCC(C)COC(=O)C	0.24862
0.19624	CCCCCCC(=O)OCC=C	0.24076	CC1CCCC(C1=O)(C)C	0.0198	CC1=CC(=C(C=C1)O)C	0.41484	CC(C)(C=C)O	0.49466	OC(C)C(=O)CCCCC	0.25102
0.19684	CCCCCCCC(=O)CCC	0.25091	CC1(C2C1C(CC(C2)C)O)C	0.0224	CC1=C(C(=CC=C1)C)O	0.4324	CC(CO)C=C	0.49729	OC(C(=O)C)CCCCC	0.25239
0.19878	COC(=O)CCC=CCCC	0.25377	C12(C(=O)CC(C1(C)C)CC2)C	0.0251	CC(C)C1=CC=CC=C1O	0.44034	C(CCCCCC=C)O	0.49947	COC(=O)CC(C)CC	0.2526
0.20022	CCCCCCC(=O)OC	0.25609	CC1=CC(C2CC1C2(C)C)O	0.0253	CCC1=CC(=CC(=C1)C)O	0.49056	C=C(CCC)O	0.51273	CCCCCCC(=O)C	0.28055
0.20029	CC(CCCCCC(=O)C)C	0.25888	CC1=CC(=O)C2CC1C2(C)C	0.0255	CC1=C(C(=C(C=C1)O)C)C	0.49518	CC(C)CCCC(C)CCCC(C)CCCC(C)(C=C)O	0.51626	COC(=O)C(O)C(C)CC	0.28207
0.20132	CCCCCCC(=O)CC	0.27203	CC1(O)C=CC2CC1C2(C)C	0.0283	CC1=C(C(=C(C=C1)C)O)C	0.51696	CC(CCCC(C)(C)O)C=C	0.51843	CC(N(C)CCCC)=O	0.28994
	CC(=O)CCCCCC=C	0.27624	CC1=CCC2C(=O)C1C2(C)C	0.0293	CC(C)C1=CC=C(C=C1)O	0.53789	C=CCCCCCCCCCO	0.52163	CCCC(C)C(=O)OC	0.30453
	CCCCCC(=O)OCCC	0.28335			CCC1=CC(=CC=C1)CO	0.86599			CC(C)CCNC(=O)C	0.31016
	CCCC(=O)CC(=O)OCC	0.29031			CCC(C1=CC=CC=C1)O	0.89011			CCC(C)CNC(=O)C	0.31146
	CCC(=O)OCCC(C)CCC=C(C)C	0.29974			Oc1ccc(cc1)OC	0.90092			N[C@@H]([C@@H](C)CC)C(OC)=O	0.3166
	CCCC=CCCC(=O)C	0.30144			CC(CO)C1=CC=CC=C1	0.91951			CCOC(=O)C(C)CC=C	0.32034
	CC(=O)CCCC=CCC	0.30914			c1(ccccc1)C(C)(C)O	0.94882			OC(C(=O)C)CCCC	0.32318
	C(C)OC(=O)CC=CCCCC	0.31214			CC(CC1=CC=CC=C1)O	0.95335			CC(CNC(=O)CC)CC	0.32654
	CC(CCC=C(C)C)CCOC=O	0.31259			c1(ccccc1)C(C)O	0.97427			CCCCOC(=O)CC	0.32816
	CCCCCC#CC(=O)OC	0.32292			C1=CC=C(C=C1)CCO	0.9912			OC(C)C(=O)CCCC	0.3297
	CCCCCCC(OC)OC	0.32458			CC1=CC(=C(C=C1)O)OC	1.03879			CC(CCOC(=O)C)CC	0.34256
	COC(=O)CC(=CCCCCCC)C	0.3289			COC1=CC=CC=C1O	1.0516			CCCC(C)OC(=O)C	0.34572
	CCC=CCCOC(=O)CC	0.33174			Cc1c(c(ccc1)C)CO	1.0985			CC(=CCCC(C)O)C	0.34589
	CCC(=O)OCC=C(C)CCC=C(C)C	0.33793			CC1=CC=C(C=C1)OC	1.14164			COC(=O)C(O)CC(C)C	0.34903
	CC(CCC(=O)CC)CC	0.33979			COC1=CC=CC=C1	1.15157			CC(C(=O)C)CCCC	0.35361
	CCCCC(=O)OCCC=CCC	0.34391			CC1=CC=CC=C1OC	1.15798			CCCCCCC(=O)OC	0.35646
	CCCCCCOC(=O)C	0.34591			CNc1ccccc1	1.16122			COC(=O)CCCC(C)C	0.35648
	CCC(C)COC(=O)C	0.34683			c1(ccccc1)CCOC	1.1648			CC(CCCC(=O)C)C	0.35945
	CC(=CC)COC(=O)C	0.3502			CC1=CC(=C(C=C1)OC)C	1.20152			CCCC=C(CC)CO	0.36183
	C(CCC=CC)OC(=O)C	0.35951			C=CC1=CC=C(C=C1)O	1.20494			CCC(=O)OCCC(C)C	0.36369
	CCC(CC)COC(=O)C	0.366			C(CC)c1occc1	1.21048				
	CCC(=O)OCC=C(C)CCCC(=C)C	0.37069			C1C(O1)C2=CC=CC=C2	1.21477				
	CCCCCC(=O)OCC(C)C	0.37486			C(CCC)c1occc1	1.22659				
	CC(CCOC(=O)CCCC)C	0.37828			c1(ccccc1)OCC	1.23506				
	O=C(OC)CCCCC(OC)=O	0.38056			c1(ccccc1)CO	1.23816				
	C(C)OC(=O)CCC(C)CC	0.39057			CC1=CC=CC=C1S	1.25784				
	CCCCCCC#CC(=O)OCC	0.3917			c1(ccccc1)C(C)N	1.27299				
	C=C(CCOC(=O)C)CCC	0.3924			C1=CC=C(C=C1)CCN	1.28624				
	CC(CC(=O)CC(=O)C)C	0.39703								
	COC(=O)CC1=CC=CC=C1	0.40361								
	COC(=O)CCC(C)CCCC	0.40475								
	COC(=O)CCC(C)CC	0.40545								
	CC(CCOC(=O)C)CC	0.40586								
	CCC=CCCOC(=O)C	0.40618								
	CCCCCOC(=O)CC	0.41297								
	CCCCC(CC)COC(=O)C	0.41704								
	CCCCCCC(=O)C	0.41736								
	C(CCCC)C(C(=O)[O-])CC	0.41885								
	CCCC=CCOC(=O)C	0.42487								
	CCCCCOC(=O)CCC	0.43247								

Table 3.2 (Continued)

Predicted Fishing Activators		Inhibitor Predictions		Inhibitor Predictions		Inhibitor Predictions		Inhibitor Predictions		Inhibitor Predictions		Inhibitor Predictions	
Or61		**Or35**		**Or25**		**Or3**		**Or76**		**Or44**		**Or66**	
Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist
CC(=O)CCCCC	0	C(=O)CCCCC	0	CC(=O)CCCCC	0	c1(ccccc1)O	0	s1cncc1	0	CCC(=O)CCCCC	0	CC(=O)CCCCC	0
C(CCCCC)O	0	CCC(=O)CCCCC	0	c1(ccccc1)COC(=O)C	0	Cc1c(cccc1)O	0	Cc1ncsc1	0	C(=O)(C)c1sccc1	0	CC(=CCCC(=O)C)C	0
CC(CCO)C	0	CC(=O)C(=O)C	0	C=CC(CCCCC)O	0	C=CC(CCCCC)O	0	C=CC(CCCC)O	0	C(=O)(C)c1ncccc1	0	C(CCCCC)O	0
C=CC(CCCCC)O	0	C(CCCC)OC(=O)C	0	C(C(C)C)OC(=O)C	0	CCC/C=C\CO	0	C=CC(CCCCC)O	0	C(CCCCC)O	0	C=CC(CCCCC)O	0
CCCC(CCC=C)O	0.0211	CCCCOC(=O)C	0.11113	C(CC(C)C)OC(=O)C	0	C(CCCC)OC(=O)C	0	C(CC)OC(=O)C	0	C=CC(CCCCC)O	0	C(CC)OC(=O)C	0
CCCC=CC(=O)C	0.0308	CC(CCC(=O)CC)C	0.13828	CCCCOC(=O)C	0.00938	CCCC=CCO	0.04339	C(C(C)C)OC(=O)C	0	CCCCC(CC)O	0.02591	CCCCCC(=O)C=C	0.17938
CCCCC=CC(=O)C	0.0361	CCCC(=O)CC	0.20649	COCCCCCC	0.00972	C1=CC=C(C(=C1)O)Cl	0.16864	CC1C(CCC1)O	0.05867	CC(CCC(=O)CC)C	0.03439	CCC(=O)C=CCCC	0.23941
CC(C)CC(=O)C	0.0377	CCCCC(=O)CC	0.21497	CC(=CCCC(C)O)C	0.01039	C1=CC=C(C(=C1)O)Br	0.24714	Cc1cocc1	0.10774	CCCCCC(CC)O	0.04433	CCCC=CCO	0.26631
C/C=C/[C@@H](O)CC(C)C	0.0392	CCCC(C)C(=O)C	0.29869	CCCCCCC(=O)C	0.01139	CCCCCOC=O	0.30732	[nH]1cncc1	0.11642	CCCC(C)C(=O)C	0.04859	CCCCCCC(=O)C	0.27547
C(CCC=CCC)O	0.0398	C(C)OC(=O)CCCCCCCCCCCCC	0.34213	CC(CCC(=O)C)C	0.01139	CC(C)CCOC=O	0.39921	C1=CSC(=C1)S	0.11789	CCCCC(C)C(C)O	0.05013	CC(C)C=CC(=O)C	0.28327
CCC(CC(C)C)O	0.0406	C(C)OC(=O)CCCCCCCCCC	0.35309	CC(CCCC(=O)C)C	0.01508	C(CCC=CCC)O	0.40594	C1CCC(C1)S	0.12239	CC(=CCO)C	0.05117	CC(=CC)CCC(=O)C	0.29221
CCCC=CCCO	0.0409	CCCCCCCCCCCCCCCCC=O	0.36475	C(=O)CC=CCCCC	0.01743	CCCCOC(=O)CC	0.40812	CC1=CC=CO1	0.12383	CC(C(=O)C)CCCC	0.05369	CC(CCCCCO)C	0.29584
CC(C(C)O)CC	0.0418	C(=O)CCCC	0.36499	CC(C)SCC1=CC=CO1	0.02413	CC(C(O)CC(=O)C)C	0.47909	CCC(C)C(CC)O	0.12774	CCCCC(=O)CC	0.05529	C(CCCCCC)O	0.30048
CC(=C)CCC(CC)O	0.0436	CCCCC(=O)C(C)C	0.36788	CCC=CCCC=O	0.02413	CCCCC(=O)OCCC	0.50851	C1CC(=O)OC1	0.13285	CC(CCCO)C	0.058	CC(CCC(=O)CC)C	0.31921
CCC(C)CCO	0.0439	CCCCC(=O)CCC	0.37223	CCCCCCC(C)O	0.02446	CCCCC(=O)OCC	0.50972	CC1=C(C=CO1)S	0.1353	CC(CCC(=O)C)C	0.05944	CC(CCC(=O)C)C	0.33167
CC(CCO)C=C	0.0453	CCCCCCCCCCCCCCCC(=O)OCC	0.37494	OC(C(=O)C)CCCC	0.02782	CCCOC(=O)C(C)C	0.51004	O1COCC1	0.1449	CCCC=C(CC)CO	0.05997	CCCC=CC(=O)C	0.33689
CC(CCC(=O)C)C	0.0469	CCCCCCCCCCCCCC(OC(C)C)=O	0.39092	CCCCCCOC=O	0.02916	CCC(=O)OCC(C)C	0.51957	CC1C(CCO1)S	0.14598	C/C=C/[C@@H](O)CC(C)C	0.06328	CCC=CCC=O	0.33917
CCC(=O)C=CCCC	0.0492	CCC(=O)CC(C)C	0.39105	CCCOC=O	0.02949	CCOC(=O)CCSC	0.5367	CCCC(C)C(CC)O	0.14858	CCCC(=O)CC(C)C	0.06331	CCCCC(=O)C=C	0.34681
C(CCCCCC)O	0.0503	CCCCCCCCCCCCOC(=O)CCC	0.39131	CCC=CCC=O	0.02983	CCCOC(=O)CC	0.55806	CC(CCCC)O	0.15585	CC(CCCC)O	0.0672	CC(CO)CCCCC	0.34934
CC(=CC)CCC(=O)C	0.0519	C(C)OC(=O)CCCCCCCCCC	0.39492	CCCCCOC=O	0.03016	CC(CO)CCCCC	0.57829			CC1=CC=CC(=C1)C=O	0.06731	CC=CCCC(=O)C	0.3583
CCC(C)C(CC)O	0.0532	CCCCCCCCCOC(=O)C(C)C	0.40389	CCCCCCC(=O)OC	0.0315	OC(C)C(=O)CCCC	0.58892			CC(C)C/C=N/OC	0.07114	CCC=CCC(=O)C	0.36115
C=CCCCCO	0.0532	CCC(C)CCC(=O)C	0.40697	CCCCOC=O	0.0315	COC(CC)CC(=O)[O-]	0.60046			C(CCC)c1occc1	0.07422	CC(=O)C=CCC	0.36858
CCC(=O)CC(C)C	0.0552	CCSC(=O)C	0.42005	OC(C)C(=O)CCCCC	0.0315	C0C1=CC=CC=C1	0.6086			CC(=C)CCC(CC)O	0.07588	CCCCC(=O)OC	0.37971
CCCC(=O)C(C)CC	0.0554	CCCCCCCCCCOC(=O)CCC	0.42264	CCCC(CCOC(=O)C)SC	0.03217	CCC=CCCO	0.61107			CCCCCC(=O)C=C	0.07651	CCCCC=CC(=O)C	0.38167
CCCC=CCO	0.0555	COC(=O)C(C)CCCC(C)CCCC(C)CC	0.45085	OC(C)C(=O)CCC	0.03217	CCC=CCCOC=O	0.63019			CC1=CC=CC=C1OC	0.07747	CCCC=CCCO	0.38698
CC(C)C=CC(=O)C	0.0583	C(=O)CCCCCCCCCCCC	0.45211	CCCC(C)C(C)O	0.03318	COC(=O)C(O)CCC	0.63655			CCCC(CCC)O	0.07818	CC(=O)C=CC=CCC	0.39861
CC(CCCC)O	0.0587	CCC(C)CC(=O)CC	0.45477			CC(=CC)COC(=O)C	0.66374			CC(C(=O)CC)CCC	0.0791	CC(C(=O)C)CCCC	0.40043
CCCC(=O)CC(C)C	0.0603	CCCCCCCCCCCCC=O	0.46023			CCCC(C)C(CC)O	0.67199			CCCCC(=O)C(C)C	0.07919	CC(CCCC(=O)C)C	0.40045
CCC(C)CCC(=O)C	0.063	CCCCCCCOC(=O)C(C)C	0.47164			CCC=CCC(=O)OC	0.67364			CCCCC(CC(C)C)O	0.07945	C(CCC=CCC)O	0.40312
CCCCC(=O)CC	0.0636	CC(C(=O)C)CCCC	0.4731			CCCCCC(OC)=O	0.67668			OC(C(=O)CC)CC	0.08212	COC(=O)C=CCC	0.40445
CC(CCCO)C	0.0639	C(CCCCCCCCCCCCC)OC(=O)CCCCC	0.48399			CCCCCCC(=O)C	0.67876			CC(C)CCC(OCC)=O	0.0827	CCC(C)CCC(=O)C	0.40842
CC(=CC=CC(=O)C)C	0.0646	CCCCCCCCOC(=O)C(C)CC	0.49233			CC(=CC)CCC(=O)C	0.68955			CCCC(C)O	0.08348	CCCC(C)OC(=O)C	0.41986
CCC(CCC)O	0.0651	C(CC)OC(=O)CCCCCCCCCCCCC	0.49309			CC(C)COC=O	0.69106			CCC(C)/C=N/OC	0.08432	CC(=CC=CC(=O)C)C	0.42038
CC(=CCO)C	0.066	C(=O)CCCCCCCCCCCCCC	0.49593			CC(=CCOC(=O)C)C	0.70445			CC(=O)CCCC	0.08491	CC=CCC=CC(=O)C	0.42075
C1(CCCCC1)O	0.0674	CCCCCCCCOC(=O)CCC	0.50032			C=CCCCCO	0.70949			CCC(C)CC(=O)CC	0.08515	CCCCC(=O)C(C)C	0.43966
CCCC=C(CC)CO	0.0674	CCCCCCCCCOC(=O)CCC	0.50138							CC1=CC=C(C=C1)C=O	0.08517	CCC(C)COC(=O)C	0.44531
CC(C)(CCO)S	0.0675	CC(C)CCNC(=O)C	0.50175							CC(=CCOC(=O)C)C	0.08538	CCC=CCCC=O	0.45984
CCC(C)CC(=O)C	0.0682	C(CCN=C(N)N)CN	0.50729							CC1=CC=C(C=C1)OC	0.0857	COCCCCCC	0.46749
CCC(C)CC(=O)CC	0.0683	CCC(C)CC(=O)C	0.50904							CCCC(CC(C)C)O	0.08582	CC=CC=CCO	0.46958
CC(CCCO)CC	0.071	CC(C(=O)CC)CC	0.51383							CCOC(=O)C=C(C)C	0.08587	CCCCC(=O)CC	0.47178
CC=CCCCO	0.0714	CCCCCCCCCCCC(OC)=O	0.5264							CCCC(=O)C(C)CC	0.0869	C(C)C(C=O)CCCC	0.4761
CC(CCCC(=O)C)C	0.0716	CC(C)CCCCCCCCC(OCC)=O	0.53636							COCCCCCC	0.08749	CC(=CC)COC(=O)C	0.47822
CC(=O)C=CC=CCC	0.0752	C(CC)OC(=O)CCCCCCCCCCCCCC	0.53795							CCCC(C)C(=O)OCC	0.08949	CC(CC)OC(=O)C	0.48924
CCC(C)C(=O)C	0.0754	CCCCCCCCCCCCCC=O	0.54558							CC(=CC=CC(=O)C)C	0.09059	C(C=CC=CCC)O	0.50674
CCCCCCC(=O)C	0.0754	SC(C)(CC(=O)C)C	0.55631									CCCCCC(OC)=O	0.50897
CCCC(CCC)O	0.0754	CCCCCCCCCCC(=O)OC(C)C	0.55633									CC(C)CC(C)CC=O	0.53763
CCCCC(=O)C(C)C	0.0778	CCCCOC=O	0.55735									CCC(=O)OCC(C)C	0.53781
CC1C(CCC1)O	0.0783	C(C)OC(=O)CCC(C)CCCCCCC	0.55764									CCC(C)CC(=O)C	0.53809
CC(CC(C)O)C	0.0802											CCCCCCC(C)O	0.55242
CC(=C)CC(C)O	0.0817											CC=CC=CC(=O)OC	0.55783
												CC(C)CCC(=O)OC	0.57366
												CCCCOC(=O)C	0.57946
												CCCC(C)C(=O)C	0.58684
												CCCC=C(CC)CO	0.58945
												CCC(C)CC(=O)CC	0.59268
												CC(CCCO)CC	0.60377
												CCC=CCC(=O)OC	0.60603
												CCCC(=O)OCC=C	0.60621

Table 3.2 (Continued)

Inhibitor Predictions		Inhibitor Predictions		Inhibitor Predictions		Inhibitor Predictions		Inhibitor Predictions		Inhibitor Predictions		Inhibitor Predictions	
Or43		**Or54**		**Or67**		**Or42**		**Or41**		**Or33**		**Or45**	
Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist	Structure	Dist
CC(=CCCC(=O)C)C	0	C(=O)CCCCCC	0	C(=O)CCCCC	0	CC(=O)CCCCC	0	C(=O)CCCCC	0	C(=O)CCCCCC	0	C(=O)CCCCCC	0
C(CCCCC)O	0	CCC(=O)CCCCC	0	CC(=O)CCCCC	0	CCC(=O)CCCCC	0	C(CCCC)O	0	C(CCCCC)O	0	CC(=O)CCCCCCC	0
C=CC(CCC)O	0	CC(=CCCC(=O)C)C	0	CCC(=O)CCCCC	0	CC(=CCCC(=O)C)C	0	C=CC(CCCCC)O	0	C=CC(CCCC)O	0	CC(=CCCC(=O)C)C	0
C=CC(CCCC)O	0	C=CC(CCCCC)O	0	C(CCCC)OC(=O)C	0	C=CC(CCCCC)O	0	C(CC)OC(=O)C	0	C=CC(CCCCC)O	0	CC(=O)C(=O)C	0
C=CC(CCCCC)O	0	C(CCCC)OC(=O)C	0	CCCCCCC(=O)C	0.10125	C(CC)OC(=O)C	0	C(CCCC)OC(=O)C	0	C(CC(C)C)OC(=O)C	0	C=CC(CCC)O	0
C(CC(C)C)OC(=O)C	0	CCCCCCC(=O)C	0.01708	CCCCOC(=O)C	0.15643	C(C(C)C)OC(=O)C	0	C(C(C)C)OC(=O)C	0	CCCCCC(C)O	0.046	CCCCCCC(=O)CC	0.1167
CC(=CC)CCC(=O)C	0.059101	COCCCCCC	0.01813	CC(=O)CCCC	0.19083	CCCCCC(=O)C=C	0.23129	C(CC(C)C)OC(=O)C	0	CCCCC(=O)CC	0.06481	CC(C=CC)O	0.15747
CCCC(C)OC(=O)C	0.068405	CCCCCCC(C)O	0.02347	CC(CCCC(=O)C)C	0.1978	CCCCCCC(=O)C	0.24032	CCCCC=CCCO	0.10879	CCC=CCCC=O	0.07322	CC(C)(C=C)O	0.17432
CCC(C)COC(=O)C	0.080603	CCC(=O)C=CCCC	0.02478	CC(CCC(=O)CC)C	0.23956	CCC(=O)C=CCCC	0.24245	CCCC=CCO	0.12302	CCCCCC(CC)O	0.08097	C(C)C(C=O)CCCC	0.17534
CCCC(CCC=C)O	0.128886	CC(CCCC(=O)C)C	0.02644	CCCC(=O)CCC	0.24237	CC(CCCC(=O)C)C	0.24837	CCCCOC(=O)C	0.12379	CCCCCCS	0.08185	C/C=C/[C@@H](O)CC(C)C	0.1858
CCCC(=O)OC(C)C	0.142767	CCCCCC(CC)O	0.0282	C(=O)CCC	0.27389	CC(C(=O)C)CCCC	0.31945	CCC=CC=O	0.1419	C(CCC=CCC)O	0.08212	CCCCCC(OC)=O	0.20332
CC(=C)CCOC(=O)C	0.1466	CC=CC=CC(=O)OC	0.03158	CC(C(=O)C)CCCC	0.27418	CCC=CCC(=O)C	0.32402	C(CCCCCC)O	0.17205	CCCCC(CC)O	0.08429	CCC(=O)C(=O)C	0.2328
CC(C(=O)C)CCCC	0.148485	CCCCCC(=O)C=C	0.03213	CCCC(=O)C	0.27753	CCCCC(=O)CC	0.33291	CC=CC=CC=O	0.17257	CCCC=CC(=O)C	0.0896	CCC(C=C)O	0.23444
OCCCCC(=O)C	0.149644	CCCCCC(OC)=O	0.03337	CCC(=O)CC(C)C	0.28534	CCOC(=O)CCSC	0.35194	CC(C)CCC(=O)OC	0.17884	CC(CO)CCCCC	0.10058	CC(C)CC(C)CC=O	0.33326
CC(CCO)C=C	0.153922	CCCCC=CC(=O)C	0.03614	CCCCC(=O)CC	0.29844	CCCCC=CC(=O)C	0.37184	CCC(C)COC(=O)C	0.19056	C=CCCCCO	0.10081	CC(=O)C(C)(C)O	0.3352
CCC(=O)CCC(=O)C	0.160348	COC(=O)C=CCCC	0.03802	CCC(C)C(=O)C	0.30936	C(CCCCCC)O	0.37637	CCC=CCCC=O	0.21783	C(=O)C=CCCCC	0.10541	CC(CC)OC(=O)C	0.33781
CC(=CCOC(=O)C)C	0.160903	CCCCCC(C)O	0.04029	COC(=O)CC(C)CC	0.31712	CCCCCCC(C)O	0.39092	CCCCC(=O)OC	0.22196	CCCCCC=CCO	0.10557	CC(C(=O)C)S	0.34124
C(C)OC(=O)CC(C)C	0.163069	CC(=CCCC(C)O)C	0.04269	CCCC(C)C(=O)C	0.31785	CC(=O)C=CC=CCC	0.3986	OC(C)(CC(=O)C)C	0.23143	CCC=CCCCCO	0.1135	OC(C(=O)C)C	0.3626
CCCC(=O)CC(C)C	0.164257	CCC=CCCC=O	0.05093	CCCC(=O)CC(C)C	0.33165	CC(CO)CCCCC	0.39876	COCC(=O)C	0.23973	C(C=CC=CCC)O	0.11601	C=C(CCC)O	0.36478
CCC(C)CC(=O)CC	0.164275	CC=CCC=CC(=O)C	0.05244	CCCCC(=O)C(C)C	0.3399	COC(=O)CC(C)CC	0.40521	OC(C(=O)C)CC	0.24103	CCC=CC=CC=O	0.11698	CC(CCCC(=O)C)C	0.39601
CCC(C)CCC(=O)C	0.167416	CC(=O)C=CC=CCC	0.06027	CCCCCC(OC)=O	0.34072	CCCCC(=O)C(C)C	0.41013	CC=CC=CCO	0.24583	CCCC(=O)C	0.11875	CCCC(C)OC(=O)C	0.39972
CCC(CCC)O	0.171184	CCCC=CC(=O)C	0.06134	CC(C)CCC(=O)OC	0.37237	CC(=CC)COC(=O)C	0.42162	C(=O)CCCC	0.24103	CCCCC(=O)C=C	0.12166	CCC(=O)C(=O)CC	0.40223
CCCCC(=O)C(=O)C	0.172674	CC(C(=O)C)CCCC	0.06476	CCC(C)CC(=O)CC	0.39231	CCCC=CCCO	0.4234	OC(C)C(=O)CCCC	0.25881	COCCCCCC	0.12233	CC(C(=O)C)CCCC	0.40593
CC(CO)CCCCC	0.17349	CC(CCCCCO)C	0.06618	CC(CCC(=O)C)C	0.3994	CC(CCCCCO)C	0.42897	CCCC(C)C=O	0.26235	CCCC=CCO	0.12436	SC(C)(CC(=O)C)C	0.40627
CC(=CC)COC(=O)C	0.177227	CCCCC(C)C(C)O	0.0667	CC(C)CC(=O)C	0.40135	CC=CCCC(=O)C	0.42936	OC(C(=O)CC)CC	0.26339	CC(=CCCC(C)O)C	0.12637	CCCCCCC(=O)C	0.41199
CC(CCC(=O)CC)C	0.179319	C(CCCCCC)O	0.07013	CCC(C)CC(=O)C	0.40249	CC(=O)CCCC	0.43969	CCCCCC(OC)=O	0.26713	CC(CCCC)O	0.12658	C(=CCCCCCC)O	0.42304
CCCC=CC(=O)C	0.18149	C(=O)C=CCCCC	0.07087	CCCCC(=O)OCC	0.40261	CC(=CC=CC(=O)C)C	0.44341	CCOC(=O)CSC	0.26768	CCCC=CCCO	0.12707	C(C)(C)OC(=O)CC	0.42318
CC(C)C/C=N/OC	0.184825	CC(CO)CCCCC	0.07106	CC1CCCC(=O)C1	0.40999	CCCC=CC(=O)C	0.44511	C(CCCCCCC)O	0.26969	CC=CC(=O)C	0.12892	CCC(C)C(=O)OC	0.43067
CCCCC(CC)O	0.18672	CCCCCOC=O	0.07429	CC(C(=O)CC)CC	0.41376	C(C=CC=CCC)O	0.44642	CC(C)CCOC=O	0.27405	CC(C)CCOC=O	0.13145	CCC(CCC)O	0.43106
CCCCCC(CC)O	0.186721	CC=CCCC(=O)C	0.07575	CCCCC(=O)OC	0.42675	C(CCC=CCC)O	0.45206	COC(=O)CCSC	0.28522	CCC=CCC(=O)C	0.13315	CC(=O)CCCCCC=C	0.43677
C=CCCCCO	0.187969	CCC(C)CCC(=O)C	0.07597	CC1CCCCC1=O	0.43005	CCC(C)CC(=O)CC	0.45242	CC(C)(C#C)O	0.29315	CC(CC)OC(=O)C	0.13586	C(=O)C=CCCCC	0.43853
CC(=CCOC=O)C	0.188353	CCCCC(=O)OCCC	0.07835	CCC(C)COC(=O)C	0.43403	CCC(C)COC(=O)C	0.45669	CSCCC=O	0.29364	C(CCCCCC)O	0.14336	CCC(C)C(CC)O	0.44015
CCCC(=O)C(C)CC	0.188926	C(CCC=CCC)O	0.07979	C(=O)CCCC	0.43419	CCCCC(=O)C=C	0.45679	COC(=O)CC(C)CC	0.29583	CCC(C)CNC(=O)C	0.14437	CCCC(=O)OC	0.44137
CCC(CC(C)C)O	0.189675	CCCCC(=O)C(C)C	0.08315	CCC(C)CCC(=O)C	0.44674	CCCCCC(CC)O	0.45857	CCC(=O)C=CC=O	0.29861	CC=CC=CC=O	0.14487	CC(=CC)CCC(=O)C	0.44203
CCCCC(=O)C(C)C	0.193014	CC(=CC=CC(=O)C)C	0.0852	CC(C)CCOC=O	0.44857	CCCCC(=O)OC	0.46275	C=CC(=O)C	0.30079	CCC=CCCO	0.14532	CCC(CC(C)C)O	0.44274
CCC=CCC(=O)C	0.197528	CCCCOC(=O)CC	0.08733	CCC(C)C(=O)OC	0.45881	CCCCOC(=O)C	0.46447	O1C(=O)C=CC1	0.30185	CC(=C)CCC(CC)O	0.14558	CCC=CCCC=O	0.44452
CC(CCCCCO)C	0.197723	CC(=CC)CCC(=O)C	0.09283	C(C)OC(=O)CC(C)C	0.46662	CCCC(CCC=C)O	0.46598	C(CCC)OCCO	0.30293	CC(=C)(C)OC(=O)C	0.14685	CC(C(C)O)CC	0.44644
CCC(CCC(C)C)O	0.198257	C(CCC)OC(=O)C=C	0.09327	CCCC(=O)C(C)CC	0.47172	CC=CC=CC(=O)OC	0.47428	CC(CO1)C=CC1=O	0.30471	CCCCOC(=O)C	0.14871	CC(CCC(=O)CC)CC	0.4465
CC(C(=O)CC)CCC	0.199389	CCC=CCC(=O)OC	0.09623	CC(C)C(=O)OC	0.47521	CCCCCC(OC)=O	0.47609	CC(=O)CC(=O)C	0.30791	CCC(C)CC(=O)CC	0.15069	COC(=O)CC(C)CC	0.44714
CCCCCCC(C)O	0.200584	CCC=CC=CC=O	0.09698	CC1CCC(=O)CC1	0.4784	CC(CC)OC(=O)C	0.48625	CCC=CCC=O	0.31165	CC(=O)C=CCC	0.15122	Oc1oc(cc1)C=O	0.44962
CCCC(CCC)O	0.201775	CC(CCC(=O)CC)C	0.09702	CC(C(=O)C)C	0.47848	CC(=CCOC(=O)C)C	0.48883	CC(C=O)CC=C	0.3128	CCCC(C)C(CC)O	0.15134	CCCCC(=O)OC	0.44998
CC(CCCC(=O)C)C	0.205507	CC(=C)CCC(CC)O	0.09898	CCCC(=O)CC	0.48637	CCC=CCC(=O)OC	0.49222	CCC=CCCOC=O	0.31781	CCC(CCC(C)C)O	0.15298	CCCCC(C)O	0.45187
CCC(C)/C=N/OC	0.205959	CCC=CCC(=O)C	0.10001	CCC(=O)OCC(C)C	0.49564	COCCCCCC	0.49308	CC1=CC=C(O1)C=O	0.32321	C(CCC)OC(=O)C=C	0.15529	CC(CC(C)O)C	0.45307
C=CCC(CC=C)O	0.20725	CC(C(=O)CC)CCC	0.10119	CC(C)CC(=O)OC	0.51387	CCCC(C)OC(=O)C	0.49839	C(CC=O)CC=O	0.32973	CCCC(=O)CC(C)C	0.15602	CCCC(CCC)O	0.45338
CCCCCC(C)O	0.207402	CCCCC(CC)O	0.1026	CCCC(C)OC(=O)C	0.51481	CC=CCC=CC(=O)C	0.49873	CCCCCC=CCO	0.3325	C(C)C(C=O)CCCC	0.1561	CCCC=CCCC(=O)C	0.45533
CCC(C)CCO	0.208306	CCC(CCC(C)C)O	0.10306	CC(C)CCNC(=O)C	0.5261	CCCC=C(CC)CO	0.50198	C(=O)CC=CCCCC	0.33264	CCC=CCC=O	0.15789	CC(=O)OC	0.45643
CC(CCCC)O	0.209289	CCCCC(=O)CC	0.10762	CC(C(=O)CC)CCC	0.55954	CCCCC(CC)O	0.50885	OC(C(=O)C)CCCC	0.33291	CCCCC=CCCO	0.15936	CCCCCCOC=O	0.45854
CCCC(CC(C)C)O	0.211751	CCCCOC(=O)C	0.10821	CCCC(C)C(=O)OC	0.56137	C(C)C(C=O)CCCC	0.51038	CC(=O)C=CCC	0.33725			CC(C)C(C)O	0.45885
CC(CCCO)C	0.21241	C(=CCCCC)O	0.10939	CCC(=O)CC	0.56309	CC(=O)CCSC	0.51225	CCOC(=O)CCSC	0.3413			CC(=O)CCCC=CCC	0.45892
CC=CCCC(=O)C	0.213161	CC(CCC(=O)C)C	0.11017	CC=O	0.56576	COC(=O)C=CCC	0.52033	C1=CC=C(C=C1)CS	0.34287			C=CCC(CC=C)O	0.46288
CC(C)CCOC=O	0.21517	CC(N(C)CCCC)=O	0.11447	CC(C)C=O	0.56717	CC(CCC(=O)CC)C	0.52496	CCCCOC(=O)CCC	0.34506			C(=O)CC=CCCCC	0.46605
CCCC(C)C(C)O	0.215426	CCCCC(=O)C(=O)C	0.1151	CC1CC(=O)CC1	0.56937	CC(CC(C)O)CCC	0.52875	CCCC(C)OC(=O)C	0.34609			CCCC=CC(=O)C	0.46703
CCCCOC(=O)C	0.215704	CC(=CC)COC(=O)C	0.11604	CCOC(=O)C(C)C	0.57175	CCCCC=CCCO	0.53877	C(C=CC=CCC)O	0.34614			CC(C)(C#C)O	0.46727
CCCCC(C)C(C)O	0.215793			CCCCOC=O	0.57178	CC(C)C=CC(=O)C	0.54876	CC=CC(=O)C	0.35381			CCCCCC=CC=O	0.47121
CC=CC=CCO	0.216517			CCCOC(=O)C(C)C	0.57674	CCCCCC(C)O	0.55137	Cc1cocc1	0.35634			CC(C)OC(=O)C	0.47159
CCC(C)C(=O)C=CC	0.220364			CC1CCCC(=O)O1	0.57956	CCC(C)/C=N/OC	0.567					CCCC(CC(C)C)O	0.47199
CC(=C)CCC(CC)O	0.221259			CCC=O	0.59242	CCCCOC(=O)CC	0.58251					CCC(C)(CC)O	0.47361
CC=CCCCO	0.221827			CCCC(=O)OCCC	0.6036	CC(=O)CCC=C	0.58476					CCC(CC)C=O	0.4761
CC(N(C)CCCC)=O	0.224469			CN(C)C(=O)C	0.60529	CCCCC(=O)OCC	0.58564					CCCC(=O)OC(C)C	0.47801
CC(CCCO)CC	0.225174			CC(CC)OC(=O)C	0.60612	CC(C)CCNC(=O)C	0.58586					CC(C)CCC(=O)C(C)C	0.47882
CC(=O)C=CC(=O)CC	0.22518			CC(N(C)CCCC)=O	0.60849	CC(N(C)CCCC)=O	0.59632					CCC(C)CCC(=O)C	0.48056
C(C=CC=CCC)O	0.228377			C(C)OC(=O)C(C)CC	0.61929	CCC(C)CCC(=O)C	0.60318					CCCCCOC=O	0.48663
CCCCC(=O)OC	0.230676			CCCOC(=O)CC	0.62947	CC(=CC)CCC(=O)C	0.61089						
CCCCC(=O)C=C	0.231865			CC(C)COC=O	0.63757	CC(C)CC(C)CC=O	0.61134						
CCC(C)/C=N/O	0.232626			CCCCC(=O)OC	0.64299	CC(=O)C=CCC	0.61689						
CC(=CC=CC(=O)C)C	0.232967			N1C(=O)CCCC1	0.65989	CCC=CCC=O	0.6202						
CCOC(=O)C=CC	0.233448			C1CCC(=O)NCC1	0.66392	CCC(C)CNC(=O)C	0.62049						
CCCC=CCO	0.233829					CCC(=O)OCC(C)C	0.62161						
CC(CC(C)O)CCC	0.233954												

Table 3.2 (Continued)

Table 3.3: Predicted odors validated as effective activators for several Anopheles Ors

Subsets of purchasable odors from prediction lists for several Ors were validated using single unit electrophysiology. The Or names, number of activators over the number of odors tested, accuracy percentage, and primary class (aliphatic or aromatic) of responding odors is listed for each validated Or. The total accuracy, as well as accuracies for aliphatic and aromatic responsive Ors, is also provided.

Ors	Active/Tested	Accuracy	Responding Class
AgOr1	5/9	56%	Aliphatic
AgOr2	1/6	17%	Aromatic
AgOr4	12/13	92%	Aliphatic
AgOr6	2/5	40%	Aromatic
AgOr8	7/9	78%	Aliphatic
AgOr10	8/14	57%	Aromatic
AgOr12	8/8	100%	Mix
AgOr20	2/15	13%	Mix
AgOr30	16/16	100%	Aliphatic
AgOr48	7/8	88%	Aliphatic
AgOr53	5/12	42%	Mix
AgOr56	13/14	93%	Aliphatic
Total		65%	

Aliphatic Only	84%
Aromatic Only	38%

Table 3.3

Table 3.4: Activity of a large panel of aromatic odors was tested against three aromaticly tuned Ors

Responses of the previously selected 37 aromatic odors are reported for Or2, Or6, and Or10. The number of odors tested, number activating at greater than 50 spikes/sec, and percentage are listed for each Or. An average activity is provided below.

General Aromatic Assay			
Odor Receptor	**Num Tested**	**Num >50 Spikes/Sec**	**Percentage**
AgOr2	37	4	11%
AgOr6	37	3	8%
AgOr10	37	3	8%
		Average	9%

Table 3.4

Table 3.5: Optimized descriptor sets for Or2, Or6, and Or10

Optimized descriptor symbols and final descriptor-activity correlations are listed for each *Anopheles* Or. Descriptors are listed in ascending order of when they were selected into the optimized set. The number of times a descriptor was selected in an optimized descriptor set represent weights.

Or	Descriptors
AgOr6	ARR
Correlation	R8e.
0.7075761	Mor07u
	BEHm7
	GATS2v
	C.040
	Mor16e
	C.024
	Infective.80
	nFuranes
	ARR.1
	H.049
	Mor30m
	R8e..1
	R3u
	Mor07e
	DISPv
	X4A
	B06.C.O.
	C.001
	TPSA.NO.

Or	Descriptors
AgOr2	ATS6m
Correlation	nCconj
0.7844095	H1v
	ASP
	EEig11r
	X4A
	nArCO
	E3p
	O.057
	Gs
	piPC09
	GATS3e
	B06.C.O.
	ESpm01d
	Mor23m
	nROH
	Infective.80
	E2e
	B04.C.N.
	B06.C.O..1
	piPC05
	EEig11d
	O.057.1
	H1m
	nRCHO
	R3e.
	nRCOOR
	MATS3e
	Gm
	Infective.80.1
	nROH.1
	Mor28e
	JGI4
	MATS7v
	nR.Cp
	PJI3
	C.006
	O.057.2
	piPC09.1
	Mor08p
	nPyridines
	R6e.
	B06.C.O..2
	ESpm01d.1
	EEig11d.1
	O.057.3
	R3e..1
	G3s
	ESpm05u

Or	Descriptors
AgOr10	Dv
Correlation	ARR
0.7818453	Infective.80
	Mor18p
	JGI6
	nBnz
	HATS6m
	X5A
	B07.C.C.
	N.075
	HOMA
	Infective.80.1
	nROH
	HATS5m
	C.006
	Mor28u
	RCI
	B07.C.C..1
	MATS3e
	JGI6.1
	piPC05
	Infective.80.2
	RCI.1
	nFuranes
	EEig14r
	RCI.2
	GATS6v
	N.075.1
	EEig11d
	ARR.1
	HATSv
	JGI7
	PCR
	Infective.80.3
	Ds
	C.043
	JGI6.2
	Mv
	EEig13d
	HOMA.1
	Infective.80.4
	X5A.1
	G3e
	PCD
	B07.C.C..2
	Mor32m
	nFuranes.1
	RCI.3
	nPyridines

Table 3.5

Table 3.6: Top 75 predicted compounds for each *Drosophila* Or

The chemical structure and predicted distance from the known training activators is listed for each odor. All distances represent the minimum distance based on optimized descriptors to the previously known strongest active compound listed in the gray cells for that particular Or.

Or2		Or6		Or10	
Structure	Distance	Structure	Distance	Structure	Distance
C(=O)(C)c1ccccc1	0	C(=O)(C)c1ccccc1	0	Cc1c(cccc1)O	0
Cc1c(cccc1)O	0	C(=O)(C)c1sccc1	0	Cc1cc(ccc1)O	0
Cc1cc(ccc1)O	0	C(=O)(C)c1sccn1	0	Cc1ccc(cc1)O	0
Cc1ccc(cc1)O	0	CCC(C1=CC=CC=C1)=O	0	C(C)c1c(cccc1)O	0
C(C)c1c(cccc1)O	0	CC1=CC=C(C(C)=O)O1	0	CC1=CNC2=C1C=CC=C2	0
c12c([nH]cc1)cccc2	0	CCC(C1=CC=CC=C1)=O	0	COC1=CC=C(C)C=C1	0
CC1=CNC2=C1C=CC=C2	0	CC(=O)C1=CC=CO1	2.814136	COC1=CC=CC=C1C	0
N1C=CC=C1	0	CC(=O)C1=CC=CN1	3.024555	COC1=CC=C(C)C=C1C	0
CC1=CC(O)=CC=C1	0	CC(=O)C1=CC=CC=C1N	3.589921	COC1=CC=CC=C1	0
SC1=CC=CC=C1	0	CC1=CC(=C(O1)C)C(=O)C	4.542537	Cn1c2c(cc1)cccc2	0.9605277
CCC1=CC(=CC=C1)O	0.51988	CC(=O)C1=CN=CC=C1	4.603079	CCC1=CC=CC=C1S	1.08787
c1(ccccc1)C(C)N	0.7157186	C1=CC=C(C=C1)C(=O)C#N	4.853174	CC1=CC=CC=C1S	1.182823
Cn1c2c(cc1)cccc2	1.037082	CC(=O)C1=CC=CC=C1O	5.11137	CC1=C(C(=CC=C1)O)C	1.216329
CCC1=CC=CC=C1S	1.090404	CC(=O)C1=NC=CN=C1	5.311086	CC1=CC(=C(C=C1)C)O	1.239482
C1=CC=C(C(=C1)O)Cl	1.202171	CC1=NC=CN=C1C(=O)C	5.32461	CC1=CC(=C(C=C1)O)C	1.239482
CC1=CC=CC=C1OC	1.261832	C1C(O1)C2=CC=CC=C2	5.368613	CCC1=CC(=CC=C1)O	1.324363
CC(=O)C1=CC=CC=C1O	1.263328	CC1=CC=C(O1)C=O	5.706266	c1(ccccc1)C(C)N	1.382239
CC1=CC(=CC(=C1)O)O	1.27613	O=C(C1=CC=CO1)OC	5.861467	CC1=C(C(=C(C=C1)C)O)C	1.445959
CC1=C(C(=CC=C1)C)O	1.279905	CC1=C2C(=O)CCN2C=C1	5.918714	CC1=C(C=CC(=C1)O)O	1.446803
CC1=C(C(=CC=C1)O)C	1.28218	C1=CC=C(C=C1)CS	5.951914	COc1c(cccc1)N	1.465265
NC1=CC=CC=C1S	1.305137	CCOC(=O)C1=CC=CO1	5.958834	CNc1ccccc1	1.483803
CN1C=CC=C1	1.318499	CC1=CC=C(C=C1)OC	6.049608	CC1=C(C(=C(C=C1)O)C)C	1.627195
CC1=C(C=C(C=C1)O)C	1.346765	COC1=CC=CC=C1	6.09279	CCC1=CC(=CC(=C1)C)O	1.716798
CNc1ccccc1	1.360506	CC1=CC=CC=C1OC	6.180274	C1=CC=C(C(=C1)N)O	1.807295
CC(=O)C1=CC=CC=C1N	1.372227	COC(=O)C1=CN=CC=C1	6.185672	CC(=O)C1=CC=CC=C1N	1.828949
CC1=C(C=CC(=C1)O)O	1.407024	C1=COC(=C1)CS	6.205407	CCC1=C(C=CC(=C1)O)O	1.857206
CC1=CC(=C(C=C1)C)O	1.426872	CC1=CC=C(C=C1)C(=O)C	6.300831	CC1=CC=CC(=C1)C=O	1.860195
CC1=CC(=C(C=C1)O)C	1.426872	O=CC1=CC=CC=C1NC	6.304842	CC(=O)C1=CC=CC=C1O	1.916893
CC1=CC(=CC(=C1)O)C	1.431107	CC1=CC=CC(=C1)C=O	6.348972	CC(C)C1=CC=CC=C1O	1.91721
COC1=CC=CC=C1	1.441395	CC1=C(SC(=N1)C)C(=O)C	6.383765	C1C(O1)C2=CC=CC=C2	2.006981
COc1c(cccc1)N	1.526856	n1cc(ccc1)C=O	6.396123	CC1=CC(=CC(=C1)O)O	2.040846
C1=CC(=CC(=C1)O)O	1.547724	CNc1ccccc1	6.452874	CC1=CC(=C(C=C1)O)OC	2.170293
COC1=CC=CC=C1O	1.549906	Nc1c(cccc1)C=O	6.516774	CC1=C(C=C(C=C1)O)C	2.424889
C1=CC=C(C(=C1)CO)O	1.562781	C1=CC=C(C=C1)N	6.516774	CC1=CC=CC=C1C=O	2.456347
C1=CC(=CC=C1O)O	1.612072	CC1=CN=C(C(=N1)C)C	6.529618	Nc1c(cccc1)C=O	2.472857
CSC1=CC=CC=C1	1.635403	CCOC(=O)C1=CC=CC=C1	6.54385	C1=CC=C(C=C1)N	2.472857
c1(c(cccc1)O)O	1.638952	COc1c(cccc1)C=O	6.641815	CC1=CC(O)=C(O)C=C1O	2.545385
C1=CC(=C(C(=C1)Cl)O)Cl	1.665124	COC(=O)C1=CC=CN1	6.65084	CC1=C(C(=CC=C1)C)O	2.54936
CC(C)C1=CC=CC=C1O	1.747473	CCC1=CC=CO1	6.67399	O=CC1=CC=CC=C1NC	2.561711
C1=CSC(=C1)S	1.760378	C(C)c1c(cccc1)C=O	6.678888	COC1=CC=CC=C1O	2.582564
C1=CC=C(C(=C1)N)O	1.773475	CC1=NC=CN=C1	6.69373	NC1=CC=CC=C1S	2.663355
c1(ccccc1)C(C)(C)O	1.77454	OCC(=O)c1ccccc1	6.706399	CC1=C(C(=CC=C1)C)S	2.664136
CC(=O)C1=CC=CN1	1.784645	O=[C]1(C(C)=O)OCCC1	6.764351	c1(ccccc1)CO	2.713267
CC1=C(C(=CC=C1)C)S	1.865646	CCC1=CC=CC=C1S	6.799576	C1=CC=C(C(=C1)CO)O	2.729375
C1=CC(=C(C=C1Cl)O)Cl	1.879134	Cc1c(c(ccc1)C)CO	6.824733	c1(ccccc1)C(C)(C)O	2.737525
CC1=CC=NC2=CC=CC=C12	1.881836	C1=CC=C(C=C1)C(C#N)O	6.85846	CSC1=CC=CC=C1	2.753418
C1=CC=C2C(=C1)C=CC=N2	1.925142	CC1C(=CCC1)C(=O)C	6.886481	CC1=CC(=CC(=C1)O)C	2.79423
C1=CC(=C(C=C1Cl)Cl)O	1.927863	CSC1=CC=CC=C1	6.892301	C1=CC=C(C=C1)S	2.800785
CC1=CC(=C(C=C1)O)OC	1.96065	COC1=CC=C(C=C1)C=O	6.905708	CC1=CNC(=C1)C(=O)OC	2.835645
C1=CC=C(C(=C1)O)Br	2.024861	CC1=CCCC(=O)C1	6.941003	COC1=CC(=CC=C1)OC	2.932035
[C-]#[N+]CC1=CC=CC=C1	2.06791	CCC1C=CC(=O)O1	6.988239	COC1=CC=CC=C1OC	2.942067
c1(ccccc1)CO	2.099141	CC(=O)C(=O)C1=CC=CC=C1	6.995212	C(C)c1c(cccc1)C=O	3.022379
CCC1=C(C=CC(=C1)O)O	2.147843	c1(ccccc1)COC	7.014262	CC(=O)C1=CC=CN1	3.047612
C1=CC(=CC(=C1)O)C=O	2.169262	c1(ccccc1)CC(=O)C	7.018756	Cc1c(c(ccc1)C)CO	3.071884
C1=CC=C(C=C1)C(C#N)O	2.171679	CC1=CC(=O)CCC1	7.043093	CC1=C(C=CC(=C1C)O)O	3.142583
s1cccc1	2.200401	CC1=NC=CN=C1OC	7.0771	COC(=O)C1=CC=CN1	3.199062
C1=CC(=CC=C1C=O)O	2.20377	CC1=CNC(=C1)C(=O)OC	7.116798	CC1=C(SC=N1)C=C	3.232885
C1=CC=C(C=C1)C(=O)C#N	2.228871	Oc1oc(cc1)C=O	7.117114	c1(c(cccc1)O)O	3.240345
COC1=CC=CC=C1OC	2.27902	c1(ccccc1)CO	7.1293	C1=CC(=CC=C1O)O	3.263079
O=CC1=CC=C(O)C=C1C	2.285233	OCc1oc(cc1)C=O	7.197465	C1=CC=C(C(=C1)C=O)O	3.287156
C1=CC=C(C(=C1)C=O)O	2.308407	o1c(ccc1)C=O	7.219843	Oc1c(c(ccc1)C)C=O	3.296966
c1(ccccc1)C#N	2.315736	CCC1CCC(=O)O1	7.222036	CC1=NC=NC2=CC=CC=C12	3.301889
Oc1c(c(ccc1)C)C=O	2.348356	CC(C)C=CC(=O)C	7.232699	C1=CC=C(C(=C1)O)Cl	3.336167
O=CC1=CC=CC=C1NC	2.357368	CC1CCCC(=O)C1=O	7.235711	C1=CC(=CC(=C1)O)C=O	3.365463
C1=CC=C(C=C1)CS	2.373006	CC(=O)C1=NCCS1	7.239078	O=CC1=CC=C(O)C=C1C	3.3787
CC1=NC=NC2=CC=CC=C12	2.393298	c1(ccccc1)CC#N	7.29472	C1=CC(=CC=C1CO)O	3.382749
[nH]1nccc1	2.398815	CC1=CC(=C(S1)C)C(=O)C	7.298992	C1=CC(=CC(=C1)O)O	3.391496
CCC1=CC(=CC(=C1)C)O	2.429883	Oc1ccc(cc1)C(=O)C	7.307233	C1=CC(=CC=C1C=O)O	3.392312
c1(ccccc1)C(=O)O	2.440054	CC1C(CCC1)C(=O)C	7.312247	C1=CC=C(C=C1)CS	3.409539
c1(ccccc1)CC#N	2.450592	CC1=CC=CC=C1C(=O)O	7.32185	CC1=CC=C(S1)C=O	3.438549
C1C(O1)C2=CC=CC=C2	2.481837	COc1c(cccc1)N	7.336948	n1c[nH]cc1CCN	3.481199
CC1=CC=CC=C1C(=O)O	2.485752	CCC1=NC(=CS1)C	7.352985	C1CC2=CC=CC=C2NC1	3.50207
Nc1c(cccc1)C=O	2.490524	NC1=CC=CC=C1S	7.35831	CC1=C2C(=CC=C1)N=CC=N2	3.505628
C1=CC=C(C=C1)N	2.490524	CCNC1=CNC=N1	7.365535	CC1=CC=NC2=CC=CC=C12	3.53354
CC1=CC=CC=C1C=O	2.533502	C1=C(NC=N1)CCN	7.365535	CC1=C(C=CO1)SC	3.541116

Table 3.6

Table 3.7: Dividing training odors into three distinct sets based upon odor structure and receptor response

Odors were divided into three different screening categories (Aromatic Activator Screen, Broad Activator Screen, and Inhibitor Screen). For each odor the name, final activity value, and inclusion in each screening set are listed. Odor activities were normalized normalized to range from -100 (maximum observed inhibition) to 100 (maximum observed activation).

Odor Name	Final Activity	Aromatic Activator Screen	Broad Activator Screen	Inhibition Screen
butanal	-85	No	Yes	Yes
pentanal	-51	No	Yes	Yes
hexanal	-32	No	Yes	Yes
heptanal	-21	Yes	Yes	Yes
octanal	-20	Yes	Yes	Yes
butanol	-25	Yes	Yes	Yes
pentanol	-41	No	Yes	Yes
hexanol	-70	No	Yes	Yes
heptanol	-38	No	Yes	Yes
octanol	-35	No	Yes	Yes
butanone	-25	Yes	Yes	Yes
pentanone	-28	Yes	Yes	Yes
hexanone	-19	Yes	Yes	Yes
heptanone	-12	Yes	Yes	Yes
octanone	-18	Yes	Yes	Yes
butyl acetate	-28	Yes	Yes	Yes
pentyl acetate	-15	Yes	Yes	Yes
hexyl acetate	-12	Yes	Yes	Yes
heptyl acetate	-8	Yes	Yes	Yes
octyl acetate	-15	Yes	Yes	Yes
butyric acid	-94	No	Yes	Yes
pentanoic acid	-26	Yes	Yes	Yes
hexanoic acid	-16	Yes	Yes	Yes
heptanoic acid	-14	Yes	Yes	Yes
octanoic acid	-21	Yes	Yes	Yes
pentane	-31	No	Yes	Yes
hexane	-25	Yes	Yes	Yes
heptane	-29	Yes	Yes	Yes
octane	-34	No	Yes	Yes
2,3-butanedione	-99	No	Yes	Yes
1-octen-3-ol	-27	Yes	Yes	Yes
Ethanol	-16	Yes	Yes	Yes
3-octanol	-14	Yes	Yes	Yes
Methanol	-14	Yes	Yes	Yes
Nonanol	-12	Yes	Yes	Yes
Eugenol Methyl Ether	-9	Yes	Yes	Yes
Acetic Acid	-7	Yes	Yes	Yes
g-valerolactone	-5	Yes	Yes	Yes
Fenchone	-2	Yes	Yes	Yes
Isoamyl Acetate	-2	Yes	Yes	Yes
Limonene	-2	Yes	Yes	Yes
Menthol	-2	Yes	Yes	Yes
E2-hexenal	0	Yes	Yes	Yes
Geranyl Acetate	0	Yes	Yes	Yes
Methional	0	Yes	Yes	Yes
Eugenol	1	Yes	Yes	No
4-methylphenol	3	Yes	Yes	No
Isopropyl Alcohol	3	Yes	Yes	No
Carvone	4	Yes	Yes	No
Phenylethanone	5	Yes	Yes	No
Anisole	6	Yes	Yes	No
Benzaldehyde	6	Yes	Yes	No
Benzophenone	8	Yes	Yes	No
Citronellal	8	Yes	Yes	No
Geraniol	8	Yes	Yes	No
Ethyl Acetate	8	Yes	Yes	No
Methylsalicylate	13	No	Yes	No
Thymol	15	No	Yes	No
Cyclohexanone	48	No	Yes	No
Indole	21	Yes	Yes	No
2-methylphenol	24	No	Yes	No
methyl pyruvate	-100	No	Yes	Yes
propionyl bromide	-88	No	Yes	Yes
propionyl chloride	-73	No	Yes	Yes
propionaldehyde	-68	No	Yes	Yes
2,3-pentanedione	-55	No	Yes	Yes
2-heptanol	-39	No	Yes	Yes
2-(propylamino)-ethanol	-39	No	Yes	Yes
butyryl chloride	-39	No	Yes	Yes
propionic acid	-32	No	Yes	Yes
2-methyl-3-heptanone	-26	Yes	Yes	Yes
3-heptanol	-16	Yes	Yes	Yes
4-(methylthio)-1-butanal	-15	Yes	Yes	Yes
4-hydroxy-2-butanone	-11	Yes	Yes	Yes
2,5-dimethylthiophene	-9	Yes	Yes	Yes
6-methyl-5-hepten-2-ol	0	Yes	Yes	Yes
1,5-pentanediol	0	Yes	Yes	Yes
1-hepten-3-ol	0	Yes	Yes	Yes
3-decanone	1	Yes	Yes	No
pyruvic acid	2	Yes	Yes	No
3-nonanone	2	Yes	Yes	No
4-heptanone	2	Yes	Yes	No
2-hexanol	2	Yes	Yes	No
1-bromohexane	3	Yes	Yes	No
1-hexanethiol	3	Yes	Yes	No
hexylsilane	3	Yes	Yes	No
phenylacetaldehyde	3	Yes	Yes	No
1-iodohexane	3	Yes	Yes	No
2,4,5-trimethylthiazole	5	Yes	Yes	No
ethyl valerate	5	Yes	Yes	No
cis-2-hexene	5	Yes	Yes	No
3-methyl-2-pentene	5	Yes	Yes	No
methoxyacetone	6	Yes	Yes	No
1-chlorohexane	8	Yes	Yes	No
cis-3-hexen-1-ol	10	Yes	Yes	No
fluoroacetone	10	Yes	Yes	No
acetophenone	15	No	Yes	No
2-acetylthiophene	31	No	Yes	No
pyridine	99	Yes	Yes	No
thiazole	100	Yes	Yes	No
2-ethyl-3,5(6)-dimethylpyrazine	8	Yes	Yes	No
2,5-dimethylpyrazine	26	Yes	Yes	No
pyrazine	-8	Yes	Yes	Yes
naphthalene	-14	Yes	Yes	Yes

Table 3.7

Table 3.8: Optimized descriptors selected for the Aromatic Activator Screen

Optimized descriptor symbols, brief descriptions, classes, dimensionality, and occurrences are listed for the aromatic activator screen. Descriptors are listed in ascending order of when they were selected into the optimized set. Weights indicate the number of times a descriptor was selected in an optimized descriptor set.

symbol	breif description	class	dimensionality	occurrence
N.075	R--N--R / R--N--X	atom-centred fragments	2	1
R3v.	R maximal autocorrelation of lag 3 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3	1
H.049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	2	1
nRCHO	number of aldehydes (aliphatic)	functional group counts	1	1
nN	number of Nitrogen atoms	constitutional descriptors	1	1
ISH	standardized information content on the leverage equality	GETAWAY descriptors	3	1
EEig07d	Eigenvalue 07 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2	1
piPC04	molecular multiple path count of order 04	walk and path counts	2	1
MATS4e	Moran autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2	1
ESpm14d	Spectral moment 14 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2	1
Mor12m	3D-MoRSE - signal 12 / weighted by atomic masses	3D-MoRSE descriptors	3	1

Table 3.8

Table 3.9: Optimized descriptors selected for the Broad Activator

Optimized descriptor symbols, brief descriptions, classes, dimensionality, and occurrences are listed for the broad activator screen. Descriptors are listed in ascending order of when they were selected into the optimized set. Weights indicate the number of times a descriptor was selected in an optimized descriptor set.

symbol	breif description	class	dimensionality	occurrence
HNar	Narumi harmonic topological index	topological descriptors	2	1
R3v+	R maximal autocorrelation of lag 3 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3	4
HATS3m	leverage-weighted autocorrelation of lag 3 / weighted by atomic masses	GETAWAY descriptors	3	1
Mor13p	3D-MoRSE - signal 13 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3	1
ISH	standardized information content on the leverage equality	GETAWAY descriptors	3	2
P1s	1st component shape directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3	1
R4e+	R maximal autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3	1
nRCHO	number of aldehydes (aliphatic)	functional group counts	1	2
JGI2	mean topological charge index of order2	topological charge indices	2	2
E1u	1st component accessibility directional WHIM index / unweighted	WHIM descriptors	3	2
MATS5m	Moran autocorrelation - lag 5 / weighted by atomic masses	2D autocorrelations	2	1
STN	spanning tree number (log)	topological descriptors	2	2
DISPe	d COMMA2 value / weighted by atomic Sanderson electronegativities	geometrical descriptors	3	1
B06.C.O.	presence/absence of C - O at topological distance 06	2D binary fingerprints	2	1
X4A	average connectivity index chi-4	connectivity indices	2	4
JGI3	mean topological charge index of order3	topological charge indices	2	1
De	D total accessibility index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3	2
Mor25u	3D-MoRSE - signal 25 / unweighted	3D-MoRSE descriptors	3	1
nRCOX	number of acyl halogenides (aliphatic)	functional group counts	1	1
B03.O.O.	presence/absence of O - O at topological distance 03	2D binary fingerprints	2	1
nHDon	number of donor atoms for H-bonds (N and O)	functional group counts	1	1
MATS3e	Moran autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2	1
RBF	rotatable bond fraction	constitutional descriptors	1	1
GATS5m	Geary autocorrelation - lag 5 / weighted by atomic masses	2D autocorrelations	2	1
C.008	CHR2X	atom-centred fragments	2	1
Mor13v	3D-MoRSE - signal 13 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3	1
R6u.	R maximal autocorrelation of lag 6 / unweighted	GETAWAY descriptors	3	1

Table 3.9

Table 3.10: Optimized descriptors selected for the Inhibitor Screen

Optimized descriptor symbols, brief descriptions, classes, dimensionality, and occurrences are listed for the inhibitor screen. Descriptors are listed in ascending order of when they were selected into the optimized set. Weights indicate the number of times a descriptor was selected in an optimized descriptor set.

symbol	brief description	class	dimensionality	occurrence
ATS1p	Broto-Moreau autocorrelation of lag 1 (log function) weighted by polarizability	2D autocorrelations	2	1
Gu	total symmetry index / unweighted	WHIM descriptors	3	6
PW5	path/walk 5 - Randic shape index	Topological indices	2	1
H-052	H attached to C0(sp3) with 1X attached to next C	Atom-centred fragments	2	4
H4m	H autocorrelation of lag 4 / weighted by mass	GETAWAY descriptors	3	3
Rtu+	R total index / unweighted	GETAWAY descriptors	3	3
HATS6m	leverage-weighted autocorrelation of lag 6 / weighted by mass	GETAWAY descriptors	3	3
B03[C-O]	Presence/absence of C - O at topological distance 3	2D Atom Pairs	2	1
nR=Cs	number of aliphatic secondary C(sp2)	Functional group counts	1	2
EEig09r	Eigenvalue 09 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2	1
Mor16m	3D-MoRSE - signal 16 / weighted by atomic masses	3D-MoRSE descriptors	3	3
X5A	average connectivity index chi-5	connectivity indices	2	1
EEig02r	Eigenvalue 02 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2	2
RDF055m	Radial Distribution Function - 5.5 / weighted by atomic masses	RDF descriptors	3	1
EEig04d	Eigenvalue 04 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2	2
B06[C-C]	presence/absence of C - C at topological distance 06	2D binary fingerprints	2	1
JGI4	mean topological charge index of order4	topological charge indices	2	1
RDF085m	Radial Distribution Function - 8.5 / weighted by atomic masses	RDF descriptors	3	1
Mor08u	3D-MoRSE - signal 08 / unweighted	3D-MoRSE descriptors	3	1
MATS5e	Moran autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities	2D autocorrelations	2	2
B02[C-C]	presence/absence of C - C at topological distance 02	2D binary fingerprints	2	1
nCrs	number of ring secondary C(sp3)	functional group counts	1	1
X4Av	average valence connectivity index chi-4	connectivity indices	2	1
R7e+	R maximal autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3	1
EEig08r	Eigenvalue 08 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2	1
E3s	3rd component accessibility directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3	1

Table 3.10

Table 3.11: Top predicted natural library compounds for the aromatic activator, broad activator, and inhibitor screens

The chemical structure and predicted distance from the known training activators is listed for each predicted odor. All distances represent the minimum optimized descriptor distance to a training set compound, which are listed in gray.

Broad Activator		Aromatic Activator		Inhibitors	
Structure	**Distance**	**Structure**	**Distance**	**Structure**	**Distance**
O=C1CCCCC1	0	C1=NC=CC=C1	0	CCCC(O)=O	0
C1=NC=CC=C1	0	S1C=NC=C1	0	CC(C(C)=O)=O	0
S1C=NC=C1	0	C1=NC=CN1	2.297704	O=C(C)C(OC)=O	0
O=C1CCCCCN1	2.037369	N1N=CC=C1	2.768945	O=C(C(C)=O)C	1.193469
O=C1NCCCC1	2.16724	CC1=CSC=N1	3.028959	NCCC(O)=O	2.028579
C1=CC=CS1	2.997682	CCCCCC1=NC=CC=C1	3.534064	CCC(C(O)=O)=O	2.579103
O=C1C=CCCC1	3.977348	CN1C=CC=C1	3.721854	CCC(N)C(O)=O	2.752901
CC1CCNCC1	4.39789	CCC1=NC(=CN=C1)C	3.758666	CC(N(C)C)=O	3.088135
C1=NC=CN1	4.619738	C=C/C=C\CCCCCCCCCO	3.778543	CC(C(OC)=O)=C	3.128039
O=C1C=CC=CO1	4.919501	CC1=CN=CC(C)=N1	3.815402	CC(OCC)=O	3.133545
N1N=CC=C1	4.943772	C1=CC=CS1	3.817309	CC(C)/C=N/O	3.16974
CN1CCCCC1	5.167015	CC1=NC=CN=C1CC	3.82998	CC(C)C(C(O)=O)=O	3.179965
CC(C)=CC1CC(C)=CCO1	5.609835	CCC1=NC=CN=C1CC	3.853768	CC(C)CC(O)=O	3.318447
CC(C)=C[C@H]1C[C@@H](C)CCO1	5.751913	C12=CC=CC=C1N=CC=N2	3.889099	C/C(C(O)=O)=C\C	3.442493
O=C1CCCC1	5.94834	CCC1=NC=CN=C1	3.914359	C/C=C(C)/C([O-])=O	3.452961
CC(C)=CC1C=C(C)CCO1	5.971966	CC1=NC=CN=C1OC	4.038081	C=C(C)CCO	3.736961
O=C1CC(C)(O)CCO1	5.996197	COC1=NC=CN=C1CC	4.03881	O=C(O)COC	3.778532
CC(C)=CC1OCCC(C1)C	6.075926	C=CC1=CN=CC(C)=N1	4.048223	CC(C)C(OC)=O	3.80734
OC1CCCCC1	6.143672	CC1=COC=C1	4.074861	CC(C)=CCO	3.874121
S1SCCC1	6.220288	CC1=NC=C(CC)C=C1	4.124326	CC(COC)=O	3.874164
C1C=CCS1	6.313776	C/C=C/C=C\CCCCCCCO	4.131876	C/C(C=O)=C\C	4.012943
OC1(C)C=CCCC1	6.407874	CC/C=C\CCCCCCO	4.139062	CCC(C)C=O	4.027003
OC1(C)CCCC1	6.471678	CC1=CC2=C(C=C1)N=CC=C2	4.187466	C=C(O)CCC	4.210918
O=C1C/C=C\CCCCCCCCO1	6.523937	CCC/C=C/C=C/C/C=C\CCO	4.235145	C=CC(CC)=O	4.220098
CC(=CCCC(=CCCC1=COC=C1)C)C	6.54101	CC/C=C\C=C\CCCCCCCCCCO	4.260037	CC(CC=C)=O	4.255649
C1=CC=CC=CC1	6.558737	CC1=CN=C(C(C)CC)C(C)=N1	4.26035	CC(C)CC#N	4.2874
CC1=NC=C(CC)C=C1	6.577361	CC/C=C\C/C=C\C/C=C\CCO	4.262767	O=C(O)C(N)CO	4.307723
N1C=CC=C1	6.588811	CC1=NC=CN=C1CC(C)C	4.265956	CC(C)CCO	4.368743
CN1C=CC=C1	6.60897	COC1=NC=CN=C1C(C)CC	4.290186	NC(OCC)=O	4.378132
O=C(NCCC(C)C)C1=CC=CC=C1	6.612483	CC/C=C\CCCCCCCCO	4.295488	CC(OC)OC	4.394431
CC1SCCN1	6.61741	CC(C1=CN=CC(C(C)C)=N1)C	4.295748	C/C=C(C)/CO	4.410392
O=C(OC(C)C)C1=CC=CC=C1	6.638106	OCC/C=C\C1=CC=CC=C1	4.312156	C=C(C)C(C)=C	4.446982
CCCC(C1=CC=CC=C1)=O	6.651088	CC/C=C\CCCCCCCCCCO	4.317797	CC(C(C)=C)=O	4.523044
O=C(N1CCCCC1)C2=CCCCC2	6.675009	CC/C=C\C/C=C\CCCO	4.328344	CCC(C)C#N	4.540592
CC(C)CC1=NC=CS1	6.701117	O=COCC1=CC=CC=C1	4.347131	O=C(O)C(CC)CC	4.545043
CCCCCCC1CCCC(O1)=O	6.717317	CC/C=C\C=C/CCCCCO	4.35801	CC(CCC)=O	4.576413
CCCCCCCC1CCCC(O1)=O	6.743187	OCCCCCC1=CC=CC=C1	4.371786	CC(C(C)CC)=O	4.608949
CC1CCCC(C)N1	6.751586	CC/C=C/C=C\CCCCCCO	4.375425	O=COCC	4.696431
O=CC1=COC=C1	6.853976	CCCCC/C=C\C=C\CO	4.38817	CC(C)(C)C(O)=O	4.786178
O=C(OCC=C(C)C)C1=CC=CC=C1	6.86057	CCC/C=C\C=C\CCCCCCCCO	4.401625	N[C@@H](C(C)C)C(O)=O	4.809623
CC1=COC=C1	6.867993	CC1C=CCC1	4.413842	CC(C)C[N+]([O-])=O	4.811683
O=C(C)CC(OCC1=CC=CC=C1)=O	6.899082	O=COC/C=C/C1=CC=CC=C1	4.423901	CC(C(O)=O)=C	4.818327
OC1=CC=CC(O)=C1	6.902425	CC1=NC=C(CC)N=C1C	4.441315	CC(C(O)CC)=O	4.835998
O=C(N(CC)CC)CC1=CC=CC=C1	6.905865	NCCC1=CC=CC=C1	4.45126	CC(C)=CC(O)=O	4.870955
CCC(N1CCCCC1)=O	6.908113	N1=CCCC1	4.455448	CC(OC(C)=C)=O	4.88651
C/C(C(OCCC1=CC=CC=C1)=O)=C\C	6.915907	C/C=C\C=C/CCCCCCCCCO	4.455697	CC(C(C(C)C)=O)=O	4.932188
CCC(OCCC1=CC=CC=C1)=O	6.94	C1(CCCC2=CC=CC=C2)=NC=CC=C1	4.48037	CCC(C)C(O)=O	4.94746
O=C1C/C=C\C/C=C\CCCCCO1	6.96282	O=CC1=CC=C(C(C)=C)C=C1	4.484765	CC(O)COC	4.960194
CCC/C=C\C1CC=CC(O1)=O	6.971452	CCCC/C=C\CCCCCCO	4.4849	CC(O)C(CC)=O	4.969207
CCCCCC1=NC=CC=C1	6.99577	CC1=C(SC=N1)CCO	4.495566	O=CN(C)C	4.971724
CCCCC/C=C\C1CC=CC(O1)=O	7.00054	CC1=CN=C(C)C(C(C)CC)=N1	4.495643	CCC(OC)=O	4.979604
CCCC/C=C\C1CC=CC(O1)=O	7.008492	CC1CCCC1	4.507732	O=C(C)/C=C/C	5.003101
CC(C)=CCC1=C(C)C=CO1	7.01097	CC1=C(C)N=CS1	4.524113	CC(C)=CC#N	5.003739
C/C=C/C(OCC1=CC=CC=C1)=O	7.011722	CCCC(C1=NC(C)=CN=C1C)C	4.534104	CC([C@H](O)C)=O	5.00945
OCCCCCC1=CC=CC=C1	7.012331	C1(C2NCCCC2)=CC=CN=C1	4.551649	OC(C)(CC)C([O-])=O	5.013806
CCCC(OCCC1=CC=CC=C1)=O	7.016306	C1([C@H]2NCCCC2)=CC=CN=C1	4.551649	CC(C)(O)C(OC)=O	5.020397
CC(CCC1=CC=CC=C1)=O	7.018781	OCCCC1=CC=CC=C1	4.559646	C[C@@H](O)CC(O)=O	5.047488
CC(C)CC(OCCC1=CC=CC=C1)=O	7.0234	CCCCCC/C=C/CCCCCCO	4.559792	CC(C(O)C)=O	5.057977
CC(SCC1=CC=CO1)C	7.030379	CC/C=C\C/C=C\C=C\CO	4.564464	CCC(CO)=C	5.059612
OC1=CC=CC=C1	7.03727	CC(C1=CN=C(C)C=N1)C	4.57072	CC(C)CC=O	5.151031
OC1=CC=C(CCC)C=C1	7.045313	CCCCCC/C=C\CO	4.573949	CC(OCC)=S	5.164894
CCCC1OCC(O1)C	7.057154			SC(C)C(O)=O	5.213845
CCCCCCCCCC(N1CCC(C)CC1)=O	7.066965				

Table 3.11

Table 3.12: Predicted odors validated as activators and inhibitors of the CO_2 receptor

The overall accuracy for CO2 predictions considering all three predicted sets. The accuracy and percentage accuracy are listed for each class.

Activity	Accuracy	Percentage
Activators (>50 spikes/sec)	25/139	18%
Activators (>30 spikes/sec)	42/139	30%
Inhibitors (<-5 spikes/sec)	25/139	18%
	Total	48%

Table 3.12

Table 3.13: Optimized descriptor sets for each Citrus Psyllid ORN

Optimized descriptor symbols and final descriptor-activity correlations are listed for each Citrus Psyllid ORN. Descriptors are listed in ascending order of when they were selected into the optimized set. The number of times a descriptor was selected in an optimized descriptor set represent weights.

Or	Desc.Symbol	Or	Desc.Symbol	Or	Desc.Symbol	Or	Desc.Symbol
RP2_B	R7m	RP2_C	ESpm10d	RP2_A	ATS5v	RP4_C	SP02
Correlation	QYYe	**Correlation**	L1m	**Correlation**	ATS4e	**Correlation**	ATS6m
0.8289435	Infective.80	0.746558	Mor30p	0.8315387	DISPv	0.807409	R5p.
	RDF065m		nHDon		R7p.		ATS5m
	JGI6		ESpm10d.1		H5v		SP12
	RDF080v		Mor27p		nHDon		Mor29e
	nHDon		E3m		As		R7m.
	Mor27e		H.051		B05.C.C.		R8u.
	H8v		Mor24p		nDB		RDF020e
	E1u		Mor27p.1		E2s		Mor23p
	R2v.		HATS8m		Mor26v		H.052
	Mor03m		Mor16u		Infective.80		R8p.
	R7m.1		R1u.		B05.C.C..1		Mor16p
	Mor32u		ESpm06d		L.Bw		BEHm7
	H6m		R4v.		piPC08		SP12.1
	H8m		Mor13p		G3p		R6m.
	Mor16u		B04.C.C.		RDF080e		EEig08x
	O.056		L3s		nR.Cp		HATS3p
	Mor27e.1		EEig08x		BELm1		G3u
	JGI6.1		nHDon.1		nHDon.1		nCb.
	EEig12x		Mor27e		PJI3		B07.C.O.
	QYYe.1		Mor30m		H8u		R6m..1
	R8p.		nR.Cs		Mor24m		H7u
			Mor30e		DISPv.1		B04.C.C.
			Mor10u		nCconj		EEig11x
			C.006		B05.C.C..2		MATS8m
			R5v.		L2s		RTu.
			Mor04m		nHDon.2		BEHm7.1
					H4m		R7m..1
					G3u		R8e.
					nR04		C.025
					Infective.80.1		H4p
					nRCO		CIC3
					RDF075e		Mor30u
					DISPv.2		Mor16m
					O.057		R6m..2
					H5e		B07.C.O..1
							BEHm7.2
							EEig11x.1
							BEHm7.3
							EEig08x.1
							R8e..1
							R7m..2
							R5p..1
							SP12.2
							BEHm7.4
							Mor16p.1
							nRCO
							H7u.1

Table 3.13

Or	Desc.Symbol	Or	Desc.Symbol	Or	Desc.Symbol	Or
RP6_A	nRCHO	RP6_B	ATS5v	RP6_C	SP01	RP7_B
Correlation	nCconj	**Correlation**	Xu	**Correlation**	Mor13u	**Correlation**
0.8951214	EEig10d	0.8758008	nHDon	0.8019783	SP11	0.8016844
	EEig03d		nDB		RDF020u	
	ESpm14d		B05.C.C.		H7u	
	H5m		Mor05e		R6m.	
	H8u		E2s		Mor16p	
	Mor27e		Infective.80		R7m.	
	nROH		PJI2		RTu.	
	F01.C.O.		piPC08		Mor22e	
	H.051		nCconj		R8e.	
	R2p		B05.C.C..1		H5m	
			L.Bw		RDF050v	
			Infective.80.1		JGI5	
			G3p		RDF020e	
			Mor15v		R8v.	
			R7p.		R6m..1	
			R6u		EEig08x	
			H8u		Mor30e	
			B05.C.C..2		HATS4u	
			DISPv		Mor16p.1	
			nHDon.1		RTu..1	
			H5e		R7m..1	
			RDF080e		CIC3	
			E1s		R6m..2	
			nR04		H7e	
			E2s.1		C.025	
			nDB.1		B04.C.C.	
			Infective.80.2		B07.C.O.	
			Mor03v		nOHt	
			nOHs		nR03	
			PJI3		RDF085p	
			C.026		R8u.	
			L.Bw.1		H5m.1	
			R6v.		RDF020u.1	
			nDB.2		RTu..2	
			B05.C.C..3		Mor16p.2	
			Infective.80.3		R6m..3	
			nR10		H7u.1	
			DISPv.1		B07.C.O..1	
			B05.C.C..4		R8e..1	
			JGI7			
			Mor28v			
			E2s.2			
			nCconj.1			
			Infective.80.4			
			B05.C.C..5			
			G3u			

Table 3.13 Continued

Desc.Symbol	Or	Desc.Symbol
Mor02p	RP7_C	nRCHO
Mor27e	**Correlation**	EEig10x
R7p.	0.8422402	nCconj
DISPm		EEig03d
Infective.80		RDF020e
JGI5		DISPv
B07.C.C.		Mor30u
HATS4e		PJI3
Mor23p		G3s
Mor32m		GATS2p
CIC2		RDF080e
R7m.		nROH
JGI4		R3p
Mor10m		RDF020e.1
X5A		EEig10x.1
EEig10r		ESpm01d
B07.C.C..1		nRCOOR
H3m		H.051
Mor03p		GATS5e
F05.C.O.		H8u
R6m.		H.051.1
JGI5.1		Mor30e
B07.C.C..2		nROH.1
Mor27e.1		EEig09x
HATS4e.1		Jhetp
E2e		Mor30u.1
B07.C.C..3		RDF020e.2
Mor03m		
B04.C.C.		
EEig10r.1		
CIC2.1		
B07.C.C..4		
JGI4.1		

Table 3.13 Continued

Table 3.14: Top predicted natural library compounds for Citrus Psyllid ORNs

The chemical structure and predicted activity are listed for the top predicted activators for each ORN. Compounds in gray are training set odors.

RP2_B		RP2_C		RP4_B	
SMILES	Pred Activity	SMILES	Pred Activity	SMILES	Pred Activity
C/C1=C\CC(C)(C)/C=C/C/C(C)=C/CC1	74.35797	CCC/C=C/C=O	45.26686	CC1=CC=C(C(C)C)C=C1	163.3939
CC(OCCCCC)=O	71.94693	CC(C)=CCCC(C=C)=C	44.17876	CC(C1=CC=CC=C1)=O	139.1738
C=C1CC/C=C(C)/CC[C@@]2([H])C(C)(C)C[C@]12[H]	71.84607	CCCC(OC)=O	42.36729	CC(C12CCC(C1C2)=C)C	135.1755
C/C(CCC=C(C)C)=C\C=O	63.22675	CCCC(OCC)=O	30.8423	CC1(C)[C@@]2([H])CC(C)=CC[C@@]12[H]	128.784
C=C/C(C)=C/CC=C(C)C	58.49042	O=CC1=CC=CC=C1	26.24606	CCCCCCCC=O	117.5805
CCCCCCCC=O	48.96719	C=CC(O)CCCCC	23.14035	CC/C=C\CC=O	116.1112
CC(CCC=C(C)C)CC=O	36.11657	CC1=CCC(C(C)=C)CC1=O	21.80823	CC(OCC(C)C)=O	112.9869
C/C1=C\CC(C)(C)/C=C/C/C(C)=C/CC1	74.35797	C/C=C/CC(OC)=O	51.43602	CCC(C1=NC=CS1)C	210.9587
CC(OCCCCC)=O	71.94693	CCCC=CC=O	45.26686	CCCC1=CC=CO1	201.5245
O=C(OC)CCCCC(OC)=O	71.28565	CC/C=C/CC(OC)=O	44.65782	CC(C)CC1=NC=CS1	195.0182
C/C1=C\CC/C(C)=C/C/C(C[C@@H]1C)=C(C)\C	71.24567	CCC(OC)=O	44.38818	CC(C1=CC=CC=C1)C	194.1631
CC1=CCCC(=CCC(CC1)C(=C)C)C	68.60787	CC(=CCCC(=C)C=C)C	44.17876	CC1=CC=CC=C1CC	193.5604
CC(C)CCC(OCC)=O	68.13362	CCCC/C=C/C=O	42.65992	COCC1=CC=CC=C1	191.4757
CCCC(OCC(C)C)=O	68.07507	CCCC(OC)=O	42.36729	CC1=CC(CC)=CC=C1	191.0865
CC(/C=C/CCCCC)=O	66.2354	C/C=C/C(OC)=O	42.26823	CCOC1=CC=CC=C1	190.5536
CC(OCCCCCC)=O	65.7931	CCC(/C=C/C=O)=O	41.84369	CCCC1=CC=CC=C1	188.6928
CC1=CC=C(O1)C(C)CC=O	63.96134	CC(/C=C/CCC)=O	40.55074	CCC1=CC=CC=C1	188.517
O=C/C=C(CCC=C(C)C)\C	63.22675	CCC/C=C/C(OC)=O	40.52789	CC1=CSC(CC)=N1	187.4761
CCCCCCC(OC)=O	63.10904	CC1CCC2=C(C=NC=C12)C	40.00807	CC1=CSC(C(C)C)=N1	184.7897
CC(/C=C/CCCC)=O	62.47671	CCCCC/C=C/C=C/C=O	39.88164	COC(OC)C1=CC=CC=C1	178.6651
CC(CCC=C)C=C(C)C	62.38316	CC/C=C/C(OC)=O	39.71849	CC(C1=CC(C)=CC=C1)C	178.4423
CCC(OCCCCC)=O	62.34478	CC/C=C/C=C/C=O	39.4017	CC1=CC=C(CC)C=C1	177.9384
C=C(C1C/C=C(C)/CC/C=C(C)\CC1)C	61.23283	C=CCC(O)CC=C	38.99514	C1(C2=CC=CC=C2)CO1	175.7152
C/C=C/C(OCC(C)C)=O	60.88536	C=C/C(C)=C/C=C/C(C)C	38.95081	CC1=CC(CC)=CC=C1C	174.1878
C=C/C(C)=C/CC1C(C)(C)O1	59.00969	CCCCC(OC)=O	38.01496	CCC1=CC=CN=C1	174.0245
C=C/C(C)=C/CC=C(C)C	58.49042	O=C(OC)CCC(OC)=O	37.8518	CCC1=CC(CC)=CC=C1	169.4623
CCC(OCCC(C)C)=O	56.55607	C=C/C(C)=C/CCC(C)=C	37.74838	CCCCCC1=CC=CO1	168.622
CCCCC1CCC(O1)=O	56.5295	C/C(C(OC)=O)=C\C	37.48334	N#CCC1=CC=CC=C1	167.7722
COC/C=C(CCC=C(C)C)/C	56.44963	C/C=C/C=C/C(OC)=O	37.47729	CCCCC1=CC=CO1	166.5448
CC1=C[C@](C(C)(C)CC[C@@H]2C)([H])[C@@](CC1)([H])C2=C	56.43485	CC/C=C/CC(OCC)=O	37.34352	CC1=CC=C(C(C)C)C=C1	163.3939
CC(CCCCCC=C)=O	55.76329	O=C(C)/C=C/C	37.33052	COCCC1=CC=CC=C1	163.2257
C=C(O1)C1(C)=CCC=C(C)C	54.60038	C/C=C/C=C/C=O	37.21464	CC(C)CCC#N	161.5202
O=C(OC)CC1=CC=CC=C1	54.0418	COC1=CC=C(C=C)C=C1	37.11306	SCCC1=CC=CC=C1	161.2887
C=C/C=C(C)/CCC=C(C)C	53.745	CCCC/C=C/C/C=C/C=O	36.75797	CC1=CC=CC=C1C(C)C	160.0841
COC/C=C(CCC=C(C)C)\C	53.73228	CCC(OCC=C)=O	36.4936	CSCC1=CC=CC=C1	159.0215
C=CC(CCC=C(C)C)C	53.71671	C=CC(CC)=O	36.43435	CCC1C2COC(O2)(C)O1	158.5089
CC(C)CC(OCC(C)C)=O	53.64558	C=C(C)/C=C/C=C(C)/C=C	36.26165	CC1=NC=C(CC)C=C1	158.0411
CCCCC(OCCC)=O	53.12195	O=C1(C/C=C/C)=CCCC1C	36.06916	CCCCC1=CC=CC=C1	156.9442
O=[N+](/C=C\C1=CC=CC=C1)[O-]	52.23921	O=C(OC)CC(OC)=O	36.01778	CSC1=CC=CC=C1	156.4088
CCCCC(OCC)=O	52.18524	CCCCC/C=C\C=C\C=O	36.00254	[C-]#[N+]CC1=CC=CC=C1	154.3357
CCCCC/C=C/C=O	52.16527	CCC(OCC/C=C\CC)=O	35.75418	CCCC1OCC(O1)C	154.2606
CC(CCCCCC)=O	52.08181	CC(C)C(OC)=O	35.71133	O=CC1=CC=C(CC)C=C1	153.2292
CCCC(C)C(OCC)=O	51.98421	C=C/C=C/C=C\CCCCC	35.50662	CC1COC(O1)CC(C)C	152.8799
CC(C)=C1CCC(C)C2=C(C(C)CC2)C1	51.6773	CC/C=C/CCC=O	35.3759	CCCC(OC)=S	152.8392
CC(C1CC=C(C=O)CC1)C	51.29109	CCC(OCCC)=O	35.34795	SC1=CC=CC=C1CC	152.6492
O=C(C)C(OCCC(C)C)=O	51.17537	CC/C=C/C=O	35.05646	COC1=CC=CC=C1	152.1006
CC(OCCC(C)CC)=O	51.10972	C=C/C=C/CCCC	35.0319	CCC1=NC=CN=C1	150.7356
CC(CCC(OCC)=O)=O	51.10906	O=COCCC	35.02602	CSCC1=CC=CO1	148.9938
CC(OC(C)(C1=CC=CC=C1OC)C)=O	51.00987	CCC/C=C/C=C/C=O	35.00791	CC1=NC=C(CC)N=C1C	144.4554
C=CCC(C)C(OCC)=O	50.96592	CC(CC=CC=CC)C	34.89028	CCC1=NC(=CN=C1)C	144.0637
O=CCC1=CC=C(C)C=C1C	50.59055	O=C(OC)CCCOC	34.68742	CC1OC2(C)C(C)CCC1O2	142.4315
COC(C=C)(C)CCC=C(C)C	50.49973	COC1=CC=C(/C=C/C)C=C1	34.60386	CC1(C)C2CCC(C)=CC12	142.1961
CC1(C)C2CC/C(C)=C\CC/C(C)=C\C12	50.36754	CCCSC	34.57437	CCCCSC#N	141.7605
O=COCCCCCC	50.33509	CCCC(O)CCC=C	34.19629	O=CC1=CC=CC(CC)=C1	141.4747
CC(C)=C/C=C/C(C)=C\C	50.29364	O=CC1=CN=CC=C1	34.17189	CC1C=CC2(C(C)C)CC12	140.7053
CC(C)CCCC(OC)=O	50.00946	C=C/C(C)=C/C	33.9544	CC(O)C1=CC=CC=C1	140.5338
C/C(CCC=C(C)C)=N\O	49.90491	C/C=C/C(OCC)=O	33.48005	CC1=NC=CN=C1CC	139.2927
CC(C)=CC1=CC=CC=C1	49.89894	CCC(/C=C/C)=O	33.43157	COC1CC=C(C)CC1	139.2339
CCCCCCCC=O	48.96719	CC(OCCCCC/C=C/CC)=O	33.40156	CC(C1=CC=CC=C1)=O	139.1738
CC/C=C\CCC=O	48.87246	CCC(OCC)=O	33.32586	CC(OC1=CC=CC=C1)=O	137.8998
CC(OC1CCCCC1)=O	48.54704	CC(OC)OC	33.32402	COC1=CC=C(C)C=C1	137.7904
CC1CCC=C2C1(C3C(C3(C)C)CC2)C	48.32759	CCCC(OCC/C=C/CC)=O	33.27204	O=C(C1=C(C)N=C(C)S1)C	137.15
CCCC/C=C/C=O	48.08597	CCC/C=C\C(OC)=O	33.02956	O=C(OC)C1=CC=CC=C1C	137.0934
O=COCCCCCCC	48.03822	C=C=C(C)CC	32.95811	CC12CCCC(O2)C(CC)O1	136.7338
O=C(OC)C1=CC=CC=C1C	47.67373	O=C(N1C)N(C)C=CC1=O	32.91734	CNC1=CC=CC=C1	136.328
CCCC(OCCC)=O	47.49852	CCCC/C=C/CC=O	32.77198	CCC(OC1=CC=CC=C1)=O	135.7478
C=C/C(C)=C\CC=C(C)C	47.18779	O=COC1CCCCC1	32.71905	N1(CC2=CC=CO2)C=CC=C1	135.7411
O=COCCCCC	47.10583	C/C=C/C=C/C(OCC)=O	32.70471	C=C1CCC2(C(C)C)CC12	135.1755
CC(CC1=CC=CC=C1)=O	46.94897	CCCCC/C=C/C/C=C/CC(OC)=O	32.552	CC(C)CC#N	134.4416
O=[N+](CCC1=CC=CC=C1)[O-]	46.86099	CCC1=CC=C(C=C)C=C1	32.41418	COC1=CC=CC=C1C	134.3196
CCCCCC(OC)=O	46.73148	CCCC/C=C/C=C/C=O	32.03659	CCCC1CCC(O1)=O	133.6848
C=C(C)C(OCCC(C)C)=O	46.5108	CC(OC1CCCCC1)=O	31.94634	CCC(C)CCC	132.8239
CCCCCC/C=C/C=O	46.33825	CC(OC(C)C)=O	31.59381	O=CC1=CC=CC=C1CC	132.788
CC(OCCC/C=C\C)=O	46.16463	CC(O/C=C/CCCCCC)=O	31.5376	CC1=CC=C(C)C(C)=C1	131.9495
CCCCC1C(CC(=O)O1)C	46.15281	CCC/C=C/CCC(OC)=O	31.51514	CC(O)(C1=CC=CC=C1)C	131.9484
C=C([C@@H]1CC=C(C=O)CC1)C	45.83685	CCCC/C=C/C=C(C)C	31.44217	O=C(OC)CC1=CC=CC=C1	131.7989
CCC(OCCCC)=O	45.48793	CC(C)=CC(CC(C=C)=C)=O	31.42432	COC1=CC=C(C)C=C1C	131.7978
C=CCCCCCC	45.46762	COC1=CC=C(OC)C=C1	31.38461	COC(O)C1=CC=CC=C1	131.719
CSSCC1=CC=CO1	45.46385	CC(C/C=C/CC)=O	31.37606	CC(C)CCCC	131.6372
CCC(C)CCC(OC)=O	45.44599	CC(=CCCC(=CC=CC(=O)C)C)C	31.37038	CC(C1=CN=C(C)C=N1)C	131.534
O=C(C1CCCCC1)OC	44.91413	CC1=CCC=CC1	31.31489	CC1=C(C)C=CS1	130.8618
C=CC(OCCCC)=O	44.851	CC(C)CCCC(OC)=O	31.30609	O=CCCCC(C)C	130.8068
CCCCC(/C=C/C=O)=O	44.74891	CC(CCC=C(C)C)C=O	31.17691	CC1=CN=C(C)C(CC)=N1	129.885
CCCCC/C=C(CC)/C([O-])=O	44.65088	CN1CCCCC1	31.17583	CCC(C)C#N	129.1444
O=CC1=CC=C(OCC)C=C1	44.49086	CC1=CC=C(C(C)=C)CC1	31.11419	O=CCCC1=CC=CC=C1	128.9196
CC(OCCCCCC)(C)=O	44.25152	C=C(C)CCC	31.06791	CCCC1CCCC(O1)=O	128.6957
CCC(OC1=CC=CC=C1)=O	44.13589	CC(C(OC)=O)=C	31.04751	CCCCCC1C(C)C1	128.5922
CCCC1=CC=CC=C1	44.07246	CC/C=C/CC/C=C/C=O	30.89184	C#CCCCCCCCC	128.5503
C=C(C1CCC(C)C12CC=C(C)CC2)C	43.77837	CCCC(OCC)=O	30.8423	CC(=O)C1=CC=C(C=C1)OC	128.3256
CCCCCC1=CC=CO1	43.65871	CCC(C(COC)=O)=C	30.83327	CC1=CCC2(C1C2)C(C)C	128.1545
CC(C)=CCC1=C(C)C=CO1	43.61321	C=CC(OCC)=O	30.83192	CC(OC(CCC)C)=O	128.0058
CC1=NC=CN=C1CC(C)C	43.43477	CC(COC)=O	30.78695	CC1N(C(C)=O)CCCC1	127.756
CCCCCCC=O	43.1834	CCC(C)CC(OC)=O	30.61718	CCC1=CC=CO1	127.5711
CCC(CC)C(OCC=C)=O	43.09237	CCCC(OCC/C=C\CC)=O	30.53002	CCC(C)(C)CCC	127.5105
CCCCCC(OCC)=O	43.08604	C/C(C(OCC=C)=O)=C\C	30.51805	CCOCOCC	127.4569
C=CCCCCCCC	43.04082	CC(CCC1)OC21CCCO2	30.41643	COC1=CC=CC=C1OC(C)=O	127.3973
O=C1O/C(C2=C1C=CC=C2)=C\CCC	42.88664	CC(/C=C/CC)=O	30.40109	CCC(C)C(OC)=O	127.2072
CCC(C1=CC=CC=C1)=O	42.71533	CC(OC)CC	30.34567	CC(C)(C)CC(C)CC	126.9944
CCCC(OC(C(C)=O)C)=O	42.6727	O=COCCCC	30.31383	CC(C1=CC=CC=C1)C=O	126.5322
CC1=CCCC(=C)C2CC(C2CC1)(C)C	42.66021	CC(/C=C/C(C)C)=O	30.29812	CC(C1=CC=CC=C1)OCC	126.26

Table 3.14

RP4_C			RP6_A			RP6_B		
SMILES	Pred	Activity	SMILES	Pred	Activity	SMILES	Pred	Activity
CCCCCCO		88.03097	CC/C=C\CC=O		55.52031	CC(C12CCC(C1C2)=C)C		134.0782
CC(C1=CCC(C)=CC1)C		81.96087	CCCCCCCC=O		50.76842	CC1=CC=C(C(C)C)C=C1		130.4358
CCCC(OCC)=O		70.8636	CC/C=C\CC=O		55.52031	CC(C1=CC=CC=C1)=O		125.9246
C=C([C@H]1CC=C(C)CC1)C		56.16668	CCCC/C=C/CC=O		55.23 115	CC(OCC(C)C)=O		123.529
C=C([C@@H]1CC=C(C)CC1)C		53.44199	CCCCCCCCC=O		51.5647	CC1(C)[C@@]2([H])CC(C)=CC[C@@]12[H]		120.9403
O=CC1=CC=CC=C1		41.21922	O=CCCCC=O		50.84559	CCCSSCCC		119.8744
CCC(/C=C/C)=O		40.3888	CCCCCCCC=O		50.76842	CC(OCCCCC)=O		106.2103
CCC(OCCC)=O		88.57896	CCC(CC)C=O		49.23892	CCC/C=C\CO		97.64069
CCCCCCO		88.03097	CCCCC(CC)C=O		46.17592	CC(C1=CC=CC=C1)C		154.2561
NCCCCCN		82.40683	CCC(C)CCCC/C=C\CCCCCCC=O		44.74221	CC(C1=CC(C)=CC=C1)C		153.5955
CC1=CCC(=CC1)C(C)C		81.96087	CCCCCCCC/C=C\CCCCCC=O		44.26983	CC1=CC(CC)=CC=C1		152.9753
CCCCCCOC		77.64128	C/C=C/CC(C)C(C)C=O		44.2 1182	CCC1=CC=CC=C1		149.1351
CCC(CCCC)=O		77.38877	CCCCC/C=C\C/C=C\CCCCCCC=O		44.03917	CCCC1=CC=CO1		148.7293
C=CC(CCCC)=O		76.88151	CCCCC/C=C\C/C=C\CCCCCCCC=O		43.72791	CCC(C1=NC=CS1)C		143.9979
CCCC(OCC)=O		70.8636	CCCCCC/C=C\CCCCCC(CC=O)=O		42.81228	CSCC1=CC=CO1		143.4794
CCCCCC(O)=O		70.27421	CC/C=C\C/C=C\C/C=C\CCCCCCC=O		41.48486	CCC1=CC=CN=C1		141.5474
CC(CCCCO)=O		69.68066	CC(C)(C)CC(C)CC=O		41.48082	CC1=CC=CC=C1CC		139.4834
CCCC(CCC)=O		68.15898	CCCC/C=C\CC/C=C\CCCCCC=O		41.43222	CCC1=NC=CN=C1		138.4367
CCCCC(C(O)=O)=O		67.27872	CC(C)CC(C)CC=O		41.34868	CC1=CN=C(C)C(CC)=N1		137.8212
CC(CCCCC)=O		66.08893	CCCCCC/C=C\CCCCCCCC=O		40.79962	CCCC1=CC=CC=C1		136.8536
NCCCCCC		65.7308	CC(C)CC=O		40.45953	CCC1=CSC(CC)=N1		135.7359
CCCCC(OCCCC)=O		65.19794	CCCCC#CCCCCCCCCC=O		40.2463	CCC1=NC=CN=C1CC		135.3627
CCCCC(OC)=O		64.4434	C=CCC(C)C=O		40.06869	CSCC1=CC=CC=C1		134.9309
CCCC(OCCC)=O		64.29792	CC/C=C/CCC=O		39.8543	CCC1=NC(=CN=C1)C		134. 1139
CC1=CC=C(C(C)=C)CC1		63.89223	C/C=C/C/C=C\CCCCCCCC=O		39.68227	C=C1CCC2(C(C)C)CC12		134.0782
CCC(OCCCC)=O		60.78221	CC/C=C\CCC=O		38.46738	CC1=CC=C(CC)C=C1		133.5617
C=C1CCC(C(C)=C)=CC1		59.32255	CCCCCCCC/C=C\CCCCCCCC=O		37.37 114	CC1=NC=CN=C1CC		132.9788
CC(C1=CN=C(C)C=N1)C		58.79435	CCC(C)C=O		37.3392	CC1=CSC(C)C)=N1		132.7151
CC(OCCCC)=O		58.25072	CCCC/C=C\CCC=O		37.00024	CCCC1=CC=CC=C1		131.8252
CSCCCCO		58.21219	CC/C=C\CCCCC=O		36.90013	CC1=CC(CC)=CC=C1C		131.1524
CCCCC(OCCC)=O		58.07201	CCC(C)C=O		36.8585	CCC1=NC=CN=C1CC		131.0653
CC(/C=C/C(CC)=O)=O		58.02782	CCCCCCC=O		36.69429	CC1=CCC2(C1C2)C(C)C		130.8624
CC(CCC(CC)=O)=O		57.7253	O=CCC1=CC=CC=C1		35.66337	CC(O)(C1=CC=CC=C1)C		130.6334
O=C(C(C)=C1)C=C(C(C)C)C1=O		56.69925	CCCCCC=O		35.30428	CC1=CC=C(C(C)C)C=C1		130.4358
C=CCCCCC		56.53 117	CCCC/C=C\CCCCCCCCC=O		34.93723	CC(SSC(C)C)C		129.5921
CC1=CC[C@@H](CC1)C(C)=C		56.16668	CCCCC(/C=C/C=O)=O		34.48002	CC1C=CC2(C(C)C)CC12		129.4238
CCCCC(OCC)=O		55.37078	CCCCC=O		34.45637	CCC1C2CCOC(O2)(C)O1		128.56 11
CC(C)C1=CCC(=C)CC1		55.27365	CC=O		34.2548	CCCSSC		126.9605
O=C(OCC)OCC		55.10922	CCCC/C=C\CCCCCCCC=O		33.70733	CC1=NC=CN=C1OC		126.0013
O=CNC1=CC=CC=C1		54.64541	CCCC=O		33.21461	CC(C1=CC=CC=C1)=O		125.9246
CCCCCC=O		54.07587	CC/C=C\CCCCCCCC=O		31.90149	O=C(C1=CC=CN=C1)C		124.8349
C/C=C/C(OCC)=O		53.94009	CCC/C=C\C/C=C\C/C=C\CCCCCCCC=O		31.59909	CC1=CC=C(C)C)C		124.8332
CC1=CCC(CC1)C(C)=C		53.44199	CC(C1=CC=C(C(C)C)C)C=C1)C=O		31.14121	CC(C)CC1=NC=CS1		123.6646
C=C([C@@H]1CC=C(C)CC1)C		53.44199	CC/C(C)=C\CC/C(C)=C/CC(C)C(C)CC=O		31.08018	C=CCSSCCC		123.6242
O/N=C/C1=CC=CC=C1		52.41753	CC(C)C1=CC=C(C=C1)CC(C)C=O		31.03749	CCC1=CC(CC)=CC=C1		123.591
CC(CC/C=C/C)=O		52.36687	CCCCC/C=C\C/C=C\CCC=O		29.83813	CC(OCC(C)C)=O		123.529
O=C(O)C1=CC=C(O)C=C1		52.15408	CC(SC)CC=O		29.71734	COC1=NC=CN=C1		123.5075
C=C1CCC(C(C)=C)C=C1		51.85418	CCC=O		29.61341	CC1C(OC(C)=O)CCC1		122.6161
C/C=C/CCO		51.58962	CCCC/C=C\CCCCCCCCC=O		29.36574	CC1=NC=C(CC)C=C1		122.4707
C/C=C/C=C/C(OCC)=O		50.88131	CCCC/C=C/CCCCCCCCC=O		29.09899	C1(C2=CC=CC=C2)CO1		122.2966
C=CCCCCO		50.85841	CC/C=C\CCCCCCCCC=O		28.85484	CCC1=CC=C(C=C)C=C1		122.2946
O=C(C)CC(OCC)=O		50.68649	CCCC/C=C\CCC/C=C/CCCCC=O		28.75354	C[C@H](C1=CC=CC=C1)CO		121.2901
C=CCC(O)CC=C		50.35582	CCCCCC/C=C\CCCCCC=O		28.32705	OCCC1=CC=CC=C1		120.9415
C=CC(OCCCC)=O		50.149	O=CCCSC		28.02997	CC1=CC=C(C)C=O)O1		120.6793
CC=C(C)C=CC=C(C)C		49.62951	CCCCCC/C=C\CCCCCCCCC=O		27.82103	CCCC1CCC(O1)=O		120.0403
CC1=CC=C(CC1)C(C)C		49.24534	CCC(C)CCCC(C)CCC=O		27.55956	CCC1=NC=CN=C1C(C)C		119.9674
CCC(/C=C/C=O)=O		49.20329	CCCCCCCC/C=C\CCC=O		27.54716	CCCSSCCC		119.8744
CCC(CCCCC)=O		49.01636	CC(CCC=C(C)C)C=O		27.33422	CC1=NC=CN=C1SC		119.8238
CC1C=C(CCC1C=O)C		48.64532	O=C(O)CCCCCCCC=O		26.61495	CC(C1=CN=C(C)C=N1)C		119.3759
CCCC(OCCCC)=O		48.47395	CC1=CCC(C1(C)C)CC=O		25.84082	CC12CCCC(O2)C(CC)O1		119.3573
C=CC(CCCCC)=O		48.31301	CC(C)C=O		25.62622	C=C(C1=CC=CC=C1)C		119.3014
CCCCCO		48.17758	O=CC(C)CC1=CC=C(C(C)C)C)C=C1		25.47008	SCCC1=CC=CC=C1		119.2656
C/C=C/C(OC)=O		48.14932	CCCCCCCC/C=C\CCCCCCCCC=O		25.15138	C=CCCC(C)C		118.642
O=CCCCCC		48.08631	CCCCC/C=C/CCCCC=O		25.14498	O=C(C1=C(C)N=C(C)S1)C		118.2089
CCCCC(/C=C/C)=O		48.05477	CC1=C(C(CCC1)(C)C)CC=O		24.59309	CC1(C)C2CC=C(C)CC21		118.1076
O=C(OC)CC(OC)C=O		47.82076	CC/C=C\C/C=C\CCC=O		23.99376	CCC1=CC(=CC=C1)CO		117.0701
CCCC(CC(OCC)=O)=O		47.81427	CC/C=C\CCCCCCC=O		23.89949	CCOC1=NC=CN=C1C		116.6756
OCCC1CCCC1		47.64478	CCC(/C=C/C=O)=O		23.35864	CC(OC(C)CC)=O		116.6452
CC(/C=C/CCC)=O		47.06617	CCC(C/C=C\C=C/CCCCCCCCC=O		22.7958	CC1=CC=C(C)C(C)=C1		116.1719
CCC/C=C(C1)C(O)=O		46.97055	O=CCC1=CC=C(C)C=C1C		22.71871	CSCC1=NC=CN=C1CO1		116.0629
CCCC1SC(C)OCC1		46.92427	CC1(C)C(CC12)CCC2C=O		22.56 11 1	CC1=NC=C(CC)N=C1C		115.2144
CC(/C=C/C=C(C)C)=O		46.84 116	CCC/C=C/C=C\CCCCCCCCC=O		21.89512	CCCCCC1=CC=CO1		114.4769
CCCCC(O)=O		46.45414	CC/C=C\C=CCCCCCCCCC=O		21.57717	CCC1=NC=CN=C1C(C)=O		114.2716
O=C(CC)CCC		45.75258	CCCC/C=C\C=C\CCCCCCCC=O		21.27533	CC1OC2(C)C(C)CCC1O2		114.1827
CC(C(CCCC)=O)=O		45.67465	CC1=CC=C(O1)C(C)CC=O		20.49682	O=CC1=CC=C(CC)C=C1		113.986
CC(C)C(CC1)C=CC1=C		45.54189	CC/C=C/C=C/CCCCCCCCC=O		20.17205	CC1=NC=C(C)N=C1C		112.9504
O=C(N1C)N(C)C=CC1=O		45.3955	CCC/C=C/C=C/CCCCCCCCC=O		19.93297	O=C(C)C1=CC=C(C)C=C1		112.6495
CCCCCCS		45.29639	CC(C1OC(C=C)C)CC1)C=O		19.87415	CC(CCSC)=O		112.624
O=C(N1C)N(C)C2=C(C=CC=C2)C1=O		45.09746	CC(CC1=CC2=C(C=C1)OCO2)C=O		19.81655	CC(CCC1)OC21CCCO2		112.1986
CCCCCCC		44.3957	CCCCCC/C=C\CCCC=O		19.53233	CC(CCC(C)C)=O		111.6846
CCC(C1=CC=CC=C1)=O		44.29157	O=CC1C(C(C)=O)CCC1C		19.21828	CCC(OC(CCC)C)=O		110.9251
O=C(O)CC1CCCCC1		44.27086	O=CC1(C)CC(C)=C(C)CC1		19.18392	CC(OC(C)COC)=O		110.8575
CCC([C@@H](C)[C@H](O)CC)=O		44.07604	CCC/C=C\C=C\CCCCCCCCC=O		19.0844	CCOC1=NC=CS1		110.2391
CCC/C(C)=C(C)O)=O		44.07124	CC1C=CC(CC1=O)C		18.8576	O=CC1=NC=CC=C1)C		109.7006
C=CC(O)CCCC		43.82648	CCCC/C=C\CCC/C=C/C=C/CCC=O		18.75997	CCCCC1=CC=CO1		109.6941
CCCCCC(OC)=O		43.76267	O=C1C(O2)C2(CC(C)C1)C)C=O		17.82453	C=CCSSC		108.6 116
CCCC(C(O)=O)=O		43.67156	O=CCCC1=CC=CC=C1		17.35709	CC(CC(C)CC)=O		107.924
O=C(N1)NC2=C(NC(N2)=O)C1=O		43.43836	O=CCCCCCCCCOC(C)=O		16.94432	CCC(C)C(OC)=O		107.4248
CC(CCC=C(C)C)=O		43.3631	CCCC/C=C/CCCCCCCCCC=O		16.89494	O=CC1=CC=CC(CC)=C1		107.4013
O=C(OCC)C1=CC=CO1		43.12843	CC/C=C/C=CCCCCCCCCC=O		16.89059	CC1=CN=C(CC)C)C=N1		107.3825
CCC(C1CCC(O1)=O		43.0524	CC/C=C/C=C\CCCCCCCC=O		16.85496	CCCC(SC)CCO		107.1016
CCC(OCC=C)=O		42.9705	CCCCC(CCCC)=O		16.82522	CCCC(C)C(OCC)=O		107.0201
CCC(N1CCCCC1)=O		42.48007	CCCCCCCCCC(C)C=O		16.71336	CCC1=CC=CC=C1CC		106.9824
C=C/C=C/CCCC		42.31857	O=C(C)C=O		16.69362	CCCS(C)SCCC		106.9291
CC(C)C(CCCC)=O		42.26556	CC(C1=CC=CC=C1)C=O		16.54825	CC1CCC2=NC=CN=C12		106.774
CC(C)=CC(OCC)=O		42.0531	C=C/C=C/CCCCCCCCC=O		16.28343	CC1(C)C2CCC(C)=CC12		106.7335
O=C(OC)CCOC		41.98418	CCCCC/C=C/C(C)CC=O		15.97413	CC(OCCCCC)=O		106.2103
CC(C)C/C=N/OC		41.96767	CC/C=C\C=C/CCCCCCCCC=O		15.75651	CCCC1OCC(O1)C		105.9842
C=CCSCC=C		41.92669	CC/C=C\C=C/CCCCCCCC=O		15.71584	CC1=CC(C(C)=O)=C(C)S1		105.7342
CCC(C(C)CCC)=O		41.67139	O=CC1(C)CCC(C)=CO1		15.27656	CCCCC1CCC(O1)=O		105.613

Table 3.14 Continued

RP6_C			RP7_B			RP7_C		
SMILES	Pred	Activity	SMILES	Pred	Activity	SMILES	Pred	Activity
CCCC(OCC)=O		86.13873	C/C1=C\CC(C)(C)/C=C/C/C(C)=C/CC1		126.4586	CC/C=C\CC=O		36.52692
CC1=CCC(C(C)=C)CC1=O		79.27182	C=C1CC/C=C(C)/CC[C@@]2([H])C(C)(C)C[C@]12[H]		119.7731	CCCCCCCC=O		30.50133
CCCCCCO		67.29345	CC(C1=CC=CC=C1)=O		95.38008	CC(OCC(C)C)=O		16.84203
CC(C1=CCC(C)=CC1)C		65.13451	CCCC(OCC)=O		76.21054	CC/C=C\CC=O		36.52692
OC1=CC=CC=C1C(OC)=O		58.16259	CCCCCCO		67.77761	CCCCCC=O		33.78351
C=C([C@H]1CC=C(C)CC1)C		54.28428	C/C1=C\CC(C)(C)/C=C/C/C(C)=C/CC1		126.4586	O=CCCSC		33.33575
CCCC(OCC)=O		86.13873	CC(C1=CC=CC=C1)=O		95.38008	CC(C)CC=O		31.71206
CCC(OCCC)=O		83.38456	CCCC(CCC)=O		92.59738	CC/C=C/CCC=O		31.37961
CC1=CCC(C(C)=C)CC1=O		79.27182	CC(CC(CC(C)C)=O)=O		91.15 114	CCC(C)C=O		31.0054
O=C(OC)CCSC		79.13522	CC(CCCC(C)C)=O		89.91984	CCCCCCCC=O		30.50133
C=CCCCCC		77.20856	CCC(C)C(OC(C)C)=O		87.99201	CCCCCCC=O		29.72137
C=CC(CCCC)=O		76.38981	CC(CCC=C(C)C)=O		87.32837	C=CCC(C)C=O		29.40163
CCCCC(OC)=O		76.16308	CC(C)CC(OC(C)C)=O		87.14336	O=CCCCC=O		28.98876
CC(C)C(CC1)C=CC1=C		74.89678	CC(C)CCC=C(C)C		83.85426	CC/C=C\CCC=O		28.42686
O=C(C(C)=C1)C=C(C(C)C)C1=O		74.12276	CC(OCCC(C)C)=O		83.3632	CC(CC1=CC2=C(C=C1)OCO2)C=O		26.77087
O=C1CC(CC)OC(CC)=C1		73.47661	CC(OC(CCC)C)=O		82.94325	CCCC/C=C/CC=O		26.07484
C=CC(O)CCCC		73.1929	O=C(C1=CC=CO1)OC		81.97053	CCCC=O		25.96475
CCC(CCCC)=O		71.946 11	CC1=CCC(C2=C1C=CC(=C2)C)C(C)C		80.79637	CCC(CC)C=O		25.59708
CC/C=C/C(OC)=O		71.24633	O=C(C1=CC=CN=C1)C		80.65999	CC(C)CC(C)CC=O		25.02162
CCC(OCC=C)=O		71.14553	CC(CC(CC(C)C)=O)C		80.58242	CCCCC(CC)C=O		24.15837
O=C1C=C(CC)OC(C)=C1O		71.05457	CC(C)C(CCC(C)=O		80.41422	CCCC(C)C=O		23.89254
C=CCSSSC		70.38179	CC(OCCC(C)=C)=O		80.22339	CC(C)C=O		23.88156
C=C1CCC(C(C)=C)C=CC1		70.27329	CC1=CCCC(=C)C2CC(C2CC1)(C)C		76.99125	CCCCC=O		23.52 119
NCCCCCC		69.99448	CC(NCCC(C)C)=O		76.5694	CC(SC)CC=O		21.55237
C/C=C/CC(OC)=O		69.2848	CC1=CCC(C=CCC(=CCC1)C)(C)C		76.23738	CC(C1OC(C=C)(C)CC1)C=O		21.16626
CC(CC/C=C/C)=O		69.24917	CCCC(OCC)=O		76.21054	O=C1C(O2)C2(CC(C)(C1)C)C=O		18.74665
C=C(C)C(CCC1C)CC1=O		68.66812	N#CC(C1=CC=CC=C1)=O		75.54794	CCC(C)C[N+](][O-])=O		18.60151
O=C1C=C(C)CCC1C(C)C		68.14354	CC(C)C(CC(CC)=O		75.52013	CCCCC/C=C\C/C=C\CCC=O		18.52774
C=C(C)C(CCC1C)=CC1=O		67.73656	CC(CCC(CC)=O)=O		75.43823	SCC1=CC=CO1		18.04152
CC(OCCCC)=O		67.64093	CC(OCC=C(C)C)=O		74.71706	O=C(OC)C1=C(C)C=CC=C1O		17.93662
CCCCCCO		67.29345	CC(/C=C/C=C(C)C)=O		74.55248	CCCCN=C=O		17.56101
C=C1CCC(C(C)=C)C=C1		66.99069	O=C(C1=NC=CC=C1)C		74.49208	O=CCC1=CC=C(C)C=C1C		17.37472
C=C/C=C/C/C=C/C		66.84646	CC(C1=CC=CC=C1)C		74.0 1133	O=CCC1=CC=CC=C1		17.32718
C/C=C/C(OCC)=O		66.64964	C/C=C/[C@@H](O)CC(C)C		73.71874	CC(C1=CC=CC=C1)C=O		17.29056
CCOC1=CC=CC=C1		65.88814	CC1CCC(C2=C1CCC2C)C(=C)C		73.5887	CC1SCCN1		17.16448
O=C(N1C)N(C)C2=C(C=CC=C2)C1=O		65.85213	CC(C)CC(OCC)=O		73.46298	CCC(C)/C=N/OC		17.12756
CC(C)C1=CCC(=C)CC1		65.65882	CCCC(O)CCC		73.3584	CC(C)CC[N+](][O-])=O		16.86213
CC1=CCC(=CC1)C(C)C		65.13451	O=C(O)C1=CC=CC=C1		73.31343	CC(OCC(C)C)=O		16.84203
CCC/C=C/C(O)=O		63.74982	CC(C)=C1CCC(C)C2=C(C(C)CC2)C1		72.78 11	SCCCCCCS		16.78584
CCCCCC(O)=O		63.63551	N#CCC1=CC=CC=C1		72.23238	O=C(O1)OCC1C		16.65321
C=CCCCCO		63.27646	C=C(C)C(=C\C/C=C(/)CCC=C(C)C		71.73195	O=C([O-])OS(C)SSC		16.58147
CCCC1CCCC(O1)=O		62.21691	CCC(CC(C)C2=C1CCC2C)C(=C)C		71.16331	CC(C)C[N+](][O-])=O		16.55789
CSSSCSC		62.04977	C=C(C1=CC=CC=C1)C		70.22063	[O-]C(C[N+](C)(C)C)=O		16.53871
O=C(OCC)CSC		61.81868	CCCCCC(O)=O		70.09967	O=COCC=C(C)C		16.48318
CC1=NC2=CC=CC=C2N=C1		61.46244	CC(O)CC(C)CCC		69.75241	CC(NCC(C)CC)=O		16.30734
CCCC1SC(C)OCC1		61.36557	CC(C)=CC(OCC)=O		69.04661	SCCCCCS		16.24069
O=C1(C/C=C\C)=CCCC1C		60.99415	CC1CCC2=C(CCC(CC12)C(=C)C)C		68.915	CCCCC/C=C/CCCCC=O		16.21917
CC1=C(CCCC)C(CC1)=O		60.61307	CCC(CCCC)=O		68.86344	CC(C)C(OCC)=O		16.10514
CSCCCCO		60.4865	CC(O)C1=CCC=CC=C1		68.28692	CCC=O		16.09149
C=C/C=C/CCCC		60.14782	CCCCCCO		67.77761	O=C/C(C)=C/CCC1=COC=C1		16.08724
CC1=NC2=CC=CC=C2N=C1C		59.98227	CC(C)CC(C)CCC		67.66886	OC1=C(OC)C=C2CC=COC2=C1		16.08161
C/C=C/CCCO		59.95597	CC(CCCCC)=O		67.50108	CC(NCCC(C)C)=O		16.05204
CC1=CC=C(C(C)=C)CC1		59.7741	CC1=CC2C(C(C)C)CCC(C)C2=CC1		67.2853	O=C(OC)CCSC		16.0168
CCCC/C=C/O		59.55189	O=C(C1=NC=CN=C1)C		67.07072	OC1=CC(C)=CC(OC)=C1C		15.84205
O=C(OC)C1=CC=CC=C1SC		59.29277	CC(OC(C)COC)=O		66.89404	C/C(CCC(O1)=C1(C)C)=C/C=O		15.74377
CC(C)C/C=N\OC		59.19607	[C-]#[N+]C1=CC=CC=C1		66.39249			
CC(/C=C/CCC)=O		59.08062	SCCC1=CC=CC=C1		66.19192			
CC1CCC(C(C0C2=O)C)C2C1		58.72099	CC1=C2C=CC=CC2=CC=C1		66.18268			
O=C1C=C(C(C)C)CCC1C		58.68 119	CC(CC/C=C/C)=O		66.06761			
OC1=CC=CC=C1C(OC)=O		58.16259	CC(C1=CC2=C(C)C=CC2=C(C)C=C1)C		65.222			
CC1=CN=C(CC)C(C)=N1		58.04722	CC(C)(C)CC(C)CCO		65.18269			
CCC(OC1=CC=CC=C1)=O		57.98341	CC1=CC=C(C=O)C=C1		64.98162			
CCC(/C=N\O)C		57.90524	CC(C)C(O)C(C)(C)C		64.94732			
C=CCSCC=C		57.6607	C=CCC(O)CC=C		64.92842			
O=C1(C/C=C/C)=CCCC1C		57.63992	C=CC1(C)C(C(C)=C)CC(C(C)=C)CC1		64.86062			
CC(C1=CN=C(C)C=N1)C		56.97493	COC(C=C)(CC/C=C(CCC=C(C)C)\C)C		64.52239			
CCCC(CCC)=O		56.07365	C/C(C(OC(C)C)=O)=C\C		64.35953			
CN1C(C2=NC=CC=C2)CCC1		55.89354	O=C1C(CC)CCCC1		64.20792			
CC(CCCCC)=O		55.65697	CCCCC(CC)CO		64.0447			
CCC(O)CCCC		55.03286	O=C(OC)COCC		63.7288			
CN1C(CCC1)C2=CC=CN=C2		54.53139	CC1=CC2C(C(C)C)CCC(C)=C2CC1		63.04465			
C12=CC=CC=C1N=CC=N2		54.33279	CCC(C1=CC=CC=C1)O		62.80598			
CC1=CC[C@@H](CC1)C(C)=C		54.28428	O=C(CC)CC		62.68223			
CC(C1CC(C(C)=C)C2)(O1)C2=O		54.25887	CC(C(O)CCCC)=O		62.66521			
O=C(N1)N=C2N=CN=C2C1=O		54.08839	CCC(C1=CC=CC=C1)=O		62.60064			
O=C1C=CC2=C(O1)C=C(O)C=C2		53.91696	CC(C)=CC(C=C(C)C)=O		62.25825			
CCC(N1CCCC1)=O		53.80537	C=CC1=CC=CC=C1		62.19846			
OC1=CC=CC=C1CCC		53.80223	O=C(N(C1=O)C)N(C2=C1N(C=N2)C)C		61.90239			
C12=CC=CC=C1N=CC=C2		53.64288	CCCCCCS		61.85532			
O=C(OCC)C1=CC=CC=C1		53.42226	C=C(C)C(OCC(C)C)=O		61.75396			
O=C(OCC)OCC		53.40861	CC(C)CC(O)CCC		61.6526			
O=C(C=C1)C=C(CC)C1=O		53.16238	C=C(C1C/C=C(C)/CC/C=C(C)CC1)C		61.64239			
O=C1OC2=C(C=CC=C2)C=C1		52.39659	CC1=CC=CC=C1CC		61.608 11			
O=C(OC)C1=CC=CC=C1OC		52.29142	CC(O)CCCC=O		61.58498			
CC(C/C=C/CC)=O		52.16263	CC(C)C(CCC(C)C)=O		61.33144			
CC1COC2C1CCC(=C2)C		51.91039	O=COCCC(C)C		61.20754			
CCC(C1=CC=CC=C1)=O		51.65885	CC1=CCC2(CC1)C(=C)CCCC2(C)C		61.06316			
O=C1C=C(C)CC[C@H]1C(C)=C		51.52354	C=C1CCCC(C)C)C](C@@]21CCC(C)=CC2		61.01872			
COC1=CC=CC=C1C(OC)=O		51.519 11	CC(O)C(C)CCCC		61.01721			
OC1=CC=CC=C1C(OCC)=O		51.470 11	COC1=CC=CC=C1C		60.96861			
CCCCCCS		51.09141	CCC(C)C(CCC)=O		60.7091			
OC1CC2CCCCC2CC1		50.99655	CC(OC1=CC(C)=CC=C1C(C)C)=O		60.625 11			
CCC(OCC=C)=S		49.9 1146	CC1=CC2C(=C(CCCC2(C)C)C)CC1		60.32029			
CCCC(C1=CC=CC=C1)=O		49.87987	CCCCCC=O		60.2745			
CCCCCCOC		49.77356	C=C([C@@H]1C[C@@H]([C@](C=C)(CC1)C)C(C)=C)C		60. 1139			
O=C(OC)C1=CC=CC=C1N		49.71818	CC(O)COCC(O)C		59.55436			
O=C(C)CC(OCC)=O		49.69014	C=CC1=CN=CC(C)=N1		59.54975			
CCC(OCC)=O		49.61974	CC(O)CCCCC		59.50504			
CSSSSSC		49.44057	CC1=CN=C(C)C(C(C)CC)=N1		59.12154			
O=C1SC2=CC=CC=C2N1		48.96835	CCC(C(C)C(CC)=O)=O		59.03413			

Table 3.14 Continued

Table 3.15: Optimized descriptor sets for each mammalian OR

Optimized descriptors occurrences, symbol, brief description, class, and dimensionality are listed. A summary of the total number of descriptors selected for the receptor repertoire is provided at the beginning. Descriptors are listed in ascending order of when they were selected into the optimized set. Weights indicate the number of times a descriptor was selected in an optimized descriptor set.

Descriptor Class Type Counts for all Ors

GETAWAY descriptors	109
atom-centred fragments	49
2D autocorrelations	48
RDF descriptors	48
3D-MoRSE descriptors	46
WHIM descriptors	43
functional group counts	33
2D binary fingerprints	26
Burden eigenvalues	23
edge adjacency indices	21
geometrical descriptors	21
topological descriptors	14
2D frequency fingerprints	13
topological charge indices	12
atomtypes (Cerius2)	11
molecular properties	11
walk and path counts	7
constitutional descriptors	6
Randic molecular profiles	5
topological (Cerius2)	4
information indices	4
connectivity indices	3
structural (Cerius2)	1
eigenvalue-based indices	1
charge descriptors	0

Dimensionality Counts (Weights Included)

Num zero dimensional descriptors:	7
Num one dimensional descriptors:	104
Num two dimensional descriptors:	176
Num three dimensional descriptors:	272

Origin (Weights Included)

Num Dragon descriptors:	546
Num Cerius2 descriptors:	13

Dimensionality Counts (Weights Excluded)

Num zero dimensional descriptors:	7
Num one dimensional descriptors:	37
Num two dimensional descriptors:	93
Num three dimensional descriptors:	155

Origin (Weights Excluded)

Num unique Dragon descriptors:	284
Num unique Cerius2 descriptors:	8

Odor Receptor Name	Weight	Symbol	Description	Class	Dimenionality
MOR1.1					
	2	Mor17m	3D-MoRSE - signal 17 / weighted by atomic masses	3D-MoRSE descriptors	3
	8	H-051	H attached to alpha-C	atom-centred fragments	1
	2	R6p+	R maximal autocorrelation of lag 6 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	4	Mor23e	3D-MoRSE - signal 23 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	5	Mor30m	3D-MoRSE - signal 30 / weighted by atomic masses	3D-MoRSE descriptors	3
	4	R5v+	R maximal autocorrelation of lag 5 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	Mor32e	3D-MoRSE - signal 32 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	JGI7	mean topological charge index of order7	topological charge indices	2
	2	E1s	1st component accessibility directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	nCt	number of total tertiary C(sp3)	functional group counts	1
	1	nArCO	number of ketones (aromatic)	functional group counts	1
	1	O-058	#NAME?	atom-centred fragments	1
	1	B07[C-O]	presence/absence of C - O at topological distance 07	2D binary fingerprints	2
	2	GATS2m	Geary autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations	2
	1	RDF110e	Radial Distribution Function - 11.0 / weighted by atomic Sanderson electronegativities	RDF descriptors	3
	1	nCconj	number of non-aromatic conjugated C(sp2)	functional group counts	1
	1	C-006	CH2RX	atom-centred fragments	1
	1	nCrs	number of ring secondary C(sp3)	functional group counts	1
	1	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	C-003	CHR3	atom-centred fragments	1
	1	MATS8m	Moran autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
MOR106.1					
	1	Mor25p	3D-MoRSE - signal 25 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	BEHe6	highest eigenvalue n. 6 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	IC1	information content index (neighborhood symmetry of 1-order)	information indices	2
	2	C-006	CH2RX	atom-centred fragments	1
	2	RTe+	R maximal index / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	piPC07	molecular multiple path count of order 07	walk and path counts	2
	1	RDF045e	Radial Distribution Function - 4.5 / weighted by atomic Sanderson electronegativities	RDF descriptors	3
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	R7v+	R maximal autocorrelation of lag 7 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	HOMT	HOMA total	geometrical descriptors	3
	1	RDF035m	Radial Distribution Function - 3.5 / weighted by atomic masses	RDF descriptors	3
	1	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	1	SHP2	average shape profile index of order 2	Randic molecular profiles	3
MOR107.1					
	15	nCq	number of total quaternary C(sp3)	functional group counts	1
	1	Mor07v	3D-MoRSE - signal 07 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	BELv3	lowest eigenvalue n. 3 of Burden matrix / weighted by atomic van der Waals volumes	Burden eigenvalues	2
	1	R4p+	R maximal autocorrelation of lag 4 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	E1u	1st component accessibility directional WHIM index / unweighted	WHIM descriptors	3
	1	JGI4	mean topological charge index of order4	topological charge indices	2
	1	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes	geometrical descriptors	3
	1	nR06	number of 6-membered rings	constitutional descriptors	0
	1	RDF040m	Radial Distribution Function - 4.0 / weighted by atomic masses	RDF descriptors	3
	1	R8e+	R maximal autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	piPC07	molecular multiple path count of order 07	walk and path counts	2
	1	L2s	2nd component size directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	piPC08	molecular multiple path count of order 08	walk and path counts	2
	1	nDB	number of double bonds	constitutional descriptors	0
	1	H-051	H attached to alpha-C	atom-centred fragments	1
	1	E1m	1st component accessibility directional WHIM index / weighted by atomic masses	WHIM descriptors	3
	1	B05[C-O]	presence/absence of C - O at topological distance 05	2D binary fingerprints	2
	1	BELp3	lowest eigenvalue n. 3 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
MOR129.1					
	1	Dv	D total accessibility index / weighted by atomic van der Waals volumes	WHIM descriptors	3
	2	HATS7e	leverage-weighted autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	F01[C-C]	frequency of C - C at topological distance 01	2D frequency fingerprints	2
	1	Mor10u	3D-MoRSE - signal 10 / unweighted	3D-MoRSE descriptors	3
	1	HATS5m	leverage-weighted autocorrelation of lag 5 / weighted by atomic masses	GETAWAY descriptors	3
	2	H1v	H autocorrelation of lag 1 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	Mor11v	3D-MoRSE - signal 11 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	B05[C-C]	presence/absence of C - C at topological distance 05	2D binary fingerprints	2
	1	RDF085e	Radial Distribution Function - 8.5 / weighted by atomic Sanderson electronegativities	RDF descriptors	3
	1	Dm	D total accessibility index / weighted by atomic masses	WHIM descriptors	3
	1	H0m	H autocorrelation of lag 0 / weighted by atomic masses	GETAWAY descriptors	3
	1	Mor10e	3D-MoRSE - signal 10 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	D/Dr06	distance/detour ring index of order 6	topological descriptors	2
	1	HATS6m	leverage-weighted autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
	1	MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	RDF035m	Radial Distribution Function - 3.5 / weighted by atomic masses	RDF descriptors	3
	1	H7e	H autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	H0v	H autocorrelation of lag 0 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	R5m	R autocorrelation of lag 5 / weighted by atomic masses	GETAWAY descriptors	3
MOR136.1					
	1	RDF030m	Radial Distribution Function - 3.0 / weighted by atomic masses	RDF descriptors	3
	1	S_dssC	S_dssC	atomtypes (cerius2)	1
	1	Mor05m	3D-MoRSE - signal 05 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	BELe1	lowest eigenvalue n. 1 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	R5u+	R maximal autocorrelation of lag 5 / unweighted	GETAWAY descriptors	3
	1	SHP2	average shape profile index of order 2	Randic molecular profiles	3
MOR139.1					

Table 3.15

	2	H1v	H autocorrelation of lag 1 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	HATS1u	leverage-weighted autocorrelation of lag 1 / unweighted	GETAWAY descriptors	3
	1	TPSA(NO)	topological polar surface area using N,O polar contributions	molecular properties	1
	1	piPC06	molecular multiple path count of order 06	walk and path counts	2
	1	H8m	H autocorrelation of lag 8 / weighted by atomic masses	GETAWAY descriptors	3
	1	GVWAI-80	Ghose-Viswanadhan-Wendoloski drug-like index at 80%	molecular properties	1
	1	R1v	R autocorrelation of lag 1 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	GATS2m	Geary autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations	2
	1	Mor24e	3D-MoRSE - signal 24 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
MOR162.1					
	1	HOMA	Harmonic Oscillator Model of Aromaticity index	geometrical descriptors	3
	1	HATS5m	leverage-weighted autocorrelation of lag 5 / weighted by atomic masses	GETAWAY descriptors	3
	1	E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	H2e	H autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	R6v+	R maximal autocorrelation of lag 6 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	P2e	2nd component shape directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	RCI	Jug RC index	geometrical descriptors	3
	1	HATS6m	leverage-weighted autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
	1	RDF035m	Radial Distribution Function - 3.5 / weighted by atomic masses	RDF descriptors	3
	1	H1e	H autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
MOR170.1					
	3	C-025	R--CR--R	atom-centred fragments	1
	1	F05[C-O]	frequency of C - O at topological distance 05	2D frequency fingerprints	2
	1	R6v+	R maximal autocorrelation of lag 6 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	2	E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	S_aaCH	S_aaCH	atomtypes (cerius2)	1
	2	RTe+	R maximal index / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	R8u+	R maximal autocorrelation of lag 8 / unweighted	GETAWAY descriptors	3
	1	ATS2p	Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	RDF045v	Radial Distribution Function - 4.5 / weighted by atomic van der Waals volumes	RDF descriptors	3
	1	P2p	2nd component shape directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
	1	R7p+	R maximal autocorrelation of lag 7 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	R5m	R autocorrelation of lag 5 / weighted by atomic masses	GETAWAY descriptors	3
	1	HOMA	Harmonic Oscillator Model of Aromaticity index	geometrical descriptors	3
MOR184.1					
	1	S_dCH2	S_dCH2	atomtypes (cerius2)	1
	1	GATS1m	Geary autocorrelation - lag 1 / weighted by atomic masses	2D autocorrelations	2
	1	H-047	H attached to C1(sp3)/C0(sp2)	atom-centred fragments	1
	1	Qindex	Quadratic index	topological descriptors	2
	1	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes	geometrical descriptors	3
	1	L2v	2nd component size directional WHIM index / weighted by atomic van der Waals volumes	WHIM descriptors	3
	1	nCIC	number of rings	constitutional descriptors	0
	1	HATS2v	leverage-weighted autocorrelation of lag 2 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	Mor07u	3D-MoRSE - signal 07 / unweighted	3D-MoRSE descriptors	3
	1	R7e	R autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	MPC06	molecular path count of order 06	walk and path counts	2
	1	EEig09r	Eigenvalue 09 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	nCIR	number of circuits	constitutional descriptors	0
	1	RDF045m	Radial Distribution Function - 4.5 / weighted by atomic masses	RDF descriptors	3
	1	BELp3	lowest eigenvalue n. 3 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	1	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	L2p	2nd component size directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
	1	HATS7p	leverage-weighted autocorrelation of lag 7 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	R4p+	R maximal autocorrelation of lag 4 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	Mor10m	3D-MoRSE - signal 10 / weighted by atomic masses	3D-MoRSE descriptors	3
MOR185.1					
	1	BAC	Balaban centric index	topological descriptors	2
	1	HATS2m	leverage-weighted autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	R8u+	R maximal autocorrelation of lag 8 / unweighted	GETAWAY descriptors	3
	1	X5A	average connectivity index chi-5	connectivity indices	2
	1	E1e	1st component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	RTe+	R maximal index / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	GATS2m	Geary autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations	2
	1	H1p	H autocorrelation of lag 1 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	HATS7p	leverage-weighted autocorrelation of lag 7 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	RDF035m	Radial Distribution Function - 3.5 / weighted by atomic masses	RDF descriptors	3
	1	SHP2	average shape profile index of order 2	Randic molecular profiles	3
	1	R2p+	R maximal autocorrelation of lag 2 / weighted by atomic polarizabilities	GETAWAY descriptors	3
MOR189.1					
	1	nCrs	number of ring secondary C(sp3)	functional group counts	1
	1	V-DIST-mag	V-DIST-mag	topological (cerius2)	2
	1	J3D	3D-Balaban index	geometrical descriptors	3
	2	Atype_C_40	Number of Carbon Type 40	atomtypes (Cerius2)	1
	1	EEig11r	Eigenvalue 11 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	R4m	R autocorrelation of lag 4 / weighted by atomic masses	GETAWAY descriptors	3
	1	Mor07p	3D-MoRSE - signal 07 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	GVWAI-80	Ghose-Viswanadhan-Wendoloski drug-like index at 80%	molecular properties	1
	1	RDF025m	Radial Distribution Function - 2.5 / weighted by atomic masses	RDF descriptors	3
	1	B07[C-C]	presence/absence of C - C at topological distance 07	2D binary fingerprints	2
	1	R4u+	R maximal autocorrelation of lag 4 / unweighted	GETAWAY descriptors	3
	1	Mor22e	3D-MoRSE - signal 22 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	O-058	#NAME?	atom-centred fragments	1
	1	EEig07d	Eigenvalue 07 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	EEig11d	Eigenvalue 11 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
MOR2.1					
	1	RDF045m	Radial Distribution Function - 4.5 / weighted by atomic masses	RDF descriptors	3
	2	GVWAI-80	Ghose-Viswanadhan-Wendoloski drug-like index at 80%	molecular properties	1
	1	BELe7	lowest eigenvalue n. 7 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	RDF050m	Radial Distribution Function - 5.0 / weighted by atomic masses	RDF descriptors	3
	1	Mor14e	3D-MoRSE - signal 14 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	EEig12x	Eigenvalue 12 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	Mor16m	3D-MoRSE - signal 16 / weighted by atomic masses	3D-MoRSE descriptors	3
	2	EEig12r	Eigenvalue 12 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	Mor26p	3D-MoRSE - signal 26 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	Mor09u	3D-MoRSE - signal 09 / unweighted	3D-MoRSE descriptors	3
	1	nCconj	number of non-aromatic conjugated C(sp2)	functional group counts	1
	1	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	R3u+	R maximal autocorrelation of lag 3 / unweighted	GETAWAY descriptors	3
	1	B07[C-O]	presence/absence of C - O at topological distance 07	2D binary fingerprints	2
	1	F02[C-C]	frequency of C - C at topological distance 02	2D frequency fingerprints	2
MOR203.1					
	6	B06[C-C]	presence/absence of C - C at topological distance 06	2D binary fingerprints	2
	1	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	1	Mor08p	3D-MoRSE - signal 08 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	Hbond acceptor	Number of Hydrogen bond acceptors	structural (Cerius2)	0
	1	RDF075u	Radial Distribution Function - 7.5 / unweighted	RDF descriptors	3
	1	R6v	R autocorrelation of lag 6 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3

Table 3.15 Continued

	3	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	2	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	MATS4m	Moran autocorrelation - lag 4 / weighted by atomic masses	2D autocorrelations	2
	1	G3s	3st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	D/Dr05	distance/detour ring index of order 5	topological descriptors	2
	1	nArCO	number of ketones (aromatic)	functional group counts	1
	1	B10[C-C]	presence/absence of C - C at topological distance 10	2D binary fingerprints	2
	1	MATS6m	Moran autocorrelation - lag 6 / weighted by atomic masses	2D autocorrelations	2
	1	BEHm7	highest eigenvalue n. 7 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	MAXDN	maximal electrotopological negative variation	topological descriptors	2
	1	L2s	2nd component size directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	BELe3	lowest eigenvalue n. 3 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	RDF095u	Radial Distribution Function - 9.5 / unweighted	RDF descriptors	3
MOR204.6					
	1	nCconj	number of non-aromatic conjugated C(sp2)	functional group counts	1
	1	S_dsCH	S_dsCH	atomtypes (cerius2)	1
	2	HATS3m	leverage-weighted autocorrelation of lag 3 / weighted by atomic masses	GETAWAY descriptors	3
	1	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes	geometrical descriptors	3
	1	R4u+	R maximal autocorrelation of lag 4 / unweighted	GETAWAY descriptors	3
	2	D/Dr06	distance/detour ring index of order 6	topological descriptors	2
	1	R6m	R autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
	1	Mor10m	3D-MoRSE - signal 10 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	EEig09r	Eigenvalue 09 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	R7e	R autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	R3u+	R maximal autocorrelation of lag 3 / unweighted	GETAWAY descriptors	3
	1	R2v	R autocorrelation of lag 2 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	F05[C-O]	frequency of C - O at topological distance 05	2D frequency fingerprints	2
	1	HATS7e	leverage-weighted autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
MOR207.1					
	3	C-025	R--CR--R	atom-centred fragments	1
	1	F05[C-O]	frequency of C - O at topological distance 05	2D frequency fingerprints	2
	1	R6v+	R maximal autocorrelation of lag 6 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	2	E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	S_aaCH	S_aaCH	atomtypes (cerius2)	1
	2	RTe+	R maximal index / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	R8u+	R maximal autocorrelation of lag 8 / unweighted	GETAWAY descriptors	3
	1	ATS2p	Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	RDF045v	Radial Distribution Function - 4.5 / weighted by atomic van der Waals volumes	RDF descriptors	3
	1	P2p	2nd component shape directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
	1	R7p+	R maximal autocorrelation of lag 7 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	R5m	R autocorrelation of lag 5 / weighted by atomic masses	GETAWAY descriptors	3
	1	HOMA	Harmonic Oscillator Model of Aromaticity index	geometrical descriptors	3
	1	R7v+	R maximal autocorrelation of lag 7 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	H1p	H autocorrelation of lag 1 / weighted by atomic polarizabilities	GETAWAY descriptors	3
MOR273.1					
	1	P2s	2nd component shape directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	2	Jhetv	Balaban-type index from van der Waals weighted distance matrix	topological descriptors	2
	1	GATS5m	Geary autocorrelation - lag 5 / weighted by atomic masses	2D autocorrelations	2
	1	HATS1u	leverage-weighted autocorrelation of lag 1 / unweighted	GETAWAY descriptors	3
	1	Mor20u	3D-MoRSE - signal 20 / unweighted	3D-MoRSE descriptors	3
	1	JGI6	mean topological charge index of order6	topological charge indices	2
	1	R1e	R autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	Mor32u	3D-MoRSE - signal 32 / unweighted	3D-MoRSE descriptors	3
	1	EEig08r	Eigenvalue 08 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	BEHm3	highest eigenvalue n. 3 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
MOR250.1					
	1	Dv	D total accessibility index / weighted by atomic van der Waals volumes	WHIM descriptors	3
	6	H-047	H attached to C1(sp3)/C0(sp2)	atom-centred fragments	1
	1	BEHm5	highest eigenvalue n. 5 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	3	C-006	CH2RX	atom-centred fragments	1
	2	MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	2	D/Dr06	distance/detour ring index of order 6	topological descriptors	2
	2	Mor24p	3D-MoRSE - signal 24 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	3	P1s	1st component shape directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	EEig07d	Eigenvalue 07 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	R7p+	R maximal autocorrelation of lag 7 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	3	E1v	1st component accessibility directional WHIM index / weighted by atomic van der Waals volumes	WHIM descriptors	3
	2	L2s	2nd component size directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	2	Mor10e	3D-MoRSE - signal 10 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	H1v	H autocorrelation of lag 1 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	4	R4m	R autocorrelation of lag 4 / weighted by atomic masses	GETAWAY descriptors	3
	1	RDF035m	Radial Distribution Function - 3.5 / weighted by atomic masses	RDF descriptors	3
	1	BELm1	lowest eigenvalue n. 1 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	2	HATS7m	leverage-weighted autocorrelation of lag 7 / weighted by atomic masses	GETAWAY descriptors	3
	1	F04[C-O]	frequency of C - O at topological distance 04	2D frequency fingerprints	2
	1	B04[O-O]	presence/absence of O - O at topological distance 04	2D binary fingerprints	2
	1	B05[C-C]	presence/absence of C - C at topological distance 05	2D binary fingerprints	2
	2	C-003	CHR3	atom-centred fragments	1
	3	H1p	H autocorrelation of lag 1 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	2	Mor11m	3D-MoRSE - signal 11 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	EEig07x	Eigenvalue 07 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	3	GVWAI-80	Ghose-Viswanadhan-Wendoloski drug-like index at 80%	molecular properties	1
	2	JGI7	mean topological charge index of order7	topological charge indices	2
	3	piPC07	molecular multiple path count of order 07	walk and path counts	2
	1	nR09	number of 9-membered rings	constitutional descriptors	0
	1	RDF060p	Radial Distribution Function - 6.0 / weighted by atomic polarizabilities	RDF descriptors	3
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	Mor10m	3D-MoRSE - signal 10 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	GATS2m	Geary autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations	2
	2	BELe1	lowest eigenvalue n. 1 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	IC1	information content index (neighborhood symmetry of 1-order)	information indices	2
	1	MATS4m	Moran autocorrelation - lag 4 / weighted by atomic masses	2D autocorrelations	2
	1	nArOR	number of ethers (aromatic)	functional group counts	1
	1	D/Dr05	distance/detour ring index of order 5	topological descriptors	2
	1	RDF060u	Radial Distribution Function - 6.0 / unweighted	RDF descriptors	3
MOR256.17					
	7	BIC	BIC	topological (cerius2)	2
	1	EEig01x	Eigenvalue 01 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	6	HATS6m	leverage-weighted autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
	5	nOHp	number of primary alcohols	functional group counts	1
	3	S_sssCH	S_sssCH	atomtypes (cerius2)	1
	3	nArCO	number of ketones (aromatic)	functional group counts	1
	3	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	3	nR=Ct	number of aliphatic tertiary C(sp2)	functional group counts	1
	3	HTm	H total index / weighted by atomic masses	GETAWAY descriptors	3
	1	Mor30v	3D-MoRSE - signal 30 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	2	GATS6m	Geary autocorrelation - lag 6 / weighted by atomic masses	2D autocorrelations	2
	1	RDF130m	Radial Distribution Function - 13.0 / weighted by atomic masses	RDF descriptors	3
	2	RDF040e	Radial Distribution Function - 4.0 / weighted by atomic Sanderson electronegativities	RDF descriptors	3
	1	nR10	number of 10-membered rings	constitutional descriptors	0
	1	B09[C-S]	presence/absence of C - S at topological distance 09	2D binary fingerprints	2

Table 3.15 Continued

	1	C-008	CHR2X	atom-centred fragments	1
	1	X5A	average connectivity index chi-5	connectivity indices	2
	1	MATS7m	Moran autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	Mor24v	3D-MoRSE - signal 24 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	1	Mor27e	3D-MoRSE - signal 27 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	EEig09d	Eigenvalue 09 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
MOR258.1					
	3	R1v	R autocorrelation of lag 1 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	BELm2	lowest eigenvalue n. 2 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	2	H1p	H autocorrelation of lag 1 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	E1u	1st component accessibility directional WHIM index / unweighted	WHIM descriptors	3
	1	HATS2m	leverage-weighted autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	JGI2	mean topological charge index of order2	topological charge indices	2
	1	R1p	R autocorrelation of lag 1 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	B06[C-C]	presence/absence of C - C at topological distance 06	2D binary fingerprints	2
	1	L2s	2nd component size directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	DISPm	d COMMA2 value / weighted by atomic masses	geometrical descriptors	3
	1	SHP2	average shape profile index of order 2	Randic molecular profiles	3
	1	H1u	H autocorrelation of lag 1 / unweighted	GETAWAY descriptors	3
	1	P2s	2nd component shape directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
MOR259.1					
	1	nCb-	number of substituted benzene C(sp2)	functional group counts	1
	1	G3s	3st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	R4m	R autocorrelation of lag 4 / weighted by atomic masses	GETAWAY descriptors	3
	1	MPC06	molecular path count of order 06	walk and path counts	2
	1	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes	geometrical descriptors	3
	1	R1p	R autocorrelation of lag 1 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	F05[C-O]	frequency of C - O at topological distance 05	2D frequency fingerprints	2
	1	R2v+	R maximal autocorrelation of lag 2 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	E1u	1st component accessibility directional WHIM index / unweighted	WHIM descriptors	3
	1	GATS2m	Geary autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations	2
	1	E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	H1u	H autocorrelation of lag 1 / unweighted	GETAWAY descriptors	3
MOR260.1					
	1	RDF110v	Radial Distribution Function - 11.0 / weighted by atomic van der Waals volumes	RDF descriptors	3
	2	R2m+	R maximal autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	JGI8	mean topological charge index of order8	topological charge indices	2
	1	Hy	hydrophilic factor	molecular properties	1
	1	O-056	alcohol	atom-centred fragments	1
	1	H-050	H attached to heteroatom	atom-centred fragments	1
	1	MATS8m	Moran autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	RDF080e	Radial Distribution Function - 8.0 / weighted by atomic Sanderson electronegativities	RDF descriptors	3
	1	HATS6m	leverage-weighted autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
MOR261.1					
	1	RDF110e	Radial Distribution Function - 11.0 / weighted by atomic Sanderson electronegativities	RDF descriptors	3
	2	C-006	CH2RX	atom-centred fragments	1
	1	RDF075v	Radial Distribution Function - 7.5 / weighted by atomic van der Waals volumes	RDF descriptors	3
	1	RDF095u	Radial Distribution Function - 9.5 / unweighted	RDF descriptors	3
	1	DISPm	d COMMA2 value / weighted by atomic masses	geometrical descriptors	3
	1	C-001	CH3R / CH4	atom-centred fragments	1
	1	BEHm8	highest eigenvalue n. 8 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	RDF110v	Radial Distribution Function - 11.0 / weighted by atomic van der Waals volumes	RDF descriptors	3
MOR268.1					
	2	ATS8e	Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic Sanderson electronegativ	2D autocorrelations	2
	1	C-006	CH2RX	atom-centred fragments	1
	1	TIE	E-state topological parameter	topological descriptors	2
	1	RDF050m	Radial Distribution Function - 5.0 / weighted by atomic masses	RDF descriptors	3
	3	B09[C-O]	presence/absence of C - O at topological distance 09	2D binary fingerprints	2
	5	B07[C-C]	presence/absence of C - C at topological distance 07	2D binary fingerprints	2
	4	H-051	H attached to alpha-C	atom-centred fragments	1
	3	S_sssCH	S_sssCH	atomtypes (cerius2)	1
	1	MATS6m	Moran autocorrelation - lag 6 / weighted by atomic masses	2D autocorrelations	2
	2	G2p	2st component symmetry directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
	1	RDF115e	Radial Distribution Function - 11.5 / weighted by atomic Sanderson electronegativities	RDF descriptors	3
	1	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	RDF130u	Radial Distribution Function - 13.0 / unweighted	RDF descriptors	3
	1	G3u	3st component symmetry directional WHIM index / unweighted	WHIM descriptors	3
	1	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	1	nArCO	number of ketones (aromatic)	functional group counts	1
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	GATS1m	Geary autocorrelation - lag 1 / weighted by atomic masses	2D autocorrelations	2
	1	nCconj	number of non-aromatic conjugated C(sp2)	functional group counts	1
	1	MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
MOR271.1					
	1	H-050	H attached to heteroatom	atom-centred fragments	1
	1	PHI	Kier flexibility index	topological descriptors	2
	3	JGI6	mean topological charge index of order6	topological charge indices	2
	2	RDF075m	Radial Distribution Function - 7.5 / weighted by atomic masses	RDF descriptors	3
	2	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	1	SPH	spherosity	geometrical descriptors	3
	2	Hy	hydrophilic factor	molecular properties	1
	1	EEig09x	Eigenvalue 09 from edge adj. matrix weighted by edge degrees	edge adjacency indices	2
	1	RTu+	R maximal index / unweighted	GETAWAY descriptors	3
	1	GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	RDF110p	Radial Distribution Function - 11.0 / weighted by atomic polarizabilities	RDF descriptors	3
	1	GATS4m	Geary autocorrelation - lag 4 / weighted by atomic masses	2D autocorrelations	2
	1	MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	R8m+	R maximal autocorrelation of lag 8 / weighted by atomic masses	GETAWAY descriptors	3
MOR272.1					
	1	BLTF96	Verhaar model of Fish base-line toxicity from MLOGP (mmol/l)	molecular properties	1
	1	DISPe	d COMMA2 value / weighted by atomic Sanderson electronegativities	geometrical descriptors	3
	1	B08[C-O]	presence/absence of C - O at topological distance 08	2D binary fingerprints	2
	1	HATS6p	leverage-weighted autocorrelation of lag 6 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	MATS6m	Moran autocorrelation - lag 6 / weighted by atomic masses	2D autocorrelations	2
	1	RDF110v	Radial Distribution Function - 11.0 / weighted by atomic van der Waals volumes	RDF descriptors	3
	1	G3s	3st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	1	H-050	H attached to heteroatom	atom-centred fragments	1
	1	R2m+	R maximal autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	HATS5e	leverage-weighted autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	H3m	H autocorrelation of lag 3 / weighted by atomic masses	GETAWAY descriptors	3
	1	BEHm3	highest eigenvalue n. 3 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
MOR273.1					
	1	P2s	2nd component shape directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	2	Jhetv	Balaban-type index from van der Waals weighted distance matrix	topological descriptors	2
	1	GATS5m	Geary autocorrelation - lag 5 / weighted by atomic masses	2D autocorrelations	2
	1	HATS1u	leverage-weighted autocorrelation of lag 1 / unweighted	GETAWAY descriptors	3

Table 3.15 Continued

	1	Mor20u	3D-MoRSE - signal 20 / unweighted	3D-MoRSE descriptors	3
	1	JGI6	mean topological charge index of order6	topological charge indices	2
	1	R1e	R autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	Mor32u	3D-MoRSE - signal 32 / unweighted	3D-MoRSE descriptors	3
	1	EEig08r	Eigenvalue 08 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	BEHm3	highest eigenvalue n. 3 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
MOR277.1					
	1	HATS3v	leverage-weighted autocorrelation of lag 3 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	4	E1e	1st component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	MATS5m	Moran autocorrelation - lag 5 / weighted by atomic masses	2D autocorrelations	2
	1	E2p	2nd component accessibility directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
	2	RDF035m	Radial Distribution Function - 3.5 / weighted by atomic masses	RDF descriptors	3
	2	R2m+	R maximal autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	R8e+	R maximal autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	3	SHP2	average shape profile index of order 2	Randic molecular profiles	3
	1	GATS6m	Geary autocorrelation - lag 6 / weighted by atomic masses	2D autocorrelations	2
	1	BELp3	lowest eigenvalue n. 3 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	1	Jhetp	Balaban-type index from polarizability weighted distance matrix	topological descriptors	2
	1	H-051	H attached to alpha-C	atom-centred fragments	1
	1	EEig09r	Eigenvalue 09 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	1	BELe3	lowest eigenvalue n. 3 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden eigenvalues	2
	1	nRCOOR	number of esters (aliphatic)	functional group counts	1
	1	E2u	2nd component accessibility directional WHIM index / unweighted	WHIM descriptors	3
	1	JGI5	mean topological charge index of order5	topological charge indices	2
	1	GATS3m	Geary autocorrelation - lag 3 / weighted by atomic masses	2D autocorrelations	2
	1	L2u	2nd component size directional WHIM index / unweighted	WHIM descriptors	3
	1	S_dssC	S_dssC	atomtypes (cerius2)	1
	1	G3e	3st component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities	WHIM descriptors	3
	1	B06[C-C]	presence/absence of C - C at topological distance 06	2D binary fingerprints	2
	1	L2p	2nd component size directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
MOR30.1					
	1	B09[C-O]	presence/absence of C - O at topological distance 09	2D binary fingerprints	2
	1	O-056	alcohol	atom-centred fragments	1
	2	B07[C-O]	presence/absence of C - O at topological distance 07	2D binary fingerprints	2
	1	RDF085v	Radial Distribution Function - 8.5 / weighted by atomic van der Waals volumes	RDF descriptors	3
	5	H-051	H attached to alpha-C	atom-centred fragments	1
	1	RDF075m	Radial Distribution Function - 7.5 / weighted by atomic masses	RDF descriptors	3
	3	C-006	CH2RX	atom-centred fragments	1
	1	L/Bw	length-to-breadth ratio by WHIM	geometrical descriptors	3
	1	EEig07d	Eigenvalue 07 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	RDF095v	Radial Distribution Function - 9.5 / weighted by atomic van der Waals volumes	RDF descriptors	3
	1	E2p	2nd component accessibility directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
	2	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	Mor30m	3D-MoRSE - signal 30 / weighted by atomic masses	3D-MoRSE descriptors	3
	1	Mor09u	3D-MoRSE - signal 09 / unweighted	3D-MoRSE descriptors	3
	1	EEig06d	Eigenvalue 06 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	1	Infective-80	Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%	molecular properties	1
	1	R6e+	R maximal autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	GATS2m	Geary autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations	2
	1	RDF045e	Radial Distribution Function - 4.5 / weighted by atomic Sanderson electronegativities	RDF descriptors	3
MOR33.1					
	1	F07[C-O]	frequency of C - O at topological distance 07	2D frequency fingerprints	2
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes	geometrical descriptors	3
	1	F04[C-O]	frequency of C - O at topological distance 04	2D frequency fingerprints	2
	2	Mor17e	3D-MoRSE - signal 17 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	H-051	H attached to alpha-C	atom-centred fragments	1
	1	C-006	CH2RX	atom-centred fragments	1
	1	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
	1	H-050	H attached to heteroatom	atom-centred fragments	1
MOR37.1					
	1	B09[C-O]	presence/absence of C - O at topological distance 09	2D binary fingerprints	2
	1	O-056	alcohol	atom-centred fragments	1
	1	B08[C-O]	presence/absence of C - O at topological distance 08	2D binary fingerprints	2
MOR40.1					
	2	RDF110v	Radial Distribution Function - 11.0 / weighted by atomic van der Waals volumes	RDF descriptors	3
	1	H-051	H attached to alpha-C	atom-centred fragments	1
	1	Mor17e	3D-MoRSE - signal 17 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	1	C-006	CH2RX	atom-centred fragments	1
	1	GATS7m	Geary autocorrelation - lag 7 / weighted by atomic masses	2D autocorrelations	2
MOR41.1					
	1	HVcpx	graph vertex complexity index	information indices	2
	1	Kappa-3	Kappa-3	topological (cerius2)	2
	4	GATS3m	Geary autocorrelation - lag 3 / weighted by atomic masses	2D autocorrelations	2
	8	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	3	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	2	Mor10v	3D-MoRSE - signal 10 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	2	RDF085p	Radial Distribution Function - 8.5 / weighted by atomic polarizabilities	RDF descriptors	3
	3	EEig08r	Eigenvalue 08 from edge adj. matrix weighted by resonance integrals	edge adjacency indices	2
	4	BIC	BIC	topological (cerius2)	2
	5	GATS8m	Geary autocorrelation - lag 8 / weighted by atomic masses	2D autocorrelations	2
	3	H6m	H autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
	1	BEHp3	highest eigenvalue n. 3 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues	2
	1	JGI3	mean topological charge index of order3	topological charge indices	2
	4	RDF085u	Radial Distribution Function - 8.5 / unweighted	RDF descriptors	3
	1	GATS2m	Geary autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations	2
	4	Mor22e	3D-MoRSE - signal 22 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
	3	nCq	number of total quaternary C(sp3)	functional group counts	1
	1	E1p	1st component accessibility directional WHIM index / weighted by atomic polarizabilities	WHIM descriptors	3
	1	RDF050v	Radial Distribution Function - 5.0 / weighted by atomic van der Waals volumes	RDF descriptors	3
	2	PW3	path/walk 3 - Randic shape index	topological descriptors	2
	1	GVWAI-80	Ghose-Viswanadhan-Wendoloski drug-like index at 80%	molecular properties	1
	1	C-008	CHR2X	atom-centred fragments	1
	1	H6v	H autocorrelation of lag 6 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	MATS3m	Moran autocorrelation - lag 3 / weighted by atomic masses	2D autocorrelations	2
	1	Atype_C_18	Number of Carbon Type 18	atomtypes (Cerius2)	1
	1	BEHv3	highest eigenvalue n. 3 of Burden matrix / weighted by atomic van der Waals volumes	Burden eigenvalues	2
	1	RDF130u	Radial Distribution Function - 13.0 / unweighted	RDF descriptors	3
	1	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	1	B05[C-O]	presence/absence of C - O at topological distance 05	2D binary fingerprints	2
	1	MEcc	molecular eccentricity	geometrical descriptors	3
	1	X2A	average connectivity index chi-2	connectivity indices	2
MOR5.1					
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	F07[C-O]	frequency of C - O at topological distance 07	2D frequency fingerprints	2
	1	F04[C-O]	frequency of C - O at topological distance 04	2D frequency fingerprints	2
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	Mor17e	3D-MoRSE - signal 17 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors	3
OR1A1					

Table 3.15 Continued

	2	HATS8u	leverage-weighted autocorrelation of lag 8 / unweighted	GETAWAY descriptors	3
	1	E1s	1st component accessibility directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	2	R5v+	R maximal autocorrelation of lag 5 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3
	1	Mor02p	3D-MoRSE - signal 02 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	DISPe	d COMMA2 value / weighted by atomic Sanderson electronegativities	geometrical descriptors	3
	3	B06[C-C]	presence/absence of C - C at topological distance 06	2D binary fingerprints	2
	2	nRCHO	number of aldehydes (aliphatic)	functional group counts	1
	1	Mor21p	3D-MoRSE - signal 21 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	JGI3	mean topological charge index of order3	topological charge indices	2
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	RDF100m	Radial Distribution Function - 10.0 / weighted by atomic masses	RDF descriptors	3
	1	MATS5m	Moran autocorrelation - lag 5 / weighted by atomic masses	2D autocorrelations	2
	1	R6m+	R maximal autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
	1	RDF095u	Radial Distribution Function - 9.5 / unweighted	RDF descriptors	3
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	ESpm06d	Spectral moment 06 from edge adj. matrix weighted by dipole moments	edge adjacency indices	2
	1	BELm5	lowest eigenvalue n. 5 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
OR2J2					
	3	E1s	1st component accessibility directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors	3
	1	F06[C-O]	frequency of C - O at topological distance 06	2D frequency fingerprints	2
	1	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes	geometrical descriptors	3
	1	JGI5	mean topological charge index of order5	topological charge indices	2
	3	H-051	H attached to alpha-C	atom-centred fragments	1
	1	H1u	H autocorrelation of lag 1 / unweighted	GETAWAY descriptors	3
	4	DISPm	d COMMA2 value / weighted by atomic masses	geometrical descriptors	3
	1	BEHv8	highest eigenvalue n. 8 of Burden matrix / weighted by atomic van der Waals volumes	Burden eigenvalues	2
	2	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	1	B06[C-O]	presence/absence of C - O at topological distance 06	2D binary fingerprints	2
	1	O-057	phenol / enol / carboxyl OH	atom-centred fragments	1
	1	Mor10u	3D-MoRSE - signal 10 / unweighted	3D-MoRSE descriptors	3
	1	H8m	H autocorrelation of lag 8 / weighted by atomic masses	GETAWAY descriptors	3
	2	F01[C-C]	frequency of C - C at topological distance 01	2D frequency fingerprints	2
	2	HATS5p	leverage-weighted autocorrelation of lag 5 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
	1	H6m	H autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
	1	S_dssC	S_dssC	atomtypes (cerius2)	1
	1	HATS0e	leverage-weighted autocorrelation of lag 0 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	R7p+	R maximal autocorrelation of lag 7 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	VEA2	average eigenvector coefficient sum from adjacency matrix	eigenvalue-based indices	2
OR2W1					
	2	B05[C-C]	presence/absence of C - C at topological distance 05	2D binary fingerprints	2
	2	HATS6p	leverage-weighted autocorrelation of lag 6 / weighted by atomic polarizabilities	GETAWAY descriptors	3
	1	RDF110m	Radial Distribution Function - 11.0 / weighted by atomic masses	RDF descriptors	3
	1	R2m+	R maximal autocorrelation of lag 2 / weighted by atomic masses	GETAWAY descriptors	3
	1	MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D autocorrelations	2
OR5P3					
	2	nCconj	number of non-aromatic conjugated C(sp2)	functional group counts	1
	1	HATS3m	leverage-weighted autocorrelation of lag 3 / weighted by atomic masses	GETAWAY descriptors	3
	3	R7e	R autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	Yindex	Balaban Y index	information indices	2
	1	R7e+	R maximal autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	2	BLTD48	Verhaar model of Daphnia base-line toxicity from MLOGP (mmol/l)	molecular properties	1
	1	R6m	R autocorrelation of lag 6 / weighted by atomic masses	GETAWAY descriptors	3
	1	H-047	H attached to C1(sp3)/C0(sp2)	atom-centred fragments	1
	1	R4m	R autocorrelation of lag 4 / weighted by atomic masses	GETAWAY descriptors	3
	3	GATS2m	Geary autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations	2
	1	DISPm	d COMMA2 value / weighted by atomic masses	geometrical descriptors	3
	2	Mor10u	3D-MoRSE - signal 10 / unweighted	3D-MoRSE descriptors	3
	1	RDF050p	Radial Distribution Function - 5.0 / weighted by atomic polarizabilities	RDF descriptors	3
	1	B04[O-O]	presence/absence of O - O at topological distance 04	2D binary fingerprints	2
	1	BELm1	lowest eigenvalue n. 1 of Burden matrix / weighted by atomic masses	Burden eigenvalues	2
	1	Mor11v	3D-MoRSE - signal 11 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors	3
	2	HATS7e	leverage-weighted autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors	3
	1	B05[C-C]	presence/absence of C - C at topological distance 05	2D binary fingerprints	2
	1	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	atom-centred fragments	1
	2	B07[C-C]	presence/absence of C - C at topological distance 07	2D binary fingerprints	2
	1	nRCOOH	number of carboxylic acids (aliphatic)	functional group counts	1
	1	Mor18p	3D-MoRSE - signal 18 / weighted by atomic polarizabilities	3D-MoRSE descriptors	3
	1	HATS2v	leverage-weighted autocorrelation of lag 2 / weighted by atomic van der Waals volumes	GETAWAY descriptors	3

Table 3.15 Continued

Table 3.16: Top 100 predicted compounds for each mammalian OR

Chemical name or Pubchem compound ID (CIDs), SMILES strings, and distances, of the top ~100 predicted compounds for each Or. All distances represent the minimum distance based on optimized descriptors to the previously known strongest active compound listed in the gray cells for that particular Or.

Mor1-1		Mor106-1		Mor107-1		Mor129-1		Mor136-1		Mor139-1		Mor162-1	
SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance
C1C2=CC=CC=C2OC1=O	0	C(=O)(C)c1ccccc1	0	C[C@@]12CC[C@@H](C1(C)C)CC2=O	0	CC1=CC[C@@H](CC1=O)C(=C)C	0	C[C@@]12CC[C@@H](C1(C)C)CC2=O	0	CC1=CC[C@@H](CC1=O)C(=C)C	0	O=C1CCOC2=C1C=CC=C2	0
C(=O)(CCCCC)O	0	C(=C)Cc1ccccc1	0	C[C@@]12CC[C@H](C1(C)C)CC2=O	0	CC1=CC[C@H](CC1=O)C(=C)C	0	C[C@]12CC[C@H](C1)C(C2=O)(C)C	0	CC1=CC[C@H](CC1=O)C(=C)C	0	C(=O)(C)c1ccccc1	0
C(=O)(CCCC)O	0	O=C1OC2=C(C=CC=C2)C=C1	0	CC1(C2(CCC1(CC2=O)N)C)C	0.2473219	C(=O)(C)c1ccccc1	0	C[C@@]12CC[C@H](C1(C)C)CC2=O	0	O=C1CCOC2=C1C=CC=C2	0	O=C1OC2=C(C=CC=C2)C=C1	0
C(=O)(CCCC)O	0	c1(ccccc1)OC(=O)C	0	CC1C2(CCC1(C(=O)C2)C)C	0.2764237	O=C1OC2=C(C=CC=C2)C=C1	0	C[C@@]12CC[C@@H](C1)C(C2=O)(C)C	0	CC1=C(CCCC)C(CC1)=O	0	CC1NC(=O)C2=CC=CC=C2O1	0.03923089
C(=O)(CCCCC)O	0	CC=COC1=CC=CC=C1	0.08891955	CC1(C2(CCC1(CC2=O)O)C)C	0.3186515	C1C(=O)CNC2=CC=CC=C21	0.1220788	CCC1(CC(OC1=N)(C)C)CC	0.05816986	C1CCC2=C(C1)CCCC2=O	0.04565114	C1=CC=C2C(=C1)C=C(NO2)C=O	0.05289857
CC1=CC2=C(C=C1)OC(=O)C2	0.04917087	CC(S)SC1=CC=CC=C1	0.1534956	CC1(C2CC(C1(C(=O)C2)C)O)C	0.3204451	CC1=CCCC(C1)(C)C(=O)C	0.131721	C1=CC(=C2C(=C1)S2)C(=S)N	0.06587855	CCC(C)C1CCC(=O)CC1	0.04807124	CC1=CN=C(C(=N1)C)C=O	0.06707111
CC1=CC2=C(CC(=O)O2)C=C1	0.06445035	C=C(C1=CC=CC=C1)C	0.1583203	CC1(C2CC(=O)C1(CC2O)C)C	0.3482935	C1C2C(=CC=CO2)C=CC1=O	0.1324226	CC(C)(C)C1CC(=O)C2C1C2	0.06816311	C1CC2CC=CCC2C(=O)C1	0.04894259	C1C=C(C2=C(O1)N=CC=C2)O	0.06713544
CC1=CC=CC2=C1OC(=O)C2	0.06478577	CC(=C)C1=CC=CC=C1	0.1583203	CC1(C2CCC1(C(=O)C2)CO)C	0.366073	CC1CC(CC=C1C#N)(C)C	0.1401867	CCC1(CCCC1=O)CC	0.0729801	CC(C)C1=CC(=O)CCC1	0.04953565	C1=CC=C2C(=C1)C=C(C(=O)O2)N	0.06748666
CC(CCCC(=O)O)N	0.0766186	CC1=C2C(=CC=C1)N2C	0.1611889	CC1(C2CCC1(C(=O)C2)CS)C	0.4426886	CC1=CCC(CC1)C(C)C=O	0.1440427	CCCCCC(CC)C(=O)C1CC1	0.07530886	CCCC(C)C1=CCCC1=O	0.05030901	C1CC2=CC3=C(C(=O)N2C1)NC=C3	0.06865916
CC1=CC(=C2C(=C1)CC(=O)O2)C	0.09134395	CC1=CC=CC=C1C=C	0.1622158	CCC12CCC(C1(C)C)CC2=O	0.4550365	C1CCC2C(C1)CCC(=O)O2	0.1447183	CC(C)C12CCC(C1)(CC2=O)C	0.07590504	CC(C)CCC1=CCCC1=O	0.05046165	CC1=CC(=C(C=C1)C=O)C	0.07170324
CC1=C(C2=C(CC(=O)O2)C=C1)C	0.09749545	CC1C(N1)C2=CC=CC=C2	0.1766318	CC12CCC(C1(C)CO)CC2=O	0.4566952	CC(C)(CO)C1=CC=CC=C1	0.1488985	CC(=O)C1CCC2C1CCCC2	0.08637492	CC1CCCC2=C1CCC2=O	0.05373275	C1=CC(=CC(=C1)NC(=O)O)C=O	0.07340692
CC1=CC(=C2CC(=O)OC2=C1)C	0.1021952	CSC(=O)C1=CC=CC=C1	0.1913783	CC1(C2CCC1(CC2=O)C)C	0.4653999	C1=CC=C2C(=C1)C=COC2=O	0.1541709	CC1CCC2(CC1)C=CC(C2=O)C	0.08638542	CC(C)CC1=CC(=O)CC1	0.05429959	C1=CC(=C(C=C1C(=O)CO)O)O	0.07503177
CC(C)(C)C(CCC(=O)O)O	0.1026351	CSC(C1=CC=CC=C1)S	0.1981317	CC1(C2CCC1(C(C2=O)O)C)C	0.4703974	COC12CCC(CC1)NC(=O)C2	0.1556671	CC1(C2CC(=O)C1(CC2=O)C)C	0.08683412	CN(C)CCC1=CC=NC=C1	0.0645497	CC1=CC(=C(C(=C1)O)C=O)O	0.07530568
C1C2=C(C(=CC=C2)N)OC1=O	0.1122016	C=COC1=CC=CC(=C1)O	0.1996824	CC1(C2CC(C1(CC2=O)C)O)C	0.5505192	C1=CC=C2C(=C1)C=CC(=O)N2	0.1562217	CCN1C2CCC1CC(=O)C2	0.0869081	CC1=C2CSCC2CC1=O	0.0662884	C1COC(=N1)C2=CN=CC=C2	0.07562215
C1C2=CC=CC=C2NC1=O	0.1200522	C=CC1=CC(=C(C=C1)S)S	0.2060344	CC(C)C12CCC(C1)(CC2=O)C	0.5732307	CC(=O)C1=CC=CC=C1N	0.1588058	CC1C=CC2(C1=O)CCCCC2	0.08745782	C1CCC2(CCC2)C(=O)C1	0.06641502	CC(=O)C1=CC2=CC=CC=C2C1	0.07665453
C1C2=C(C=CC=C2OC1=O)N	0.1221153	C=CC1=CC(=CC=C1)N=C=S	0.2098719	CC1(C2CCC1(C(=NC)C2)C)C	0.5802225	CC1CC(=CC(C1CO)C)C	0.1637881	C1CCC(=O)NCCCC(=O)C1	0.0884375	C1C(=O)COC2=CC=CC=C21	0.06771594	CC1=C(NC=C1C(=O)C)C	0.07707266
C=CCCCC(=O)O	0.1245319	C#COC1=CC=CC=C1	0.2130395	CC1(C(CC2C1(C2)C)CC=O)C	0.6171489	CC1CCC(=CC1=O)C(C)C	0.1638838	CCC1CCCC(=O)CCC1CC	0.09197357	CC1CCCC2=C1C(=O)CC2	0.06775099	C1=C(NC(=C1)C2=CN=CO2)C=O	0.0776176
CC(C)(C)CCCC(=O)N	0.1311838	C1CC1CC2=CC=CC=C2	0.2152684	CC1C2CCC(C1=O)(C2(C)C)C	0.629438	C1=CC=C2C(=C1)C(=O)C=CN2	0.1644921	CC(C)OC1=NC=CC=CN1	0.09294477	CCCCC1CCC(=O)C1=C	0.07151075	C1=CC=C2C(=C1)C(=O)C(=CO2)N	0.07804614
C(CC(N)N)CC(=O)O	0.1339592	COC=CC1=CC=CC=C1	0.2170264	CC1(C2CCC1(C(=O)C2)C=C)C	0.6401904	CC1CC(CC=C1C=O)(C)C	0.1653725	C1CC2CC(C1)CC(=O)C2	0.09388228	CC1(C2C1C(=O)C(=C)CC2)C	0.07206624	C1=CC2=C(NC(=O)C=C2)N=C1	0.07999517
CC(C1=C2CC(=O)NC2=CC=C1)N	0.1356993	CC1=C(C2=C(O2)C=C1)C	0.2178679	CC1(C2(CCC1(CC2=O)OC)C)C	0.6463336	C1=CNC2=CC(=O)C=CC2=C1	0.168101	C1CC2COCC(C1)C2=O	0.1089249	CC1=CCCC(=C(C)C)C1=O	0.07226345	C1=CC2=C(C=CC(=C2O)C=O)C=C1O	0.08200772
CC1=C2CC(=O)OC2=CC=C1	0.136907	c1(ccccc1)CC#N	0.2180725	CCCC1(CCC(C(=O)C1)(C)C)C	0.6494132	C1=CC=C2C(=C1)C(=O)C=CO2	0.1688119	CC1(C(=O)CC23C1(CCC2)CCC3)C	0.1093057	CC1CC(=O)C2=C1CCCC2	0.07233297	C1=CC=C2C(=C1)C=C(C=N2)O	0.08304375
CC1=CC=CC2=C1NC(=O)C2	0.1403033	CC1=CC=CC=C1OC#C	0.2181859	CC1(CCCC12CC=NC2)C	0.6903515	C1C=CC2=C(C1O)N=CC=C2	0.1727835	C1CCC(=O)C2CCCC(C1)C2	0.1097336	C1CCC(=C2CCCC2=O)C1	0.0727393	CC1=C(NC2=C1C=C(C=C2)C=C)C	0.08360804
C(=O)CCCC	0.1404621	C=COC1=CC=CC=C1	0.2188387	CC1(C2CCC1(C(=O)C2)C=O)C	0.7002376	CC1=CCC(CC1O)C(=C)C	0.1737252	CCCCCC1(CCCC1=O)CC=C	0.1102119	C1CC2CCC(=O)C=C2C1	0.07644502	C1=CC(=CC=C1C=O)C2=NN=CO2	0.08368816
C1C2=C(C=C(C=C2)O)OC1=O	0.1431714	CSC1=CC=CC2=C1C=C2	0.219095	CC1(C(C1(C)C)C(=O)NC2CC2)C	0.7104875	C1C2=CC=CC=C2C(=O)CN1	0.1787667	CN(C)C(=NS(=O)O)N(C)C	0.1104801	CC(=CC1=CC(=O)CCC1)C	0.07737008	CC(=O)C1=CC=C(C=C1)N	0.08371712
CCC(C)(C)CC(=O)N	0.1436566	C1=CC=C(C=C1)C(=O)NO	0.2207775	CC12CC3C1(C(=O)CC2C3)C	0.7142688	CC1CCC(CC1C)(C)C=O	0.1795582	CC(C)(C=C)C1CCCC1=O	0.1145671	CC1CC2C(C2(C)C)CC1=O	0.07785543	CC(=O)C1=CN=CC=C1	0.08373728
CC(CCCC(=O)O)O	0.1452177	C1=CC=C(C=C1)N(C(=S)S)O	0.2218528	CCOC(=O)C1C2(C13CC3)CC2	0.7263507	C1CC(CC=C1)CC#N	0.1819388	C1=CC=C(OC=C1)NCO	0.116571	CCCCCC1=CCCC1=O	0.07816834	C1=CC=C2C=C(C=CC2=C1)C=NO	0.0837472
C1C2=C(C=C(C=C2)N)NC1=O	0.146264	C1=CC=C2C(=C1)C(=O)SC2=O	0.2247712	CC12CCCC(=O)C1(COC2)C	0.7281732	CC(C)(C)C1=CC=C(CC1)O	0.1839154	C1C2CC3CC1C(C3=O)C=C2	0.1167317	CC1=CC(=O)C(CC1)C(=C)C	0.07862932	COC(=O)C1=CC=C(C=C1)O	0.08383989
CC(C)(C)C1=CCC(=O)O1	0.1538394	C1=CC=C(C=C1)S(=O)(=O)N=C=S	0.226246	CCC12CCC1C(CC2=O)(C)C	0.7489101	c1(ccccc1)C(C)O	0.1866107	COC1CCC(=O)C12CCCC2	0.1170204	CCC(C)CC1=CCCC1=O	0.0793086	C1C=CC2=CC=CC=C2SS1	0.08392745
CCCCCC(=O)N	0.1546248	CCC1=C2C(=CC=C1)N2	0.2282767	CC1(C2CC(=O)C1(C=C2)C)C	0.7799842	C1C(=O)C=C2C=CC=CC2=N1	0.1886765	CC1C2CCCN1CC2=O	0.1171827	C1CCC2(CC1)CCC2=O	0.08011041	CN(C)C1=NC=C(C=C1)C(=O)O	0.08405942
CC(C)(C)CC(CC(=O)O)O	0.156154	C1=CC=C(C=C1)C2OS2(=O)=O	0.2323088	CC1(C2CC(C1=O)(CC2=O)C)C	0.7833572	C1CC=CC(C1)CC=O	0.1891273	CCN1CC2CCCCC2C1=O	0.1181509	CCOCOC1=CC=CC=C1	0.08063265	CC1=CC2=C(C=C1)C(=O)C(CO2)N	0.0859875
C1=CNC=C1CCC(=O)N	0.1571392	C1=CC(=CC=C1OO)OO	0.2342308	CC1(CC(=O)C12CCCC2)C	0.7897701	C1C2=CC=CC=C2C(=O)C=N1	0.1903747	C1=CC=C(OC=C1)N	0.1191393	CCCC1CC(CC(=O)C1)C	0.08151939	CC1=CC(=C(O1)C)C(=O)C	0.08635016
CC1=C2CC(=O)NOC2=CC=C1	0.1582407	C1=CC=C(C=C1)NNC#N	0.2359517	CC12CCC(C1(C)C=O)CC2=O	0.8161139	C1CCC(CC1)CC#N	0.1905308	CC12CCCCCC1CCC(=O)C2	0.1197761	CC1=C(CCCC1=O)C2CC2	0.08158758	C1=CC=C2C(=C1)C=C(NO2)N	0.08729785
CCCCC(=O)NO	0.1589801	C1=CC=C(C=C1)S(=O)NN	0.2381539	CC1=CC2CCC1(C2(C)C)C	0.8186718	C1=COC2=CC(=O)C=CC2=C1	0.190531	CCC(CC)C1CC(=C)C(=O)O1	0.1218735	CC1C2C(CCC2=O)CC1=C	0.08208043	CC1=NC(=O)C2=C(N1)N=CC=C2	0.08837931
CC1=C(OC2N1C(=O)C2)C(C)(C)C	0.1594522	C1=CC(=CC=C1NN)NN	0.2398691	CC1C2(C1(C(=CC(=O)C2)C)C)C	0.822287	C=C(C1=CC=CC=C1)O	0.1906969	COC1CCCC=C1CCC(=O)O	0.1246557	CC(C)CC1CC=CC1=O	0.08326684	C1CC(C1)C2=CC=NN2	0.0884257
CC(C)(C)C(CC(=O)O)O	0.1599369	C1=CC=C(C=C1)C(=O)NN	0.2401607	CC1(CCC(CC1=O)(C)CC=C)C	0.8313264	CC(=O)C1(CCCCC1)C=C	0.1945836	CC1CC(=O)C2CCCC1C2=O	0.1248081	C=C(C)C(CCC1C)=CC1=O	0.0836955	CC1=CC(=C2C(=C1)CCC2=O)O	0.08848855
CC1=CC(=C2CC(=O)NC2=C1)C	0.160678	C1=CC=C(C=C1)C(=NN)N	0.2411232	CCC(=O)OC1C(C(=O)C1(C)C)(C)C	0.8347707	CC1=CC=C(C=C1)C(C)C=O	0.1960448	C1CC2CC(=O)CC(C1)S2	0.1274795	CCCCC1=C(C(=O)CC1)C	0.0857632	CC(=O)C1=C(C2=CC=CC=C2C1)O	0.08866954
C1=C(C=NN1)CCC(=O)N	0.1613953	COC(C1=CC=CC=C1)N	0.2429121	CC1(C2CCC1(C(=C)C2)C)C	0.8380462	CC(=O)C1=CCCC=C1	0.2001595	CCN1C(N(C(=C1C(=O)O)C)C)C	0.1277573	CCC(CC=C(C)C)C(=O)C	0.08589869	C1=CC2=C(C=CC(=C2)O)C=C1C(=O)O	0.0893506
C1C2=C(C=CC=C2N)NC1=O	0.1646353	CC1=CC=CC=C1SOC	0.2430258	CC1(C2CCC1(C(C2)C(=O)O)C)C	0.8430182	CC(=O)C1=CC=C(C=C1)N	0.2017445	CCCCC1(CCCC1=O)CCC	0.1280622	CC1CC(=O)CCC1CC=C	0.08719162	CC1=C(C=C2CN=C(C2=C1)N)N	0.09001244
C1C2=C(C=CC(=C2)O)OC1=O	0.1649221	CS(=O)(=O)NC1=CC=CC=C1	0.2450717	CC1(CCCC1(C)CC=O)C	0.8437473	CC1=CC=C(C=C1)C(=C)CO	0.2049152	C1CC=CC2(C1)CC3CCC2C3=O	0.1305261	CC1CCCC(C(=O)C1)C(C)C	0.08738357	C1=CC2=C(C=CC(=C2)C(=O)O)C(=C1)O	0.0901394
CCCC1=NNC(=O)O1	0.1654599	CCC1C2=CC=CC=C12	0.2464661	CC1C2C(C2(C)C)C(=O)C1(C)C	0.846785	CC1CC=C(C(C1)(C)C)C=O	0.2082587	C1CC2COC(=C2C(=O)C1)O	0.1340426	CC1=CC(=O)N2N1C=CC=C2	0.08743294	CC1=CNC2=C1C=C(C=C2)C=C	0.09042043
C1C2=C(C(=CC=C2)O)NC1=O	0.1658537	C=CC1=CC=CC=C1O	0.2474448	CCC(=O)OC12C(C1(C)C)C2(C)C	0.854632	C1CCC(CC1)(C2=COC=C2)O	0.2082939	CCCN=C(C(=O)C)N(C)C	0.1344627	CC1CC(=C(C1=O)C(C)C)C	0.08865448	CC1COC2=CC=CC=C2C1=O	0.0908897
C1C2=C(C=CC=C2O)OC1=O	0.1659459	C1=CC=C(C=C1)NNC(=S)S	0.2475107	CC1(C2CCC1(C(=O)C2)C(CO)O)C	0.8606029	C1=CC=C2C(=C1)C(=CC(=O)O2)N	0.2099995	CC(C)(C1=CC2COCC2C1=O)O	0.1349313	CC1C2CCC(=O)C2CC1=C	0.08898883	CNC(=O)C1=CC=CC=C1	0.09103495
CCC1=C(C=CN=C1)CC(=O)O	0.1674463	CSC1=CC(=CC=C1)SC	0.2488941	CC1(CCCC(C1)(C)CC(=O)O)C	0.8670466	C1CC(=O)OC2=CC=CC=C21	0.2107984	CCC(C)C1CCC(=O)CC1	0.1359457	CC1CC2C(C1=C)CCC2=O	0.08951675	C1C(=CC2=CC=CC=C2O1)C=O	0.09167367
CCC(=CCC(=O)O)CC	0.1676498	C#CCC1=CC=CC=C1	0.2489198	CC(CO)NC(=O)C1C(C1(C)C)(C)C	0.8696038	CC1=CC=C(C=C1)C(C)CO	0.2129872	C=CCC1(CCCC1=O)CC=C	0.1369123	CCCN=CN1CCCCC1	0.0896981	C1=CC=C2C=C(C=CC2=C1)C(=O)C=O	0.09234656
C=CCCCC(=O)N	0.1687806	CC(C1=CC=CC=C1)ON	0.2490619	CC1(C(=C)C2CCC1(C2N)C)C	0.8697859	CC(C)C1=C(C=C(C=C1)OC)O	0.2140027	CC1C(=O)N(C(=NC(C)C)O1)C(C)C	0.1375833	COC1C2CC(C1OC)C=C2	0.09006127	C1=CC(=CC=C1C(=O)O)NC(=O)CO	0.09244114
C(CCC(O)(O)O)CC=O	0.1693899	C1=CC=C2C(=C1)C(=O)NC(=O)N2	0.2492309	CC1(CC2C(C1=O)C2(C)C)C	0.8710852	CC1CC(=C(C=C1)C(=O)C)C	0.2145463	CC1(CCCC1=O)C2CCCC2	0.1380242	CCC(C)(C)C1CCC(=O)C1	0.0911234	C1=CC=C2C=C(C=CC2=C1)C(=O)O	0.09244616
CC(C)(C)CCCC=O	0.1697975	C=CCC1=CC(=CC=C1)S	0.2494043	CC1(C2CC(C1(C=C2)C)O)C	0.8742333	C1CCC2C(C1)CCC(O2)O	0.2148902	CCC(C(=C)C(=O)N(C)C)N	0.140408	CC1=C(CC(CC1)C(=O)C)C	0.09152521	C1CNCC2=C3C=CC=CN3C=C21	0.0925999
CCC1=NNC(=O)N1N	0.1703478	C=C=CCC1=CC=CC=C1	0.2503525	CC1(C2CCC1(C(=NO)C2)C)C	0.8757104	CCC1(CCCC=C1)CCO	0.2154841	CC1(C=CC2C13C2(C(=O)CC3)C)C	0.1409102	COC1C2CCC(=C1OC)C2	0.09224474	C1CC(=O)C=C1C2=CC=CC=C2	0.09316978
CCC(CCC(=O)O)O	0.1706556	COOC1=CC=CC=C1	0.2505793	CC1(C(C1(C)C)C(=O)NCCO)C	0.8793016	C1CCC(CC1)CC=O	0.2156243	CC(=C)C(=O)N(CN)C(C)(C)C	0.141526	CCCCC1=CCC(C1=O)C	0.09285003	C1=CC=C2C(=C1)C=C(C(=O)O2)O	0.09340879
CC1=C(C=C2CC(=O)OC2=C1)O	0.1710742	C1=CC=C(C=C1)OS(=O)(=O)N	0.2518608	CC(CNC(=O)C1C(C1(C)C)(C)C)O	0.8800709	CC1(CCC(=O)C=C1C=C)C	0.2162661	C=C(C1=CC=CO1)C(CO)O	0.1419081	CC1CCC2=C(C1=O)CCC2	0.09343191	CC1=NC=C(C=C1)C2=NC=CO2	0.09409661
C(CC=O)CC(N)N	0.171738	C1=CC=C(C=C1)C2NS2(=O)=O	0.252967	CC1(CC(CC1(C)CC=O)O)C	0.8885944	C1C=CNC2C1C(=CC=C2)O	0.2167523	CC(=O)CC1(CCCC1=O)C	0.1448626	C1CCC(=C2CCC(=O)C2)C1	0.09347315	C1C=CC(=O)N1C2=CC=NC=C2	0.09444769
CN1C(=O)CC2=C1C(=CC=C2)O	0.1744477	C1=CC=C(C=C1)NNN=O	0.2539633	CCCC12CCC(C1(C)C)CC2=O	0.8896106	CCC1=C(C(=CC=C1)O)O	0.2177856	C1CC2CCC(=O)C(C1)C2	0.1450279	CC1=CCC=C(C1=O)C(C)C	0.09380767	COC(=O)C1=CC=C(C=C1)N=O	0.09478938
COC1=CC=CC2=C1OC(=O)C2	0.1752014	C1=CC=C(C=C1)C(=N)SO	0.256078	CC1(C2CCC1(C2C=O)C)C	0.8899731	C1=CNC2=C(C(=O)C=CC2=C1)O	0.2213244	C=C1C2CCCC=CC2OC1=O	0.1455592	CC1=C2C(=O)COC2=CC=C1	0.09398124	C1C(=CC(=O)O1)C2=CC=CC=C2	0.0949507
CC1=C(C=CC2=C1CC(=O)N2)N	0.1758284	CC(=S)NC1=CC=CC=C1	0.2560913	CC1(C2CCC1(C2C(=O)O)C)C	0.8950171	CC(C1=CC=CC=C1)(O)O	0.2219223	CC1C2CCC(C1=O)CC2	0.1474644	CC1=C(C(=CCC1)C(=O)C)C	0.09431535	CC(=O)N1CC2=CC=CC=C21	0.09504737
CC1=CC(=CC(=C1O)C)CC=O	0.1761056	C1CC2=C1C=C3CC3=C2	0.2562868	CC1(C(=C)C2CCC1(C2)CN)C	0.8953548	CCC(C#N)C1CCCCC1	0.2228901	CC1CC(=O)CCC12CCCC2	0.1484672	CC1=C(CCC1=O)C(C)C	0.09441255	C1=CC2=C(NC(=O)C=CN2)N=C1	0.09509792
CSC1=CC=CC2=C1CC(=O)N2	0.1763796	C1=CC=C(C=C1)OC(=O)N	0.2566164	CCC(=O)C1=CCC(C1(C)C)(C)C	0.8968087	C1CCC2C(C1)C(CCN2)O	0.2237265	CC1(C2CCC1(C(=O)C2=NO)C)C	0.1486547	CC1(CC2CC(=O)CC2C1)C	0.09480944	C1C2=CC=CC=C2C(=O)C1=CN	0.09522506
CCCCC(=O)N	0.176911	CS(=O)(=O)OC1=CC=CC=C1	0.2569961	CCC(=O)C1(CCCC1(C)C)CC	0.8995239	C1CC2=C(CC1=O)C=CN=C2	0.2249067	C1CN(C1)C(=O)C2CCNCC2	0.1509058	CC(C)CN=CN1CCCC1	0.09534618	CC1=CN=C(C(=N1)C)C(=O)C	0.09555824
CC1=C(C(=C2C(=C1C)CC(=O)O2)C)C	0.1771246	C=CC1=CC(=CC=C1)O	0.2592065	CCOC(=O)C1CC1(C)C(C)(C)C	0.8996372	CC1=CC(=O)C(CC1)C(=C)C	0.2251019	CC1(C2CCC1C(=O)CC2)C	0.1512115	CCCCN=CN1CCCC1	0.095609	C1=CC2=C(C=C(C=C2)C=O)C(=C1)O	0.09556262
C1C2=C(C=CC3=C2N=CN3)NC1=O	0.1793646	C1=CC=C(C(=C1)N(O)O)O	0.2597032	CC1(CC=CC(C1)(C)CC=O)C	0.9015818	C1=CC=C(C=C1)C2C=N2	0.2266077	C1CNC2CC1CC(=O)C2	0.1528717	CC1C2CC(C1=O)CC2(C)C	0.09566855	C1C=COC2=C1C=CC(=C2)O	0.09568683
CC(=C)CCCC(=O)O	0.1794674	C=C1C=CC2=CC=CC=C12	0.259743	CC1(C2CCC1(C(=O)C2O)C)C	0.9055545	CCC1=CC(=O)CCC1(C)C	0.2272797	C1CC2C3CC(C=C3)C(C1)C2=O	0.1531701	C1CC=C2C(C1)CCCC2=O	0.09583053	C1CN=NC2=CC=CC=C21	0.0957617
C1C2=C(C=CC=C2O)NC1=O	0.1804486	C1=CC=C(C(=C1)NC#N)O	0.261642	CC1(C2CCC1(C(=O)C2N)C)C	0.907136	C1(=O)(=O)C(=CCC(C1)C(=C)C)C	0.2278994	CCOC(=O)C1CSC(=S)N1	0.1547841	CCC1=C2N(CCO2)N=C1C	0.09605521	C1CC(=C=O)S(=O)C2=CC=CC=C21	0.09654614
CC1=C(C=C2CC(=O)NC2=C1)O	0.1835288	CC1=CC=CC(=C1)C=C	0.2640661	CC(C)C1(CC(C1)(CC(=O)O)CN)C	0.9138853	CN1CC(=O)CC2=CC=CC=C21	0.2283002	CC1(C2(CCC1(C(=O)C2)C)C)C	0.1548371	CC1=C(C(=O)CC=C1)C(C)C	0.09616375	CC1=CC(=C(C=C1)C=O)O	0.09664747
CCCC(CC(=O)O)O	0.1837773	C1=CC=C(C=C1)OS(=O)(=O)O	0.2641598	CCC12CCC(C1(C)C)CC2=C	0.9185449	Oc1ccc(cc1)C(=O)C	0.2284998	C1CCC2(CC1)C(=O)CCCN2	0.1551447	CC1CC(C2CC2C1=O)(C)C	0.09639987	CC1=CC(=C(C=C1C=O)C)C	0.0970306
C1=CNC=C1CCC(=O)O	0.1843949	C1=CC(=CC=C1NC#N)O	0.2644259	CC1(CCCC(C1)(CC(=O)O)CN)C	0.9207203	C1CC2=C(C(=O)C=CC2)OC1	0.2286939	C1CC(=O)CC2(C1)CCCO2	0.1554643	CC1=CC(CCC1=O)C(C)C	0.0966804	CC1=NC=C(C(=C1)N)C=O	0.09713527
C1=CC(=CC=C1CC=O)O	0.1859167	COC(c1ccccc1)O	0.2644933	CC1(CC2CCC1(C2)C=O)C	0.9210848	C1=CC2=CNC=CC2=CC1=O	0.2299888	CC1CCCC(=O)C12CCCC2	0.1558911	CC(=O)C1C=C1C(C)(C)C	0.09712166	CC1=C(C2=C(C=C1)C(=CC=C2)C=O)O	0.09772749
C1C2=CNC3=C2C(=CC=C3)NC1=O	0.1864415	C1=CC=C(C=C1)NNNN	0.2649921	CC(=N)OC1CC2CCC1(C2(C)C)C	0.926675	CC1=C(C(=O)C(CC1)C(C)C)O	0.2306438	CCC(CC)(C(=C)OCC)O	0.1567107	CC1CC(C(=O)C(C1C)C)C	0.09741224	C=CC1=NC(=CC=C1)N	0.09823491
CC(=CCCC=O)C	0.1876119	C1=CC=C(C=C1)OOC(=O)S	0.2652418	CCOCC12CCC(C1(C)C)CC2=O	0.9352263	CCC1=C(C(=CC=C1)O)N	0.2321495	CC1(CC2C1C(=O)CCC2)C#C	0.1575193	CC(=O)C1CC2CC1C=C2	0.09770139	COC(=O)C1=CC=C(C=C1)NN	0.09836613
CCC1=C2CC(=O)NC2=CC=C1	0.190325	CC(C1=CC=CC=C1)NO	0.2653897	CC1(C2CCC1(C(=O)C2)CN(C)C)C	0.9484006	C1=CC(=O)C2=CNC=CC2=C1	0.2324381	CCCCCOC1CCCC(=O)C1	0.1576065	CCOC1=CC=C(C=C1)C=O	0.0978754	CC1=CC=C(C=C1)C2=NC=NO2	0.09855551
CC1=CC2=C(CC(=O)N2)C(=C1C)C	0.1918109	C1=CC(=CC=C1N)NC#N	0.265507	CC1(C2CCC1(C(=O)C2)C(=O)O)C	0.9566291	C1CC(CC=C1)CC=O	0.2333096	CC1(OC2C(O1)C(=C)OC2O)C	0.1576705	CC1=CC(=O)C(=C(C)C)CC1	0.09827508	C1=CC=C2C=C(C(=CC2=C1)C=O)O	0.09861097
C1C2=C(C=CC(=C2)S)NC1=O	0.1924935	C1=CC=C(C=C1)NN=O	0.2655501	CC(C)NC(=O)C1C(C1(C)C)(C)C	0.9615399	CC1C=C(CCC1(C)C)C=O	0.2336705	CC(C1CCCCN1)C(=O)C2CC2	0.1584858	CC1(CC(=O)C2=C1CCC2)C	0.09828917	CC(=S)C1=CC=CC=C1	0.09867964
CC(=C(C)CCC(=O)O)C	0.1928007	C1=CC=C(C=C1)ONN=O	0.2664382	CC1(C2CC(C1=O)C(C2=O)(C)C)C	0.9626683	CCC1C=CCCC1C(C)O	0.2339059	CCCC1(CCCC1=O)CCC	0.1596546	CC#CCCC1=CCCC1=O	0.09857885	C1C(=CC2=CC=CC=C2O1)C(=O)O	0.09888103
CN1C(=O)CC2=CC=CC=C21	0.1943768	CN1C2=CC=CC=C2C1=O	0.2666446	CC1(C(CCC1(C)C#N)C(=O)OC)C	0.9676483	C1=CC=C2C(=C1)C(=O)C(=CO2)N	0.2359213	C1CC(=O)C(=C1)C(C2=CC=CO2)O	0.1599201	C1CC(=O)OC2=CC=CC=C21	0.09870844	CC1=CC=CC(=C1)C2=CC(=NO2)C	0.09953516
CC(C)(CCC(=O)O)C=C	0.1956975	C1=CC2=C(N2N)C(=C1)N	0.2669408	CC1(C2CCC(C2)C1(C)CN)C	0.9678022	C1=CC=C2C(=C1)C=CNC2=O	0.2377815	COCC1CCCCCC1=O	0.1601283	CCC12CCC=CC1CC2=O	0.0991675	C1=CC=C2C(=C1)C=CC(=O)N2	0.09962077
CC1=CC(=CC(=C1)O)CC(=O)O	0.1966269	C1=CC=C2C(=C1)NSN2	0.267303	C(#N)C1C2(C1(C2C#N)C#N)C#N	0.9682853	C1COC(=O)C2=CC=CC=C21	0.2380478	C1CC(=O)CC2(C1)OCCS2	0.1608221	CC1CCCC2C1=CC(=O)C2	0.0993776	CC(=O)C1=CC=C(C=C1)NC(=O)C	0.0997688
C1C2=C(C=CC(=C2)O)NC1=O	0.1979319	COC1C2=CC=C1C=C2	0.2674565	CC1(C2CCC1(C(C2)C=C)C)C	0.9687564	CC1=C(CC(CC1=O)C(=C)C)O	0.2384784	CC1(C2CCC(=O)C1C=C2)C	0.1608377	CCCC1=CCCCC1=O	0.1001967	CC1=C(C2=CC=CC=C2C1)O	0.09988057
CCCCC(=O)OO	0.1992974	C1=CC=C2C(=C1)NC(=S)N2	0.2675165	CCCC1(C(=O)CCC1(C)C)CC	0.972691	CC(C)C1CCCC(C1=C)O	0.2388732	CC1CCOC12CCCC(=O)C2	0.1610245	C1CCN2C=NCCCC2C1	0.1002489	CC1CC(=O)C2=CC=CC=C2O1	0.09993025
CC1=C(C=C2C(=C1)CC(=O)N2C)C	0.2002687	C=C(C1=CC=CC=C1)O	0.2682041	CC(C)(C)C1CCC(=O)C2(C1)CC2	0.9783082	C1=CC=C(C=C1)C(=C=N)C#N	0.2391575	CC12CCCC(C1CCC2=O)N	0.1618876	CC1CCCC(C1=O)CC=C	0.1006476	C1=CN=CC=C1C2=CNC(=O)O2	0.100433
CC1=CC2=C(CC(=O)N2C)C=C1	0.2005693	C1=CC(=CC(=C1)OC#N)O	0.2683056	CC(C)C12CCCC1(C2)C(=O)O	0.9785211	CC1=C(CCCC1)C(C)C=O	0.2393376	COC1C(=CCO1)C=O	0.1622084	CC1C2C=CCCC2(C1=O)C	0.1007632	CC1=NC(=NC=C1C=O)C	0.1004676
CC(C)(C)CCC(=O)O	0.2014808	CC#CC#CC1=CC=CC=C1	0.2685751	CCC1(CC(C1)(CC(=O)O)CN)C	0.9793228	CC1=CC(C(CC1)C(=C)C)O	0.2417573	CCC1(CCC(=O)CC1C)C	0.1654169	CCC(C)(C)C1=CCCC1=O	0.1007962	CC1=C(C=CC(=C1)N(C)C)C(=O)O	0.1006659
CC(C)(C)CCCC(=O)O	0.2016944	CC1=CC=C(C=C1)C=C	0.2689521	CC12CCC3(C1C3C(=O)C2)C	0.9794213	CC1CC(=O)CC(C1)(C)C=C	0.2418123	C1CC2CC1C3C2C(=O)C=CC3=O	0.1669775	CC1CC=C(C1=O)C(C)C	0.1026656	C1=CC(=CC=C1C(=O)O)NCC(=O)N	0.1011126
CC1=C(C=CC=C1O)CC(=O)O	0.2017215	C1CC1C(C2=CC=CC=C2)O	0.2702288	CC1(CC2(CC1CC2=NO)C)C	0.9803292	CCC1=C(C=C(C=C1)O)O	0.2418811	O=C(O)C[C@]1(C)[C@H](C(C)=C)CC1	0.1672923	CCC1=CCC(CC1)C(=O)C	0.1032739	C1C(C(=O)C2=CC=CC=C2O1)N	0.1012488
CC(C)(C)CCC(=O)N	0.2020651	C1=CC=C2C(=C1)OC(=O)S2	0.2702813	CC1(CC(=O)C2CCC2(C1)C)C	0.9807019	C1CNC=C2C1=CC(=O)C=C2	0.2421116	CCN(C)C(=O)C(=C)C(C)N	0.1684682	CC1=CC=CC2=C1OCC2=O	0.1044986	CNC(=O)C1=CN=CC=C1	0.1015008
C1C2=C(C=NC=C2)NC1=O	0.2040879	CC1=CC=C(C=C1)C(=S)C	0.2707252	CC1(CCCC(C1)(C)CC(=O)N)C	0.9810874	CCN1C(CC(CC1(C)C)N)(C)C	0.2425739	CC(=O)C1(CCCCC1)C	0.1689092	CC1CC=CC12CCCC2=O	0.1053045	CC1=CC=C(C=C1)C2=CC(=O)CC2	0.1015277
CC(C)(C)CCC=O	0.2041772	C1CC1C2=CC=C(C=C2)O	0.2720757	CC12CCC(=O)C1(CCCC2O)C	0.9853798	C1CCC2=C(C1)C=CC(=O)N2	0.2439061	C=C1CC2(CCSCC2)OC1=O	0.1704567	CC(=O)C1CC=CC2(C1)CC2	0.1062828	C1C(=CC(=O)O1)C2=CC=C(C=C2)O	0.101807
CCCC1=NNC(=O)N1C	0.2047837	C1CC(=O)OC2=CC=CC=C21	0.2732323	CC1(C(C1(C)C)C(=O)NC(=O)NC)C	0.9866892	CCC1(CCCCC1)C(C)O	0.2457081	CC(C)CC(=O)C(C(C)C)SC	0.1719915	CC1=C(SC(=S)S1)C=O	0.1064463	CC1=CC=CC(=C1O)C(=O)C	0.1018771
CC1=C(C=CC(=C1)CC(=O)O)O	0.205401	C1CC(=O)OC2=CC=CC=C21	0.2732323	CC1(C2CCC(C2)(C1NC)C)C	0.9871263	CC(C)(CO)C1=CC=C(C=C1)N	0.2469503	CCC(C(=NC(C)(C)C)C#N)OC	0.1725743	CCCCC1=CCCC1=O	0.1064606	CC1=CC(=CN1C)C(=O)C	0.1019132
C1=CN=C(N1)CC(C(=O)O)O	0.2054279	CSC1=CC=C(C=C1)SC	0.2738047	CC1(C2CCC1(C(=O)C2=O)C)C	0.9894577	C1=CC=C2C(=C1)C(=O)C=C(N2)O	0.2473247	CCCCC1(CCC(C1=O)(C)C)C	0.1726562	CC(=CC1CCCC1=O)C	0.1069071	CC1CC2=CN3C=CC=CC3=C2CN1	0.1019769
CC(CC(CC(=O)O)N)N	0.2056608	C1CC1C2=CC=CC=C2O	0.2746365	CC1(CCC(C=C1)(C)CCO)C	0.9909359	COC1=CC(=CO1)C2=CC=CC=C2	0.2476564	C1CC(=O)C2CC=CC1S2	0.1728521	CC1=CCC(CCC1=O)(C)C	0.1081057	CN1C=NC2=CC=CC=C2C1=O	0.1020677
C1OC2=CC=CC(=C2O1)CC(=O)O	0.2065537	C1CC2=CC=CC=C2C3C1O3	0.2754811	CC1(C(=O)CCC1(C)CO)C	0.9914543	CCC1=CC(=CC=C1)O	0.2476921	C1C2=C(C(=O)O1)NC(=O)SC2	0.1780171	CC1=CC=CC(=O)C(=C1)OC	0.1081808	CC1=CC=C(C=C1)C(=O)C	0.1021774
CC1(C(=O)CC1(C)C=C)C	0.2073416	C1=CC(=CC=C1O)OC#N	0.2758167	CC1(C2CCC1(C(C2)O)CO)C	0.993922	C1C=CC2=CC=CC(C2=N1)O	0.248533	CC1C2C(C2(C)C)C(=O)C1(C)C	0.1787916	CC1=C2CCCC2(CC1=O)C	0.1084658	CC1=NC2=CC=CC=C2C(=O)N1C	0.1023452
CC1(CC(=O)OC2=C1C=C(C=C2)O)C	0.2085055	CC=CC1=CC=CC=C1	0.2763366	C1CC12CC2(CCC#N)C#N	0.9964612	C1=CNC2=C(C(=O)C=CC2=C1)N	0.2485976	CC12CCCCCCC1CC(=O)CC2	0.1790382	C(=O)C=CCC=CCCCC	0.1087649	CC(=NNC(=O)C1=CN=CC=C1)C	0.1024015
C1=CN=NC=C1CC=O	0.2085959	CC(=C)OC1=CC=CC=C1	0.2769261	C1CCC23CCCC2(C3)C(=O)C1	0.9983489	O=C1C=C(C)CCC1C(C)C	0.2488628	CC(=C)C(CC1CO1)S(=O)(=O)O	0.1791792	C1CCC2=C(C1)CCC(=O)C2	0.1090536	C1CC2=C(C1)C=C3C(=C2)CCC3=O	0.1026672
CC1=C2CC(=O)N(C2=CC=C1)C	0.2088309	CC1=CC=C(C=C1)CC#N	0.2785339	CC1(C(C1(C)C)C(=O)N(C)CCO)C	0.9984891	C=CC1=CC(=CC=C1)O	0.249357	CC1=NC(=O)OC12CCCCC2	0.1792016	C1CCC2C(C1)OCO2	0.1094799	CC(=O)C1=COC2=CC=CC=C2C1	0.1027005
C=CCCCC=O	0.2088725	COC#CC1=CC=CC=C1	0.2786407	CCC1(C(=O)CC12CC2)CC	1.001484	CC(C)(C1=CC=C(C=C1)OC)N	0.2495643	COC1C=CC=CC(=N1)OC	0.1796928	CC1(C)C(CC1C2C)CC2=O	0.109544	C1CCN(C1)C2=CN=CN2	0.1029465
CCC(C)(C)CC(=O)OO	0.210236	C1=CC(=CC=C1C(=O)O)O	0.2787757	CC1CCC(C(=O)C1(C)C)(C)C(=O)C	1.005706	C1CC=CC(C1)CC#N	0.2495819	C1CC(=O)C2(C1=O)CCNCC2	0.1797304	CC1=CC(=O)CC(C1)C(=C)C	0.1095623	COC(=N)C1=CC=CC=N1	0.1029702
CCCC1=NC(C(=O)N1)C	0.2106007	C1=CC=C(C=C1)N=N	0.2787806	CC1(C2CCC1(CC2O)C)C	1.008828	O=C(OC1)CC2C1CCCC2	0.2505763	CCCN1CC2CCCC(C1)C2=O	0.1803329	C=C1CC2C1C(=O)CCC2	0.1099035	C1C(=O)NC2=C(N1)C=C(C=C2)C(=O)O	0.1030325
CC1=C(NC=C1)CC(=O)O	0.2113321	C1=CC=C(C=C1)ON=C=O	0.2806496	CC1(C2CCC1(C(C2C(=O)N)O)C)C	1.009302	C1=CC2=C(C(=O)C(=O)C=C2)N=C1	0.2515999	CC1(CCCC2(C1CCC2=O)C)C	0.1804905	CCC1=C(C(=O)CCC1)CC	0.1100155	CC(=O)C1=CC(=CN1)C(=O)C	0.1030509

Table 3.16

Mor170-1

SMILES	Distance
O=C1CCOC2=C1C=CC=C2	0
C(=O)(C)c1ccccc1	0
O=C1OC2=C(C=CC=C2)C=C1	0
C1=CC=C2C(=C1)C=C(NO2)C=O	0.06619393
CC1=NC(=O)C2=CC=CC=C2N1	0.0745505
CN1C=NC2=CC=CC=C2C1=O	0.08105557
CC1NC(=O)C2=CC=CC=C2O1	0.08323528
COC(=N)C1=CC=CC=C1	0.09673947
C1=CC=C2C(=C1)C(=O)N=CC=N2	0.1024509
C1C2=CC=CC=C2ON=C1C=O	0.1030488
C1=CC=C(C=C1)C(=O)CCN	0.1099358
CNCC(=O)C1=CC=CC=C1	0.1100806
CC1COC2=CC=CC=C2C1=O	0.1146739
C=CN1C=NC2=CC=CC=C2C1=O	0.1150157
C1=CNN(N=C1)C2=CC=C(C=C2)C=O	0.1202606
CC1=NC2=CC=CC=C2C(=O)N1C	0.1237537
C=NNC(=O)C1=CC=CC=C1	0.1254566
CN1NC(=O)C2=CC=CC=C2O1	0.1294035
CC1=NC2=CC=CC=C2C(=O)N1N	0.1318522
CNNC(=O)C1=CC=CC=C1	0.1335452
C1NC(=O)C2=CC=CC=C2S1	0.1342063
CC1=CC(=O)C2=CC=CC=C2O1	0.134347
CN1C(=O)C2=CC=CC=C2N=N1	0.1366252
C1=CC=C2C(=C1)C=C(C(=O)O2)C#N	0.1383896
C1=CC=C2C(=C1)C(=O)C=NS2	0.1389853
CC(=O)NN=CC1=CC=CC=C1	0.1393494
C1=CC=C2C(=C1)C=CC(=O)N2	0.1397765
CN1C(=O)C2=CC=CC=C2N=C1N	0.1422149
CCOC(=N)C1=CC=CC=C1	0.14513
C1=CC=C2C(=C1)C=C(NO2)O	0.1455027
CNC1=CC(=O)C2=CC=CC=C2N1	0.1477967
C1=CC=C2C(=C1)C(=O)SN2	0.1491949
C1C(C(=O)C2=CC=CC=C2O1)N	0.1500831
CC1=NNC2=CC=CC=C2C1=O	0.1531366
C1=CC(=CC=C1C(=O)CS)S	0.1534062
C1=CC=C2C(=C1)C(=O)C=C(O2)N	0.1549527
C1C(=CC2=CC=CC=C2O1)C=O	0.1551537
CCCCOC(=N)C1=CN=CC=C1	0.1592257
CC1=COC2=CC=CC=C2C1=O	0.1595478
C1=CC=C2C(=C1)C(=O)OC=N2	0.1598404
C1=CC=C2C(=C1)C(=O)N=C(S2)N	0.1613346
CC1CC(=O)C2=CC=CC=C2O1	0.1615579
C1C(=O)C2=CC=CC=C2ON1	0.1619112
C1=CC(=CC=C1C=O)S	0.1621826
C1=CC=C(C=C1)C=NNC=O	0.1636204
C1=CC=C2C(=C1)C=C(C(=O)O2)N	0.1664853
CON1C2=CC=CC=C2C1=O	0.1687667
C1=CC=C2C(=C1)C(=CC(=O)O2)N	0.1698968
C=CCOC(=N)C1=CC=CC=C1	0.1706281
C1=CC=C2C(=C1)C(=O)C(=CO2)N	0.171085
C1COC(=N1)C2=CN=CC=C2	0.1718966
C1=CC=C(C=C1)C=CNC=O	0.1723568
C1=CC=C2C(=C1)C(=O)N=C(N2)NC#N	0.1726051
C1=CC=C2C(=C1)C(=O)N=CN2N	0.1731622
C=C1COC2=CC=CC=C2C1=O	0.1745533
CN(C)C1=CC=C(C=C1)C=O	0.1748691
CC(C(=O)C1=CC=CC=C1)N	0.1751408
CCCC(=O)C1=CN=CC=C1	0.1754592
CC1=NC2=CC=CC=C2C(=O)O1	0.175988
C1=CC=C2C(=C1)C(=O)C=CN2	0.1761884
CCC1COC2=CC=CC=C2C1=O	0.1763109
CN1C(C2=CC=CC=C2N=N1)O	0.1763695
CC1=CC(=O)C2=CC=CC=C2N1	0.1785427
CCCOC(=N)C1=CN=CC=C1	0.1786124
CCC(=O)C1=CC=CC=C1	0.1811816
C1OC(=O)C2=CC=CC=C2S1	0.1812771
C1=CC=C2C(=C1)C(=O)C=C(N2)NN	0.18309
COC1=CC=C(C=C1)C=NC#N	0.1834068
C1=CC=C(C(=C1)C2=CO2)O	0.1838348
COC1=C2C(=CN=N1)C=CC=N2	0.1843654
CN1C=NC(=O)C2=CC=CC=C21	0.1845829
C1=CC=C2C(=C1)C(=O)N=NS2	0.1848174
C1=CC=C(C=C1)C(=O)NCCS	0.1850502
CNC(=O)C1=CC=C(C=C1)S	0.18571
CNC(=O)C1=CC=CC=C1	0.1867364
CC1=NC(=O)CC(=N1)C2=CN=CC=C2	0.1881487
C1=CC=C2C(=C1)C(=O)C=CO2	0.1892903
C1=CC(=CC(=C1)N)C(=O)CN	0.1901027
C1=CC=C2C(=C1)C(=O)C=C(N2)N	0.1909013
COC1=CC=C(C=C1)C2=NNCN=N2	0.1920341
C1C(=O)C2=CC=CC=C2OS1	0.1925363
CN1C2=CC=CC=C2C1=O	0.1926621
C1C=CN2N1NC2C3=CC=C(C=C3)O	0.192789
C1=CC=C(C=C1)C=NCC=O	0.193573
C1CC1C(=O)NN=CC2=CN=CC=C2	0.1939808
CC1=CC2=CC=CC=C2OC1=O	0.1943789
C=C(C1=CC=CC=C1)O	0.1993172
C1CC(=O)OC2=CC=CC=C21	0.2012817
COC1=CC=CC(=C1)C=C=C	0.2012826
CCSC(=O)C1=CN=CC=C1	0.2013818
C1=CC=C2C(=C1)C(=O)N=CN2	0.2033238
C1C(=O)COC2=CC=CC=C21	0.2044286
C1CC1NC(=O)C2=CC=CC=C2	0.2047669
CN(C)C1=CC=C(C=C1)C(=O)C#N	0.2070898
C1C2=C(O2)C3=CC=CC=C3O1	0.2073703
C=CC1COC2=CC=CC=C2C1=O	0.2074198
C1=CC=C2C(=C1)C(=O)N(N=N2)CS	0.2074352

Mor184-1

SMILES	Distance
CC1=CC[C@@H](CC1=O)C(=C)C	0
C[C@@H]1CC[C@H](CC1=O)C(=C)C	0
CC1=CC[C@H](CC1=O)C(=C)C	0
CC1CCC(C(=C1)O)C(=C)C	0.2379638
CC(=C)C1CC=C(C(=O)C1)O	0.3728224
CC1=CCC(CC1=O)C(=C)C=O	0.3772072
CC1(CCCCC1=O)CC=C	0.3977689
CC1CC(=O)C(C(=O)C1)CC=C	0.4097808
CC1=CC(=O)C(CC1)C(=C)C	0.4227823
CC(=C)CC1(CCCCC1)O	0.4271719
C=CCC1CCC(=O)NC1=O	0.4313385
CC(=C)C1CCC(CC1)C=O	0.4430801
CC1CCC(C(=O)C1)C(=C)C	0.4515747
CC1C(CCC1C(=O)O)C=C	0.4557655
CC1=CCC(CC1O)C(=C)C	0.4625187
CC(=C)C1CCC(=CC1)C(O)O	0.4628184
CC(=C)C1CCC(=CC1)C(=O)N	0.472155
C=CCCC1C(=O)CCCC1=O	0.4769287
CC(CC=C)C1CCCCC1=O	0.4788082
CC1CCCC1(CC=C)CC=O	0.48704
C=CCCC(=O)C1(CCCCC1)O	0.4917173
C=CCCC1(CCCCC1)C=O	0.4941705
CC(=C)CC1=CCCCC1=O	0.4981998
C=CC1CCC(CC1)C(=O)OO	0.5022362
CC(=O)C(=C)C1CCCCC1	0.5079129
CC(=C)CC1(CCCCC1=O)C	0.5103795
CCCCC(=C)C1(CCCCC1)O	0.5108372
CC1CCC(C(C1)O)C(=C)C	0.5122089
CC(=C)C1CCC(=CC1)C=NO	0.513459
CC(=C)C(C1=CCCCC1=O)O	0.5150736
C=C(CC1CCCC(=O)C1)C=O	0.5160087
CC1CCCC(=O)C1CCC=C	0.5162618
CC(=C)CC1(CCCCC1)C	0.5202251
C=CCC1CCC(C1)C(=O)O	0.5209791
CC1C(=O)CC(CC1=O)C(=C)C	0.5217194
CC1(CCCCC1=C)CCC=O	0.522958
CC(=O)C1(CCCCC1)C=C	0.5243087
CC1CCC(CC1O)C(=C)C	0.5249237
CC1=CCC(CC1)C(=C)CC(=O)C	0.531335
CC(=C)C1CCCC(=O)C1	0.5317809
CC1=CCC(CC1)C(=C)C=O	0.5349245
CC(=C)CC1CCCCC1=O	0.5361614
CC(=C)C1CCC(CC1)CO	0.5376741
CC1=C(CC(CC1=O)C(=C)C)O	0.5382746
CC1CCC(CC1)C(=C)CO	0.5403689
CC(=C)C1CCC(CC1)O	0.5430105
CC1CCCC(=O)C1CC=C	0.5452694
C=CCCCCC1(CCCCC1)O	0.5483312
CC(=C)C1CCC(=CC1)CCC=O	0.5522314
CC(=C)CC(=O)CC1CCCC1	0.5534906
CC(=C)CC1CCCCCC1=O	0.5544272
CC1C(C(OC1=O)C)CCCC=C	0.5577659
CC(=C)CCC1(CCCCC1)O	0.5586655
CC1(C(=O)CCCC1=O)CC=C	0.5609808
CC1=C(C(C(CC1)C(=C)C)O)O	0.5635416
C=CCC1(CCCCC1)C=O	0.5649692
CC1=CC(C(CC1)C(=C)C)O	0.566907
C=C1CC1CC(C(=O)O)N	0.5676443
CC(=C)C1CCC(CC1)(C)O	0.56807
C=CC1(CCC(=O)CC1)O	0.5702246
CC1=CCC(C(C)=C)(O)CC1	0.5746414
CC1(CCCCC1=C)CCO	0.5754528
CC(=C)C1CCC(=O)CC1	0.579186
C=CCCC1(CCCCC1)O	0.5797328
CC(=C)C1=CC(=NC=C1)C(=O)C	0.5798283
CC1=CCC(CC1)(CC=C)O	0.580173
C=CCC1(CCCCC1=O)O	0.586857
CC1CCC(C(C1)O)C(=C)O	0.5898029
CC1CCCC(C1=O)CC=C	0.5907102
C1(=O)(=O)C(=CCC(C1)C(=C)C)C	0.5975424
C=CC1(CCC(=CC1)C=O)C=O	0.5976542
CC(=O)CCCC1CCCCC1=C	0.5979454
CC(=C)C1CCC(=CC1)C(=O)O	0.5988352
C=CCC1CCC(=O)C=C1O	0.6003185
CC(=C)C1CCC(C1)(C)C(=O)O	0.6020785
CC(C)C1=CCC(=C)CC1	0.6037065
CC(=C)C1CCC(=C(C1)O)CO	0.6041441
CC1=CCC(CC1)C(C)=C	0.6078608
CC(CCC1C(O1)(C)C)C=C	0.608671
C=CC(=O)C1(CCCCC1)O	0.6094027
CC1=CCC(CC1=NO)C(=C)C	0.6102408
CC(=C)C1(CCCCC1)O	0.6115156
CC1(CCCCC1)C(=C)CCC=O	0.6115773
CC(=C)C(=O)C1CC(=O)OC1	0.6123015
C=C1CC=C(C(=C1)O)C(=O)O	0.6128066
CC(=C)C1CCCCC1=O	0.6135868
CC(=C)CCC1CCCCC1=O	0.6165297
CC(=C)CC1CCOC(=O)C1	0.6166073
CC1(O)CCC(C(C)=C)C=C1	0.6211655
CC(=O)OC1(CCCCC1)C=C	0.6216972
C=C1CCC(=NNC(=O)N)CC1	0.6224033
CC1=CC=C(S1)CC(=C)C(=O)O	0.6224149
C=CCCC1CCCC(=O)C1	0.6226769
CC1=C(C(=O)CCC1)CCC=C	0.6232394
CC(=C1CCC(O1)(C)C=C)C	0.626257
CC1(CCCC(=O)O1)CC=C	0.6275748
C=CCC1(CCCC1=O)C#N	0.6277897

Mor185-1

SMILES	Distance
C(=O)(C)c1ccccc1	0
O=C1OC2=C(C=CC=C2)C=C1	0
C1CC2=COC=C2CC1=O	0.02606647
C1=CC2=C(C=CC(=O)N2)N=C1	0.03118136
C1CC(=O)C2=C1C=CC=N2	0.0323404
C1CCC2=C(C1)CCC(=O)O2	0.03555811
CCC1CCCC(C1)N	0.0418893
CC(=C)C1=COC=C1	0.04258475
CCC1=CCC(CC1)O	0.04279004
CCC1CCC(CC1)N	0.04383469
C1CC(=O)C2(C1)CC2	0.04396265
C1CC(=O)C2=C1N=CC=C2	0.04402903
C1CN2CC(=O)OCC2CN1	0.04424295
C1C(CC2C1CNC2)N	0.046368
C1CCC2C(C1)CC2=O	0.04657459
C1CN2CCOC(=O)C2CN1	0.04905777
C1C2C=CC=C2C(CN1)O	0.04945669
C1CCC2(CC1)CC2O	0.04991706
C1CCC2=C(C1)C(=O)CN2	0.05012532
CCC1=C(CCCC1)O	0.0513685
CC(=O)C1=CCCC=C1	0.05278158
C1CC(=O)C2=CN=CN=C21	0.05282474
C1=CC2=C(NC=CC2=O)N=C1	0.05299349
CONCCN1CCCC1	0.05360856
CCC1CCCCC(C1)O	0.0536915
CC1=CC(=CCC1)OC	0.05373938
CCC1=CC=CC=C1N	0.05481358
C1CC(C2C1CNC2)N	0.05518074
CC(=C)C1=CNC=C1	0.05549717
CC(C)C1CNCC=C1	0.05600957
C1CC2CCC(=O)C3C2(C1)O3	0.05616981
C1CCC(C2CNCC2C1)N	0.05652076
C1CC2=C(C=C1)NNC(=O)C2	0.05701812
CC(N1CCCCCC1)O	0.05795114
C1CC(=O)C2=C1CC=CC=C2	0.05836502
C1C(=O)CNC2=CC=CC=C21	0.05836631
C1CCN2CCCOC(=O)C2C1	0.05923841
C1CC2=C(C(=O)NC1)N=CN2	0.05952994
COC1=CC(=O)NCCN1	0.0595711
C1CCN2CCNC(=O)CC2C1	0.05968883
C1=CC2=C(NC(=O)C=CN2)N=C1	0.06004616
C1CCC(CC1)C2CCC2=O	0.06023724
C1CC2=C(CC1=O)C=CN2	0.0610729
C1CC2CC(=O)CCN2C1	0.0621596
CCC1=C(C=NC=C1)C	0.06224206
CC(=O)N1CC2CC1CN2	0.06255557
C1=CC2=C(C=C1S)C=NN2	0.06260877
CC1=C(C=NC=C1)C=C	0.06314167
C1CC2CCC(=O)C=C2C1	0.06318773
C1=CC2=C(C=C1C(=O)O)NC=CO2	0.06328246
CN=C1CCCC(=O)C1	0.0636295
C1=CC2=CC=CC(=O)N2C=C1	0.06390911
CC1C2C(COCO2)OCO1	0.06411597
CN1CCNCCC1CO	0.06413717
C1C=CCOC2N1C(=O)C2	0.06419423
CC1CCC(=C1)SC	0.0645993
C1=CC=C2C(=C1)C(=O)N=CO2	0.06465735
COC1CC(CCC1O)C2CO2	0.06489509
C1C(CNC1CO)S	0.0650043
C1CCC2C(C1)CCC(=O)O2	0.0658593
CC(=O)C1=CC=NC=C1	0.06617431
CC(=O)C1C2C1COC2	0.06695876
COC1=CC2=C(C=C1)C(=O)CC=C2	0.06724022
C1=CC2=C(C(=O)N=CN2)N=C1	0.0674308
C1CC(=O)C2=NNCN2C1	0.06755074
CN1CC2=C(C1)CNC2	0.06756412
C1CCC2C(C1)CCC(O2)O	0.06782162
C1C=CC(=O)C2=C1CNNC2	0.06802411
C1CC=C(C(=O)C1)C2=COC=C2	0.06802758
C1=CN=CC2=C1C(=O)NN=C2	0.06806372
CC1=C2C=C(C=CC2=NC=C1)OC	0.06819661
C1C=CC=C2C1CC(=O)C=C2	0.0682767
C1CC(=O)C=CC=C1	0.06843499
CC(C1CCCC=C1)N	0.06849173
CCC1=C(C=CN=C1)C	0.06864301
C1=CNC2=NC=NC=C2C1=O	0.0686631
CC1=C(CCCC1)OC	0.06879747
CC(=O)C12CC(C1)C2	0.06883269
C1C(=O)N=NN1C2=CC=CC=C2	0.06891363
COC1=CCC(=N)C=C1	0.06926642
CC1=CCOC2=C1C=CC=N2	0.0693349
COC1=CC2=C(C=C1)C(=O)NCCN2	0.06966745
CN1C=NC2=C1CCNC2	0.06982317
COc1c(cccc1)N	0.07058182
CC(=O)N1CCCC=C1	0.07091169
C1CCC2=C(C1)N=C(N=N2)N	0.07095861
C1CCC2=C(C1)CCCC2=O	0.07114383
C1CC(=O)C2=C1C=CC=CC2	0.0712115
C1CC2=C(CC=C1)C(=O)CC2	0.07122314
C1=C(C(=CS1)O)C=O	0.07147205
CC1=CC=CC(C1)OC	0.07168902
C1CC2=CC3=C(C(=O)N2C1)NC=C3	0.07198178
CC(=O)N1CC2CC1CO2	0.07210112
C1CN(C1=O)C2=CC=CO2	0.07225106
C1CCC2C(C1)CCCC2O	0.07241177
COC1=CC=C(C=C1)C(=O)N2CC2	0.07245435
COCOCC1CCCO1	0.07253032

Mor189-1

SMILES	Distance
C[C@@]12CC[C@@H](C1(C)C)CC2=O	0
CC1=CC[C@@H](CC1=O)C(=C)C	0
C[C@]12CC[C@H](C1)C(C2=O)(C)C	0
C[C@@]12CC[C@H](C1(C)C)CC2=O	0
CC1=CC[C@H](CC1=O)C(=C)C	0
C[C@@]12CC[C@@H](C1)C(C2=O)(C)C	0
CC1(C2CCC1(CC2=O)C)C	0.04668916
CC(=C)C1CC=C(C(=O)C1)O	0.06676451
CC1=CC(=O)C(CC1)C(=C)C	0.1183577
C=C(C)C(CCC1C)=CC1=O	0.1347063
CC1CCC(=CC1=O)C(C)C	0.1680434
CC1=CCCC(C1=O)(C)C#C	0.1763185
CC1=CC(=O)CC(C1)C(=C)C	0.1962587
CC1=CC(CCC1=O)C(C)C	0.1976955
CC1C2(CCC1(C(=O)C2)C)C	0.2061102
CC1CC=C(C(=O)C1)C(C)C	0.2072536
O=C1C2C(CCC(C2)C1C)=C	0.2078362
CC1(CC(C(=O)C1)CC=C)C	0.2092703
CC(C)(C12CCC(=O)C1C2)O	0.2097237
CCC1=C(C(=O)C(CC1)C)C	0.2140007
CC1=CCCC(C1=O)(C)C=C	0.2180427
CCCC1(CCC(=O)C=C1)C	0.2194855
CC(C1C=CCCN1)C(=O)C	0.2209319
CC1=CCC(CC1)C(=C)C=O	0.2250367
CC1(CCC(=O)C=C1C=C)C	0.2310372
C1CN(CCC1CCN)C=O	0.2406093
O=C1C=C(C)CCC1C(C)C	0.241701
C=C1C=CCCC1CCC=O	0.2420684
CCC1=CC(=O)CCC1(C)C	0.2445011
CC1C(=C)C2CCC(C2)C1=O	0.2468526
CCCC1CCC(=O)C(=C1)C	0.2519408
CC1C(C(=O)C=CN1C)(O)O	0.2525677
CC1(CC(=C(C(=O)C1)O)O)C	0.2532433
CC(=C)C1CCC(=CC1)C=O	0.2564928
CC1C2CC2(CC1=O)C(C)C	0.2579151
CC1(CCC=C(C1=C)C=O)C	0.2583599
CC(C)(C)C1CC(=O)C2C1C2	0.2593795
C[C@@H]1[C@H]2C[C@]2(CC1=O)C(C)C	0.2594466
CC1(C2C(C1=O)CCCO2)C	0.2597198
CC1(CCC(=O)C(=C1)OC)C	0.2611222
CC1(C2CCC1C(=O)C=C2)C	0.270247
CC1(CC=C(C(=O)C1)CO)C	0.2773754
CC1(CCN(CC1)C(=O)O)N	0.2797628
CC(=CC1CCC=CC1)C=O	0.2807091
CC1CCCC=C1C(=O)C=C	0.2841142
CC1C=C(CCC1(C)C=O)C	0.2842825
CC1CC(C2CC2C1=O)(C)C	0.2861447
CC1CC(=O)C2(C1C2)C(C)C	0.2869487
CCCC1(CCCOC1)C=O	0.2904827
CC(C)CC1=CC(=O)CCC1	0.2909898
C1C(C(C(C1=O)(CO)O)O)N	0.2924971
C1CC2C=CC=CC2C(=O)C1	0.2940839
CC1C=C(CCC1=O)C(C)C	0.2948304
CC(=C)CC1=CCCCC1=O	0.2954905
CN1CC2CCCC(C1)C2=O	0.2970333
CC(C)(C1CCC=CC1=O)O	0.2971604
CC1(C2CC=CC(C2)C1=O)C	0.2978292
C=CC1CC(C(C1)C=O)C=C	0.2980856
CC1CC2CCC(=O)C2C=C1	0.2987179
CCC1=C(CC(CC1=O)C)C	0.2989803
CC1=C(C(=O)CCC1)C(=C)C	0.2999092
CC1=CC(CC(C1=O)(C)C)O	0.3007561
CC(=O)C1CC(CNC1)NC	0.3024095
CC(C)(C)C1CC(=O)C=CO1	0.3030642
CC1(CC(=CC(=O)C1)CO)C	0.305218
CC(=CC1=CC(=O)CCC1)C	0.3065182
CC1=CCCC(=C(C)C)C1=O	0.3090123
CC1=C(CC(CC1=O)(C)C)O	0.3092388
CCCC1CC(CC1=O)(C)C	0.3119717
CC1CCC(C1C=O)C(=C)C	0.31434
CC1=CC(=O)C(=C(C)C)CC1	0.3163057
O=C1C(C)=CCC(C(C)C)C1	0.317051
CC1CNC(CN1C(=O)O)C	0.3172796
CC1(CCN(CC1)C(=O)O)O	0.3178126
CC1(CCN(CC1)C(=O)O)C	0.3199334
CC1(CC(=CC(=O)C1)NC)C	0.320232
C1CN(CCC1CCO)C=O	0.3209767
CC12CCCC=C1OCC2=O	0.325901
C=C(CC1CCC=CC1)C=O	0.3269808
CC(C1CC=C(C=O)CC1)C	0.3273557
CC1CC(CC(=C1)C)C(=O)C	0.3298992
CC1(CCCN(C1)C(=O)O)C	0.3308055
C1CC2=CC=CC(=O)C2NC1	0.3320085
CCN1CCC(=O)C(C1C)C	0.3344908
C1CC(CNC1)CC(C=O)N	0.3346634
CC1CC(CC(=C1C)C)C=O	0.3363059
CC1CC(=CCN1C)C(=O)C	0.3367353
CC1=CC(C(C(C1=O)O)O)O	0.3367478
CN1C=CC(=O)CC1CC=C	0.3369189
CC12CCC(C1)C=CC2=O	0.3379139
CC1CC(=CCC1C(=O)C)C	0.3388549
C1CC=C(C(C1)CCC=O)O	0.3434852
CCC(=O)C1(CCOCC1)C	0.3450098
CC(C)(C)C1CC(=O)C=CN1	0.3470305
CC1CC=CCC1(CO)C=O	0.3483415
CC1(CCC(=CO1)OC)C=O	0.3505473
C1CNCCC1CC(C=O)N	0.3510096

Mor2-1

SMILES	Distance
CC[C@@H](C1=CC=CC=C1)C(=O)O	0
CC[C@H](C1=CC=CC=C1)C(=O)O	0
C(=C)COC(=O)Cc1ccccc1	0
CC1C(=O)N(C1=O)C2=CC=CC=C2	0.1582261
CCCOC(=O)CC1=CC=CC=C1	0.1927919
C#CCOC(=O)CC1=CC=CC=C1	0.2135794
CC1CCC2=C(C1)SC(=N2)CC#N	0.2261997
C1CC1(C2=CC=CC=C2)OCCS	0.2307089
C=CCOC(=O)CC1=CC=C(C=C1)O	0.2685227
COC(=O)Cc1ccc(cc1)OC	0.2720388
CCC(C1=C(C=NC=C1)CO)OC	0.2840643
COC(=O)CC1=CC=CC(=C1)C#N	0.2858069
C1CCS(=O)(=O)C2=CC=CC=C2C1	0.2878607
C1CC1(C2=CC=CC=C2CO)O	0.2910175
CC(C)OC(=O)CC1=CC=CC=C1	0.2962586
C1=C(C(=C(N1)CN)CC#N)CCC#N	0.2973288
CC(CC1=CC=CC=C1)N=C=S	0.2995831
CCC(COCC1=CC=CC=C1)S	0.2998371
C1CC2=CC=CC=C2NC(=O)C1	0.3024153
C1CC(C2=C(C1)C=CS2)NC(=O)N	0.3030766
CC1(CN(C2=C1SC=C2)C=O)C	0.3041082
CCC(C1=CC=CC=C1N)C(=O)O	0.3045982
C1CC(C2=CC=CC=C2SC1)O	0.3073687
C1CC(=O)CCC2=C(C1)NC=C2	0.3082988
C1CCC(C(=O)CC1)C2=NC=CN2	0.3085543
CC(C1=CC=CC=C1N)C(=O)OC	0.3109102
C1=CC(=CC(=C1)CN)CC(=O)O	0.3115366
CCOC1=CC=C(C=C1)CC(=O)O	0.3116775
CCOC(=O)CC1=CC=C(C=C1)O	0.3117947
CCN1C(OC2=C1C=C(C=C2)C)C	0.3119305
CCCOC1=CSC2=CC=CC=C21	0.3127963
C1C2C(NCN1)SC3=C2C=NC=C3	0.3141638
CCC(C1=CC=CC=C1)C(=O)OC	0.3146093
COC(=O)CC1=CC=C(C=C1)NN	0.314688
CC1CC2=C1C=CC=C2S(=O)(=O)O	0.3164913
CC(C=C)N(CC1=CC=CC=C1)O	0.3168748
CCOC(=O)C(C)C1=CC=CN1C	0.3178926
CC#CCOC(=O)CC1=CC=CC=C1	0.3181635
C1=CC=C(C(=C1)CC(=O)O)CN	0.3225541
CN1CC2=C(CC1C(=O)O)SC=C2	0.3238227
CCOC(=O)C(C)C1=CC=CC=C1	0.3280601
CN1C=CC=C1C2CCC(=O)N2C	0.329925
CC1=C2CCCN2C(=O)CCC1	0.3307236
C1=CC=C(C=C1)C(CO)C(=O)O	0.3332723
CC(C1=CC=CC=C1)C(=O)OC	0.3341447
C1CC2=C(CC1C(=O)O)SC=C2	0.3348343
CC(C)N1CSC2=CC=CC=C21	0.3356953
CCOC(=O)CC1=CC=CC=C1	0.3379899
CCC1C2C(CCN2C(=O)O)NC1=O	0.3382412
CCOC1(C2=CCCCN2C(=O)O1)C	0.3422228
CCOC(=O)CC1=CC=C(C=C1)N	0.342778
C1CCC2=C(C(C1)N)N=CC=C2	0.3433717
CCN1C(OC2=CC=CC=C21)C	0.3446326
CN(CC1=CC=CC=C1)CC(=S)N	0.3455022
CCC1C(COC1=O)C2=NC=CN2C	0.3489334
C1CC2=C(C=CC(=C2)O)NC(=O)C1	0.349914
COCC(C1=CC=CC=C1)C(=O)O	0.3515843
CN=C1CNCC2=CC=CC=C2S1	0.353092
CC1=CCC2C(C1)N2C(=O)OC	0.3530949
C=CCC1=CC=CC=C1OCCN	0.3552634
CC(C1=CC=CC=C1)C(=O)OC=C	0.3563263
CC1=C(C=CC=C1O)C(C)C(=O)O	0.3579266
COCC1=CC=CC=C1COC	0.3583624
CC1CCNCC2=C1ON=C2OC	0.3585697
CC(COCC1=CC=CC=C1)S	0.3591302
CC(N(C)CC1=CC=CC=C1)S	0.3594296
C1CCC2=C(C=CS2)C(C1)NC=O	0.3603129
C1=CC(=C2C=CSC2=C1)CCO	0.3640987
CCCN1C2=CC=CC=C2N(C1=O)N	0.3647515
CC1=C2C(=CC=C1)NC(=O)CCS2	0.3676669
CC1(OC2=CC=CC=C2O1)CCN	0.3676686
CC1=CC=CC=C1N(C)C(=O)C	0.3680748
CC1=C(C(=CC=C1)CO)OCC#N	0.3685367
CC1(CC(CCO1)O)C2=CN=CS2	0.3691189
C1CC2=C(C=C(C=C2)O)NC(=O)C1	0.3697989
CC1=C(C(=CC=C1)C)OCCC#N	0.369876
CCCOC1OC2=CC=CC=C2O1	0.3698959
CC(C1=CC=CC=C1)C(=O)OC#N	0.3707031
CC(C1=C(C=NC=C1)N)C(=O)OC	0.3712612
C1COC2=CC=CC=C2C(C1N)O	0.3719231
CC(C#N)C1=CC=C(O1)CN(C)C	0.3727306
CC1=CC(=C(C=C1)OC)CC(=O)O	0.3738466
CC(C1=CC=CC=C1CN)C(=O)O	0.375163
CC1=C(C(=CC=C1)C)OCC(=N)N	0.3760855
CC(=C)CC1=CC=CC=C1SC	0.3764092
COC1=CSC2=C1C=C(C=C2)N	0.3772695
C1=CC=C(C=C1)CC(=O)OCCN	0.3776953
COC(=O)CC1=CC(=C(C=C1)O)N	0.3783322
COC1=CC=CC=C1CC(=O)O	0.3796353
CCN1C(CC2=CC=CC=C21)CN	0.3803442
CC1C(C2=C(CN1)C=CS2)OC	0.3805739
COC(=O)C(CO)C1=CC=CC=C1	0.3815324
CC(C)C(=C)C1C(N=CO1)C(=O)O	0.3823562
CNC(C1=CC=CC=C1)C(=O)O	0.3838065
CCC1(C(=O)N(C(=O)N1C=C)C=C)C	0.3864021
CC1=C(C(=CC=C1)C)OCC(=S)N	0.3872609
CC1=CC=CN1C(CO)C(C)C	0.3875966

Mor203-1

SMILES	Distance
CC(=O)CCCCC	0
CC(=O)CCCCCCC	0
CC(=O)CCCCCC	0
CCC(=O)CCCC	0
CCC(=O)CCCCC	0
CCCCC(=O)CCC	0.09870324
CCC(=O)CCCC(C)C	0.1234426
CCC1CCC(CC1)C(=O)CC	0.1310437
CC(CCCCCO)C	0.1429018
CCC1CC(C1)C(=O)C	0.1522912
CCC(=O)C=CCCC	0.1569658
CCCCCC(=O)CCC	0.1636356
CC=CC=CCCO	0.1710732
CC(=O)CC1CCCC=C1	0.1711094
C=CC1=CC=C(C=C1)CCCO	0.1778983
CCCCCCC(CCC)O	0.1785371
C#CC1=CCC(CC1)CCO	0.1786504
CC(CCCC=C)O	0.1865765
CC(=C)CCCC(=O)C	0.1901121
CCCCCC1CC(=O)C1	0.1926162
CCC(CCC=C=C)O	0.1948856
CC(C)CCCC(=O)C=C	0.1949919
CCCC=C=CCO	0.1959073
CCCCC(=O)NC	0.196173
C1CC(=CC=C1)CCCO	0.1965939
C(C=CC=CCC=CCC)O	0.1966562
CC(C)CCNC(=O)C	0.1977104
CC(C)C(CCCC=C)O	0.1995084
C1CC1=CCCCCO	0.2000587
CC(C)CC=CC(=O)C	0.2005665
CCCC=CC(=O)C	0.2069004
CCCCCC1=CC(=O)C1	0.2083482
CC(=O)C=C1CCCCC1	0.2084282
CCCCC(=C)CCO	0.2097396
CCCCC(=O)C=C	0.2100252
CC(C)C1=CC=C(CC1)CO	0.2100582
CC(CCC(=O)CC)CC	0.2106138
CC1CCC(CC1)C(=O)C	0.2122727
CC(C)OC1=CCCC=C1	0.2141855
CC1(CO1)CCCC=C	0.2146121
CCC(=O)CCC(=C)C	0.215481
CC(=CC)CCC(=O)C	0.2160098
CC(CCC(CC)O)CC	0.2169532
C1CC=C(C=C1)CCCO	0.2170895
CCC(C1CCCCC1)O	0.2172078
CC(=C=CCC(=O)C)C	0.2175421
CCCC=CCCO	0.2190903
C=CCCCC#CCO	0.2196209
CC(=O)C=CC1CCCCC1	0.2214962
CCCCCCCC(C)O	0.223367
CCCCCCC(C)O	0.2233692
CC(C1=CC=C(C=C1)C=C)N	0.2238223
CC(=O)CCC1CC1=C	0.2242527
C#CCCCC1(CC1)O	0.2246465
CCC(C)CC(=O)C=C	0.2258775
CC(=O)CC1CCCCC1	0.2267393
CCCCC(CC)N	0.2276189
CC(C)OC1=CC=CNC1	0.227668
CC(C)OC1=CC=CCC1	0.2277479
CC(C)OC1=CCC=CC1	0.2279199
CC1=CC=C(C=C1)OCC=C	0.2279726
C=CCCCC1=CC=C(C=C1)O	0.2280567
C1CC1=C2CC2CCCO	0.2295331
CCCC1CCC(=O)CC1	0.232204
CC(C)OC1CCCC=C1	0.2322558
CCCCC1CCS(=O)CC1	0.2322805
CCCC(=O)CCCCCC	0.2335926
CCCC#CCCO	0.2340755
CC(=C(C)CCC(=O)C)C	0.2341524
CCCCNC(=O)C=C	0.2346179
CCC(CCC#CC)O	0.2349258
CC(C)C1=CCC(CC1)CO	0.2357801
CN=S(C)C1=CC=CC=C1	0.2363829
CC1(CCCCC1)OC=C	0.2367655
C1CCC(CC1)C2=CCC(=O)CC2	0.2372365
C1CCC(=C=CCO)CC1	0.2381111
C(CCC=CCC)O	0.2382502
CCCCC1=C(C1=O)C	0.2382673
CCCCCCC(=N)N	0.2385028
CCC1CCC(CC1)CO	0.2386696
CC(C=C)OC1=CNCC=C1	0.2388276
CC(CC/C=C(C)\C)=O	0.2393879
CCC(=O)C=C1CCCCC1	0.2394289
C=CCCCCCO	0.2399206
CN(C)C(=O)CCC=C=C	0.2417149
CC(C)C1=CCC(=CC1)CO	0.2422732
CC1(C(O1)CCCC=C)C	0.243974
CC(=CCCC(C)O)C	0.2440488
C=CCC1CCC(CC1)O	0.2444802
C1CC=C(C=C1)NCCO	0.2455177
CCCCCCC(C(C)C)O	0.245861
CC(CCCC#C)OC	0.2468195
CC1=CC=C(C=C1)COCC#C	0.2472258
C1CCC(CC1)CCCO	0.2483911
CC1=CC=C(C=C1)OCC#C	0.2485717
CC(=O)CC1=CCCCC1	0.2488646
CC(=CC1CC=CCO1)C	0.2489077

Mor204-6

SMILES	Distance
CC1=CC[C@@H](CC1=O)C(=C)C	0
CC1=CC[C@H](CC1=O)C(=C)C	0
O=C1OC2=C(C=CC=C2)C=C1	0
CC1(CCCC=C1C(=O)O)C	0.1592494
CC1CCC(=CC1=O)C(C)C	0.1767846
CC1CCC(=CC1=O)C(C)C	0.1767846
C1=CC=C2C(=C1)C=COC2=O	0.1926777
C1=CC(=O)NC2=C1C=NC=C2	0.2034395
CC1=CC(=O)C(CC1)CCO	0.2121746
C=CC1=CC=C(C=C1)C(=O)O	0.2283565
COC(=O)C1=CCCC(C1)O	0.2532418
C1=CC2=C(C(=C1)O)C(=O)NC=C2	0.2600318
C1=CC2=C(NC(=O)C=C2)N=C1	0.2607831
C1=CC(=O)NC2=NC=NC=C21	0.2758744
CN1C2C1C(CC(=C2)C(=O)OC)O	0.2813197
C1=CC2=C3C(=C1)NC=C3C(=O)C=C2	0.2828428
C=CC1=CN=C(C=C1)C(=O)O	0.2854501
CC(=C)CC1=CCCCC1=O	0.31266
CC1=CCC(CC1=NO)C(=C)C	0.3135219
C1C(=CC1=O)C2=CC=CC=C2	0.3180221
COC(=O)C1=CCC(CC1)SC	0.3191871
C1CC(=O)C2CC1C(=O)C=C2	0.3195927
CCC(=O)C1=CCCC(S1)C	0.3235687
C1CCC2(C1)CCC=C(C2=O)O	0.3305075
C1C(=CC(=O)N1)C2=CC=CC=C2	0.3309083
C1=CC=C2C(=C1)C=CC(=O)N2	0.3342322
CC1CC=C(C(O1)C)C(=O)O	0.3351069
CC1CCCC=C1C(=O)O	0.337837
CC1(CCC(=CC1)C(=O)OC)C	0.3397238
CC(=C)C1CCC(=CC1)C=NO	0.3409651
CC1=CC2(CCCCC2)OC1=O	0.3414605
COC(=O)C1=CCCC(=O)C1	0.3433653
CC1(CCC(=O)C=C1)CCC#N	0.3434153
CC1=CC=C(C=C1)C=CC(=O)N	0.3442434
C1=CC2=C(C=CN2)C3=C1C(=O)C=CN3	0.344622
CCC(=O)C=C1CCCCC1	0.3462852
C1CC(=O)C=C1C2=CC=CC=C2	0.346597
C1C(=O)C=CC2=CC=CC=C21	0.3564579
C1=CC=C(C=C1)C2=CC(=O)NN2	0.3590267
CC(=O)C1=CC=CC=C1C=C	0.3619179
C1=CN(C(=CN1)C(=O)O)C(=O)O	0.3655498
CCC1CC(=CNC1)C(=O)C	0.3720756
COC1=CC(=NCN1)OC	0.3752879
COC(=O)C1=CCCCC1	0.376413
COC(=O)C1=CNCCC1	0.3808107
C1=CC2=C(C=CC(=O)N2)N=C1	0.381177
C1=CC=C2C(=C1)C(=O)C=CN2	0.3838855
CC1=CN(C(=O)N(C1=O)C)C	0.388122
C1=CC2=C3C(=C1)C=CN3C(=O)C=C2	0.3906061
C1C=C(C(=O)N1)C2=CC=CC=C2	0.3926395
CC1=CC(=O)CC1C2=CC=CC=C2	0.3936016
CCOC(=O)C1=CNCCC1	0.3956648
CC1=CCC(=C(N1)C)C(=O)OC	0.3977997
C1=CN(C(=O)NC1=O)CC=O	0.3981899
CC(=O)C=C1CCCCC1	0.400185
CC(=O)C1=CC(CCC1)O	0.405628
CC1(CC(C(=O)C(=C1)O)(C)C)C	0.4085377
C1N=C(C=CN1C(=O)O)CN	0.4105892
C1=CC2=C(N=C1)N3C(=C2)C=CNC3=O	0.4109866
C1=CNC(=O)C2=NC=CN=C21	0.4112343
CC1=CC=CC(=C1)C(=O)C=C	0.4130462
C1=CC=C2C(=C1)C3=C(N2)C=CNC3=O	0.414148
C1=CC2=C(NC=CC2=O)N=C1	0.4156558
CC1=CC(C(CC1=O)C(=O)C)O	0.4163652
C=CC1=NC(=C(C=C1)C(=O)O)N	0.4177307
C=CC1=CC=CC=C1C=O	0.4198651
CC1=CC=C(C=C1)C=CC(=O)C	0.4208217
CC1=CC=C(C=C1)C2(CC=CC2=O)C	0.4220633
C1CC2(CC1C=C2C(=O)O)C(=O)O	0.4237751
CC1C(=CCCO1)C(=O)O	0.424504
CC1CC(C=C(C1=O)O)(C)C	0.4245487
C1=CC=C(C=C1)CN2C(=O)C=CS2	0.4277974
CCOC1=CC(=O)C(C(C1)C)C	0.4301193
C1C(=O)C=CC(O1)C2=CC=CO2	0.431292
CC1=CC2=C(C(=C(S2)C)C)C(=O)O1	0.4326049
C1=CC2=C(C=C1O)C(=O)C=CN2	0.4352579
C#CCN1C=CC(=O)NC1=O	0.4362954
CC1=CC(COC1=O)(C)C	0.4374064
CC1=C(C(=C(C=C1)O)C)C=C	0.4384387
CN1C(=O)C=C(N(C1=O)CC=C)N	0.4402955
CC1=C(C=CC(=C1)C=C)C(=O)O	0.4415632
CCN1C(=O)C(=CN(C1=O)C)C	0.4423475
CCC1=CN(C(=O)N(C1=O)C)C	0.4439708
CC1=CC(=O)N(C2=NC=NN12)CC=O	0.4457211
C1=CC=C2C(=C1)C(=NO)C=CO2	0.4466131
C1=CN(C(=O)NC1=O)C(=O)O	0.446993
C1C=C(NC2=CC=CC=C21)C=O	0.4476903
CC1=CCC(CC1=O)C2(CO2)C	0.4487036
C#CCC1=CNC(=O)NC1=O	0.4487506
COC(=O)C1=CC(C2C(C1)O2)O	0.4488419
CC1=C(NC(N=C1)(O)O)C	0.4506406
C=CC1=CC(=CC=C1)C=O	0.451164
C1=CN=CC=C1C2=CC(=O)NN2	0.4513892
CC1=C(C=C2C=C(CC2=C1)C(=O)C)C	0.452722
CC1=CCC(CC1=O)(C(=C)C)O	0.4539375
CC(=O)C1=C(CCCC1)CC=C	0.4540705
C1=CC2=C(C=CNC2=O)C(=C1)O	0.4541359

Table 3.16 Continued

Mor207-1		Mor223-1		Mor250-1		Mor256-17		Mor258-1		Mor259-1		Mor260-1	
SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance
O=C1CCOC2=C1C=CC=C2	0	C(=C)COC(=O)Cc1ccccc1	0	C(=O)(C)c1ccccc1	0	C1C2=CC=CC=C2OC1=O	0	C1C2=CC=CC=C2OC1=O	0	C1C2=CC=CC=C2OC1=O	0	C(=O)(CCCCCCCCC)O	0
C(=O)(C)c1ccccc1	0	c1(ccccc1)COC(=O)C	0	O=C1OC2=C(C=CC=C2)C=C1	0	CCC(=O)C(=O)CC	0	O=C1OC2=C(C=CC=C2)C=C1	0	c1(ccccc1)C(=O)c2ccccc2	0	C(CCCCCCCC)S	0
O=C1OC2=C(C=CC=C2)C=C1	0	c1(ccccc1)OC(=O)C	0	c1(ccccc1)OC(=O)C	0	C(CCCC)OC(=O)CCCCC	0	C1(=O)CCCCC1	0	O=C1OC2=C(C=CC=C2)C=C1	0	C(=O)(CCCCCCCC)O	0
C1=CC=C2C(=C1)C=C(NO2)C=O	0.06933906	C#CCOC(=O)CC1=CC=CC=C1	0.09948513	CCC(=O)OC1=CC=CC=C1	0.3793798	C(=O)(CCCCCCCCC)O	0	C1=CC=C2C(=C1)NC(=O)C=N2	0.04626341	C1=CC=C2C(=C1)NC(=O)C=N2	0.04197753	COC(=O)CCNC1=CC=CC=C1	0.04581804
CC1=NC(=O)C2=CC=CC=C2N1	0.08035855	CCOC(=O)CC1=CC=CC=C1	0.1235873	CCC(=O)OC1=CC=CC=C1	0.3793798	C(=O)(CCCCCCCC)O	0	CC1=NC2=CN=CN2C=C1	0.06964932	C1=CC=C2C(=C1)NC(=N2)C=O	0.09240541	C=CC(=O)CCCCCCCC(=O)O	0.04753727
CC1NC(=O)C2=CC=CC=C2O1	0.08366296	CCCOC(=O)CC1=CC=CC=C1	0.1371426	C1=CC=C2C(=C1)C=CC(=O)N2	0.384888	CCCCCCCCNC(=O)C	0.1277119	C1=CC2=C(N=CN2C=C1)C=O	0.07254368	C1=CC2=C(NN=C2C=C1)C=O	0.0990186	C#CCCCCCC=CC(=O)O	0.04908247
CN1C=NC2=CC=CC=C2C1=O	0.09183173	C=CC(=O)OCC1=CC=CC=C1	0.160449	C1=CC=C(C=C1)OC(=O)N	0.4700359	CCCCCCC(=O)OCCCC	0.1383745	CC1=CC2=C(C=C1)C=C(S2)N	0.07338813	C1=CC=C2C(=C1)C(=CC2=O)O	0.10483	CCOC1=CC=C(C=C1)OC(=O)NC	0.05223202
COC(=N)C1=CC=CC=C1	0.1088847	COC(=O)CC1=CC=CC=C1	0.2023979	CCC(=O)C1=CC=CC=C1	0.4719471	C(CCCCN)CCCC(=O)O	0.1551848	C1CC(=CC=C1)C=C=O	0.08079227	CC(=O)NC1=CC=CC(=C1)C(=N)N	0.1129747	CCCCC(CC)COCC(=O)O	0.06334015
CNCC(=O)C1=CC=CC=C1	0.1134785	CCC(=O)OCC1=CC=CC=C1	0.2040253	C1C(=O)C=CC2=CC=CC=C21	0.4828623	CC1(CCCSC1)O	0.1810965	CC1CC=NC=N1	0.08086793	C1=CC2=COC(=C2C=C1)C=O	0.1146544	C1CCC(C1)C(=O)CCCCC(=O)O	0.06629472
CC1COC2=CC=CC=C2C1=O	0.1147292	C1=CC=C(C=C1)CCCOC=O	0.2240541	CC(=O)NC1=CC=CC=C1	0.4926594	CCCCCCCCNC(=O)N	0.201845	C1=CSC(=C1)C2=NN(C=C2)O	0.08123408	C1=CC=C2C(=C1)NC(=O)O2	0.11715	CCCCNC1=CC=C(C=C1)OC	0.06633087
C1=CC=C(C=C1)C(=O)CCN	0.1167073	C=C=CC(=O)OCC1=CC=CC=C1	0.228408	C1C=CC2=CC=CC(=O)C2=C1	0.5147033	CCCCCCOOC(=O)C	0.2354623	C1CNC(C=C1)C=O	0.08133689	C1C2=CC=CC=C2ON1C=O	0.1209336	CCCCNC(C)CC(=O)OC	0.0667401
C1C2=CC=CC=C2ON=C1C=O	0.1201505	CCC(=O)OC1=CC=CC=C1	0.2532228	C1=CC=C(C=C1)NOC=O	0.534839	CCCCCCCC(=O)OCCC	0.2373345	CC1=NC2=C(C=C1)NC=N2	0.08243389	C1=CC=C2C(=C1)NC(=O)NN2	0.1401864	CNC(=O)COC1=CC=C(C=C1)C#N	0.07077628
C=CN1C=NC2=CC=CC=C2C1=O	0.124919	CC=C=CC(=O)OCC1=CC=CC=C1	0.2709531	C1=CC2=CC=CC(=O)N2C=C1	0.5353036	CCCCCC(=O)OOCCCC	0.2398072	CNC1=CC=NC=C1	0.08274814	C1=COC(=C1)C2=CC=C(C=C2)O	0.1428006	CCCCCCCCC(=O)C(=O)O	0.07088963
CC1=NC2=CC=CC=C2C(=O)N1C	0.1258814	C=CC(=O)OC1=CC=CC=C1	0.2744787	C1C=C2C=CC=CC2OC1=O	0.5594665	CCCCCCCCCC(=O)N	0.2537211	CC1=NC(=NC=C1C#N)C	0.08415395	C=C1C2=CC=CC=C2OC1=O	0.149781	CCCCCCCCCOC(=O)O	0.07169935
C1=CC=C2C(=C1)C(=O)N=CC=N2	0.1260327	CC(=O)OCCCC1=CC=CC=C1	0.2756416	C=C1C(=O)OC2=CC=CC=C2N1	0.572254	C(CCCC(=O)O)CCCN	0.256625	CC1=CCC(=O)CC1	0.08472221	C1=CC(=CC=C1C(=O)O)NCCC#N	0.1550444	CCCCCC1CCC(CC1)C(=O)O	0.07283703
CN1NC(=O)C2=CC=CC=C2O1	0.1322961	C=CCC(=O)OCC1=CC=CC=C1	0.2810861	C1=CC=C(C=C1)ON=C=O	0.5755313	C(C)OC(=O)CCCCCCCC	0.2626695	C1=COC(=C1)C2=NN=NC=C2	0.08486507	CN=CC1=CC=C(C=C1)C(=O)OC	0.1553915	CCNCCC1=CC2=C(C=C1)OCO2	0.07466881
CNNC(=O)C1=CC=CC=C1	0.1338931	C#CC(=O)OCC1=CC=CC=C1	0.2843047	CC(=O)C=C1C=CCC=C1	0.5820786	C(CCCCC)OC(=O)CCCC	0.2654262	CC1=CC2=NC=NN2C=C1	0.08532597	C1=CC=C2C(=C1)C=COC2=O	0.1566442	CC(=O)CCC(=O)NC1=CCCCC1	0.07735112
CC1=CC(=O)C2=CC=CC=C2O1	0.1355292	C1=CC=C(C=C1)CCCCOC=O	0.2902787	C1=CC=C(C=C1)N(C=O)N	0.5914498	C1=CC=C(C=C1)CCCCCCN	0.2666345	CC(=O)OCSC	0.08674062	C=CC(=O)NC1=CC=C(C=C1)C(=N)N	0.1586167	CCCC1CCC(CC1)CCC(=O)O	0.07825345
CC1=NC2=CC=CC=C2C(=O)N1N	0.1374166	CCCC(=O)OC1=CC=CC=C1	0.299426	C1=CC=C(C=C1)NC(C=O)O	0.5925996	CC(=O)NCCCCCCN	0.2891708	C1COC2=C1C=C(C=C2)C(=O)O	0.08754144	C1=CC=C2C(=C1)C=NN2C=O	0.1601506	COC(=O)C=CCNC1CCCCC1	0.0817986
C1=CC=C2C(=C1)C=CC(=O)N2	0.1397979	CC#CCOC(=O)CC1=CC=CC=C1	0.3071297	CC1=CC=C(C=C1)C(C)O	0.5991708	CCCCCCCOC(=O)CCC	0.2895618	C1=CC=C2C(=C1)C=CN2N=O	0.08813593	CN=NNC1=CC=C(C=C1)C(=O)OC	0.1621242	CCOC(=O)CCNC(=O)CC=C	0.08188413
CC(=O)NN=CC1=CC=CC=C1	0.1408284	CCCCC(=O)OC1=CC=CC=C1	0.3087009	C1=CC=C(C=C1)ONC=O	0.6023895	CC1(NCCCN1)S	0.2910104	C1=C2N=C(N=CN2N=C1)C#N	0.08876163	C1=CC=C2C(=C1)C(=CO2)N=O	0.1633878	CNCC1=CC=C(C=C1)N2CCCC2	0.0845592
C1=CC=C2C(=C1)C=C(C(=O)O2)C#N	0.1409447	COC(=O)CCC1=CC=CC=C1	0.3129654	C1=CC=C(C=C1)OC(=O)NS	0.6065903	CCCCCCCCOC(=O)N	0.2941937	C1=CC2=C(N=C1)SC(=C2)C=O	0.088843	COC1=CC=C(C=C1)NC(=O)C=C	0.1651134	CCCCCCOC(C)CS	0.08794286
CCOC(=N)C1=CC=CC=C1	0.1460253	COC(=O)CCCCC1=CC=CC=C1	0.3463383	c1(ccccc1)NC=O	0.6157112	CCCCCCCCOC(=C)C	0.2947368	C1CC2=C(C=C1)NNC(=O)C2	0.08899985	CC(=O)NC1=CC=C(C=C1)N(C)C	0.1651899	CCCCCC(=O)OC(C)S	0.08888077
CN1C(=O)C2=CC=CC=C2N=N1	0.1465203	CCCCOC(=O)CC1=CC=CC=C1	0.3527429	C1=CC=C(C=C1)NC(=O)N	0.620347	C1=CC=C2C(=C1)NC(=O)O2	0.2981597	C1CSN(N1)COC(=O)N	0.0890455	CCC(=O)C1=CC=C(C=C1)OC	0.1667638	CCCCCCOC(=O)CC(=O)O	0.08920102
C=NNC(=O)C1=CC=CC=C1	0.1469559	CC(=O)OC=CC1=CC=CC=C1	0.3589161	CCC(C1=CC=C(C=C1)N)O	0.6529591	CCCCCCNCC(=O)O	0.2988469	C1CC(SC1)CO	0.09011572	C1=CC=C2C(=C1)C=C(C=N2)O	0.1670003	CCNCC1=CC=C(C=C1)C(=O)OC	0.08954888
C1NC(=O)C2=CC=CC=C2S1	0.1485208	COC(=O)CCCC1=CC=CC=C1	0.3614484	CC1=CC=CC=C1C(C)O	0.659726	C1CCC(CC1)(O)S	0.3023335	CC1=NN(CS1)C	0.09045361	CCNC1=CC=C(C=C1)N(O)O	0.1673194	CCCCOC1=CN=C(C=C1)C(=O)O	0.08968592
C1=CC=C2C(=C1)C(=O)SN2	0.1501398	C1=CC=C(C=C1)C=CCOC=O	0.3668489	C1=CC=C(C=C1)OC#N	0.668126	C1=CC=C(C=C1)CCCCCNO	0.3041483	C1CC2=C(C1)SC(=N2)N	0.09100104	CCC(=O)C1=CC=C(C=C1)NO	0.1704537	COC(=O)CNCCC1=CC=CC=C1	0.09056786
C1C(C(=O)C2=CC=CC=C2O1)N	0.1512307	C=COC(=O)CCC1=CC=CC=C1	0.3705043	CC(C1=CC=C(C=C1)N)O	0.670643	CC(=O)CCCCOC(=O)C	0.3058697	C1C=CC2=CN=NC2=C1	0.09216601	C1=CC=C2C(=C1)OS2(=O)=O	0.1718171	CCCCCCCC(=O)NCC	0.09156643
CN1C(=O)C2=CC=CC=C2N=C1N	0.1514078	C1=CC=C(C=C1)CCCCCOC=O	0.3737975	C=CC1=CN=C(C=C1)C=O	0.6741813	C1=CC=C2C(=C1)NC(=O)N2	0.3085566	CC1=NC2=C(O1)C=CN=N2	0.09277338	C1C2=CC=CC=C2S1(=O)=O	0.1760533	CCCCCCNCN1C=NC=N1	0.09158454
C1=CC=C2C(=C1)C=C(NO2)O	0.1543114	CC(=O)OCC=CC1=CC=CC=C1	0.3767563	C1=CC=C(C=C1)C(=O)N	0.6754651	C1=CC2=CNON2C=C1	0.3135061	C1CN(C1)C2=NC=CO2	0.0931637	CCNC1=CC=C(C=C1)C(=O)OCC	0.1825772	CCCCOC1=NC=C(C=C1)C(=O)O	0.09173586
C1C(=CC2=CC=CC=C2O1)C=O	0.155389	CC(=C)C(=O)OC1=CC=CC=C1	0.3781045	CSC(=O)C1=CC=NC=C1	0.679814	CCCCCCCOC(=O)N	0.3146737	C1=CC2=C(N=C1)N=C(N2)C=O	0.09397485	C1=CC=C2C(=C1)C=CS2(=O)=O	0.1861117	CC1=CC=C(C=C1)C=NNC(=O)C	0.09274995
C1=CC(=CC=C1C(=O)CS)S	0.1595061	CC(=O)OC#CC1=CC=CC=C1	0.3811758	CC1=C(C=CC(=C1)C=C)C=O	0.6824018	CCC1=NN=C(S1)NC	0.3184107	CC#CC1=C(C=NC=C1)N	0.09415977	C1=CC=C2C(=C1)N=CS2(=O)=O	0.1953836	CC1=CC=C(C=C1)NC(=O)COC	0.09366909
CNC1=CC(=O)C2=CC=CC=C2N1	0.1596583	COCC(=O)OC1=CC=CC=C1	0.3818025	CC(C1=CC=CC=C1N)O	0.6834855	CCCCCCOC1=CC=CC=C1	0.319478	C1CN2C=CC(=O)N2C1	0.09510532	COC1=CC=C(C=C1)C(=O)CN	0.1961028	C=CC1=CC=C(C=C1)CCCC(=O)O	0.09456447
CC1=NNC2=CC=CC=C2C1=O	0.1603431	C=CC(=O)OCCCC1=CC=CC=C1	0.3857962	C1=CC=C(C=C1)OC=O	0.6870004	CNCCCCCCCC(=O)O	0.3250441	CC1=NC2=C(O1)N=NC=C2	0.09528534	CC1=CC=C(C=C1)NC2=NON=C2C	0.1965068	CCCCC1=CN=C(C=C1)C(=O)NC	0.0951854
C1=CC=C2C(=C1)C(=O)OC=N2	0.1609765	CCC(=O)OCCCC1=CC=CC=C1	0.4011092	C1=CC=C(C=C1)NOC=N	0.6919384	CCCCCCCCNC(=O)OC	0.3351572	C1C=NOC1CC=O	0.0955993	C1=CC=C2C(=C1)C(=O)OC=N2	0.1998679	CC1=CC=C(O1)C=NNC(=O)OC	0.09543358
CC1CC(=O)C2=CC=CC=C2O1	0.1618874	CCCCCC(=O)OC1=CC=CC=C1	0.4036705	CC(NC1=CC=CC=C1)(O)O	0.6920565	C#CC1=C(C(=CC=C1)C#C)N	0.3361486	CC1CC=COC1	0.09607712	CC1=CC=C(C=C1)N2CCC2=O	0.2033951	CCCCCCCOC(=O)O	0.0954759
C1=CC=C2C(=C1)C(=O)C=C(O2)N	0.1627933	C1=CC=C(C=C1)COC=O	0.4068996	C1=CC=C(C=C1)NC(=O)C=O	0.6959268	CCCCCCCC(O)(O)O	0.3384916	C1C2=CC=CC=C2NN1C(=O)O	0.09783618	C=NC(=O)C1=CC=C(C=C1)C(=O)N	0.2043851	CCOC(=O)NN=CC1=CC=C(O1)C	0.09705195
CCCCOC(=N)C1=CN=CC=C1	0.1628804	CC(C)OC(=O)CC1=CC=CC=C1	0.412378	CN(C1=CN=CC=C1)N=O	0.7001146	COCCOC1=CC=CC=C1	0.3389186	C1=CSC(=C1)C2=NNN=C2	0.09784775	C1=CC=C2C(=C1)N=CN2N=O	0.2051726	CCCCCCOOOC(=O)O	0.09713763
CC1=COC2=CC=CC=C2C1=O	0.1637411	CC(=O)OCCCCC1=CC=CC=C1	0.4139083	C=CC(=O)OC1=CN=CC=C1	0.706339	CCCCCCCC(=O)COCC	0.3397872	C1=CC2=C(C=C1O)C=C(S2)C(=O)O	0.09796812	COC1=CC=C(C=C1)NC(=O)N	0.2064523	CCOC1=CC=C(C=C1)CCNC	0.09923997
C1C(=O)C2=CC=CC=C2ON1	0.1681853	COC(=O)CCC#CC1=CC=CC=C1	0.4179807	C1=CC=C(C=C1)C(=O)S	0.7083084	CCCCCCCCCNC(=O)NC	0.3415637	CC1=NC2=NC=NC=C2N=C1	0.09842545	CCOC(=O)C1=CC=C(C=C1)N(C)C	0.2291434	CC1=CC=C(C=C1)NN=CC(=O)OC	0.1005384
C1=CC=C2C(=C1)C=C(C(=O)O2)N	0.1686474	COC(=O)COCC1=CC=CC=C1	0.4257721	C1=CC=C(C=C1)C(=O)NN	0.7111502	CCCC(=O)CCC(=O)OC	0.3432483	C1=CN(N=N1)CC(=O)N	0.09886127	C1=CC2=C(SC(=C2C=C1)O)O	0.2299909	CCCC(=O)NC1=CN(N=C1)COC	0.1007118
C1=CC(=CC=C1C=O)S	0.1688499	COC(=O)COC1=CC=CC=C1	0.4340718	C1=CC=C(C=C1)C(=O)C#N	0.7128538	CNC1=NN=C(S1)NC	0.3545416	C1CNCC(=C1)C=O	0.09949614	C1=CC(=CC(=C1)N)C2=NC(=O)NN2	0.2319916	CCCCCCC1COCCN1	0.1009531
C1=CC=C(C=C1)C=NNC=O	0.1700087	CCC(=O)OCC=CC1=CC=CC=C1	0.441583	C1=CC=C(C=C1)CC(N)O	0.7133358	CCCCCOC(=O)NCCCC	0.360758	C1C=CC2=C1C=C(C=C2)C(=O)O	0.09987804	CN(C)C1=CC=C(C=C1)NC(=O)N	0.2394643	COC1=CC=C(C=C1)NC2=CC=CC=C2	0.1010144
C1=CC=C2C(=C1)C(=CC(=O)O2)N	0.1704098	CCOC(=O)CCC1=CC=CC=C1	0.4427735	C1=CC=C(C=C1)S(=O)N	0.7146281	CCCCCCCOC(=O)C	0.3633213	CC1=CC2=C(C=C1)C=NC=C2	0.1003145	COC(=O)C1=CC=C(C=C1)NN=C	0.2401632	CC1=CC=C(C=C1)NCC2=CC=NC=C2	0.1020273
C1=CC=C2C(=C1)C(=O)N=C(N2)NC#N	0.1726924	C=COC(=O)C=CC1=CC=CC=C1	0.4457966	C1=CC=C(C=C1)NC(=O)NS	0.716907	CC(C)(C)OOCCCCC#C	0.364702	C1CS(=O)CCO1	0.1008242	C1=CC=C2C(=C1)C=C(S2)O	0.2412083	CNCC1=CC=C(C=C1)C2=CC=CC=N2	0.102463
C1=CC=C2C(=C1)C(=O)N=CN2N	0.173784	COC(=O)CC=CC1=CC=CC=C1	0.4491596	C=C1C2=CC=CC=C2ONC1=O	0.7190709	CCCCCCCCOC(=O)NC	0.3713083	CC1=NC2=C(C=C1)NN=C2	0.101283	C1C(=O)C=CC2=CC=CC=C21	0.2420575	CCCCCCCCC(=O)NC	0.1030238
CC1=NC2=CC=CC=C2C(=O)O1	0.1760244	CCCC(=O)OCC1=CC=CC=C1	0.4493704	C1ON(O1)C2=CC=CC=C2	0.7222355	CCCCCCCCOC(=O)C	0.3728931	C1=CC2=CC(=NN2C=C1)C(=O)O	0.1014415	C1=CC=C2C(=C1)C=CN2C=O	0.2441588	CCCCCCC1CC(=CC(=O)C1)O	0.1042897
CCC1COC2=CC=CC=C2C1=O	0.1763525	c1(ccccc1)C(C)OC=O	0.4531518	C=C1CC=CC=C1N=C=O	0.725016	C#CCCCCCCC(=O)O	0.374313	CC1=CC2=C(C=C1C)N=C(N2)N	0.1014849	C1=CC=C2C(=C1)N=CN2C=O	0.2570274	CCCCCC1=CC=C(C(=S)C1)O	0.1057732
C1=CC=C2C(=C1)C(=O)C=NS2	0.1770091	C=CC(=O)OCC=CC1=CC=CC=C1	0.453392	COC1=CC=C(CC1)C(=O)N	0.7262696	CCCCCCCCC1=CC=CO1	0.3753883	C1CCOC(=N)C1	0.1015078	C1=CC=C2C(=C1)C=NC(=O)N2	0.2599685	CCCCOC1=CN=C(N=C1)C(=O)O	0.1075921
C=C1COC2=CC=CC=C2C1=O	0.1777917	CCCC(=O)OCOC1=CC=CC=C1	0.4594144	CCCC(=O)C1=CC=CC=C1	0.7266735	C1CNC2=C1SC=C2	0.3809527	CC1CN1C(=O)SC	0.1016095	CC1=CC=C(C=C1)C2=NOC(=N2)C	0.2603975	CCOC1=CNC2=C1C=CC(=C2)OC	0.1079465
C1COC(=N1)C2=CN=CC=C2	0.1792888	CC(=C)OC(=O)CC1=CC=CC=C1	0.4606335	C=CC(=O)C1=CN=CC=C1	0.727193	CC(C)(CNC1=CC=CC=C1)N	0.3838605	C1CC(OC=C1)C=O	0.1016398	CC=CC(=O)NC1=CC=C(C=C1)OC	0.2610168	CCCNCC1CCN(CC1)C	0.1080846
C=CCOC(=N)C1=CC=CC=C1	0.1793112	CC(C)COC(=O)CC1=CC=CC=C1	0.4610775	C1=CC=C(C=C1)CC(NO)O	0.7281083	CCCCCCCCC(=O)N	0.3880282	C1=CC=C(C=C1)SCO	0.1019654	C=C1C=CC(=O)C2=CC=CC=C12	0.2620487	CCCCCCCN(C)C(=O)O	0.1095993
C1=CNN(N=C1)C2=CC=C(C=C2)C=O	0.1794657	COC(=O)C=CC1=CC=CC=C1	0.4628345	C1=CC=C(C=C1)C(=O)SN	0.7379573	CCCCCCOOC(C)(C)C	0.3887468	CCS(=O)(=O)NC1CC1	0.1035476	CC1=NC=C(C2=CC=CC=C12)O	0.2637201	C=CCCCCC(=O)CCCC(=O)O	0.1100669
C1=CC=C2C(=C1)C(=O)C=CN2	0.1796827	CC(C)C(=O)OC1=CC=CC=C1	0.4660403	CC(=CC=O)C1=CC=CC=C1	0.7390813	C(CCN)CCNCC(=O)O	0.3888208	C1C=CC(O1)OC=O	0.1043858	C1=CC(=CC=C1NCC(=O)NC=O)O	0.2638623	CC(=O)CCCCCCCS	0.1104009
CC1=CC(=O)C2=CC=CC=C2N1	0.1798026	C=COC(=O)COC1=CC=CC=C1	0.4667059	CC(=C)C1=CC=C(C=C1)N=C=O	0.7400931	C1=CC=C(C=C1)OCCNN	0.3898445	CN1CCCC(=C1)C=O	0.1044181	CC1=NC(=CO1)C2=CC(=CC=C2)N	0.2643008	CNCC1=NOC(=C1)C2=CN(N=C2)C	0.110401
CN1C(C2=CC=CC=C2N=N1)O	0.179857	C#CC(=O)OCCCC1=CC=CC=C1	0.468857	C1=CC=C(C=C1)NC(=O)S	0.7418715	C1CNC2=CC(=C(C=C21)O)O	0.3946607	CC1=CC2=C(C=C1)OC=C2	0.1044385	C1=CC(=CC(=C1)NC(=O)CCS)N	0.266331	CC1=CC=C(C=C1)C=CC(=O)NC	0.1106139
C1=CC=C2C(=C1)C(=O)C(=CO2)N	0.1820228	c1(ccccc1)COC(=O)C=CC	0.471485	CC1=CC=C(C=C1)C(=O)C	0.7424576	C1CCC(CC1)(N)S	0.3964639	CC1CC=CS1	0.1048361	COC1=CC=C(C=C1)C=NNC(=O)N	0.2674834	CC1=CC=C(C=C1)NCC(=O)OC	0.1107641
CON1C2=CC=CC=C2C1=O	0.1825544	CCOC(=O)CCCC1=CC=CC=C1	0.4735791	CCC(=O)NC1=CC=CC=C1	0.7432415	CCCC1=COC(=N1)C2=CC=CC=C2	0.4047688	CC1CCC(=O)CO1	0.1051196	C1=CC(=CC=C1C=O)C2=NN=CO2	0.273599	CCCCCCN1CCCNC1=O	0.1116488
C1=CC=C(C=C1)C=CNC=O	0.1826441	CN(C)CC(=O)OC1=CC=CC=C1	0.4854457	CC(=O)NOC1=CC=CC=C1	0.752393	CC1=CC=C(C=C1)NCC#N	0.4052417	CNC(=O)N1CC1C#N	0.105348	C1=CC=C2C(=C1)C(=O)NNC2=O	0.2753972	CCOC(=O)NC(=O)CCCC=C	0.111656
CCC(=O)C1=CC=CC=C1	0.1836267	CC(=CC(=O)OC1=CC=CC=C1)C	0.4868151	C1=CC=C(C=C1)OC(=O)NN	0.7559961	CC=CCC1=CC=C(C=C1)O	0.405968	COC1=CC2=NNN=C2C=C1	0.105717	COC1=CC=C(C=C1)C(=O)N	0.2774421	CCC1=CC=C(C=C1)C=NNC(=O)C	0.1128842
CN1C=NC(=O)C2=CC=CC=C21	0.1849264	c1(ccccc1)COC(=O)CCCC	0.4880825	C1CC1(C2=CC=CC=C2)O	0.7599564	CCC1=CC=C(C=C1)OC(=O)CC	0.4075889	C#CCSCC(=O)N	0.1058726	CC1=CC=C(C=C1)N2CCNC2=O	0.2778931	CCCCCOCCC(C)S	0.112892
CCCOC(=N)C1=CN=CC=C1	0.1852108	CCOC(=O)COCC1=CC=CC=C1	0.491219	C1=CC=C(C=C1)OC=N	0.764867	C1CCN(CC1)C2=CNC=C2	0.4078831	C1=CC=C2C(=C1)N=CN2N=O	0.1058831	C1=CC=C2C(=C1)C=CC(=O)N2	0.2794787	CCCCCCOC(CC)CS	0.1143021
CNC(=O)C1=CC=C(C=C1)S	0.186622	CCCOC(=O)CCC1=CC=CC=C1	0.4976125	C1=CC=C(C=C1)NCN=O	0.7665449	CCOC(=O)OCCCCCN	0.4085535	C1(=O)C=CCCC1	0.1063838	CCC(=O)C1=CC(=CC=C1)OC	0.2808203	CCCCCCCCCC(=O)NC	0.114313
CNC(=O)C1=CC=CC=C1	0.1868345	CC(=O)OCC#CC1=CC=CC=C1	0.4982229	C1=CC=C(C=C1)NC(C=O)N	0.7680675	CCCCCCCCC(=O)NCO	0.4212588	C1C2=CC=CC=C2N=C1OO	0.1065172	CC1=CC=CC(=C1)C2=NC(=O)NN2	0.282005	C=CCOCC1=CC=C(C=C1)C(=O)O	0.1151577
CCCC(=O)C1=CN=CC=C1	0.1878621	CC(=O)OCNCC1=CC=CC=C1	0.5059386	CS(=O)(=O)C1=CN=CC=C1	0.775816	COC(=O)C(=O)OC	0.4240297	C1CC2=C(C=CN2C1)C=O	0.1070297	C1C2=CC=CC=C2S(=O)(=O)O1	0.2878304	C1=CC(=CC=C1CCCC#N)C(=O)O	0.1158386
COC1=C2C(=CN=N1)C=CC=N2	0.188291	C1=CC=C(C=C1)OCCOC=O	0.5067189	C1=CC=C(C=C1)C(=O)SO	0.7768097	CCOC1=NC(=NC(=N1)N)N	0.4248055	C=C1CCC(=O)C=C1	0.1071082	C=C1C2=CC=CC=C2NC1=O	0.2883823	C1C=C1CCCCCCC(=O)O	0.1166089
C1=CC=C2C(=C1)C(=O)C=C(N2)NN	0.1885667	C1=CC=C(C=C1)CCOC=O	0.5094009	C#CC(=O)NC1=CC=CC=C1	0.7769156	CCC(=O)CCCCCCC(=O)CC	0.4252747	C1=CN=C2C(=N1)C=C(S2)C(=O)O	0.1071924	C1=CC=C2C(=C1)C=C(N2)O	0.2911803	C=CCNC1=CC=C(C=C1)C#N	0.1167684
CC(C(=O)C1=CC=CC=C1)N	0.1887906	C=C(O)OCC1=CC=CC=C1	0.5139242	CC(=O)C1=CC=C(C=C1)N	0.7775055	CN=C(CCCCCCC(=NC)N)N	0.4256652	C1=COC(=C1)C(=O)CC#N	0.1072729	C1=CC=C2C(=O)C=CC(=O)C2=C1	0.2951977	CC1=CC(=CC(=C1)OCCCNC)C	0.1169137
C1=CC=C2C(=C1)C(=O)N=NS2	0.1907101	CC(=O)OCCCOC1=CC=CC=C1	0.514507	C1=CC=C(C=C1)N=CNO	0.78395	CC#CCCC1(OCCO1)C	0.4264494	CC1=NC2=CN=NC=C2N=C1	0.1072834	C1=CC(=CC=C1C=O)C2=NC=NO2	0.3006473	CN(C)CCCNCC1=CC=CO1	0.1186123
C1=CC=C2C(=C1)C(=O)C=CO2	0.1923703	CCOC(=O)C=COC1=CC=CC=C1	0.5176604	CN(C1=CC=NC=C1)N=O	0.7847648	COC1=C(C(=CC=C1)OC)SC	0.4275318	CC(C=O)C1CC(=O)N1	0.1073354	C1=CC=C2C(=O)C=CC(=O)C2=C1	0.3007127	C1=CC=C(C=C1)NC2=NC=C(C=N2)C#N	0.1212142
C1C(=O)C2=CC=CC=C2OS1	0.1925518	C=C(COC1=CC=CC=C1)O	0.5204479	C=C1CCC2=CC=CC=C2C1=O	0.7882237	CCC1=CSC(=N1)CC	0.4276709	C1CNC(CN1)C#N	0.1076403	C1=CC(=CC=C1C(=O)S)C(=O)S	0.3034644	CCCC1=CC(=CC=C1)NC(=O)C=C	0.1214704
CN(C)C1=CC=C(C=C1)C=O	0.1930319	COC(=O)C#CC1=CC=CC=C1	0.5225184	C1=CC=C(C=C1)N(C(=O)N)N	0.7885484	CC(=O)OCC1=CC=C(C=C1)N	0.4281004	CNC1=NN(C(=O)C=C1)C	0.1077243	CC1=CC=C(C=C1)C2CC(=O)C2	0.3036508	CCOC(=O)CC1=CC=C(C=C1)NC	0.1216725
COC1=CC=C(C=C1)C=NC#N	0.1938174	CCOC(=O)COC1=CC=CC=C1	0.524365	CC(O)SC1=CC=CC=C1	0.7890455	CCCCCCCNC(=O)OCC	0.4287692	CSCCC(=O)O	0.107799	CC(=O)NN=CC1=CC=C(C=C1)OC	0.3040234	C1=CC=C(C=C1)C(=O)CCCC(=O)O	0.1234908
CC1=CC2=CC=CC=C2OC1=O	0.1944149	C1=CC=C(C=C1)CCNCC=O	0.529635	C1=CC=C(C=C1)ONC(=O)N	0.7902555	CCCNC(=O)CCCCCCN	0.4297029	CC(=O)NN1CC1	0.1078811	C1=CC=C2C(=C1)C(=NN2)NC=O	0.3040574	CCCCCCC1=CC=C(C=C1)O	0.1238225
C1=CC=C(C=C1)C=NCC=O	0.1947007	CC(C)CC(=O)OC1=CC=CC=C1	0.5340142	C1=CC=C(C=C1)NCOC=O	0.7942325	CC1=C(C=CC=C1C#N)C#N	0.4309587	C1N=CSN1CC=O	0.1080445	C1=CC=C2C(=C1)C(=CN2)C=O	0.3041203	CCC(CCC1=CC=C(C=C1)O)C#N	0.1249956
CC1=NC(=O)CC(=N1)C2=CN=CC=C2	0.1948448	CC(O)OCC1=CC=CC=C1	0.537694	C1=CC=C(C=C1)C2=CC(=O)C2=O	0.7950116	CC(=O)OCCCCCCNC	0.4316256	CCC(=O)C1=NC=NC=C1	0.1080824	C1C2=CC=CC=C2CN1C=O	0.305162	CCCOC1=CC=C(C=C1)C=NO	0.1269258
C1=CC=C(C(=C1)C2=CO2)O	0.195021	c1(ccccc1)COC(=O)CC=CCC	0.5378829	C1=CC=C(C=C1)NC=C=O	0.7952095	CCCCCCCCOC(=O)CC	0.4323462	C1=CC2=C(C=CC(=C2)S)N=C1	0.1081034	C1=CC=C2C(=C1)C(=CN2)NN=O	0.3061267	CC(C)CCC(=O)C1=CC=C(C=C1)O	0.1297072
COC1=CC=C(C=C1)C2=NNCN=N2	0.195902	COCC(=O)OCC1=CC=CC=C1	0.5390166	c1(ccccc1)C(=NO)C	0.7958007	CCCCCCCCNN=C(C)C	0.4329099	C1=CN(N=C1C#N)CO	0.1083373	C1=CC(=CC=C1COC#N)COC#N	0.3071959	CC(=O)C=C(C=CC1=CC=CC=C1)O	0.1299016
C1=CC=C2C(=C1)C(=O)C=C(N2)N	0.1975684	CCOC(=O)C=CC1=CC=CC=C1	0.5399668	C1=CC=C(C=C1)CC(NN)O	0.7958035	C1CNNN(C1)CC2=CC=CC=C2	0.43297	CN1CCN2C=NC=C2C1	0.1084303	C=CC(=O)NC1=CC=CC(=C1)N	0.3123869	CCCCCCOC(C)OC(=O)O	0.1308236
C=C(C1=CC=CC=C1)O	0.1999709	CNCCNCC1=CC=CC=C1	0.5415792	CSC(=O)C1=CN=CC=C1	0.7958813	COC1=CC=C(C=C1)CCC#C	0.4353019	CSNC(=O)N1CC1	0.1084479	C1C=CC(C=C1)C2=CC=CC=C2O	0.3130324	CCNCCOCCOCC	0.1319032
C1CC1C(=O)NN=CC2=CN=CC=C2	0.2017375	COCCOC(=O)CC1=CC=CC=C1	0.5419474	C1=CC=C(C=C1)N(C=S)O	0.797922	CCCCCCOC(=O)NCCC	0.4356453	C1=CN2C=C(N=CC2=C1)C(=O)O	0.1084723	C1=CC(=CC=C1N)N2C=CNC2=O	0.3138424	C1=CC=C(C=C1)NC2=CC=C(C=C2)C#N	0.1326711
CN1C2=CC=CC=C2C1=O	0.2018451	CCOCC(=O)CC1=CC=CC=C1	0.5442348	C1=CC=C(C=C1)CC=NO	0.7983707	C1N2C=CC=CN2C(=O)O1	0.4356482	C1=CC=NC(=C1)N2C=NN=C2	0.1086097	CC(C1=CC=C(C=C1)C(C)OC)OC	0.3148852	CC(=O)NC1=CC=C(C=C1)OC(=O)C	0.1342538
CCSC(=O)C1=CN=CC=C1	0.2028423	CC(=O)CNCC1=CC=CC=C1	0.5453544	C1=CC=C(C=C1)C(=O)CC=S	0.7988777	CCCCCCCCCC(=O)C=C=C	0.4372975	CC1=CC2=C(C=C1)OCO2	0.1087269	C1=CC(=CC=C1C(=O)O)N=N	0.3160575	CC(CCCCOCCC#N)C(=O)O	0.1348869
C1CC(=O)OC2=CC=CC=C21	0.2043426	CCCOC(=O)C=CC1=CC=CC=C1	0.5453824	CC(C1=CC=C(C=C1)O)O	0.7990128	C=CCN(CC=C)N=O	0.438614	C1CC(C1)CC(=O)N	0.1087848	C1=CC=C2C(=C1)OC(=O)O2	0.3162928	CCCCCOCC(CC)S	0.1349091
C1C=CN2N1NC2C3=CC=C(C=C3)O	0.2044652	COC(=O)CNCCC1=CC=CC=C1	0.5456976	C1=CC=C(C=C1)CC(=O)NN	0.7992844	C1=CC(=O)NC=CC1=O	0.4392266	C1CC1NCC(=O)N	0.1091177	C1C2=CC=CC=C2COO1	0.3197115	CC1=CC=C(C=C1)C2=CC(=C(C=C2)C)O	0.1351912
C1C(=O)COC2=CC=CC=C21	0.2061299	COCC(=O)CC1=CC=CC=C1	0.5494509	C1CC(C1)(C2=CC=CC=C2)O	0.7995076	CCCCCCCCN1CNC=C1	0.440694	C1CONC1C=O	0.1091268	C1C=C(NC2=CC=CC=C21)C=O	0.3234246	CNC(=O)CCCCCCCC=C	0.1351973
C1=CC=C2C(=C1)C(=O)N=C(S2)N	0.2066889	C#CCNNCCC1=CC=CC=C1	0.5496824	C1=CC2=C(C3C(O3)C=C2)N=C1	0.8015428	C(CCC(=O)OCCCN)CCN	0.4415777	CC(C#N)OCC=C	0.1092127	C=CC(=O)NC1=CC=C(C=C1)O	0.3263573	CCCCCCCCOC(=O)O	0.1352406
C=CC1COC2=CC=CC=C2C1=O	0.2080846	C=CCN(CC1=CC=CC=C1)O	0.5506925	C1=CC=C(C=C1)C(=O)C=O	0.8015852	CCCCCCOC(=O)OOCC	0.4417036	COC1CCC(=O)O1	0.1097492	CC(C)C1=CC=C(C=C1)NC(=O)C	0.3264308	CC=CC=NNC(=O)C1=CC=CS1	0.1361757
C1=CC=C2C(=C1)C(=O)N=CN2	0.2087961	CN(C)N=NCC1=CC=CC=C1	0.5549933	CNC(=O)C1=C(C=C(C=C1)N)N	0.8021905	CC(=O)OCCC(C)(C)O	0.4420935	C1CC=C(C=C1)C(=O)N	0.1097619	C1=CC2=CN(N=C2C=C1)C=O	0.3301505	CCCCCC1CCC(=NO)CC1	0.1368982
C1C2=C(O2)C3=CC=CC=C3O1	0.2098703	C1=CC=C(C=C1)CCCN=C=O	0.5563128	C=C1C(=O)CCC2=CC=CC=C12	0.8025518	CC(=N)CCCCCC(=N)C	0.4421728	CC1=NC(=C2C=NN=C2)NN1	0.1103146	C1=CC=C2C(=C1)C=C(C=N2)C=O	0.3326617	CCC(CC)NCCOC(=O)C=C	0.1369147
C1=CC(=CC=C1C(=O)N)S	0.2099183	COC(=O)CNCC1=CC=CC=C1	0.5566175	C1=CC=C(C=C1)C=NO	0.803635	C(CCCNN)CCC(=O)N	0.4453138	CC1=NC2=NC=NC=C2N1	0.1103637	C1=CC=C2C=NC(=CC2=C1)C=O	0.3337109	CCCCCC(=O)CCCC(=O)O	0.1369289
COC1=CN=NC2=CC=CC=C21	0.2099267	C1=CC=C(C=C1)CCCCOO	0.5581519	C1=CC=C(C=C1)OC#CO	0.8060032	COCOCC1=CC=CC=C1	0.4466413	CC(=O)C1CSCN1	0.1105087	C1=CC=C2C(=C1)NC(=CN=O)O2	0.3351018	CCC1=CC=C(C=C1)NC(=O)C(C)C	0.138562
COC1=CC=CC(=C1)C=C=C	0.2121813	COC(=O)CCNCC1=CC=CC=C1	0.5596073	C1=CC=C(C=C1)C(=N)SO	0.8067357	CCC1=CC=C(C=C1)CC(=O)C	0.4470504	C1C=CC=C2C1=C3C(=C2)NON3	0.1107127	C1=CC=C2C(=C1)C(=O)NC2=O	0.3367244	COC1=CC=C(C=C1)CCCC(=O)O	0.1394413
C1C(C2=CC=CC=C2OO1)N	0.212798	C1=CC=C(C=C1)COCC=O	0.5598583	CC1=CC=C(C=C1)OC(=O)C	0.8078429	CCCCCCCCCN=C(N)N	0.4470558	CC1=C(CC(=O)CC1)C	0.110907	C1=CC=C2C(=C1)C=CC(=N2)C=O	0.3400559	CCOC1=CC=C(C=C1)NCCC#N	0.1395212
CN(C)C1=CC=C(C=C1)C(=O)C#N	0.2130659	CC(=O)CCOC1=CC=CC=C1	0.5607005	C(=C)(C)c1ccc(cc1)C=O	0.8079962	CNC(=O)C(=N)OC	0.4480569	CC1CC2=CNC=C2OO1	0.1109538	CC(C1=CC=C(C=C1)C(C)O)O	0.3422283	CC(=CC1=CC=C(C=C1)CC(=O)O)C	0.1397337
CCCC(=O)C1=CC=CC=C1	0.2133856	C#CC(=O)COCC1=CC=CC=C1	0.5624856	CC(C1=CC=C(C=C1)C=O)O	0.8080923	CCCCCCCOCC(=O)C	0.4523235	CN1C2=C(C=C(C=C2)S)C=N1	0.1110351	CC1=CC(=CC=C1)N2C(=O)CC=N2	0.3442423	CCCC1=CC=C(C=C1)OCC(=O)O	0.1407029
C1=CC=C2C(=C1)C(=O)NS2	0.2141586	CCOC(=O)C(C)C1=CC=CC=C1	0.5667265	CC(C1=CC=CC=C1)N=C=O	0.8091252	CC(C)(CCCC(=O)N)O	0.4527769	CC1=NN(C=C1C#N)C	0.1110403	C1=CC=C2C(=C1)C=C(O2)C=O	0.3453929	C1=CC=C(C=C1)OCCCCC(=O)O	0.141031

Table 3.16 Continued

Mor261-1		Mor268-1		Mor271-1		Mor272-1		Mor273-1		Mor277-1		Mor30-1	
SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance	SMILES	Distance
C(CCCCCCCC)O	0	C(CCCCCCC)O	0	CC(=O)CCCC	0	CC(=O)CCCCC	0	C[C@]12CC[C@H](C1)C(C2=O)(C)C	0	C[C@@]12CC[C@@H](C1(C)C)CC2=O	0	C(=O)CCCCCCCC	0
C(CCCCCCC)O	0	C/C(CC/C=C(C)\C)=C\CO	0	CC(=O)CCC	0	CCC(=O)CCCC	0	C[C@@]12CC[C@@H](C1)C(C2=O)(C)C	0	C[C@@]12CC[C@H](C1(C)C)CC2=O	0	C(=O)(CCCCCCCCC)O	0
CCCCCC(CCO)O	0.1725635	C(CCCCCCCC)S	0	CCC(=O)CCCC	0	c1(ccccc1)COC(=O)C	0	CC1=C(CCCC)C(CC1)=O	0	O=C1CCOC2=C1C=CC=C2	0	C(=O)(CCCCCCCC)O	0
CC(=O)SCC1=CC=C(C=C1)C=C	0.2079287	C=C(CCO)CC1=CC=CC=C1	0.08680266	CCC(=O)C(=O)CC	0	CC1=C(CCCC)C(CC1)=O	0	CC(C)C(OCC)=O	0	C1(=O)CCCCC1	0	C(=O)(CCCCCCCCCC)O	0.3122994
CCCCC(CCCCO)O	0.2205779	C1=CC=C(C(=C1)CCCCO)S	0.09776434	c1(ccccc1)COC(=O)C	0	CC(C)C(OCC)=O	0	CC1(C2CCC(C2)(C1=O)N)C	0.05501643	CC1(C2CC(=O)C1(C=C2)C)C	0.1480439	C#CCCCCCCC(=O)O	0.3287775
CCNC1=CN(N=C1)C2=CC=CC=C2	0.2643034	C1=CC=C(C=C1)CC2=C(OC=C2)CO	0.1173957	CC(OC/C=C(C)\C)=O	0	CC(OC/C=C(C)\C)=O	0	CCC1(C2CCC(C2)C1=C)CC	0.0736611	CC1C2(CCC1(C(=O)C2)C)C	0.1501646	C=CCCCCCCCC(=O)O	0.3395462
CCCCCCCCNN	0.2799805	CC1=CC=CC=C1C2=CC(=NO2)CO	0.1177446	CC(=O)CCC(=O)C	0.00696499	CC1CCC(=O)C1CCC#C	0.03189391	CC1CC2=CC=CC=C2N1	0.07383829	C1CC(CC(=O)C1)N	0.173522	CC#CCCCCCC(=O)O	0.3545879
CC1=CC=C(C=C1)NCC(=O)NN	0.2822771	C=CCC1=CC=CC=C1C#CCO	0.12333	CC(=C(C)N=NC)C	0.02227481	CC1=C(C/C=C\C)C(CC1)=O	0.0385644	CCCCC1=C(OC=C1C)C	0.07413368	C1C(C2=CC(=O)C(=O)CC2=N1)O	0.1843483	C(=O)CCCCCCCCC	0.3841893
CC1=CC=C(C=C1)N2C=C(N=C2)CO	0.2832535	CCCC1=CC(=C(C=C1)CN)N	0.124973	CC=C(C)C(=O)OC	0.02303129	CC1COC(O1)C(C)(C)C	0.03930837	CC1CNC2=CC=CC=C12	0.07595554	CC1CCCC(=O)C1	0.1871075	C=CCCCCCCC(=O)O	0.4070303
CCCOC(=O)C1=C(C=C(C=C1)N)O	0.2852718	CC1=CC=CC=C1CCCCO	0.1258311	CCC(=C)C(=O)OC	0.03826216	CC(C)(C)C1OCCCO1	0.04139548	CC(C)C(=O)OC=C	0.0776473	C1CC(=O)C2CC=CC1S2	0.2020606	C(=O)CCCCCCCCCC	0.4113624
CCCCN1N=C2C=CC=CC2=N1	0.2984866	C1=CC=C(C(=C1)CCCCN)N	0.132615	CN(C)C(=O)CC=C	0.03853879	CCCCCC(=O)C=C	0.04630508	CCC(C)NC	0.08056587	CCC12CCC(C1(C)C)CC2=O	0.2177276	CCCC#CCCCC(=O)O	0.4176973
CCCCNC(=O)OC1=CC=CC=C1	0.3016491	CCCCC1=CC=CC=C1CCN	0.1390437	COC(=C)C(=C)OC	0.05227276	CCC(=C)COC(=O)C	0.04645563	CCSNC(C)C	0.08292255	C1CNC2=C(C1=O)C=CC=N2	0.2429412	CCCCCC#CCC(=O)O	0.4183262
CCCC(=O)CCCCCO	0.314819	C1=CC=C(C=C1)N2C=C(C(=N2)N)CO	0.1412993	CC(C=C)C(=O)OC	0.05905649	CC1=NC(OC1)C(C)C	0.04785742	CCCCC1=C(CCCC1=O)O	0.08494866	CC1COCC(C1=O)C(C)(C)C	0.2469828	CC#CCCCCCCCC=O	0.4361534
CCCCCOC1=CCCC=C1	0.323052	CCCCCCN(C)C(=S)N	0.1478017	C1CCC(CC1)C2C=CCOO2	0.05928864	CCCC(=O)CCC	0.04868388	CCCC(C)OC	0.08782356	C1CC(=O)C2=C(NC1)N=CC=C2	0.2562849	CCCCCCCCCCCC=O	0.4486903
CCCCSC1=CC=C(C=C1)N	0.3336667	ClCCCCCCCCC	0.1517669	CCCC1SCCS1	0.059703	CC(=O)CCCCCC	0.05098426	CC12CCCC1=CC=CC2	0.08803756	C1CC(=O)C2=CC=CC=C2OC1	0.2580454	C#CC#CCCCCC(=O)O	0.4608185
CCOC(=O)CC1=CC(=C(C=C1)N)O	0.3378758	C=C=CCC1=CC=CC=C1CCO	0.1577928	CCOCC(=O)C(=C)C	0.06042718	CCCCCCC(=O)C	0.05098426	CC1CC1C2=CC=CC=C2	0.08885662	CC(C)C12CCC(C1)(CC2=O)C	0.2605051	CC(C)CCCCCC=O	0.4667896
CC1=CN=C(C=C1)NCCC(=O)O	0.3419532	CN1C(=CC(=N1)C2=CC=CC=C2)CO	0.1577991	COC(=C(C#N)C#N)C1=CC=CO1	0.06070587	CCC1=NC(CO1)(C)C	0.05112795	CC1CCC(S1)C	0.08906697	CC1(C2(CCC1(C(=O)C2)C)C)C	0.2605627	CCC#CCCCC(=O)O	0.4774187
CCOC1=CC=C(C=C1)NC(=O)NN	0.349067	C1=CC=C(C=C1)NNC(=O)CN	0.1627573	CCCC1(C(O1)(C)C)C(=O)OCC	0.06170662	CC1(C(=O)C(S1)(C)C)C	0.05313001	CN(C)C(=O)COC	0.08911168	C1C2CC3C=CC2CC1C3=O	0.2618416	CC(=C)CCCCCC(=O)O	0.478635
CCCCC=CC=CCO	0.3594656	C1=CC=C(C(=C1)CCCCO)N	0.1671111	CCN1C(=C(C(=C(C1=O)C)C)C)C	0.06559493	C(C)OC(=O)C(=C)C	0.05386116	CC1C2=CCC(C2)C1(C)C	0.08918327	CCC1(C(=O)CC12CC2)CC	0.2665882	CCCCCC1CC(=O)O1	0.4816979
CCCCCCC(CCO)O	0.3622535	CC1N=C(NN1C2=CC=CC=C2)CO	0.1706577	C=CC(=O)OC1CCCCCC1	0.06979654	CCCC1C=C(C(=O)O1)C	0.05455699	CC1CC=C(C1)C(=O)C(C)C	0.09082242	C1CC(CC(=O)C1)S	0.2680335	C=CCCCCCC=O	0.4847203
CCOC(=N)CCCCC#N	0.3642063	CCCCCCC1(CNC1=O)C	0.1736143	CCN=C1N2CCCC2CCS1	0.07527722	CC(C)(C)C1OCC(=C)O1	0.05557511	CC1CC2=C(C(=C(C12)C)C)C	0.09087068	CC1CCCC12CCC2=O	0.268307	C#CCCCCCC=O	0.4893661
CCCOC(=O)SC1=CC=CCC1	0.3687816	C1CCC(=CCCCO)CC1	0.1749776	CC(C)CC(=O)SC	0.07652444	C=C1CC(CO1)C2=CC=CC=C2	0.05582976	CC1CC=CC12CCCC2=O	0.09092754	CC1CCCC(=O)C1(C)C(C)C	0.2696565	C=CCCCCCCCCC=O	0.4941443
CCCOC1=CC=C(C=C1)CC(=N)N	0.3695899	CC(=CCCC(=CCN)C)C	0.1790916	CC(C(=O)C=C)OC	0.07826366	CC(C)(C)C(=O)N1C=CC=C1	0.05717043	CC1CCCC=C1C(=O)N(C)C	0.09201093	CC1CCCC(=O)C1(C)C	0.2719524	C1C=C1CCCCCC(=O)O	0.4958194
CCCCCSC1CCCC1	0.3784109	C1=CC=C(C(=C1)C#CCCO)N	0.1844991	CSC1CCCCO1	0.0787251	CC1=NCCN1CC2=CC=CC=C2	0.05790857	CC1=C2C(=NC=NC2=NN1)N	0.09262628	CC1(CC(=O)CC2(C1O2)C)C	0.2734493	C=CCCCC1CC(=O)O1	0.4999771
CCOC1=CC(=NO1)C2=CC=CC=C2	0.3805914	CC(=C)CCCCCCO	0.1861627	CC#CC(=O)N(C)C	0.0789155	CC(C)C(=O)OC=C	0.05836228	CC1=CC(C(CC1)C(C)(C)O)O	0.09308545	C1C=CNCC1=O	0.2802712	C(=O)(CCCCCCC=C)O	0.5103754
CC1=CC=C(S1)CNC2CCCC2	0.3893983	C1=CC=C(C(=C1)CCCCN)S	0.192109	C=CCC(CC1CCCCC1)C#N	0.07892005	CC(=C)C(=O)OC(=C)C	0.05855966	CN1C=NC2=C1C(=NC=N2)N	0.0936681	CC(C)C1CC(=O)C=CN1	0.2822056	C1C=C1CCCCCCC(=O)O	0.5167787
CC1=CN(N=C1)CCCC(=O)O	0.3914392	C(CCCCCC=C)O	0.1949365	COC(=O)C(=C1CCCCC1)C#N	0.079376	CC(=O)CCCCC=C	0.05864676	CC(=NOC(=O)C(C)(C)C)C	0.09476478	C1(=O)C=CCCC1	0.2823245	CC(C)CCCCCCCC=O	0.5233348
CCCCOC1=CC=C(C=C1)NC	0.400401	CC1=C(C=C(C=C1)CCCN)C	0.1985303	C=C=CCN1C(COC1=O)C#C	0.08053451	CC(=O)OC1CCCC=C1	0.05871441	CC(C)C1CCCCC1=O	0.09521714	CC(C)(C)C1CC(=O)C=CN1	0.2876126	C(=O)CCC=CCCCC	0.5329278
CC(CCOC1=CC=C(C=C1)N)O	0.4045965	C1CCC(C1)(NCC2=CC=CC=C2)O	0.2063455	CCC(=O)C1C(=O)CC(CC1=O)(C)C	0.08254048	CCOC(=O)C1(CC1)C	0.06032142	CC1=C(C(=O)N(C1=O)NC)C	0.09626298	CC1CC(=O)C=CN1	0.2899973	CCCCCCC(=O)C=O	0.5376703
CC(CCCCCC=O)(CO)C=O	0.4069427	CCCC1=CC2=C(C=C1)NC=C2CO	0.2131039	CC1C=CN(C=CC1=C)C(=O)C	0.08378967	CC(=CC)COC(=O)C	0.06154408	C1CC1=CC2=CC=CC=C2	0.09676587	CC1CCC(=O)N1C	0.2921599	CC(C)CCCCCC(=O)O	0.5519592
CCCCCC(CCCCO)O	0.413368	CC1=CC=C(C=C1)N2C=CC=C2CO	0.2132813	CC1=CC=CC(=C1C#N)C(=O)OC	0.08382356	CCC(C)COC(=O)C	0.06223368	CC(=CC1=CC(=O)OCC1)C	0.09695268	C1CCC(=O)C(C1)N	0.2935391	CCCCCCCCC(=O)N	0.5530371
CCCOC1CC(C1)C2CC(=O)C2	0.415854	CCC1=CC=C(C=C1)C(CN)N	0.2154896	CC1CC(N(C1)C(=O)C)C(=O)C	0.08455622	CC(=C)CCOC(=O)C	0.06278437	CCCC(=C)OC	0.0971561	CC1(CC(=O)C12CCCC2)C	0.2936078	CCCC=C=CCCC=O	0.5570654
CCCOC(=O)C1=CC=C(C=C1)N	0.4177516	COC1=CC=CC(=C1)CCCO	0.2155236	CC1=C(C=C(C=C1)C(=O)N(C)C)C	0.08463229	C1=CC=C(C=C1)N2C=CC=CC2=O	0.06322297	CC1(C(=C)C2CCC1(C2)N)C	0.09770459	CC12CCCNC1CCCC2=O	0.2944516	CCCCCCCC(=O)C=O	0.5580794
CC1=CC=C(C=C1)COC2=CC=CC=N2	0.4196408	CC1=CC=CC=C1C2=CC(=NO2)CN	0.2175655	CC1=C(C(=O)CCC1)C(=O)OC	0.08483176	CC(C)(C)OC1CCCO1	0.06380388	CC1CC1(C)CN	0.09781662	CC1CCC(C(=O)C1C)C	0.2966442	C#CCCCCCCCC=O	0.5587924
CC(=C)COC1=CC=C(C=C1)CC#N	0.4198459	C1=CC=C(C=C1)N2C=C(OC2=O)CN	0.2179772	CC(C)CC(=O)C(C)C(=O)C(C)(C)C	0.08727654	CCCC(C)C1=CCCC1=O	0.06429856	CC1CC(SC1)C	0.09835188	C1CC2=C(C=CC=N2)C3C1O3	0.300252	CCCCCC=CCC(=O)O	0.5771456
CCCC(=O)SCC(=O)NC	0.4250377	C=CCCCC1=CC=CC=C1CO	0.2221567	CC(=O)CC1CCC=C1C(=O)C	0.08730956	CN(C)C(=O)C1CCCC1	0.06497417	CN1C2=C(C=CC=C2O)N=N1	0.09842831	C1CC(=CC(=O)C1)O	0.3008896	CCCCCCCC(=O)C=C	0.5798101
CC1=CC=C(C=C1)OCCCC(=O)O	0.4253378	C1=CC=C(C(=C1)CN2C=CC=C2C#N)N	0.2266102	CC(=O)C1CC(CC1C=C)C=C	0.08757486	CC1=COC(=N1)C23CCN(C2)CC3	0.06522684	CCCCC1=C(C(CC1=O)O)C	0.09856076	CC1C=CC(C(=O)C1C)C	0.3021769	CCCCCC#CCC(=O)O	0.5906563
CC1=CC=C(C=C1)CNC2=NC=CN2	0.4326364	CCCCCCN(C)C(=S)S	0.2280024	CCC(=O)C(C1CCCC1)C(=O)C	0.08885495	CC=CCCC(=O)C	0.06550376	CCC(C)COC	0.1010865	CC1(C2CCC1(C(=O)C2)CO)C	0.3036347	C=CCCCCCCC(=O)S	0.6068316
CCCCC(=O)CC(=O)COC	0.436914	C1CC(C1)(CCCCCN)N	0.2290869	CCC(=C=C)C(=O)C	0.08887637	CN1CCN=C1CC2=CC=CC=C2	0.06601764	CC(=O)C1C(=O)C(OC1=O)(C)C	0.101607	CC12CC3C1(C(=O)CC2C3)C	0.3038942	C(=O)CCCCCC=CC=CC	0.6092983
CC1=C(C=CC(=C1)NCCC(=O)O)N	0.4417228	C1=CC=C(C=C1)N2C(=C(N=N2)CO)N	0.2329444	CC(C)C#COC(=O)C	0.08966827	CCSC(=O)C(=C)C	0.06620419	C1CC(C=C1)C2=CC=CC=C2	0.1021592	C1CCC2C=CC(C1)C2=O	0.3052857	C#CCCCCCCCCC=O	0.6107675
CCC(CCCCCN=C=O)N=C=O	0.4484982	CCCC1=CC(=CC=C1)CCN	0.2341136	CC1=C(N(C(=O)C(=C1)SC)C)C	0.09027337	CC1C(=O)OC2(O1)CCCCC2	0.06675006	CC1CCC(=O)C1CCC#C	0.102564	CC12CCCCC1CCCC2=O	0.3052975	CCCCCCCCC(=O)C=O	0.6194306
CC1CN(C1=O)C2=CC=C(C=C2)OC	0.4526726	CCCCNC(=S)NN(C)C	0.2404278	CCC(C)CC(=O)C=C	0.0908688	CC1=CC2=C(C=C1)C3(C(O2)(O3)C)C	0.06725323	CC1(C(=O)C(C(=O)O1)(C)O)C	0.102655	CC(C)C12CCC=CC1CC2=O	0.3066741	CCCCC(=O)OOC=O	0.6209503
CC1=CC(=O)OC(=C1)CCCCO	0.4534946	CN(CC1=CNC2=CC=CC=C21)N=O	0.2405676	CC(C1CCCCC1)OC(=O)C	0.09215821	CC#CCC1CCCCC1=O	0.06775508	CC1=NC2=NC(=C(N2C=C1)C)C	0.1028573	CC1(C2(O1)CCCCC2=O)C	0.3074888	CCCCCCCC(=O)C(=O)O	0.6224536
CCCCCOC(=O)C1CCC1	0.4572937	C1CN(C(=C1C#N)N)C2=CC=CC=C2	0.2421394	C=CN1CN(CN(C1)C=C)C=C	0.09251169	COC1CC1C2=CC=CC=C2	0.06775833	CC1=C(C2(CC2(C1=O)C)C)C	0.1036801	CC(C)(C)N1CNCCC1=O	0.3105013	CCCCCCC(=O)C(=O)O	0.6228972
CCNNC(=O)NC1=CC=CC=C1	0.4648017	CCC1=C(C=CC(=C1)OC)CCN	0.2432191	CC1=CC=CC=C1C(=O)C(C)C	0.09255168	CCSC(=O)C(C)C	0.06797524	CCC1(C(C2CCC1C2)C)C	0.1039269	C1CC(=O)C2=C1C=CC(=C2)N	0.3113474	C(=O)(CCCCCCCCCC)O	0.6309238
C(C=CC=CCC=CCC)O	0.4658642	CCCCN(C1=CC=C(C=C1)C)N	0.2447534	CC(C)(C)OC1COCCO1	0.09261868	CCCCC1=C(OC=C1C)C	0.06908714	CCCN=NC(C)C	0.1048958	CC1(C2C(=O)C(CC1C2)C)C	0.3153669	CCCCC(=O)OC=O	0.6320382
CCCCCCNC(=O)N(C)C	0.4659477	CCCCCC1=CC=CC=C1CN	0.2476232	CC(C)(C)C1=CC=CC=C1OC#N	0.09276885	CC1(OC1CC=C(C)C=C)C	0.06958327	CC1=CC=NC2=NC=C(N12)C	0.1049399	C1CC=NC2=CC=CC=C2C1	0.316439	C=CC(=O)CCCCCCCC	0.6328618
CCOC(=O)NN=C1CCCC1	0.4663815	C1=CC=C(C=C1)C2=CC(=NO2)CO	0.2495896	CSC(=O)C1=CC=CCC1=O	0.0932461	CC1(C=CC2=CC=CC=C2O1)OC	0.07000891	CC1CCCC(C1C(C)C)O	0.1061147	CC1CC=CC(=O)C1	0.3171871	C#CCCCCCCCCC=O	0.6425371
C1CC1CNC(=O)CCC(=O)N	0.4669312	CC1(CCC(=CC1)CCCO)C	0.2515845	CC(C)C1CCCC(C1)N=C=O	0.09368631	CC(=O)OC1CCCCC1	0.07050657	CCC(C)(CC1=CN=CC=C1)O	0.106278	CC1(C2(CCC1(CC2=O)N)C)C	0.3199595	CC#CCCCCCCC(=O)O	0.6461503
CC(=O)NCC(=O)CCCC(=O)O	0.4687083	C1=CC=C(C=C1)CCN(C(=S)N)N	0.2522303	CC1(CC(=CC(=NC)C1)OC)C	0.09407415	CC(=C)CC(=O)N(C)C	0.07099795	CC1C(=O)OC(=O)C1(C)C	0.1067386	C1C(C2=C(C1=O)C=NC=C2)O	0.3215719	CCCCC(C)CCC=O	0.6513144
CCOC1=CC=C(C=C1)C(=O)C2CC2	0.4703171	CCCCCCNC(=S)NC	0.2531695	CCC(=C)CC(=O)CC	0.09492467	CCCC1CCC(C(=O)O1)C	0.07149991	C1C=CCC1C2=CC=CC=C2	0.1076653	CC1CC(=O)CC1N	0.3224006	C#CCCCCCCCCC(=O)O	0.6543915
CC(=O)NCC1=CC=C(C=C1)NOO	0.4711151	CCCCCCCN=NC	0.2544345	CC1=NC(=NC(=N1)CC(C)C)C	0.09494454	CC(C)(C#CC=C=C)OC	0.07224971	CCC1=C(C(=O)C(=C(O1)C)O)C	0.1078519	C1C=NNC2C1C=CC=C2	0.32246	CCCCCCC(=O)OS	0.6560432
CCCC(COC(=O)C=C)O	0.4713456	CC(C)NCC(C1=CC=CC=C1)O	0.2556588	CC(C)CN1CCCC1C(=O)OC	0.09520044	CCCC(=O)C1C=CC=C1	0.07282247	CC1CCC(C(=O)C1)C(=O)C	0.108035	CCCC12CCC(C1(C)C)CC2=O	0.3242648	CCC(C)CC(=O)OCNN=O	0.6562133
CC1=C(CCCC1SCC#C)C=C	0.4736828	CNC(=O)COC1=CC=CC=C1	0.2557549	CC(C)C1C(=O)C=CC(O1)OC	0.09530087	CC(CCCC#C)OC	0.07284866	CNC(=O)CC1=CC=CC=C1O	0.1081257	CC1(CCC(=O)C12CC2)O	0.3244342	CCCCCCC(=O)SC	0.6671683
CC1=CC(=NO1)NC(=O)CCN	0.4738877	CC(CO)C#CC1=CC=CC=C1	0.2569379	CCC(=O)C1C(=O)CC(CC1=O)C	0.0953478	CC(=C)C(=O)OC1CC1	0.073144	CC1=CCC2=C1C(=NC=N2)N	0.1087274	C1CC2=C(C(=O)C1)OC=C2	0.324941	CCCCCC(C)CC(=O)O	0.6676361
CCCCN=NNC1=CC=CC=C1	0.4748378	C#CC#CCCCCO	0.2574479	CSN1CCCCC1	0.0954107	CCCCC=CC(=O)C	0.07463951	CC1CCNC2=CSC=C12	0.1087613	CC12CCCCC1(NCCC2=O)C	0.3299895	CCC=CCCCC=O	0.6731271
CCCCCCCCC=CCO	0.4758962	CCCC1=CC=C(C=C1)CCCN	0.2579055	CC1=C(C(=O)OC(=C1)C=C(C)C)C	0.09559824	CCCCC1CCOC1C	0.07467752	CC1C(=O)C(C1(C)C)(C)OC	0.1088047	CC12CCC(C1)CCC2=O	0.3306673	CCCCCC(=O)C(=O)O	0.6741228
CCOC(=O)SC1=CC=C(C=C1)O	0.4778154	CC12CCC(CC1)(CC2)CCCN	0.2582053	CC1=CC(=O)N(C(=O)N1C)CC=C	0.09594982	CCC1N(O1)C(C)C	0.07472067	CC1C2CCC1(C(=O)C2(C)C)C	0.1088199	CN1CC=CCCC1=O	0.331401	CCCCCCCC1=COC=N1	0.6761256
CCOC1=NNC(=C1)C2=CC=CC=C2	0.4782514	C[C]1(CC/C=C(C)/C)=C(O1)CO	0.2619843	CCOC(C1=CC=CO1)OCC	0.09596565	CCC(C)C(=C)C#COC	0.07474929	C=C1CCSCC1	0.1089672	C1CNC=CC1=O	0.3316735	CCCC#CCCC=O	0.6762332
CCOC1=NOC(=C1)C2=CC=CC=C2	0.4799803	CCOC(=O)NC(C)C1=CC=CC=C1	0.2631698	CC1=CC(=CC=C1)CN(C)C	0.0959908	CCCCC1=C(C(=O)CC1)C	0.07480496	CC1(C2(C1(CCC2)CN)C)C	0.1091173	CC12CCCCC1(C(=O)CCC2)C	0.3363524	CCCCC1CC1CC=O	0.6780394
CC(=O)OCCCCC1C(O1)C=O	0.4819859	CC1=CC=C(C=C1)N2CCC(=C2N)C#N	0.2652138	CC(=C)CCC(=O)C(C)(C)CC=C	0.0963259	CN(C1=CC=CC=C1)C(=O)C=C	0.07483113	CC1=CCC(C1)C(=O)C(C)C	0.1094183	C1COC2=CC=CC=C2N=C1	0.3382365	CCCCCCC(=O)N(C)N=O	0.6823976
CCCCC(=O)NOCC=C	0.4849413	COC(=C)CC#CCCO	0.2653182	CC(=C1CCCCC1)C(=O)OC	0.09635255	CCCCCC(C)(OC)OC	0.0750319	CCCCC1=C(NC=NC1=O)C	0.1101226	CCC1(CCCCC1=O)C	0.3392663	C=CCCCCCCC=O	0.6839698
CCC1=CC=C(C=C1)OCCC=O	0.4937379	CN(CC1=CC=CC=C1)C(=N)SC	0.2675158	CC(=O)OC1CC=CCCC=C1	0.09654689	CC1=CC(OC1=O)CC=C	0.07526437	CC1=CN=CN2C1=NN=C2C	0.1102247	CC1(CCOCCC1=O)C	0.3437433	CCCCCC#CC(C)CC=O	0.6918118
CCCCCC1CC(C1)CO	0.4974671	CC1=CC(=C(C=C1CCN)C)OC	0.2679668	CC1CCC(CC1N2CO2)C(C)C	0.0969064	COC1(CCC1C2=CC=CC=C2)OC	0.0756035	CCCCC1=C(SC(=N1)N)C	0.1112301	CC(C)C1CC=CC(=O)C1	0.3515443	CCCCCCCC(=O)N	0.6936034
CCCCCCCCC(=C=C)CN	0.4981083	CCCCCCNC(=O)S	0.2681372	CC1CC(C(=O)C=C1)OC(=O)C	0.09691212	CC1CN(CO1)C2CCCCC2	0.07568156	C1=CC=NC(=C1)CC(C=O)C#N	0.1112544	CN1CCCCCC1=O	0.3534443	CCCCCCCCC(=O)C#C	0.6950354
CC#CCN=CC1=CC=CO1	0.4984045	CC1=C(C=C(C=C1)CO)OCOC	0.2683241	CC1=C(C(=CC=C1)C(C)(C)C)OC	0.09705795	CCC(=O)SC(C)C	0.07589343	CC1C=CC2=C1N=CC=C2	0.1112787	CC1(CC(=O)C2C1CC(=O)CS2)C	0.3553887	CCCCCC(=O)N(C)N=O	0.6951424
CC(CO)NNC1=NC(=NC(=N1)N)N	0.5005156	CCNC(=O)NC1=CC=CC(=C1)C	0.2705592	CC1=C(C=C(C=C1)CC#N)OC	0.09736501	CC(=O)OC1CC2CCN1CC2	0.0761043	CCSC1=CC(=O)NC(=N1)N	0.111704	C1CC=CNC(=O)C1	0.3559682	CCCCCCCC(=O)NCO	0.7057065
CCOC(=N)CCCCCC#N	0.5011074	CCC1=CC2=C(C=C1)NC(=C2)CO	0.2735424	CC(C)C1=CC=CC(=O)CC1=O	0.0974126	COC1=NC2=CC=CC=C2CC1	0.07620976	CC1CCSCC1	0.1119254	CC1CC2=C(C=CS2)C(=O)N1	0.356523	CC(=CCCCCC=O)C	0.7099205
CC1=CC(=NO1)CN=CN(C)C	0.5020328	CNC(=S)N(C)CC1=CC=CC=C1	0.2742432	CCC(=O)OC1CCC2C1CCC2	0.0976064	CC1=CC=C(C=C1)N(C)C(=O)C	0.07641043	CC1=CC(=C(CO1)C(=O)C)N	0.1119393	CCC12CCC=CC1CC2=O	0.3569258	CCCCCCC(=O)S	0.7155588
CCCCC(=O)CCCCCO	0.5021533	CCCCNC(=O)N1C=CNC1=S	0.2748742	CC1=CC=CC(=C1C)C(=O)N(C)C	0.09771011	CC1(CCC(CC1)CC#N)C	0.07655777	CCCCC1=C(C=CCC1)O	0.1121559	C1CCC(=O)C2(CC1)CC2	0.3574388	CCCCCCCCC(=O)OC=C	0.7175887
CCC(=O)OCC1=CN=C(C=N1)OC	0.5045965	CC(C)(CCNC1=CC=CC=C1)O	0.2776271	CCN1C(=CC(=C(C1=O)C#N)C)C	0.09777768	COC(CC1CCCCC1)OC	0.07689791	CCCCC1=C(C(=O)CC1)C	0.112561	CC(C)(C)N1CCCC(=O)N1	0.3597735	CC(=CCCCCCC=O)C	0.7217996
CC(CCCOC(=O)C=C)O	0.5050363	C=C=C(CCC1=CC=CC=C1)CN	0.2781998	CCC1=CC(=CC=C1)N(C)C#N	0.09792107	CCN(CC1=CC=CC=C1)C(=O)C	0.07696524	CN1C=NC2=C1C(=NC=C2)N	0.1126043	C1COC=CC1=O	0.3609559	CCCCCCCC(=O)C#N	0.721917
CCCCCC(=O)CCCCO	0.5062308	CNC1=C(C=CC(=C1)CCN)OC	0.2783838	COC(=O)C1=CC=CN1CC#C	0.09852679	CC1=CC=C(C=C1)C2CCC2=O	0.07752899	CC1=COC(=C(C1=O)C)CO	0.1126406	C1CCC(=O)C=CC1	0.3616321	C(CCC=O)CCC(C=O)N	0.7255953
CCNC(=O)NC(=O)C1=CC=CC=C1	0.5086512	CC(C)(COC)NC1=CC=CC=C1	0.2785638	CC(C)(C)C1=C(C(=O)C(=O)C=C1)OC	0.09882807	CCCCC1CCOO1	0.0776566	CCCCC1=CCC(C1=O)C	0.113008	C1CSC2=NC(=O)CC(=O)N21	0.3617356	C=CCCCCC(=O)C(=O)O	0.7375877
CCCCCCC(CO)O	0.5106431	CC1=CC=C(C=C1)CC2(CCCO2)C	0.278616	CCC(C)(C)OOC(=O)C(C)(C)C	0.09896188	CCCCC(=C)C(=C)OC	0.07798966	CC(C)C(C)OC	0.1133053	CC1(CCCC(=O)C1(C)C)C	0.3619967	CCCCCC1(CC(=O)O1)C	0.7379237
CC(=O)OCCCC1=CC=C(O1)C=O	0.5116236	CCCCCSSC(C)(C)C	0.2798165	CC(=O)CN(C)C	0.0992457	CC1CN=C(O1)C(=C)C	0.0780852	CC(C)NC(C)S	0.1134074	C1CNCC(=O)C2=C1N=CC=C2	0.3658213	CCC(CCCCCC(=O)O)N	0.7410716
CCC1CCCCC1OCCCC#N	0.5138626	CC1=C(C=C(C=C1)OCC(=S)N)C	0.2804527	CC1=CC=CC=C1C(=O)C(C)C#N	0.09929142	CC1(CC1(OC)OC)C2=CC=CC=C2	0.07844056	CC1(CCC(CC1=O)(C)N=O)C	0.1136064	CCC1(CCCCC1=O)CC	0.3676169	CCCC#CC(C)CCC=O	0.7413049
CC1=CC(=CC=C1)NC(=O)OCC#C	0.5145964	CCCCCCNN(C)C	0.2813479	COCC#CC(=O)C1=CC=CO1	0.09933654	CCCCC1=COC=C1	0.07852358	CC1=C(C(=O)NC1(C)O)C(=O)OC	0.1136162	CC1CCC(=O)C1(C)C	0.3676319	CC(CCCCCC=O)O	0.7442524
COC(=O)C1=NN=C(O1)NCCC=C	0.5150404	CC1=C(C=CC=C1OCOC)CO	0.2836483	CC1=C(C(=O)CN(C1)C(C)(C)C)C	0.09970819	CC1=CCC(=CC1)CCOC	0.07865082	CCCCC1(CCCC1=O)C	0.1137967	CC(C)(C)C1CCCCC(=O)N1	0.3678083	CC(CCCCCC(C)C=O)C=O	0.7452166
CCCCCCN1CNC=C1	0.5158175	CC1=C(C(=CC=C1)C)CCNC(=O)N	0.2857546	CC(=O)OOC1=CC=CC=C1	0.1005747	CCC1=NC(=NO1)C	0.078747	CC1=NC(=O)C(=C(N)NO)C=C1	0.1141668	CCC12CCC(C1)CCC2=O	0.3681195	CC(=CCCCCCCC=O)C	0.7480642
CC(C#CC#CC#CCCCCO)O	0.5169611	CNC(=O)CN(C)C1=CC=CC=C1	0.2865872	CC(=NN(C)C)C1=CC=CC=C1	0.1009896	COCCCCCC#C	0.07875913	CC1=CC(C(CC1)C(C)C)O	0.1145154	CCN1C(CC1=O)C	0.3695459	CC#CCCCCCCC=O	0.7485604
CC1=CN=C(C=C1)NCCCC(=O)O	0.5184499	CNC(=O)N(C)CC1=CC=CC=C1	0.2870481	CC(C)(C)C(=O)C1=CCCCO1	0.1011319	C1CC=C(C1)N2CCCC2=O	0.07899711	C1=CC=NC(=C1)C2C=CC=C2	0.1146218	C1CC(CC(=O)C1)O	0.3700499	C(=O)CCCC=CCCCC	0.7490338
CC1CC2=C(O1)C=CC(=C2)CCCN	0.5207977	CN(CC1=CC(=CC=C1)OC)N=O	0.2872887	Cc1cc(=O)oc(c1)CC(C)C	0.1012257	CC(C)(C)C(=O)C1=CC=CO1	0.0790988	CC1=CN=C2C=CC=NC2=N1	0.1149353	C1C2=CC=CC=C2OSC1=O	0.373351	CCCCCCCCC(=O)NC	0.749357
C1CCC(CC1)CCCCON	0.5209594	CCC1CN(C(=O)O1)C2=CC=CC=C2	0.2877766	CCCCC1(CCCC1)C(=O)OC	0.101226	C1CC(=O)N(C1)C2=CC=CS2	0.07941554	CC1(C(=O)CCC1=O)CCC=C	0.1149958	CC1(CC(N(C(=O)C1)C)O)C	0.3737635	CCCC(=O)OCN(C)N=O	0.752977
CC1=CC=NN1CCC(=O)NC2CC2	0.5243223	CCC(C#N)NCC1=CC=CC=C1	0.2889701	CCOC(C)OC1CC2CC1C=C2	0.1014261	CC1CN(N(C1=O)C)C2=CC=CC=C2	0.07944688	CC1C(SC(=N1)NC)C(=O)C	0.115477	C1CNC(=O)N=C1N	0.3754864	CCCCCCCC(=O)N1OO1	0.7540557
CC(=O)OOCC1=CC=C(C=C1)C#N	0.5263125	COC1=CC2=C(C=C1)NC(=O)CO2	0.2890458	CC(=C=C)C(=O)OC	0.1015023	CCCCCC(=C)OC	0.07983313	CC1(C2CCC1(CNC2)C)C	0.1156816	CC1(CC(=O)C=C(N1)O)C	0.3759531	CC(=C)CCCCC(=O)O	0.7549846
CCCCCC1C(O1)OCC	0.5271171	CCCCC#CC(=O)OCC	0.2894268	CCN=C1CC(=O)CC(C1)(C)C	0.1015976	C1CCC(CC1)CC2OCCO2	0.07992526	CC1CC2=C(C1O)C=C(S2)C	0.1159147	C1CNCCC2=C1C=CC=N2	0.3765176	C(CCC=O)CCC(C=O)(N)N	0.7589454
CCSC1=NC2=C(N1)C=C(C=C2)O	0.5276184	CC(C)(C)OOCCCC=C	0.2900059	CCOC(=O)CC1=COC(=N1)C	0.1019528	CCCCCC(=NC)C	0.08015769	CC1=C(SC=C1)C(=O)C(=O)OC	0.1162126	CCC1(CCCC1=O)CC	0.3774611	CCCCCCCCC(=O)C#N	0.762509
CC1=CC=C(C=C1)CNC2=NC=CC=N2	0.5278663	CCC1=CC=C(C=C1)CCO	0.2908347	CC(C)CC1(OCCO1)CCC=C	0.1020167	CC1=CC(=C)COC1	0.0802109	CC1(C(=O)N(S(=O)(=O)N1)C)C	0.1165003	C1C2CC3CC1C(C3=O)C=C2	0.3792784	CCCCC(=O)NC(=O)N	0.7634153
CC(COC(=O)C1=CC=C(C=C1)N)O	0.5294888	CC1(CN1C(=O)NC2=CC=CC=C2)C	0.2919941	CC(=O)OC1=CC(=O)CC(C1)(C)C	0.1020428	CCCCC(C)C(=C=O)C	0.08037191	CCCCC1=C(C(=O)OC1=O)C	0.1165076	CC1(CC(=CC(=O)C1)C=O)C	0.3806841	CCCCCCCCC(C)CC=O	0.7760989
CCCCNNC(=O)C(=C)C	0.5299055	CCN(C)C(=O)OC1=CC=CC=C1	0.292043	CC1(CN(CCN(CCN1C)C)C)C	0.1023316	CC1(N(OCO1)C)CC2=CC=CC=C2	0.0806848	CCC(=O)C(=O)OCC(C)C	0.1174102	CC1CC(=O)C=CO1	0.381999	C1CC(=CCCCC(=O)O)C1	0.7785154
CCCNC1=NC=NC2=C1N=CN2C	0.5305081	CC1=CC=CC=C1NC(=O)CN(C)C	0.2929717	CCC1=C(C=CC(=C1C#N)C#N)C	0.1026593	CC1=CC(OOC1)(C)C	0.08122989	CCCCC1=C(NCC1=O)C	0.1179219	C1CC2=C(N=C1)N=CC=C2	0.3826149	C1=CC(=CC#C1)CCCCC(=O)O	0.7821369
COC1=CC=C(C=C1)OCCCC=C	0.5307103	CCC(C)(C#C)NCC1=CC=CC=C1	0.2930994	CCOC(=O)C(=C1CCCC1)C#N	0.1027356	CCCCCC1=CC=CN1C	0.08133578	CCCC(NC)O	0.1180119	CC1(CC2=C(C=NN2)C(=O)C1)C	0.3832653	CC(C)CCCCC(=O)C(=O)N	0.7838638
CCOC(=O)C#CC1=CC=CO1	0.532319	CCC=CCCCCO	0.2937533	CC1(CC(=O)C(=CN(C)C)C(=O)C1)C	0.1031703	C(C)(C)OC(=O)CC	0.0814416	CC1=CCC(C(C)(O)C)C(O)C1	0.118048	CC1C2CCC(C1=O)CC2	0.3841242	CCCCCCCC(=O)OC=C	0.784092
CC1=CC=C(C=C1)NC(=O)CSC	0.5333873	CC1=CC=C(C=C1)C2(COC(=O)N2)C	0.2941163	COC1CCC2CCCCC2C1	0.1034725	COC(=O)CC1=CC2CC1C=C2	0.08146606	CC1=C(C(=CC=C1)O)C2CC2	0.1180667	C1COC2=C(C1=O)C=CC=C2O	0.3842237	CC(C)CCCCCCCC=O	0.787559
CCCCONC(=O)C(C=C)N	0.5338072	CCCCCCCCN	0.2947833	CC1CN(CC(=O)C1C)C(C)(C)C	0.1034904	CC1=NN=NN1C2CN3CCC2C3	0.08151715	CC1(CCC1)C2=CC=CC=C2	0.1181134	CC1C(SC2N1C(=O)C2)(C)C	0.384781	C1C(CC1C#CC=O)CC(=O)O	0.7878075
CCOC(=O)OC1=CC=C(C=C1)C=O	0.5350052	CCC1=CC=C(C=C1)NC(=O)CS	0.2955538	CC(C)OC1=C(C(=O)C1=O)C(C)(C)C	0.103625	CC1(COC(=N1)C=C)C	0.08170016	CC1(C=CC2C1C2(C)C)C	0.1184948	C1CC(=O)C2=CC=CN2C1	0.3851864	C(=O)(CC=CCCCCC)O	0.7878635

Table 3.16 Continued

Mor33-1

SMILES	Distance
C(=O)(CCCCCCCCC)O	0
C(=O)(CCCCCC)O	0
C(=O)(CCCCCCCC)O	0
C(=O)(CCCCCCC)O	0
C1C=C1CCCCCCC(=O)O	0.03288254
C(=O)(CCC=CCCCCC)O	0.1022495
C(=O)(CC=CCCCCCC)O	0.1268822
CCCCCCC=CCCCC(=O)O	0.1272125
C(C)C(CCC(=O)O)CCCC	0.143634
CC(C)N1C=C(C=N1)CCC(=O)O	0.149314
C=C(CCCCC#N)C(=O)O	0.1541986
CC#CCCCCCCCC(=O)O	0.1626916
CC(CCC(=O)O)N=NC(C)(C)C#N	0.1652641
CC(C)CCCCCC(=O)O	0.181324
CCCC(=C)CCC(=O)O	0.1901408
CC(C)C1CCC(=CC1)CCC(=O)O	0.1912924
CCCCC=CCC(=O)O	0.1939872
CCCCC#CC#CCCCC(=O)O	0.2014499
C1C=CC=CC1CCC(=O)O	0.2309548
C1C=C1CCCCCCCC(=O)O	0.2314075
C(=O)(CCCC=CCCCCCCCC)O	0.2344658
CC#CCCCCC(=O)O	0.2381936
CCC1=CC(=C(C=C1)CCC(=O)O)C	0.247278
CCC1=CC=C(C=C1)CCC(=O)O	0.253656
CN(C)C1=CC=C(C=C1)CCC(=O)O	0.2546323
CC1(CCC(=CC1)CCC(=O)O)C	0.255894
CCC(C)OC(=O)CC(=O)O	0.2640391
CCCC(CC(C)C(=O)O)C#N	0.2650706
C#CCCCCCCC(=O)O	0.2760099
CCCCC(=C)CCCC(=O)O	0.2793052
CC1=CC=C(C=C1)CCC(=O)O	0.2904756
CCC1CCC(CC1)CCC(=O)O	0.2971379
CCC(C)CCCC(=O)O	0.3132652
CC(=CCCC(=C=CCC(=O)O)C)C	0.3153411
C(=O)(CCCC=CCC)O	0.3153935
C(=O)(CCCCCCCCCC)O	0.3214094
CCCC#CCCCC(=O)O	0.3303376
CCCCC(C)(CC)CCC(=O)O	0.3325188
CC#CCCCC(=O)O	0.3338235
C1C(C1C(=O)O)C2=CC=CC=N2	0.3462018
CC=C=CCCC(=O)O	0.3498346
CC#CCCCCCC(=O)O	0.3504527
CCC=C=CCC(=O)O	0.3538152
CC(C)C=C=CCCCC(=O)O	0.3580099
CCCCC(CCC(=O)O)C(C)C	0.3634014
CCCCCC(C)CCC(=O)O	0.3656946
CCCCCC=CCC(=O)O	0.3664888
CC(C)C=CCCCCC(=O)O	0.3686413
CCCCC1C(O1)CC(=O)O	0.3689044
C=CC=C=CCC(=O)O	0.3778463
C1CCC2=C(C1)C=NC2CC(=O)O	0.3817782
C(=O)(CCCC=CCCCCC)O	0.3828978
CCCC(C)C(C)CCC(=O)O	0.3883416
CCC1CCC(O1)CC(=O)O	0.3901875
CCCCC(C)CCC(=O)O	0.3981465
CC(C)C1CCC(CC1)CCC(=O)O	0.4008439
C=C=CCCCC(=O)O	0.4018696
CN1C=C(C=N1)CCC(=O)O	0.4109742
CCCCCCC#CCC(=O)O	0.4210856
C=CCC=C=CCC(=O)O	0.4244024
C(C)C(CCC(=O)O)CCCCC	0.4248932
CCCCCC#CCC(C)C(=O)O	0.4249177
CCCCC(CC)CC(C)C(=O)O	0.4331169
CC(CCCC(=O)O)CC(C)(C)C	0.4360629
C1CC=CC(C1)CCCCCC(=O)O	0.4402217
CNC1=CC=C(C=C1)CCC(=O)O	0.441039
CCCCCCC(C)CCC(=O)O	0.4418524
C1CC(SC1)CCCCC(=O)O	0.4447367
C1=CC=C(C=C1)CC(C#N)C(=O)O	0.4459466
CCCC1CC(C1)C(=O)O	0.454442
C1C=CCC(=C1)CCC(=O)O	0.4545794
CCCCC(C)CC(C)C(=O)O	0.4582201
C=CC1=NOC(=C1)CC(=O)O	0.4614348
CCC1=CC=C(C=C1)CC(C)C(=O)O	0.4644323
C(=O)(CCCCCCCCCCCCC)O	0.4647544
C1CCC(C1)C=C=CCCC(=O)O	0.4650765
CC(C)CCCCCCC(=O)O	0.4708209
C#CCCCCCCCCCC(=O)O	0.4710753
CCCCCC=C=CCCC(=O)O	0.4779645
C1=CN(C=C1)CCCCCCC(=O)O	0.4791871
CC(C(=O)O)CCCCCCC	0.4886482
CCCCCC(C)CCCC(=O)O	0.4892858
CCN1C=C(C=N1)CCC(=O)O	0.4926835
CCCCC1CC(C1)C(=O)O	0.4952327
COCOC(=O)CC(=O)O	0.5012216
CC1CCC(O1)CCC(=O)O	0.5016139
C1CC(=C1)OC(=O)CC(=O)O	0.5023661
C1=CC=C(C=C1)CCC(=O)O	0.5043305
CC(C(=O)O)CCCCCCCC	0.5046819
COC1=CC=C(C=C1)C2CC(=O)N2	0.5065262
CCC(CCCCC(=O)O)S	0.5096068
C1C(C1C(=O)O)C2=CN=CC=C2	0.510366
COC(=O)CCCC1=CNC=C1	0.5108529
CCCCCCC(CC=O)O	0.5135363
CC(C(=O)O)CCCCCC	0.5140023
C(=O)(CCCCCCCCCCC)O	0.5148502
C(=O)(CCC#CC#CCCC)O	0.5160469

Mor37-1

SMILES	Distance
C(=O)CCCCCCCCC	0
C(=O)(CCCCCCCCC)O	0
C(=O)CCCCCCCC	0
C(=O)(CCCCCCCC)O	0
CCCCCCCCCCCCCC(OC(C)C)=O	0
CCCCCCCCC=O	0
CCCCCCCCCC=O	0
O=C(C)CC/C=C(CC/C=C(C)/C)\C	0
C(=O)(CCCCCCCC)O	0
C(=O)(CCCCCCCCC)O	0
C(=O)(CCCCCCCCCCC)O	0
C(=O)(CCCCCCCCCCCC)O	0
C(=O)(CCCCCCCCCCCCC)O	0
C(=O)(CCCCCCCCCCCCCC)O	0
CC1=C(C=C(C=C1)[C@H](C)CCC=C(C)C)O	0
CC(OCCCC/C=C\CCCC)=O	0
CC/C=C/CCCCCCCCC([H])=O	0
CCCCCCCCCC[C@H](OC(C)=O)[C@@](O1)([H])CCCC1=O	0
O=C(CCCCCCC)N1CCC(C2=CC=CC=C2)CC1	0
O=C(CCCCCCCC)N1C(CC)CCCC1	0
O=C(CCCCCCCCC)N1C(C)CCCC1	0
O=C(N1CCC(C)CC1)CCCCCCCCC	0
O=C(CCCCCCCCC=C)N1CCCCC1	0
O=C(CCCCCCCCC=C)N1C(CC)CCCC1	0
O=C(CCCCCCCCC=C)N1CCC(C2=CC=CC=C2)CC1	0
O=C(CCCCCCCCC=C)N1CCC(C)CC1	0
O=C(CCCCCCCCCC)N1CCCCC1	0
O=C(CCCCCCCCCCC)N1C(C)CCCC1	0
O=C(CCCCCCCCCCC)N1CC(C)CCC1	0
O=C(N1CCC(CC2=CC=CC=C2)CC1)CC3CCCCC3	0
O=C(N1CCC(CC2=CC=CC=C2)CC1)CCC3CCCCC3	0
CC(OCCCCCCCCC/C=C\CCCCCC)=O	0
CC(OCCCCCCCC/C=C\CCCCCCCC)=O	0
CCCCCC/C=C\CCCCCCCCC(O)=O	0
CCCCCC/C=C/CCCCCCCCC(OC)=O	0
C(=O)CCCCCCCCCCC	0
C(CCCCC)OC(=O)c1c(cccc1)O	0
C(C)C(COc1c(cccc1)C(=O)[O-])CCCC	0
C(CCCCC)C(=CC=O)c1ccccc1	0
C(=O)(CCCCCCCCCCCCCC)O	0
C(=O)(CCCCCCCCCCCCCCC)O	0
C(C)OC(=O)CCCCCCCC	0
C(C)OC(=O)CCCCCCCCC	0
C(C)OC(=O)CCC(C)CCCCCC	0
C(=O)(CCCCCCCC=C)O	0
C(C)OC(=O)CCCCCCCCCC	0
C(C)OC(=O)CCC(C)CCCCCCC	0
CC(C)CCCCCCCCC(OCC)=O	0
C(C)OC(=O)CCCCCCCCCCC	0
C(C)OC(=O)CCC(C)CCCCCCCC	0
CC(C)CCCCCCCCCC(OCC)=O	0
CC(C)CCCCCCCCCCC(OCC)=O	0
C(C)OC(=O)CCCCCCCCCCCCC	0
C(C)OC(=O)CCCCCCCC=CCCCC	0
C(C)OC(=O)CCCCCCCCCCCCCC	0
C(C)OC(=O)CCCCCCCC=CCCCCCC	0
C(=O)(CCCCCCCCCCCCCCCCC)O	0
C(=O)C=CCCCCCC	0
C(=O)CCCCCCCCCCCC	0
C(=O)CCCCCCCCCC	0
CCCCCCCC/C=C/CCCCCC(OC)=O	0
O=C(OC)CCCCCCCCCCCC1CCCC1	0
CCCCCCCCCC(OC)=O	0
CCC(C)CCCCCCCCC(OC)=O	0
CCCCCCCCCCC(OC)=O	0
CCCCCCCCCCCCCCCC(OC)=O	0
COC(=O)CCCCCCCC(C)CCCCC	0
CCCCCCCCC(OC)=O	0
CCCCCCCCCCCCCCC(OC)=O	0
CCCCCCCCCCCCCC(OC)=O	0
CCCCCCCCCCCC(OC)=O	0
CCCCCCCCCCC(OC)=O	0
CCCCCCC1CCC(O1)=O	0
CC(C(=O)O)CCCCCCC	0
CC(C(=O)O)CCCCCCCC	0
C(C)C(CCC(=O)O)CCCCC	0
C(=O)(CCCCCCCCCC)O	0
C(C)C(CCC(=O)O)CCCCCC	0
CCCCCC/C=C/CCCCCCCC(OC)=O	0
C(=O)(CCCCCCCCCCCCCCC)OC	0
CC12CCC(=O)C=C1CCC3C2CCC4(C3CCC4=O)C	0
C[C@@]12CCC[C@@]1([H])[C@]3([H])CC(C4=CC(C=C[C@]4(C)[C@@]3[illegible]	0
C[C@@]1(CC[C@]2([H])[C@@]34C)CCC[C@@]1([H])[C@]2([H])CCC4C[illegible]	0
C[C@@]12CCC[C@@]1([H])[C@]3([H])CCC4CC(C(C[C@]4(C)[C@@]3([H[illegible]	0
C[C@@]12CCC[C@@]1([H])[C@]3([H])CCC4CC(CC[C@]4(C)[C@@]3([H[illegible]	0
C[C@@]12CCC[C@@]1([H])[C@]3([H])CCC4=CC(CC[C@]4(C)[C@@]3([H[illegible]	0
C[C@@]12CCC[C@@]1([H])[C@]3([H])CCC4CC(C=C[C@]4(C)[C@@]3([H[illegible]	0
C[C@@]12CCC[C@@]1([H])[C@]3([H])CCC4=CC(C=C[C@]4(C)[C@@]3([H[illegible]	0
CC12CCC3C(C1CCC2C(=O)C)CCC4=CC(=O)CCC34C	0
CC([C@H]1CC[C@@]2([H])[C@]3([H])CC(C4CC(CC[C@]4(C)[C@@]3([H[illegible]	0
CC([C@H]1CC[C@@]2([H])[C@]3([H])CCC4=CC(C=C[C@]4(C)[C@@]3([H[illegible]	0
CC([C@H]1CC[C@@]2([H])[C@]3([H])C=CC4=CC(CC[C@]4(C)[C@@]3([H[illegible]	0
C[C@@]12C=CC[C@@]1([H])[C@]3([H])CCC4=CC(CC[C@]4(C)[C@@]3([H[illegible]	0
C[C@@]12C=CC[C@@]1([H])[C@]3([H])CC[C@@]4([H])CC(CC[C@]4(C)[illegible]	0
C[C@@]12C=CC[C@@]1([H])[C@]3([H])CCC4=CC(C=C[C@]4(C)[C@@]3([H[illegible]	0
C(=O)(CCCCCCCC=CCCCCCC)O	0
C(=O)(CCCCCCCC=CCCCCCCCC)O	0

Mor40-1

SMILES	Distance
C(=O)(CCCCCCCCC)O	0
C(=O)(CCCCCCCC)O	0
C(=O)(CCCCCCC)O	0
C1=CC=C(C=C1)CCC2=NN=C(C=C2)N	0.02341789
C(=O)(CCCCCCCC=C)O	0.03694565
COC1=CC=C(C=C1)CCCCC#N	0.04434623
C=CCCCCCCCC(=O)S	0.04698026
CNC1=CC=C(C=C1)CCC(=O)OC	0.04732122
C(CCCC(O)O)CCCC(=O)O	0.04909072
CCCCC(=O)C1=CC=C(C=C1)OC	0.05102408
CCCCCC(=O)C1=CC=C(S1)O	0.05385012
CCCCCCC(CC(=O)O)O	0.0581458
CCCCCCCC(=O)OS	0.0582002
CCCCC#CC(=O)CCCC	0.05901835
CC(=C)CCCCCCC(=O)O	0.06104174
CCCCCCCC(=O)C#N	0.06314685
CCCCCCCC(=O)C=C=C	0.0637365
COC(=O)CCCCCC=CC=C	0.06446715
CCCCCCC(=O)OC(C)S	0.06476652
COC(=O)C1=CC=C(C=C1)CCC=O	0.06633289
CCCCCCCC(=O)NO	0.06904818
COC(=O)CCC1=CC=C(C=C1)C=C	0.07307333
CN(C(=O)CCCC1=CC=CC=C1)O	0.0740539
CCCCCCCC(=O)N(C)N=O	0.0767582
CCCC(=O)C=C=CC1=CC=CC=C1	0.07690233
CCCC(CCCC(=O)C=C)O	0.07774642
COC(=O)CCCCCCC#C	0.07783589
COC(=O)CCC1=CC(=CC=C1)NN	0.07920649
CCCCCCC(=O)OC	0.07977409
CCCC#CCCCC(=O)OC	0.08015927
CCC(=O)C=CC1=CC=C(C=C1)OC	0.08487605
CCCCC1=NC(=NN1)C(OC)OC	0.08730716
COC1=CC=C(C=C1)CCC(=O)OC	0.08783883
CC(CCCCCC(=O)O)O	0.08817417
CC1=CC=CC=C1CCCCC(=O)O	0.08946261
COC(=O)C1CCC(C1=O)CCC=C	0.0895777
CCCCCC(=O)C1=CC=CS1	0.09154678
COC1=CC=C(C=C1)CCCC(=O)N	0.0930181
C1=CC=C(C=C1)CCCCC(=O)N	0.0934213
CCCCCCC1=NC=C(C=C1)N	0.09413435
CC1=CC(=CC=C1)CCCCC(=O)O	0.0947023
CC(C)CCCC1=CC(=NN1)C(=O)O	0.09581364
COC1=CC=CC(=C1)CCCC(=O)O	0.09727548
CCCCCCC(CC=O)O	0.09927231
COC(=O)NN=CCC1=CC=CC=C1	0.0998653
CCCCCCC(=O)NC	0.100106
CCCC1=CC(=NO1)NC(=O)OC	0.1011953
CCC(=O)C=CN(C)C1=CC=CC=C1	0.1014411
CCCCCCC(CCC(=O)O)O	0.1022738
C1=CC(=CC=C1CCC#N)N2C=CN=C2	0.1045008
CNC(=O)CCC1=CC=C(C=C1)OC	0.104926
CCCCC(CC(=O)OC)O	0.1053018
CCCCCCC(=O)N=C	0.1057203
CCCCCC=CCC(=O)O	0.1063752
CCCCCCC1=CC=CC=N1	0.1074308
COC(=O)CCCC1=CC=C(C=C1)N	0.1087037
CCCCCCC1=NC=CC(=C1)N	0.1091682
C1=CC=C(C=C1)CCCCC(=O)O	0.1092112
CCCCCCC1=C(C=CC=N1)N	0.11081
C1=CC=C(C=C1)CCC(CC(=O)O)O	0.1120773
COC(=O)CCCC1=CC=C(C=C1)O	0.1131115
CC(C)(CCCCCC(=O)O)OC	0.1144698
C1=CC(=CC=C1CCCCC(=O)O)N	0.1149288
C1=CC(=CC=C1CCCC(=O)N)C#N	0.1149956
C=CC(CCCCCCC(=O)O)O	0.1151096
CC1=CN=C(C=C1)CCCCC=C	0.1154437
CC(=C)CCCCCCCC#N	0.1158624
C1C(=COC1=O)CCC2=CC=CC=C2	0.1162124
C1C=C1CCCCCCC(=O)O	0.1172653
CCCCCCCC(=O)NCO	0.117728
COC1=CC=C(C=C1)CCCC(=O)O	0.1188917
CC(C)CCCCC(CC(=O)O)O	0.1198341
C=CCCCCCCCCC(=O)O	0.1198667
CCCCC#CC1=CC(=O)CCC1	0.1204789
CCCCCCC(=O)OC#N	0.1205388
CCCC(=O)C=CC1=CC=CC=C1	0.1225009
CC(CCCCCCC=O)O	0.1230657
CCCCCCCC1=NNN=C1	0.1244597
CCCCCCC1=NN=C(O1)N	0.1245485
CCC=CCCCCCC=O	0.1251914
CCC(CCCCCCC(=O)O)N	0.1252176
[illegible]C(=O)CC(=O)NC	0.1252651
[illegible]C(=O)O	0.1255971
[illegible]NC(=CC(=O)N1)O	0.1264059
[illegible]CCCCCC=O)OC	0.126566
[illegible]C=C(C=C1)CCC(=O)OC	0.1270303
[illegible]C(=O)C(=C(O)O)C	0.1280732
[illegible]CCCCCCC=C	0.1287103
CCCCCCC1=NOC(=C1)C=O	0.1299349
[illegible]C1)OC	0.1303238
[illegible]=O)ON)N	0.1309691
[illegible]OCC1	0.1323051
[illegible]C=C1CCCC(=O)O)C=O	0.1331276
[illegible]C=C(C=C2)CC#N	0.1332993
[illegible]C1=CC=NC=C1	0.1357224
CCCCCCCCCC(=O)N	0.13617
CCCCCCC(=O)NN	0.1363529

Mor41-1

SMILES	Distance
CC1=CC[C@@H](CC1=O)C(=C)C	0
CC1=CC[C@H](CC1=O)C(=C)C	0
CCC(=O)C(=O)CC	0
O=C1CCOC2=C1C=CC=C2	0
C(=O)(C)c1ccccc1	0
O=C1OC2=C(C=CC=C2)C=C1	0
C1C(=O)C2=CC=CC=C2ON1	0.1690913
C1=CC=C2C(=C1)C=CC(=O)N2	0.19021
C1CC2=C(C=C1)OC(=O)C=C2	0.195202
CCCCC1=C(NNC1=O)C	0.1966665
CC=C1CCC(=O)CC1	0.2004243
C1C=CC2=C(C1=O)C=CC(=O)O2	0.2108212
CCCCC1=C(OCC1=O)O	0.2185848
C1NC(=O)C2=CC=CC=C2S1	0.2228905
C1C(=O)C2=CC=CC=C2OS1	0.2251387
C1=CC2=NNN=C2C=C1C(=O)N	0.2380402
CC1CC1(C2=CC=CC=C2)O	0.2385417
CC1CCC(=CC1=O)C(C)C	0.2447182
C1=CC=C(C=C1)S(=O)N	0.2462735
C1CNC2=CC=CC=C2C1=O	0.2512114
CC1(C(=O)O1)C2=CC=CC=C2	0.252103
C1=CC=C2C(=C1)C(=O)NNN2	0.255054
C1CSC2=CC=CC=C2C1=O	0.2633634
C1=CC2=NSN=C2C=C1C(=O)N	0.2647783
CS(=O)C1=CC=CC=C1	0.2651372
C1C(=O)C2=CC(=C(C=C2S1)O)O	0.2656982
C1C(C(=O)N1)C2=CC=CC=C2	0.2658363
C1=CC2=NON=C2C=C1C(=O)N	0.2685111
CC(=C1CCCCC1)O	0.2698948
C1=CC(=C(C=C1O)C(=O)NO)O	0.2742529
CN(C1=CC(=O)CCC1)O	0.2797649
C1=CC(=C2C(=C1)N=NS2)C(=O)S	0.2811271
C1NC2=CC=CC=C2C(=O)O1	0.2819032
C1=CC(=C(C=C1N)CCO)O	0.2819074
C1OC(=O)C2=CC=CC=C2S1	0.2822413
C1=CC=C2C(=C1)C(=O)C3=C(O2)O3	0.2849512
CC(=O)C1=CC=CN2C1=NN=N2	0.2867693
C=C(C1=CC=CC=C1)O	0.2875154
C1C=CCC2=C1C=CC=C2O	0.2879232
CCC(=O)C1=CCCCC1=O	0.2883527
C1C(=O)C2=CC=CC=C2NS1	0.2913763
C1CC(=O)C1C2=CC=CC=C2	0.2953117
C1CC=C2C(=C1)C=CC(=O)O2	0.2956927
CC1=CC(=NNC1=O)C(=O)NN	0.2962124
CC(=O)C1=CC=CC2=NN=NN21	0.2980169
CCN(C1=CC=CC=C1O)O	0.2994812
C1=CC2=NNN=C2C(=C1)C(=O)N	0.2996982
C1COC(=N1)C2=CC=CC=C2	0.300755
CC1CC=C(NC1=O)C(C)C	0.3042253
C1=CC=C2C(=C1)C(=CS2)C(=O)O	0.3051727
CC1CC(=O)OC(=C1)C(C)C	0.3061929
CC1=CC(=C(C=C1)OOO)C	0.3102744
CC1=NC2=CC=CC=C2N1O	0.3107354
CC1=CC(=CN1C)C(=O)O	0.3107893
C1CC(CNNC1)CO	0.3139191
C1=CN2C=C(C=CC2=C1)C(=O)O	0.3163863
C1=CC2=C(O2)C(=C1C(=O)O)C(=O)O	0.3172029
CSC1=CC=C(C=C1)N=O	0.3180102
CC1=CC(=CC(=C1)O)NC	0.3180681
CCC(C)C1=CC=CC=C1O	0.3205691
CC(=O)C1C(=O)C(OC1=O)(C)C	0.3208731
COC1=CC2=NCN=C2C=C1	0.3224311
C1C(=O)C2=CC=CC=C2OO1	0.3226159
CC(=C)C1=CC=CC(=O)O1	0.3234103
CC1(C2(CCNCC2)OC(=O)N1)O	0.3244962
C1=CC=C2C(=C1)C(=S)C(=O)NO2	0.3248371
CC1=C(C=C(C=C1)O)NNC	0.3252422
CCCCC1=C(C(=O)CC1)O	0.3260718
CCC(=O)N1CCNC(=O)C1=O	0.3262205
C1CC2=C(C(=O)NC(=O)O2)NC1	0.3285799
C1=CC=C2C(=C1)N=NC(=N)N2O	0.3286808
C1COC2=C(N1)C=CC(=C2)O	0.3299594
C1=CC=C(C=C1)C(=O)S	0.3310767
CCC1=CC(=C(C=C1)C#N)O	0.3312191
c1(ccccc1)C(=O)O	0.331318
CC1=C(C(=CC=C1)CCO)O	0.3332377
C1=CC=C(C=C1)C(=O)N	0.3332575
COC(=O)C1CCNC(=O)C1=O	0.3334085
C=C1COCC2=CC=CC=C12	0.337791
C=CC1=CC(=CC(=C1)O)O	0.3379981
CCCC1=C(C=CC(=C1)O)O	0.3400948
CSC1=CC(=CC(=C1)O)O	0.3402908
C1=CC2=CC(=O)NN=C2C=C1	0.3403887
C1COC2=C(N1)C=C(C=C2)O	0.3415567
C1=CC=C2C(=C1)C(=O)N=NS2(=O)=O	0.3428754
CC1=C(CCCC)C(CC1)=O	0.3431011
C1CCN2CCC(=O)NC2C1	0.3431383
C1=CC(=C(C(=C1)S)O)C(=O)NO	0.3433193
C1=CC=C2C(=C1)C(=O)OS(=O)N2	0.3440329
C1=CC2=C(C=CO2)C=C1C(=O)O	0.3451264
C1=CC2=NSN=C2C=C1C(=O)O	0.3460875
CCCCC1=C(C(=O)CC1)C	0.3471947
CC(=C1C=CC(=O)C=C1O)NO	0.3473508
C1=CC=C2C(=C1)C(=O)NS(=O)(=O)O2	0.3477493
CCCCC1=C(NCC1=O)C	0.3479773
CC1=CC2=C(C=C1)NS(=O)N2	0.3488483
C1=CC=C(C=C1)N2C(=CC2=O)O	0.348864

Mor5-1

SMILES	Distance
C(=O)(CCCCCCCCC)O	0
C(=O)(CCCCCC)O	0
C(=O)(CCCCCCCC)O	0
C(=O)(CCCCCCC)O	0
CCCCC(C)CC(C)C(=O)O	0.00047992
CCCN1CNC=C1CC(C(=O)O)N	0.00047992
CCCCC=CCC(=O)O	0.00047992
CC1=CC=C(C=C1)CC(C(=O)O)N	0.00047992
C1CCC2=C(C1)CCNC2CC(=O)O	0.00047992
CCC(C)(C)C(C)C(CC(=O)O)N	0.00095984
C1CCC(C1)C=C=CCCC(=O)O	0.00095984
CCSCCCC(=C)C(=O)O	0.00095984
C1CCC2=C(C1)C=NC2CC(=O)O	0.00095984
CCCCN=C(C)CC(=O)O	0.00095984
C(CC(C(=O)O)N)CNCC(=N)N	0.00143976
C1CC(CC=C1)CCC(=O)O	0.00143976
C1=CN(C=N1)CCCCCCC(=O)O	0.00143976
CC(=C=C1CCCCC1)C(=O)O	0.00143976
CCCCCCC=CC(=O)O	0.00143976
CC1=C(C=NN1C(C)C)CCC(=O)O	0.00143976
C1=CC=C(C(=C1)CCC(=O)O)CN	0.00143976
CCN1C(=C(C(=N1)C)CCC(=O)O)C	0.00191969
CCC(CCCCC(=O)O)S	0.00191969
CCC1=CN=C(C=C1)CCC(=O)O	0.00191969
C(CCCC(=O)O)CCCS	0.00191969
CCC1=CC=C(C=C1)N(CC(=O)O)N	0.00191969
CSCCCC(C(=O)O)N	0.00191969
C1=CC(=CN=C1)CC(CN)C(=O)O	0.00191969
CCCCC1NC(CS1)C(=O)O	0.00191969
CC(C)CC1=C(C=NN1)CCC(=O)O	0.00191969
CCCN1C(=C(C=N1)CCC(=O)O)C	0.00191969
CCCCCOOCC(=O)O	0.00191969
CC(C)(CCC(=O)O)SSC	0.00239961
C(C)C(CCC(=O)O)CCC	0.00287953
CCC1=CC=C(C=C1)C(CC(=O)O)N	0.00287953
CCSCCCC(=O)O	0.00287953
CCCCN1CC1C(=O)O	0.00287953
C1=CNC(=C1)CCCC(C(=O)O)N	0.00287953
CCC(=N)C(=N)C(C(=N)C(=O)O)N	0.00287953
CCCN1C=C(C=N1)CCC(=O)O	0.00287953
C1C(C1C(=O)O)C2=CC(=CC=C2)N	0.00287953
C(C)C(CCC(=O)O)CCCC	0.00335945
C1CCN(CC1)CC(C(=O)O)O	0.00335945
C1C=C1CCCCCCC(=O)O	0.00335945
CCCCCC#CC(=O)O	0.00335945
CCCCC#CCCC(=C)C(=O)O	0.00335945
CNNC(CC1=CC=CC=C1)C(=O)O	0.00335945
CCCC(C)C(C)C(CC(=O)O)N	0.00383937
C(CCCCC(=O)O)CCCCN	0.00383937
CC(C(CCCCCC(=O)O)N)N	0.00383937
CC(CC1=CN=CC=C1)(CN)C(=O)O	0.00383937
CCC1=CC=C(C=C1)NCC(=O)O	0.00383937
C(CCN)CCNCC(=O)O	0.00431929
CC1=CC=CC=C1CCC(=O)O	0.00431929
C1C2=CC=CC=C2C(NN1)CC(=O)O	0.00431929
CCC(C)CCC(=C)C(=O)O	0.00527914
CC1=CC(=O)N(N1)CC(C)C(=O)O	0.00527914
CCCCCC(C)(C(=O)O)N	0.00527914
CCC1=C(N(C=C1)C)CCC(=O)O	0.00527914
CC(C)CC(CSCC(=O)O)N	0.00527914
CCCCCSCC(=O)O	0.00575906
CCCCCCC(C)(C(=O)O)N	0.00575906
CCCC(CC)CC(CN)C(=O)O	0.00575906
C1CCC(=CC1)C(CC(=O)O)N	0.00575906
CCCCSC(C)C(=C)C(=O)O	0.00575906
C1=CC(=CC=C1CCC(=O)O)CN	0.00623898
C1=CC=C(C=C1)CC(C(=O)O)NN	0.00623898
CSCCCCCC(C(=O)O)N	0.00623898
CCCN1C=C(C(=N1)C)CCC(=O)O	0.0067189
CCCC(CC(C)C(=O)O)C#N	0.0067189
CCC#CCC(C)(C)C(=O)O	0.0067189
CCCCCC#CC#CCCC(=O)O	0.00719882
C=CC1CNCCC1CCC(=O)O	0.00719882
CC(C)(C(=O)O)SC1=CC=CC=N1	0.00767874
C1CCNC(C1)NCC(=O)O	0.00767874
CC(C)C1=CC=C(C=C1)C=CC(=O)O	0.00815866
CC1=CC=CC=C1CC(C(=O)O)(N)N	0.00815866
C(CCC(C(=O)O)N)CCN	0.00815866
CC(C=C1CCCCC1)C(=O)O	0.00815866
CC1=CC=CC=C1C=CC(=O)O	0.00815866
CC(C)(CCN1C=CN=C1)C(=O)O	0.00815866
CC1=NN(C=C1CCC(=O)O)C(C)C	0.00863859
CCN1C(=CC=N1)CCC(=O)O	0.00863859
CCCOOC(=O)C(=O)O	0.00911851
CCCCOC(=O)C(C(=O)O)N	0.00911851
CC#CCCCC(=O)O	0.00959843
CCCCCCC=CCCCC(=O)O	0.01007835
CCCC(C)CC(CN)C(=O)O	0.01007835
C1CCC(CC1)C=CC(=O)O	0.01055827
C1CCC(=CC1)CCC(=O)O	0.01055827
CN1CCCC(C1)OCCC(=O)O	0.01055827
CC(C)N1C=C(C=N1)CCC(=O)O	0.01055827
CCCCNCC(=O)O	0.01103819
C1C=CC=CC1CC(C(=O)O)N	0.01103819
CCCC#CCN1CCC(C1)C(=O)O	0.01103819
CN1CCC(NCC1)CC(=O)O	0.01103819
C1CC=CC(C1)CC(C(=O)O)N	0.01151811

Or1A1

SMILES	Distance
CC1=CC[C@@H](CC1=O)C(=C)C	0
C[C@@H]1CC[C@H](CC1=O)C(=C)C	0
CC1=CC[C@H](CC1=O)C(=C)C	0
CCC(=O)CCCC	0
CCC(=O)CCCCC	0
C(=C)COC(=O)Cc1ccccc1	0
CC1=C(CCCC)C(CC1)=O	0
CCCCC(=O)CCC	0.06049592
C=CCCC(=NO)C1=CC=CC=C1	0.07078234
CCCOCCC(=O)C=C	0.0717614
C1CCCN(CC1)C(=O)N2CC2	0.07369093
C1CCC2=C(C1)C=CC=C2CC(=O)N	0.07424287
CCC(=NO)C1=CC=C(C=C1)C	0.07771264
CC1CC=C(NC1=O)C(C)C	0.07826973
C1CC(=O)N(N=C1)CC2=CC=CC=C2	0.07849599
CC1(OCC(CO1)C(C=C)O)C	0.07977273
CCCC(CC1=CC=CC=C1)C(=O)N	0.08055253
CC1=C2C=C(C=CC2=NC=C1)C(=O)N	0.08126738
C1CC(C2=CC=CC=C2C1)CC(=O)N	0.08161454
COCC1=CC=CC2=C1CCCC2=O	0.08240595
CC(CC=C)OC(=O)C1=CC=CC=C1	0.08262245
CC(=O)C1=CC2=C(C=C1)OCCNC2	0.08372653
C1C=CC=C(C1=C=O)CC2=CC=CC=C2	0.0853223
CCOC(=O)CC(C)C1=CC=NC=C1	0.08749861
CC1=NC2=CC=CC=C2C=C1C(=O)N	0.08760062
CCC(C(=O)C)NC1=CC=CC=C1	0.08768825
CCCNC(=O)C1CCCN1C	0.0881405
C1C(NC(=O)CO1)C2=CC=C(C=C2)N	0.0887641
CC1=C(N=NC2=CC=CC=C12)C(=O)C	0.08925951
CCCCC(C1=CC=CC=C1)(O)O	0.08964993
CC1(OCCO1)CC2=CC=CC=C2	0.0898149
CC1(CCCNN1)C2=CC=C(C=C2)N	0.09071153
C1COCC(=O)N1C2=CC=C(C=C2)N	0.09096969
CCCCC(CN)(C1=CC=CC=C1)O	0.09309083
CC(=O)OC1=CC=CC(=C1)CCN	0.09478365
CC(C)CC(=O)CC(OC)OC	0.09483221
CN(C)N1CC2=CC=CC=C2NC1=O	0.09529523
CC(C)NC(=O)N1CC=CC=C1	0.09536646
CC1=CC2=C(C=C1)NC(N2)(C)C	0.09542733
C=C(C)C(CCC1C)=CC1=O	0.09640036
COC1=CCC2=C(C1)CCCC2=O	0.0971561
CCCCOC1(CC1C)CO	0.09737174
CC(C1=CCCCCC1)C(=O)NC	0.09754798
CC(CN1CC1)C(=O)C2=CC=CC=C2	0.09827942
C=CCC1C(=O)NC2=CC=CC=C2N1	0.0983881
CCC(C)C(C(=O)C1=CC=CC=C1)N	0.09844325
CNN(C)C(=O)NCC1=CC=CC=C1	0.09845384
CCCC(C)COC(=O)CC	0.09879688
CCCOC(=O)C(C)COCC	0.0989054
C1C(N=C(NC1=O)N)C2=CC=CC=C2	0.09943917
CCC(=O)C1=CCCN(C1)C	0.09954512
CC1CCC(=O)N1CC2=CC=CC=C2	0.09992609
CCN(CC1=CC=CC=C1)C(=O)CN	0.1001827
CC(C)C1=NC2=CC=CC=C2NC1=O	0.1005161
CC(C)C1=CC(=CC=C1)OC(=O)C	0.1021269
CC(=O)C1CCC(CC1)(C)N	0.1023978
CCCCC1=C(CCC1=O)OC	0.1026563
CC1(CNCC(O1)C2=CC=CO2)C	0.102937
CC(C)CNC(=S)NCC=C	0.1033386
CCCCC1=C(C(CC1=O)O)C	0.1033504
CC(=O)OC1CCC2=CC=CC=C2C1	0.1035213
CC1=C(C(=O)NC2=C1C=CC(=C2)N)C	0.1035421
CCCCC1(OCC(O1)CO)C	0.1038422
CC(=CC(=O)N1CCCCCC1)C	0.1038683
CC1=CC(=CC=C1)OC(=O)C(=C)C	0.1039132
CC1=CC=CC(=C1C)C(=O)C2CCC2	0.104029
CCCCC(C1=CC=CC=C1)(N)N	0.1044268
CC(=O)NC1CCCC2=CC=CC=C12	0.1052389
CC(CN1C=CC2=C1CCCC2=O)O	0.105567
CC1=CC=CC2=C1NC(=O)C(=C2C)C	0.105693
C1=CC=C(C=C1)CNC(=O)C(CN)O	0.1059311
C1C(O1)CCOC(=O)C2=CC=CC=C2	0.1062612
CC1=NC2=C(CCC2)C=C1C(=O)C	0.1062984
CC(C)C(C)C(=O)C1=CC=CC=C1	0.1064016
CN1CCCC1C(=O)NCC=C	0.1064649
CC1=CC=C(C(C)(O)C)CC1	0.1065836
CC(=O)NC1C2=CC=CC=C2CCN1	0.1067861
CC1=CC(=CC2=C1OCCC2)C(=O)C	0.107163
CC1=C(CCC1(C)C)CC(C)N	0.1072895
CC(C)NN1C(=O)CC2=CC=CC=C21	0.1072913
C1CC(N(C1)C2=CC(=O)CCC2)CO	0.1074226
CC(=O)CCC1CCCOC1OC	0.1078314
C1CC2=C(CC1C(=O)N)C=CC=C2N	0.107852
CC1=C(C(=CC=C1)N)C(=O)OCC=C	0.1082888
CC(=O)NC1=CC=CC2=C1CCNC2	0.1085607
C1=CC=C(C=C1)CC2=CC=CN=NC2=O	0.1089713
CC1=CC(=C(C=C1)C)C(=O)C2CCC2	0.1090924
CCCCC(=O)OC(CCC)O	0.1093729
CC(=O)NC1CCC2=CC=CC=C12	0.1094877
CC1CC=C(C1)C(=O)C(C)C	0.1096098
C1C(C1C2=CC=CC=C2)CNC(=O)N	0.1098267
CNC1=CC2=C(C=C1)C(=O)CC2	0.1101241
CCCCN(C)CCC(=S)N	0.1101301
C1=CC=C(C=C1)CN2C=CC=CC2=O	0.1106218
C#CCCOC(=O)C1=CC=CC=C1	0.1110363
CC(C1=CC=CC=C1)C(=O)NC(C)N	0.1111602
CCCC1=C2C(=CC(=N1)C)CCC2=O	0.1112196

Or2J2

SMILES	Distance
C(CCCCCCCC)O	0
C(CCCCCCC)O	0
O=C1OC2=C(C=CC=C2)C=C1	0
CCCCC(CCCCO)O	0.1878163
CC(C)CCCCCCO	0.2001934
C(CCCCN)CCCCON	0.2449217
C=CCCCCCC1CO1	0.2829139
CCCCCCNCC(CC)O	0.2913393
CCCCCCCCN(C)O	0.3036347
CC(C)CC1COC(N1)CCO	0.3482781
CCCCCCCCONC	0.3506971
CCCCCCC(C)NCCO	0.35406
CCCCCN1CCC(C1)CO	0.3679643
CCCCCOC(C)CCO	0.3976565
CCCCCCCC(C)CO	0.4018389
CCCCCCCCNOC	0.4019279
C(CCCCNCCO)CCCN	0.4055974
CC(C)CCCCNCCO	0.4075006
CCCC1CC(C1O)O	0.4174384
CCCCC(C)(CC(C)O)O	0.4285046
CCCC1CCN(CC1)CCO	0.4289913
CCCCCCCCCCCNO	0.4353896
C=CCCCC#CCO	0.4379203
CCCCCCNC(C)CO	0.4486621
CCC(C)CCCCCCO	0.4517683
CCCCCCC1CCNO1	0.4523566
CCCCC(CC(C)(C)O)O	0.459996
C1=CC=C(C=C1)C(=O)NCN=C=O	0.462984
CCCCCCCCON(C)C	0.466401
C=CCCCCC1CO1	0.4667611
CCCCCC(CCO)N	0.4675419
C=CCCCCCCC1CO1	0.4706314
C(C=CC=CCC=CCC)O	0.4720574
CCCC(C)CCCCO	0.4736975
CCCCCCCCC(CO)N	0.4767931
C=CCCCOCC1CO1	0.4773465
CC(C)CCC(CCO)O	0.4806107
C1C=CN(C2=CC=CC=C21)N=O	0.4849279
CCCCCCC(C)(O)O	0.4903722
C(CC=CCCCCC)O	0.4922504
CCCCCCCCCC(N)O	0.4956697
CCCCC=CCCO	0.4960279
C1CCC(CC1)N2CCNC(C2=O)N	0.4971952
C1=CC2C=CC(=O)OC2C=C1	0.4988832
CC(C)CCNCC(C)CO	0.4993408
CCCC(CC)C1C(O1)O	0.4995297
C1CCC(CC1)C(CO)O	0.4998208
CCCCCCCCOCCCN	0.5008098
CCCC1CC(OC1(C)C)O	0.5038617
C(CCC=CCCCC)O	0.509194
CC(C1=CCCC=C1)(C(=O)N)O	0.5155014
CCCCCCNCCO	0.516244
CCCC(C(CC)(CO)O)O	0.5178109
C1CCC(CC1)N2CC(C2)O	0.5185901
CCCC(C)CCCO	0.5187504
CCCCCCC(C)NC(C)O	0.5210652
C=C1C2=CC=CC=C2NC(=O)O1	0.5213711
CC(C)(C)CCOCCCO	0.523494
C=CCCCCOCC1CO1	0.524275
CCC1(C2CCC(O2)(CO1)O)C	0.5243557
C=CCCCCCC(O)OC=C	0.5275431
C1=CC=C2C(=C1)C=C3C(=N2)NC(=O)N3	0.5281304
CCCCC(C)CC(O)O	0.5286994
CCCCCOCC(CN)O	0.5299531
CC(C)CCCC(C)CO	0.5312353
CCCCC(CC)(CCO)O	0.536699
CCCCCCC(C)(C)OO	0.5369612
CNC1CCCC(C1)COO	0.540551
CCC(C)CC(CO)O	0.5408992
C1C2=CC=CC=C2C(C(=O)O1)O	0.5412624
C#CCC1=CC=CC(=O)C1=O	0.5443957
CC1=CC2=C(C=C1)NC(=O)NCC2	0.5458412
CC(C)CCOCCCNCCO	0.548697
CCCCCCCCCOC	0.5509221
CCCCCNCC(=O)NCCN	0.552574
CC1=CC(=C(C(=C1)C)N=O)C	0.552945
CCCCCCC(CO)OC	0.5533438
CCCCCCCCOC(CC)N	0.5552987
CCC(CC)CCCCO	0.5561801
C(CCC=CCCCCC)O	0.5569862
CCCCCOC(C)(CCO)O	0.5570045
CCCCOCCCCC(CO)O	0.5578998
C1=CC=C2C(=C1)C=CC(=O)N2	0.5598503
CCCCCCC(CO)O	0.562415
CCCCCCC(CC)(O)O	0.5631078
CCCCCCOC(CC)CO	0.563913
CCCCC(C)CCO	0.5668626
CCCCCCC(C)(C)C(=O)N	0.5676388
CCCCCC(C)NC(=O)CN	0.5745265
CCCCCCCCCCCOC	0.5751148
CCCCCC(CCO)O	0.5752954
C1=CC(=CC=C1CN)C(=O)NN	0.5760339
CCCCCCCC(C)NCCO	0.5762397
CCCCCCCCCCN(C)O	0.5786482
CCCCCCCC=CCO	0.5792734
CCCCCC(CC(C)O)O	0.5793733
CCCCC(CCCO)O	0.580335

Table 3.16 Continued

Or2W1		Or5P3	
SMILES	Distance	SMILES	Distance
C[C@@H]1CC[C@H](CC1=O)C(=C)C	0	CC1=CC[C@@H](CC1=O)C(=C)C	0
C[C@@H](CCC=C(C)C)CCO	0	CC1=CC[C@H](CC1=O)C(=C)C	0
C(CCCCC)O	0	O=C1OC2=C(C=CC=C2)C=C1	0
C(CCCCC)O	0	C=CC(=O)C1CCCC1	0.3110759
C(CCCCCC)O	0	CC1=CC(=O)C(CC1)C(=C)C	0.3586025
CC(=O)CCCCCCC	0	C1=CC=C2C(=C1)C=CC(=NO)O2	0.3721933
CCC(=O)CCCCC	0	C1=CC=C2C(=C1)C=CNC2=S	0.3723315
C[=C]COC(=O)Cc1ccccc1	0	C1=CC=C2C(=C1)C=COC2=O	0.3945556
C/C(CC/C=C(C)C)C=C\CO	0	C=C1C2=CC=CC=C2ONC1=O	0.3977094
C[=O)CCCCC	0	C1=CC=C2C(=C1)C=CC(=O)N2	0.4009083
C(CCCCC)OC(=O)C	0	CC1=C(C(=O)CC1)CC=C	0.4052338
CCCC(CCC=C(C)C)O	0.0004910	C1=CC(=CC2=C1C=CC[=O]O2)S	0.4053821
CC(CCC(=O)CC1C)C=C	0.0006946	CC1=CCC(CC1=NO)C(=C)C	0.4125247
C1CCC(CC1)C=NC(=O)CO	0.0006946	CC1(C2C1C(=O)C(=C)CC2)C	0.4507506
C1C=CNNC2=CC=CC=C21	0.0006946	CC1C=CC(=O)C12CCCCC2	0.452584
CCCOC1=CC=C(C=C1)C	0.0008507	CCC12CCC(=O)C=C1CC(C2)O	0.4690015
CCCCCC1CCC=CO1	0.0008507	C1=CC=C2C(=C1)C=C(CC[=O]O2)O	0.476153
CC(C)NCCCNCCCN	0.0009821	C1CCC2(CC1)CCC=CC2=O	0.48111
CC(=CCCC(C=C)C=O)C	0.0009821	C=C1CC2CCCCC2C1=O	0.4838933
C=C=CCCCCO	0.0009821	C1=CC=C2C(=C1)C[=S)C=CN2O	0.4840491
CCCCCOCC#N	0.0011803	C1CCC2(CC)CC=CC(=O)C2	0.4924455
CC(C)CCCC(C)C1CO1	0.0012030	CC[=C](=O)CCCC#C	0.4929412
CC(=O)C1CCC(=C)C(=C)C1	0.0012030	C=C1CC2(C1=O)CCCCC2	0.5054476
CC(C)(C)C(=O)OCCCO	0.0012030	CC1=CC2CCCC(=O)C2=CC1	0.5090888
CC(C)C(=C)C(=O)NC1=CC=CC=C1	0.0012784	C1CC=CC2(C1)CC=CC(=O)NC2	0.5100904
CC(C)NC(=O)CC(CCN)N	0.0012941	CC1=C2(C=C)C(=O)NC2=CC=C1	0.5151608
CDCCDCCDCDC1=CC1	0.0012971	C1=CC2=C(C(=C1)O)OC(=O)C=C2	0.516556
CCCCC(=C)N	0.0013893	CC[=O)C1=CCC2(C1)C(=C)CCC2=O	0.5272902
CCCCC=C=CC(C)(C)O	0.0013893	CC12CC(=O)C=C1CC3C2CC=C3	0.527377
CDC(=O)CC1=CC=C(C=C1)C=C	0.0014313	C=C1CCCCCCC1=O	0.5311343
CC(C)NCC1=CC=C(C=C1)NC	0.0014528	C1CC2=C(CC=C1)C(=O)CC2	0.5329122
CN(C)CC#CCCCC#C	0.0014732	CC1=CC2CCCON2C1=O	0.5338993
C=CCCCCO	0.0014736	C1=CC=C2C(=C1)C=C(C(=N)O2)C#N	0.5340002
CC#CC(=O)C1CCCCC1	0.0014736	C=CCC1=CC(=O)CCC1	0.5391452
CCCNIC=C(C=N1)C(C)NC	0.0014736	C1=CC=C2C(=C1)C=CC(=NS2)N	0.5407013
CC1(CC1C(=O)NC2=CC=CC=C2)C	0.0015355	C1CC2(CCC(=O)C=C2)CC=C1	0.542789
CCCC1C(O1)CCCC=C	0.0016288	C1CC(=CC(=O)C1)C2=CC=CS2	0.5451384
CCC(C(=O)OCC(C)C)NN	0.0016288	C1=CC=C2C(=C1)C=CN(C2=O)O	0.5462879
CCCCC1CCC(CC1)N	0.0016288	CC1=CSC2=C1C(=O)C(=C)CC2	0.5476168
CCCCCC#CCC(CC)O	0.0016288	C1=CC=C2C(=C1)C=C(CC(=O)O2)N	0.5512919
CC(C)COC1=CC=C(C=C1)CO	0.0016784	C1=CC(=O)OC2=CC(=C(C=C21)O)N	0.555675
CC1=CC(=CC(=C1)NC(=O)C(C)C)C	0.0016855	CCC1=C(C2=CC=CC=C2C1=N)N	0.5572225
CCN=CC1=CC=C(C01)C	0.0016855	C1=CC=C2C(=C1)C=CC(=O)N2O	0.5589366
CN1C=CC(=NC1=O)NCC=C	0.0016972	CC[=CCC(=O)C=C)C	0.5595898
CCC#CCCCN	0.0017014	C=C1C(=O)OC2=CC=CC=C2N1	0.5600652
CC(C)(CO)CCCCCC=C	0.0017014	C1=CC=C2C(=C1)C=CC(=O)NO2	0.5618963
CC(C)C#CCCCN	0.0017014	CCC(C)C1=CC(OC1=O)C	0.5639112
C1CC(OC1)CCCCN	0.0017555	C1=CC=C2C(=C1)C3=C(N2)NC(=O)C=C3	0.5684292
CCC1=NC(=NO1)C(C(CC)N)O	0.0017671	C1=CC=C2C(=C1)C=CC3=NNC(=O)N32	0.5728021
CCCCN1C=C(N=N1)C=C	0.0017693	C=CC1=CC=C(C=C1)C(=O)N	0.5738116
CC(CCC1=CC=C(C=C1)O)C)NC	0.0018275	CC1=CC(=O)CC1CC=C	0.575584
CDC(=O)COCC#C	0.0019408	C1CC(=CC(=O)C1)C2CCC=C2	0.5765915
C1CCC(CC1)NCC2=CC=CN2	0.0019950	CCC1=CC2CCCN2C1=O	0.5776696
C1CC(OC1)CDC2=CC=CC=C2	0.0019950	C1=CC=C2C(=C1)NC(=O)C=CS2	0.5831019
CC(=CCCC1CC1C=O)C	0.0019643	CC1=C2C=CC(=O)NC2=CC=C1	0.5841847
CCCCN1CCN(CC1)N	0.0019643	C1CC(=O)C2=C(C=C1)C=CC(=C2)O	0.5848835
CCC(C)(CCCC(=C)C)O	0.0019643	C1=CC2=CN=C(C=O)N=C2C=C1)N	0.5849539
CN(C=O)NCCCCCCO	0.0020250	C1=CC=C2C(=C1)C=CN3C2=NNC3=O	0.5850339
CCCCCC(CC(C)C)O	0.0020250	C=C1CC2=CC=CC=C2OC1=O	0.5862692
C1CCDC(C1)OCCCO	0.0020250	CCC(=C)CC1=CCCC1=O	0.5878639
CCCCCCCC=N	0.0020250	CC1=CC(=O)OC2=C1C(=C(C=C2)O)N	0.5881438
C1CC2=CC=CC(=O)C2NC1	0.0020614	CC12CC=CCC1CC(=O)C=C2	0.5900818
CCCC=C=CC(C)C)C=O	0.0020835	C=CCC1CCCC(=O)C=C1	0.5902211
CCC(=O)CCC#CC	0.0020835	C1CC2(CCC=CC2=O)CC=C1	0.5931621
CC(=CCCC(=C)C=C)CO	0.0020840	CC1=CC(=O)NC2=C1C(=CC=C2)N	0.5952647
CCCCCC(CC)O	0.0020840	CC1=C(CCC1(C)C)C(=O)OC	0.5982892
CCO[=O)N1C2CCC(O1)C=C2	0.0021289	CC1(CC1C(=O)C=C)C	0.5994571
CCCOCCC(=N)OC	0.0021289	C1C2=CC=CC=C2C=C(C1=O)O	0.6030376
C1=CC(=CC=C1CCCO)N	0.0021191	CC1=CCC(CC1=O)C2(CO2)C	0.6032878
C=CCCCCC1=NCCC1	0.0021191	CC1C=C(C(=O)O1)C2=CC=CS2	0.6051044
CCCCC1=CN=C(C=C1)C	0.0021411	CC1=CC(=O)CCCC1CC=C	0.6074293
CCCCC1=CC=C(C=C1)N	0.0021411	CC[=C]C1CCC=C1C[=O)C	0.6076714
CCCC1C(O1)C(=O)NC2CCC2	0.0021411	CC1=CC(=O)OC1C2=CC=CC=C2	0.6080884
CC(=CCCC(CO)C=C)C	0.0021411	CC1=CC(=S)C2=CC=CC=C2O1	0.6085333
CC(C)CC(=O)C=C(C)C=C	0.0021411	C1C2=CC=CC=C2C(=O)C1=CN	0.60878
C1CC(CC1)CCCCD	0.0021411	C1=CC2=C(C=CC(=O)O)2C=C1N	0.6088277
CCCCN1C=C(C(=N1)C)N	0.0021411	C1CC2(CC3CC2C=C3)C=CC1=O	0.6095803
C#CCCCCN	0.0021411	C=CCC1CC(=C)C(=O)O1	0.610739
CC(=CCCC1=NCCC1)C	0.0021411	C1=CC(=O)OC2=C1C3=C(C=C2)NC=C3	0.6113136
CCCCC(=O)C(C)(C)CC	0.0021411	CCC1=C(N2CCCN2C1=O)C	0.6116292
C1CC1NCC(=O)C2=CC=CC=C2	0.0021683	CC[=C]C1=CC=C(C=C1)O	0.6117204
C=CCOC1CC(C1)C2CC(=O)C2	0.0021683	C=C1CCC2(CC1)CC(=O)C=C2	0.6117881
CCC1=COC(=N1)NC(=O)C	0.0022269	C1=CC=C2C(=C1)C[=CC(=O)O2)N	0.6118856
CC1=CC=C(C=C1)C2=NC=C(N2)C	0.0022386	CC1=CC(=O)CC(CC1)C)C	0.6126708
CCOC1CNCCO1	0.0022560	CC1CCC(=C1=O)C(C)C	0.6138947
CC(=C)CCCC(C)(C=C)N	0.0023039	CC12CCC(=O)C=C1CCCC2O	0.6158562
CC(=CCCC=C[C)CO)C	0.0023039	CC[=O)C1=CCC(C(C1)O)C)C	0.6161952
CC#CCC1(CCCCC1)O	0.0023039	CC1=C(C2=CC=CC=C2C1=N)N	0.6163228
CDCCCCC#C	0.0023039	C1=CC=C(C=C1)C2=CC(=O)NN2	0.6175588
CC(C)C1C(O1)CCCC=C	0.0023039	CCC12CCC(=O)C=C1CCCC2=O	0.6185925
CCC1C(O1)C#CC2=CC=CC=C2	0.0023328	CCC1=C(C(=O)O1)C[C)C	0.6204279
CC(=CCC1=C/C(=CC=C1)N)N)C	0.0023441	C1=CC=C2C(=C1)C=C(C(=O)N2)O	0.6213582
C[CNCC=O)CC=O	0.0023606	CC1=CC2=C(C=C(C=C2)N)OC1=O	0.6237157
CCOC1CCC(=CC1	0.0024060	C1=CC2=C(C=C1O)C=C(C(=N)O2)C#N	0.6242207
CCCCCC1=C(NC=C1)C	0.0024060	C1=CC(=CC2=C1C=CC(=O)O2)N	0.6243975
C=CCCCC#CCO	0.0024060	C1=CC=C2C(=C1)C[=S)C=CN=N2	0.6244215
CCOC[=O)CNCC#C	0.0024111	C1CC(C(=O)C=C1C)C	0.6247424

Table 3.16 Continued

CHAPTER IV:

A New Generation of Safe DEET Substitutes that are Strong Olfactory and Gustatory Repellents of Mosquitoes and Flies

INTRODUCTION

Blood-feeding insects transmit deadly diseases such as malaria, Dengue, Filariasis, West Nile Fever, Yellow fever, Sleeping sickness and Leishmaniasis to hundreds of millions of people, causing untold suffering and more than a million deaths every year. Insect repellents can be very effective in reducing vectorial capacity by blocking the contact between blood-seeking insects and humans, however they are seldom used in disease-prone areas of Africa and Asia due to high costs and need for continuous application on skin.

N,N-Diethyl-m-toluamide (DEET) has remained the primary insect repellent used for more than 60 years in the developed world. DEET is a solvent, melting several forms of plastics, synthetic fabrics, painted and varnished surfaces(Krajick, 2006). Additionally, DEET has been shown to inhibit mammalian cation channels and human acetylcholinesterase, which is also inhibited by carbamate insecticides commonly used in disease endemic areas (Corbel et al., 2009). These concerns are enhanced by the requirement of direct and continuous application of DEET to every part of exposed skin in concentrations that can be as high as 30-100 percent. Several instances of increased resistance to DEET have also been reported in flies (Reeder et al., 2001), *Anopheles albimanus* (Klun et al., 2004), and *Aedes aegypti* (Stanczyk et al., 2010). Moreover, mosquito strains with resistance to pyrethroid insecticides, the main line of defense against mosquitoes in developing countries, are spreading (Butler, 2011). There is

therefore an urgent need to develop safer and improved alternatives to DEET. The other major barrier in developing new repellents is the time and cost of development. It has been suggested that >$30M and several years may be required for identification (Gupta, 2007) and adequate human-safety analysis of new repellent chemistries.

A major limitation to finding effective substitutes is that the molecular targets in adult mosquitoes through which DEET causes repellence are unknown. Recent studies have put forward a few competing theories about mechanism of action, but demonstration of a causal relationship is lacking. Some reports show that pure DEET causes inhibition (Ditzen et al., 2008) or mild electrophysiological modulations of neural responses to weakly activating odors in *Drosophila* antennal olfactory neurons (Pellegrino et al., 2011), but whether such effects cause repellency in either *Drosophila* or mosquitoes are unknown. On the other hand a DEET-responsive receptor, Or42a, has been identified in the maxillary palps of *Drosophila*, however Or42a is also known to be a broad-spectrum detector for several food-associated odorants that elicit attraction (Syed et al., 2011). Other studies in *Drosophila* have shown that DEET is also detected through bitter taste neurons causing contact-avoidance (Lee et al., 2010; Weiss et al., 2011). Although the three proposed mechanisms confound appropriate target choice for high-throughput screening, *Drosophila* might provide a tractable model for discovery of new DEET substitutes via direct behavioral screening of compounds.

Here, we describe identification of novel repellent odors that are safe, inexpensive, and effective at repelling *D. melanogaster* and *A. aegypti*. We first demonstrate that *D. melanogaster* and *A. aegypti* may utilize their gustatory system and olfactory system to different extents to avoid DEET, with mosquitoes relying more heavily on olfactory pathways. We then use a novel *chemical informatics* approach that

uses supervised training from known repellents to identify important structural features that are responsible for avoidance behavior. Using these features we predict novel repellents from a very large untested odor space comprising a large purchasable odor set and a natural odor library, the latter providing many chemicals that are safe and already approved for human use by both United States and European governmental safety organizations. We select four odors from the 150 natural odor library predictions and demonstrate that all four are able to cause strong avoidance in both *D. melanogaster* and *A. aegypti*. Due to the large number of candidates we are able to select new repellents with ideal properties; safe for human consumption, do not dissolve plastics and mild and pleasant aroma. These results suggest that our integrative approach of computational predictions and behavioral analysis can revolutionize the discovery of safe and effective repellent odors that could be very useful in our struggle against the increasing spread of insect disease vector species. Although it has been several years since odor receptors were identified in vertebrates and invertebrates, very few odor receptor targets have been identified for known behavior modifying odorants. This study therefore has broader implications since the approach presented can be used for identification of improved behavior modifying odorants for any organism, even if the target odor receptor is unknown. Moreover upon identification of target receptors, the same methodology can be easily adapted for a receptor-activity based approach.

RESULTS AND DISCUSSION

Contributions of Olfaction and Gustation in DEET Avoidance in *Drosophila*

In order to select a reliable behavior assay, we first tested for repellency in the well-established T-maze. We found that *Drosophila* do not avoid the arm containing DEET (10%) in this assay (Figure 4.1A), perhaps due to its low volatility. By introducing a well known repellent CO_2 (0.66%) in the other arm, we asked whether we could observe an effect by forcing flies to enter the DEET arm closer to the DEET source. We find a marginal, but not significant, reduction in avoidance to CO_2, suggesting that DEET is not an effective olfactory repellent for *Drosophila* in this assay (Figure 4.1A). *Drosophila* do however show strong avoidance when they are allowed to come in close contact to DEET in a plate-based 2-choice trap assay as described previously(Syed et al., 2011), where they can sense DEET-treated filter paper positioned at the entrance of a trap with both the olfactory and gustatory system (Figure 4.1B). *Orco* mutant *Drosophila* that lack functional odor receptors belonging to the *Or* family show ~30% reduction in avoidance to DEET (Figure 4.1B). The simplest interpretation of these results is that although olfactory receptor pathways participate in the avoidance behavior, the response mediated by the gustatory system is sufficient to generate strong avoidance of *Drosophila* to DEET.

Aedes aegypti Detect and Avoid DEET Primarily using Olfaction

To test whether the gustatory system of mosquitoes plays a substantial role in DEET repellency as well, we developed a modified hand-in-glove assay (described in Supplementary Methods) (Figures 4.1C, 4.1D, and 4.2). The assay allows us to

quantitatively analyze the repellent effects of DEET on mosquitoes attracted to a human arm, without being able to bite. Female *Aedes aegypti* mosquitoes show an equally strong avoidance behavior to DEET in both the contact and non-contact versions of the assay (Figure 4.1D). For rare landings, the time spent on the net before escape is marginally lower, but not significantly different, when direct contact with DEET was permitted (Figure 4.1E). In order to test whether the reduction in attraction is caused by a direct detection or a masking/scrambling of skin-odor detection, we measure attraction of female *Aedes* to humidity and warmth and demonstrate a significant dose-dependent reduction in presence of DEET (Figure 4.1F). These results indicate that the female *Aedes* primarily use the olfactory system to sense DEET directly and avoid it. This is consistent with previous observations that *Culex quinquefasciatus* (Syed and Leal, 2008) and *A. aegypti* (Turner et al., 2011) avoid DEET presented as a volatile stimulus directly. In fact a DEET-sensitive neuron type has been identified in *C. quinquefasciatus* (Syed and Leal, 2008) and *A. aegypti* (Stanczyk et al., 2010), however it is not known whether these neurons are responsible for repellency, or which odor receptors they express. A broadly tuned larval odor receptor AgOr40 has been shown to respond to DEET (Liu et al., 2010; Xia et al., 2008), however its role in avoidance behavior in adults has not been determined.

A Computer can be trained to Predict Repellent Behavior from chemical structure

Since behavior assays in *Drosophila* and mosquitoes appear to differ in requirements for gustatory and olfactory exposure for repellency (Figure 4.1G), we would require use of both systems to reliably identify improved DEET substitutes. However volatile chemical space that can be exploited to find such improved DEET

substitutes is vast (potentially >400,000) thus using behavior assays alone is unfeasible from the perspective of time and cost of chemical purchase. Moreover, since a protein target for DEET action is unknown, high-throughput screening, or sophisticated computational protein-ligand docking based approaches to identify new ligands are also not possible.

To circumvent these problems so as to enable selection of compounds for behavior assays, we developed a novel chemical informatics approach. We hypothesized that the unknown target protein is recognizing specific structural features of DEET and other known structurally related repellents. By identifying structural features that are shared amongst DEET and the known repellent compounds, we can utilize them to rapidly screen an extremely large number of compounds *in-silico* to identify novel repellents, thus greatly reducing both the cost and time required. We assembled a training set of known repellents that included: the two approved ones DEET (a carboxamide) and picaridin; and also 34 N-acyl piperidiens (Katritzky et al., 2008) that are structurally related to picaridin; eucalyptol, linalool, alpha-thujone, beta-thujone (Kline et al., 2003; Klocke et al., 1987; Syed and Leal, 2008) and a structurally diverse panels of odors that are not expected to elicit repellency via similar target receptors (Carey et al., 2010; Hallem and Carlson, 2006). The study where the 34 n-acyl piperidiens were identified also showed that a chemical-structure-based approach could be successfully applied to predicting repellency (Katritzky et al., 2008). For our analysis, compounds from different sources were approximated into a single metric of “protection duration” as a rough indicator of repellency. The non-repellent diversifying training set of odors were assigned protection times of zero, while the approved repellents DEET and Picaridin were assigned the highest value since these would have the properties

important for regulatory approval. Compounds were clustered using Euclidean distance and hierarchical clustering based on differences in repellency values, and a set of 5 compounds with the highest activity that clustered together was classified as “training repellents”.

We expect that only specific structural features of the repellent odors will interact with target proteins to elicit repellent responses, and not the entire molecule. We assumed that identification of structural features that are shared across repellent odors would enable a search for these features within a large chemical space, potentially identifying novel repellents (Maldonado et al., 2006). We decided to focus on a descriptor-based computational approach that is effective in structure analysis and is efficient in subsequently searching a large chemical space rapidly. We calculated mathematical values for 3,242 molecular descriptors, that describe the 3-D structure of a chemical, for our 201 compound training set and using a Sequential-Forward-Selection method (Haddad et al., 2008) we incrementally identified a unique subset of 18 descriptors that were highly correlated with repellency (correlation of 0.912)(Figures 4.3A, 4.3B, and 4.3C). As expected the repellent odors cluster together in the training set if the optimized descriptor subset is used to calculate Euclidean distance amongst them (Figure 4.3B). These 18 molecular descriptors represent a collection of predominately 2D and 3D descriptor types. Inspection suggests that 6 member rings, carbon-nitrogen distances, tertiary amides, and oxygen placement are prominent in the optimized subset. Interestingly, although repellents are topically applied chemicals, the Ghose-Viswanadhan-Wendoloski drug-like index is selected, which is an aggregate descriptor that usually suggests similarity of chemical features important for drugs (Ghose et al., 1999).

In order to improve the predictive ability of the chemical informatics approach we used the optimized descriptor set to train a Support Vector Machine (SVM), which is a well-known supervised learning approach (Cortes and Vapnik, 1995)(Figure 4.3A). To validate the predictive ability of our approach, we performed 5-fold cross-validation using SVMs on the training set. Each cross-validation run excluded ~20% of the repellents as a test-set, while the remaining repellents were used to train a SVM. We predicted repellency for the withheld test-set odors using the trained SVM. This operation was repeated 5 times, each trial performed excluding a different subset of the training set, and the whole process was repeated 20 times for consistency. A mean Receiver-Operating-Characteristic (ROC) analysis curve representing the prediction accuracy was generated, and the Area-Under-Curve (AUC) value was determined to be 0.994, indicating that the *in-silico* approach was extremely effective in predicting repellents from compounds that are excluded from the training set (Figure 4.3D).

High-Throughput *In Silico* Screen identifies Safe Natural Odors as repellents

We use the 18 optimized-descriptors and SVM method to screen *in-silico* a large chemical library consisting of >440,000 chemicals from a database called eMolecules of putative volatiles. Upon inspection, we find the top 1,000 predicted compounds (0.23% of hits) represent a structurally diverse group of chemicals that retain some structural features of the known repellents (Figures 4.4A and 4.4B). We computed logP values of the 1,000 compounds to identify ones that are predicted to be lipophilic (logP >4.5) thus allowing for selection of other compounds that are less likely to pass through the skin barrier in topical applications (Walker et al., 2003) (Figure 4.4B). In addition, we

computed the predicted vapor pressures of these chemicals since volatility may predict ability of long-term protection vs. increased spatial domain of action (Figure 4.4B). Taken together the results of the screen present a very large collection of novel predicted repellents with desirable properties, identified via a computationally guided search of odor space.

Although the *in-silico* screen bypasses the challenge of not knowing the protein target, the most significant challenge lies in identifying effective repellent substitutes for DEET that are affordable and safe and that can be rapidly approved for human use. In order to identify compounds that fit these criteria, we applied our *in-silico* screen to an assembled natural odor library consisting of >3,000 chemicals identified as either originating from plants, insects, or vertebrate species or compounds already approved for human use as fragrances, cosmetics or flavors (See Methods). Using the trained SVM and optimized descriptor set on the natural library, we identified the top 150 ranked predicted repellent compounds. Predicted repellents share similarity in some parts of the structure while providing a diverse set of compounds (Figure 4.4C). For example, several anthranilates and pyrazines were identified that represent novel groups of safe and natural compounds, although such compounds were absent from the training set. These 150 compounds were arranged by predicted logP and vapour pressure values to provide a high-priority list of candidates for behavioral testing (Figure 4.4C).

Candidate Natural Odors Strongly Repel *Drosophila* and mosquitoes

In order to test for repellency we first used *Drosophila melanogaster* in the 2-choice trap assay that previously showed DEET repellency (Reeder et al., 2001; Syed et al., 2011). We selected 4 affordable compounds from the list, (Methyl N,N-Dimethyl

anthranilate, Ethyl anthranilate, Butyl anthranilate, 2,3-dimethyl-5-isobutyl pyrizine) the first 3 of which have been thoroughly tested for human use, have excellent safety profiles, a very mild and pleasant aroma like grapes, and have already been approved for human consumption/oral inhalation by the FDA, World Health Organization and European Food Safety Authority (EFSA, 2008; JECF, 2007) (Figure 4.4D). The pyrizine is a natural compound produced by ants as a trail pheromone (Tentschert et al., 2000). All 4 of the predicted compounds showed a very strong repellent effect on *Drosophila* across multiple doses tested using the trap-based 2-choice assay (Reeder et al., 2001; Syed et al., 2011) (Figure 4.4E). The effect was consistent at two different time points, 48 hours and 72 hours after the start of the assay.

One of the major disadvantages of DEET is its property of solubilizing plastics and synthetic materials (Krajick, 2006), which affects its usefulness by the military, and at home. We tested the ability of the 4 repellents to dissolve a 3 x 3mm square of vinyl. The plastic square disappeared completely in DEET within 6 hrs; however in the 4 predicted repellents there was no significant difference in the weight of the vinyl squares after 6 hrs, and even at the 30hr time point (Figure 4.5A).

In order to test whether these promising candidates were in fact repellent to mosquitoes, we performed behavior trials using the modified hand-in-a-cage assay. Notably, we find that all 4 compounds applied at 10% concentration demonstrate substantial repellency. Both the fraction of mosquitoes present on the net window over the arm throughout the duration of the assay, and the cumulative number of mosquitoes present on the net window were substantially lower in the test compounds as compared to solvent controls (Figures 4.5B, 4.5C, 4.6). For the mosquitoes that did land on the

repellent treatment, the escape index, as measured by the frequency of take off, was substantially higher as well (Figure 4.7).

A New Generation of Safe and Effective Repellent Odors

Taken together we have identified a number of affordable and safe compounds with repellent properties, which establishes a significant advance toward the identification of DEET substitutes that are excellent candidates for regulatory approval for human and animal use (Figure 4.5D). This was made possible by a computational screening strategy that identified shared structural features of existing repellents to use for rapid identification of structurally related novel repellent compounds. Apart from the 4 compounds we have behaviorally identified as repellents, we have also used the screen to identify ~1000 novel compounds and ~146 additional natural compounds, many approved for use in human food and cosmetics, that may lead to several additional effective repellents for deadly insects. The repellency strategy can also have great potential if used in combination with other behavior control strategies such as masking of CO_2-mediated attraction behavior or population control by trapping as a part of an integrated pull-mask-push approach (Turner et al., 2011; Turner and Ray, 2009). Since several of these compounds are affordable, repellent for fruit flies and approved for human consumption, they may have implications for control of agricultural pest insects as well that cause billions of dollars in crop loss as well as in protecting animals and pets that have tendencies to lick their skin. Moreover such substitutes may also have implications for control of DEET-resistant insects. We expect such repellents that are safe, affordable and do not solubilize plastics to pave the way for formulations that can

be used in control of insect-human contact in disease endemic areas of the world and to provide an important line of defense against deadly diseases.

Figure 4.1: Contribution of olfaction and gustation in DEET avoidance

(A) Preference index of *Drosophila* in a T-Maze to DEET (10%), CO_2 (0.66%), and opposing gradient of both compounds. N = 8-39 trials, ~40 flies/trial, error bars = s.e.m., ** = p-value < 10^{-4}. **(B)** Preference index of *Drosophila* wild-type and *Orco-/-* flies to DEET (10%) in a two choice trap assay measured at 48 hrs and 72 hrs after start. N = 5-10, 10 flies/trial, error bars = s.e.m. ** = p-value < 10^{-3}. **(C)** Photograph of the hand-in-glove assay to measure repellency in mosquitoes. **(D)** Relative attraction of female *Aedes aegypti* in hand-in-glove assay measured as a percentage of mosquitoes spending >5-sec on the net covered window of glove, measured at 1 min intervals. (Left) DEET (10%) treated netting placed at the top level allowing contact, or (Right) at the lower level not allowing contact. N=5 trials for each, 10 mosquitoes/trial. **(E)** Average time spent per mosquito on net for each landing event. **(F)** Schematic of assay and mean number of female *Aedes* mosquitoes at net probing a heat and humidity source >5-secs. Mean calculated across 10 time-points, every 30-sec interval. DEET (10%) treated or Acetone (Solvent) treated netting is placed at a distance that mosquitos can not contact. N=3 trials for each, 10 mosquitoes/trial, error bars = s.e.m. p-value ** = < 10^{-3}, *=<0.05. **(G)** DEET mediated repellency in *Drosophila* is mediated by both olfaction and gustation in *Drosophila* while primarily by olfaction in *Aedes aegypti*.

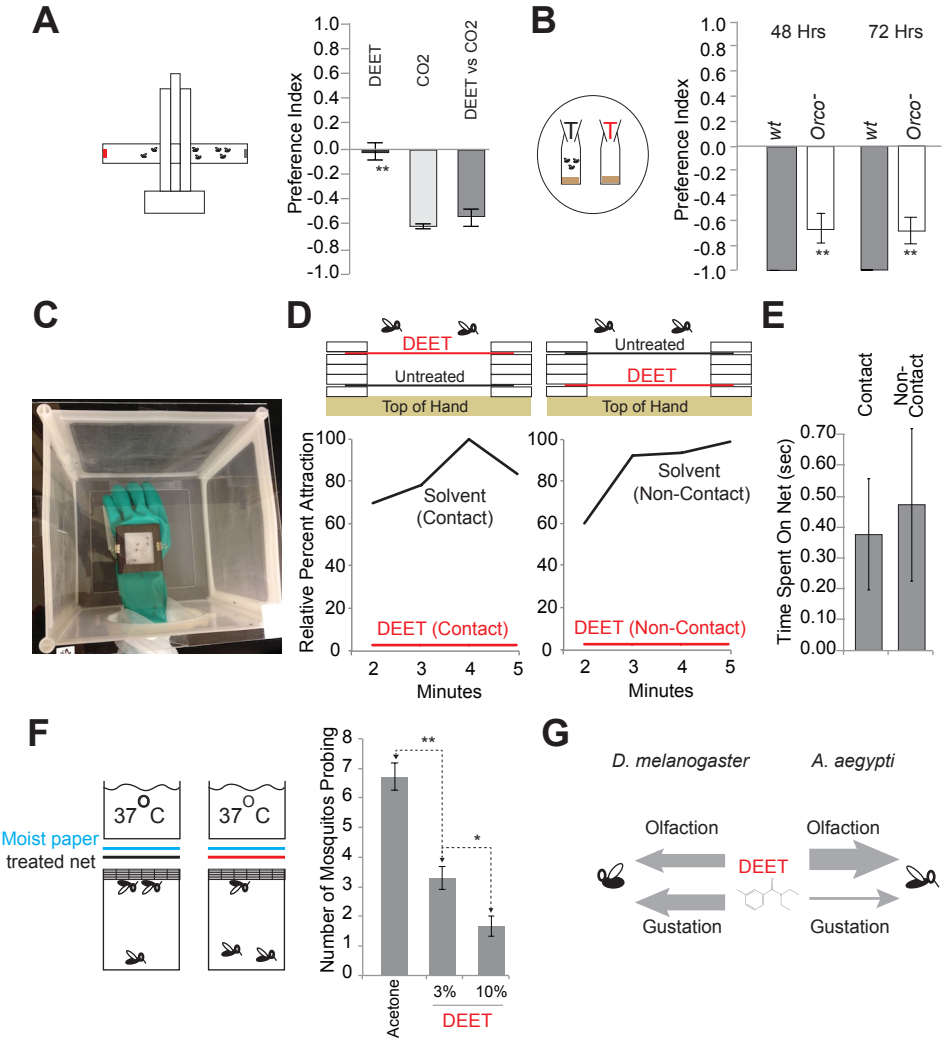

A
Preference Index
DEET
CO2
DEET vs CO2
B
48 Hrs
72 Hrs
wt
Orco-
C
D
DEET
Untreated
Top of Hand
Relative Percent Attraction
Solvent (Contact)
DEET (Contact)
Solvent (Non-Contact)
DEET (Non-Contact)
Minutes
E
Contact
Non-Contact
Time Spent On Net (sec)
F
37°C
Moist paper
treated net
Number of Mosquitos Probing
Acetone
3%
10%
G
D. melanogaster
A. aegypti
Olfaction
Gustation

Figure 4.1

Figure 4.2: Mosquito behavioral assay glove setup

The assay glove is assembled in the following order: rubber glove with window cut into the hand, magnet glued around the cut window, control or test odor treatment mesh, three spacer magnets that prevent mosquitoes from biting through to the hand, untreated mesh to prevent mosquitoes from touching the treated mesh, and finally a top magnet. One metal clip is then used on each side of the stack to further reinforce the arrangement of magnets and mesh.

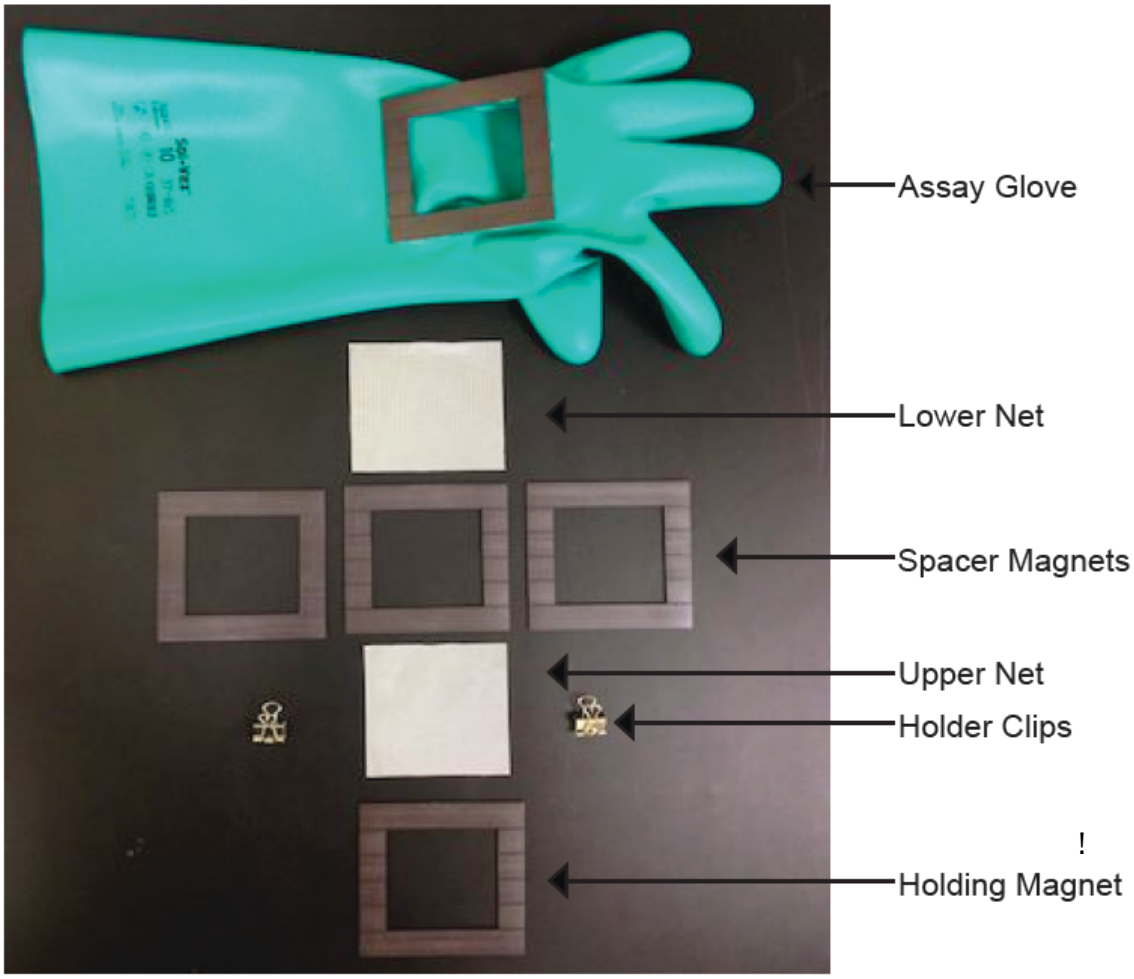
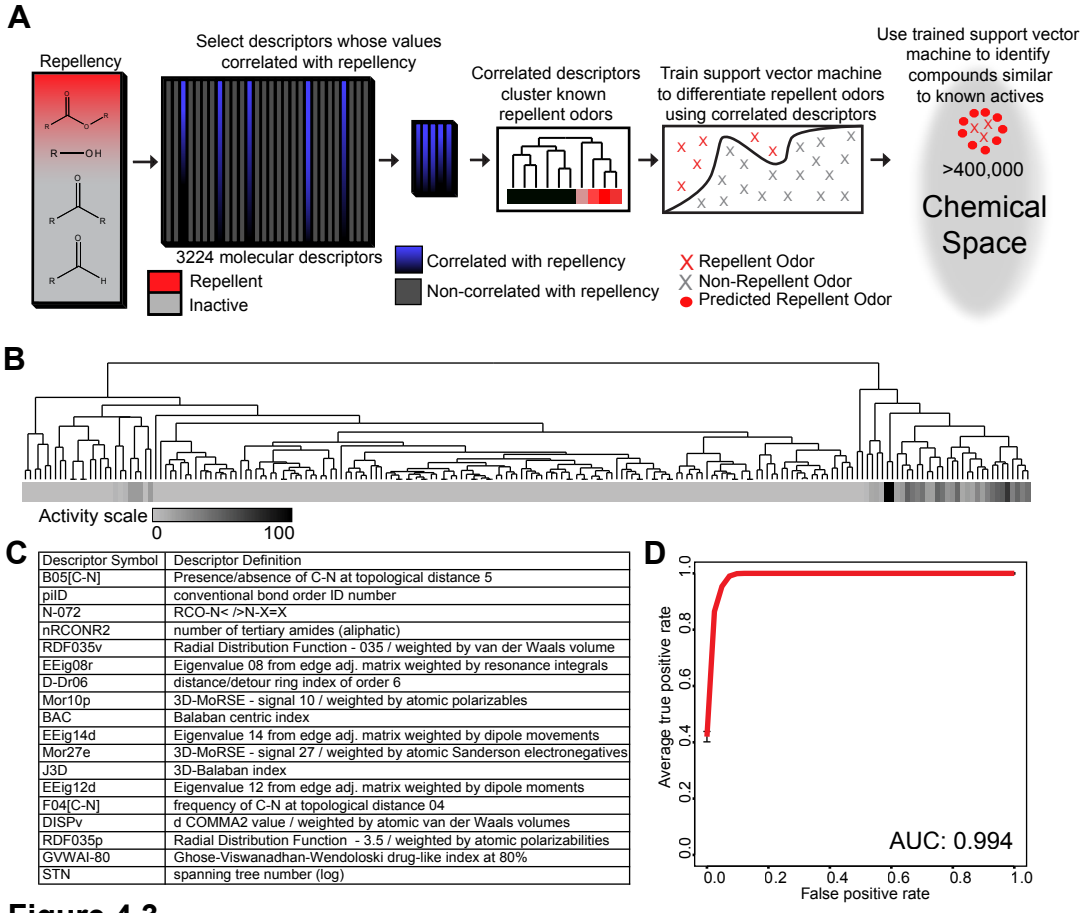

Assay Glove
Lower Net
Spacer Magnets
Upper Net
Holder Clips
!
Holding Magnet

Figure 4.2

Figure 4.3: A chemical informatics method to predict repellency

(A) Overview of the cheminformatics pipeline used to identify novel DEET-like ligands from a larger chemical space. **(B)** Hierarchical cluster analysis of the 201 odorants of the training set using the optimized descriptor set to calculate distances in chemical space. **(C)** Repellency-optimized descriptor symbols and brief descriptions arranged according to order in which they were selected for the optimized set. **(D)** Receiver-operating-characteristic curve (ROC) representing computational validation of repellent predictive ability. The mean true-positive value from 20 independently run 5-fold cross validations is plotted, where ~20% of the dataset was left out of training-set as a test-set for each run. The mean area under the curve (AUC) is provided.



Descriptor Symbol	Descriptor Definition
B05[C-N]	Presence/absence of C-N at topological distance 5
piID	conventional bond order ID number
N-072	RCO-N< />N-X=X
nRCONR2	number of tertiary amides (aliphatic)
RDF035v	Radial Distribution Function - 035 / weighted by van der Waals volume
EEig08r	Eigenvalue 08 from edge adj. matrix weighted by resonance integrals
D-Dr06	distance/detour ring index of order 6
Mor10p	3D-MoRSE - signal 10 / weighted by atomic polarizables
BAC	Balaban centric index
EEig14d	Eigenvalue 14 from edge adj. matrix weighted by dipole movements
Mor27e	3D-MoRSE - signal 27 / weighted by atomic Sanderson electronegatives
J3D	3D-Balaban index
EEig12d	Eigenvalue 12 from edge adj. matrix weighted by dipole moments
F04[C-N]	frequency of C-N at topological distance 04
DISPv	d COMMA2 value / weighted by atomic van der Waals volumes
RDF035p	Radial Distribution Function - 3.5 / weighted by atomic polarizabilities
GVWAI-80	Ghose-Viswanadhan-Wendoloski drug-like index at 80%
STN	spanning tree number (log)

Figure 4.3

Figure 4.4: Identification of repellents using *in-silico* screening of a large chemical space

(A) Examples of two approved repellents DEET and Picaridin, and two unapproved repellents (Katritzky et al., 2008). **(B)** (Left) Representative predicted repellent odors from the odor library of >400,000. Computationally determined (Middle) LogP values and (Right) vapor pressure values for the top ranked 1000 predicted repellent compounds. **(C)** (Left) Representative structures from the top 150 predicted repellent compounds from the natural odor library of >3,000. Computationally determined (Middle) LogP values and (Right) vapour pressure values for the top ranked 150 predicted repellent compounds. Color arrowheads indicate values for DEET and odors selected for behavior experiments from the natural library indicated in **D. (D)** Preference index of Drosophila adults to predicted repellents at three different concentrations in a two choice trap assay measured after 24 hrs and after 48 hrs. N = 7-10 trials each treatment at 48 hrs and 3-10 at 24 hrs (trials with <30% participation were excluded), 10 flies/trial, error bars = s.e.m

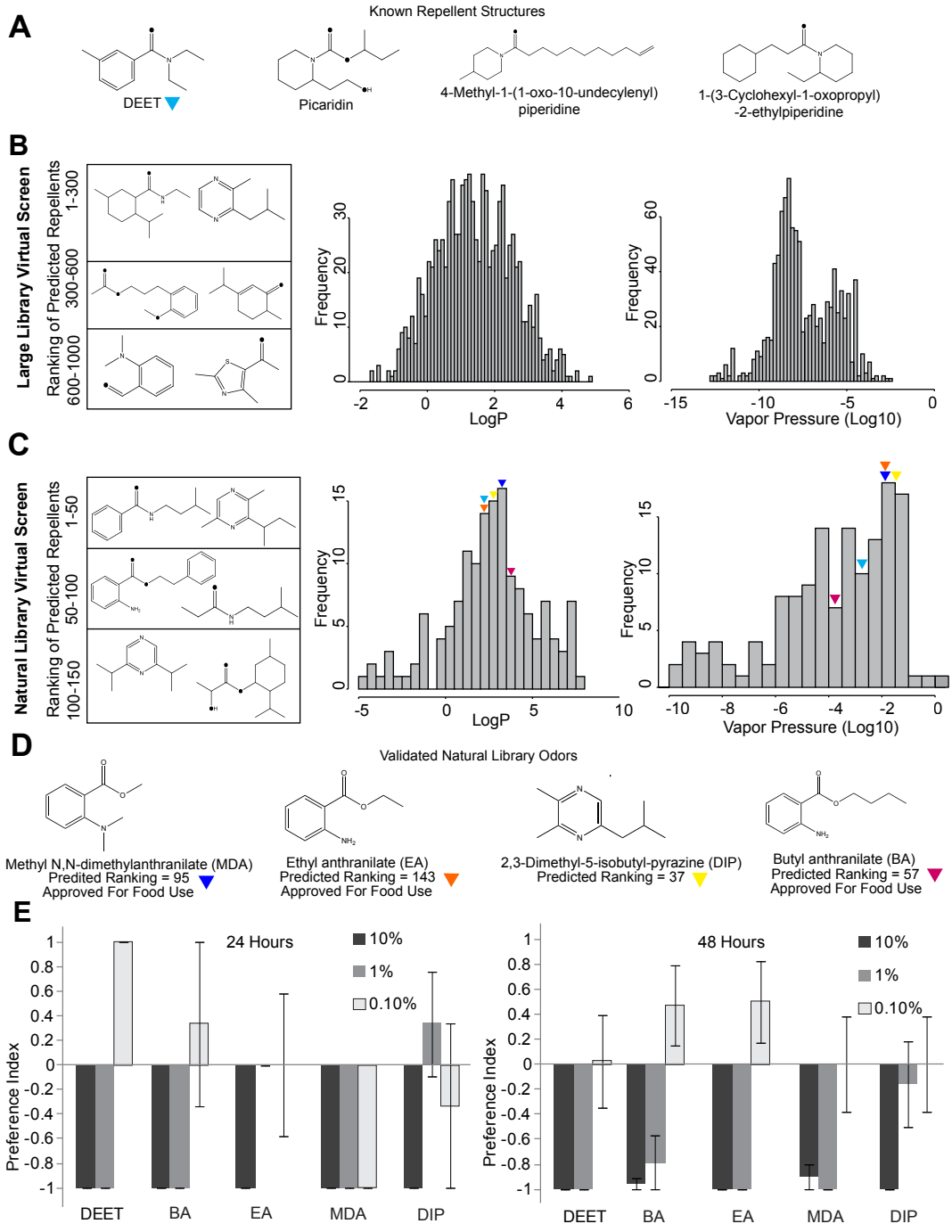
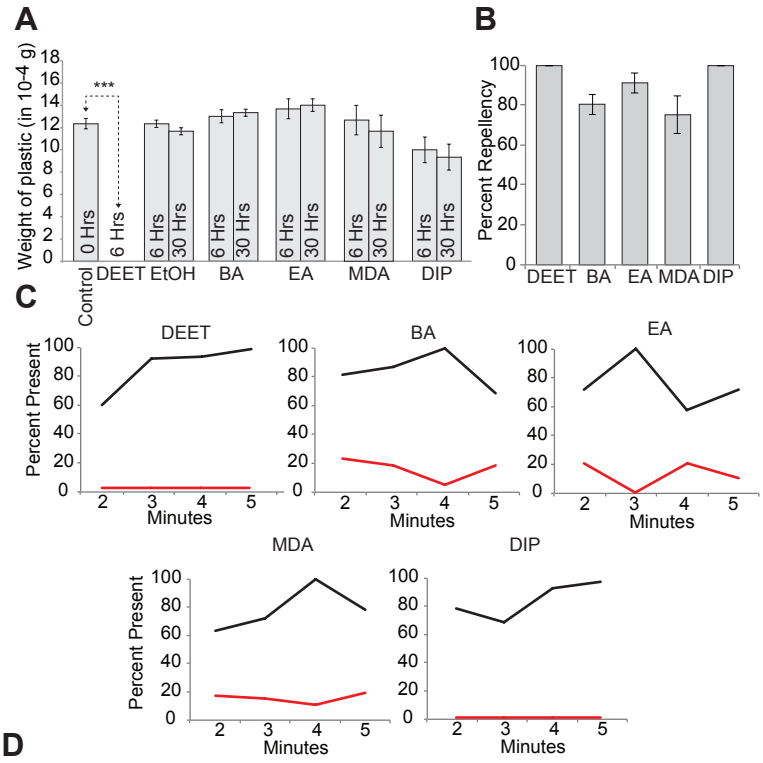

A
Known Repellent Structures
DEET
Picaridin
4-Methyl-1-(1-oxo-10-undecylenyl) piperidine
1-(3-Cyclohexyl-1-oxopropyl) -2-ethylpiperidine
B
Large Library Virtual Screen
Ranking of Predicted Repellents
1-300
300-600
600-1000
Frequency
LogP
Vapor Pressure (Log10)
C
Natural Library Virtual Screen
Ranking of Predicted Repellents
1-50
50-100
100-150
Frequency
LogP
Vapor Pressure (Log10)
D
Validated Natural Library Odors
Methyl N,N-dimethylanthranilate (MDA)
Predited Ranking = 95
Approved For Food Use
Ethyl anthranilate (EA)
Predicted Ranking = 143
Approved For Food Use
2,3-Dimethyl-5-isobutyl-pyrazine (DIP)
Predicted Ranking = 37
Butyl anthranilate (BA)
Predicted Ranking = 57
Approved For Food Use
E
24 Hours
48 Hours
10%
1%
0.10%
Preference Index
DEET
BA
EA
MDA
DIP

Figure 4.4

Figure 4.5: A new class of mosquito repellents with desirable safety profiles

(A) Mean weight of vinyl pieces following submersion in indicated compounds for indicated amount of time. N = 3, error bars = s.e.m., *** = p-value < 10^{-5}. **(B)** Cumulative percentage of repellency across minutes 2,3,4 and 5 of indicated treatment (10%), in comparison to the appropriate solvent control. N=5 trials/treatment, 40 mosquitoes/trial. **(C)** Mean percentage landing as measured by mosquitoes spending at least-5 secs on the protective window of glove, measured at different time-points in the 5-min assay. **(D)** Summary of desirable properties of new insect repellents reported in this study.



D

	Repellent To Drosophila	Repellent To Aedes	Dissolves Plastic	Approved For Consumption
DEET	**Yes**	**Yes**	Yes	No
MDA	**Yes**	**Yes**	**No**	**Yes**
BA	**Yes**	**Yes**	**No**	**Yes**
EA	**Yes**	**Yes**	**No**	**Yes**
DIP	**Yes**	**Yes**	**No**	No

Figure 4.5

Figure 4.6: Mosquito escape index

An index representing the number of mosquitoes that touched and left the behavioral assay glove mesh vs the number that touched the mesh.

Escape Index = (Average Number of mosquitoes that touched and left the mesh during a five second window over the following time points: 1 minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes) / ((Average Number of mosquitoes that touched and left the mesh during a five second window over the following time points: 1 minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes) + (Average Number of mosquitoes that stayed in the mesh during a five second window over the following time points: 1 minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes)).

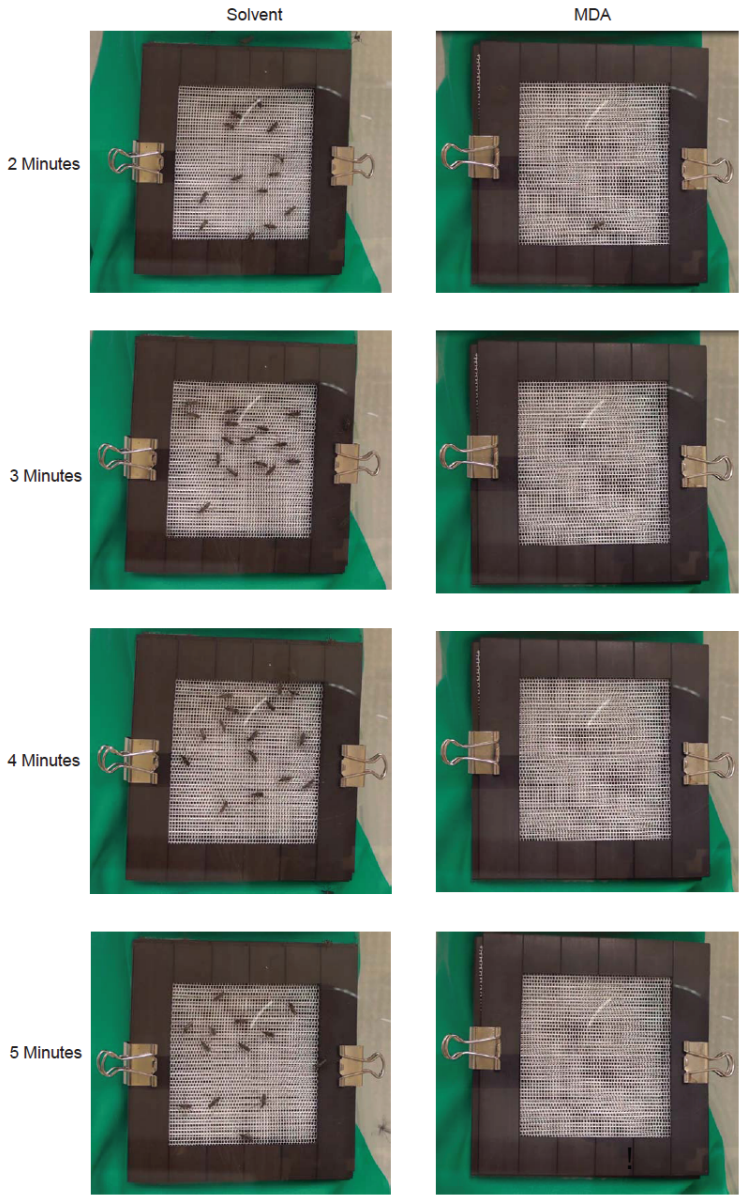


Figure 4.6

Figure 4.7: Natural compounds are effective at repelling *aedes aegypti* in the hand in glove assay

Representative still photographs from specific time-points of video assaying landing of female *Aedes aegypti* on solvent treated and MDA treated netting in the hand-in-glove assay.

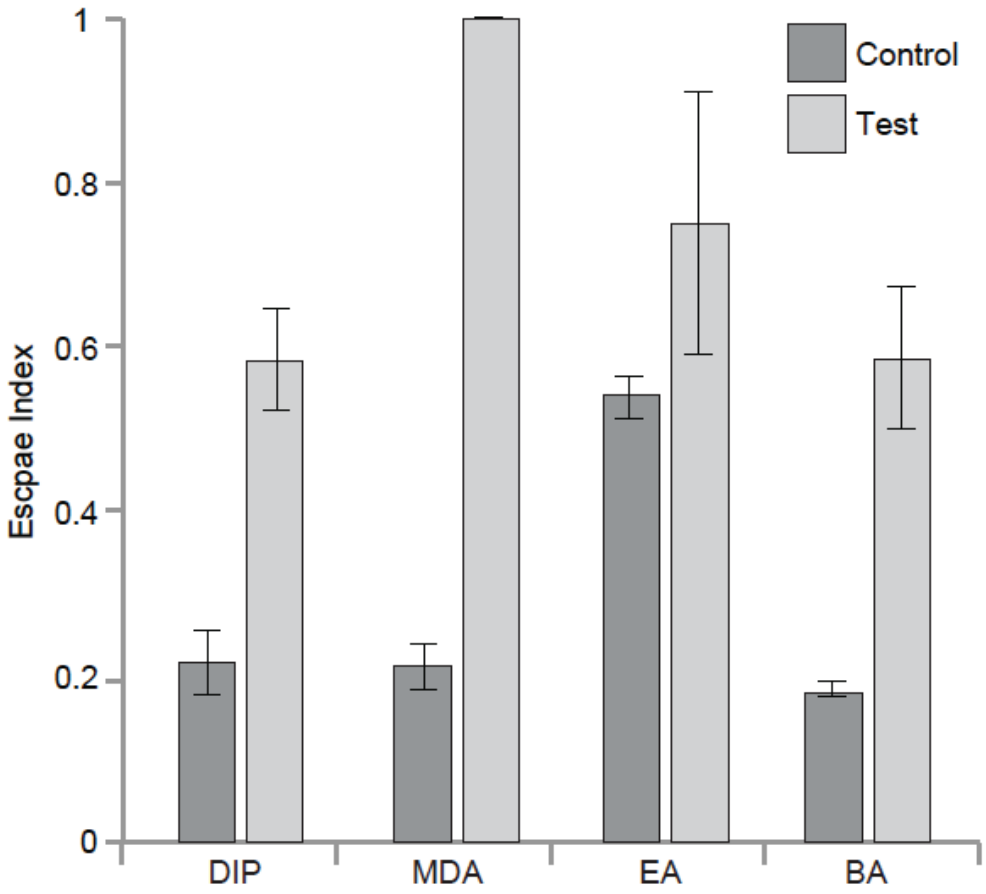


Figure 4.7

CHAPTER V:

Analyzing Termination Dynamics of Prolonged Activating Odors and their Effects on Receptor-Mediated Olfactory Behavior

INTRODUCTION

Successful interpretation of a volatile chemical environment allows an organism to make more informed decisions based upon what could be critically important information. The olfactory system is often utilized by insects and mammals for tasks such as finding food, determining oviposition sites, avoiding predators, and identifying mates (van Naters and Carlson, 2006). These high priority tasks are achieved in part through recognition of odors by large families of Odor receptor (Or) genes that often encode 7-transmembrane proteins that are expressed in the Olfactory Receptor Neurons (ORNs) (Buck and Axel, 1991; Clyne et al., 1999; Dahanukar et al., 2005; de Bruyne and Baker, 2008; Vosshall et al., 1999). Generally, a single functional Or gene, along with an obligate co-receptor Orco, is expressed per odor receptor neuron class, causing neuronal responses to be a direct result of the activity of a single odor receptor protein (Couto et al., 2005; Fishilevich and Vosshall, 2005; Larsson et al., 2004; Vosshall and Hansson, 2011). In this fashion, neurons are able to send environmental information from the periphery to higher brain centers for interpretation and behavioral decisions. Using this system an organism can decode and analyze either single compounds, such as pheromones, or complex blends, such as rotting fruit in the environment.

Activation of ORNs can lead to distinct behavioral responses, attraction and repellency, which can be robustly and reproducibly measured in *Drosophila* larvae (Aceves-Pina and Quinn, 1979; Ayyub et al., 1990; Fishilevich and Vosshall, 2005;

Kreher et al., 2008; Monte et al., 1989). While the degree of behavioral response varies across odors, it is dependent on both the odor stimuli and odor concentration (Kreher et al., 2008). Recently, large numbers of Ors have been decoded using the empty neuron system, providing a wonderful knowledgebase to draw upon for analysis (Hallem and Carlson, 2006; Kreher et al., 2008). By testing a panel of 27 odors against all 21 Ors expressed in *Drosophila* larvae one group was able to analyze the relationship between Or activity and behavioral responses on a systems level for the first time (Kreher et al., 2008). This analysis found that attraction and avoidance can be caused by activation of either multiple Ors or, perhaps more interestingly, by single Ors. It was also shown that odors that activated similar sets of Ors are able to behaviorally mask the presence of each other.

One Or in *Drosophila* in particular, Or42b, has been identified as an odor receptor whose response is explicitly linked to an innate behavioral response (Semmelhack and Wang, 2009). Activation of this Or in the adult, and thus its associated ORN ab1A and glomerulus DM1, by the well-known *Drosophila* attractant apple cider vinegar causes a robust attractive response (Semmelhack and Wang, 2009). It has since been demonstrated that the Or42b associated ORN is a starvation modulated neuron (Root et al., 2011). Starvation increases transcription of short neuropeptide F Receptor (sNPFR1) in ab1A, increasing the neurons response to odor stimuli (Root et al., 2011). This cellular response to starvation was shown to be necessary and more importantly sufficient, for increased food search behavior to apple cider vinegar, suggesting that Or42b and associated ORN ab1A may be hardwired for appetitive choices and critically important for *Drosophila* food searching (Figure 5.1A).

This system provides a powerful tool to investigate whether periphery stimulation by odors alone is sufficient to modify the strengthened food search behavior of starved flies.

Traditionally, odors have been thought to interact with receptors causing one of two unique responses, either activation or inhibition of the neuron in which it is expressed (Dobritsa et al., 2003; Hallem et al., 2004). When considered at a systems level, these two categories of responses allow some odors to activate and others to shut down odor receptor neurons in a combinatorial fashion. Very recently, a third type of response to a receptor-odor interaction has been discovered wherein a receptor is activated for a greatly extended period of time (Montague et al., 2011; Turner et al., 2011). In one analysis a special class of odors called pyrizines were found to cause prolonged activation with a maximum response of ~160 spikes/second that lasted for minutes before returning to baseline in DmOr33b and DmOr59b (Montague et al., 2011). In the other, our lab identified CO_2 detecting ORNs that express members of the gustatory receptor gene family in mosquitoes that demonstrate ultra-prolonged termination kinetics, which can disrupt detection of CO_2 and behavior for several minutes (Turner et al., 2011). Additionally, in a few instances, ORNs have been shown to have a tonic response that lasts up to 10 sec beyond the end of the 0.5 sec stimulus period (Hallem and Carlson, 2006; Hallem et al., 2004). The currently understood effects of ORN activation on behavioral responses have emerged from studying the relationship between a very limited set of receptor-odor interactions (~0.05% of putative and ~5% of known volatiles tested) and associated behavioral responses. As a result, there is potential for new behavioral responses to be observed, especially in regard to the newly identified UP activators, as very little is know about what role differences in temporal aspects of receptor activity will have on behavior.

Here we identify the first UP activators of several *Drosophila* odor receptors, including for the behaviorally important ab1A ORN, and demonstrate that these odors have profound behavioral effects on *Drosophila* attraction. We first applied a novel chemical informatics approach to identify structural features that are important for activation of the ab1A neuron. We then utilized these features to perform an *in silico* screen of a large chemical space, predicting a set of odors which are active on ab1A. We validated the predictive approach computationally and tested a number of compounds using single unit electrophysiology, both showing high success rates. This large number of new ligands enabled us to identify a small set of odors that represent the UP mode of signaling. Odors that cause this UP signaling leave a short-term memory trace, where for minutes after exposure of ab1A to an UP activator the neuronal response to other ab1A ligands is dramatically changed. We demonstrate that even a brief exposure of ab1A to these odorants trigger activity in the neuron in a manner that it looses the ability to subsequently detect other ligands in the environment, which ultimately effects the ability of *Drosophila* to track towards highly attractive ab1A activating odors. In addition to UP activators, we also discover that an inhibitory odor of ab1A can cause an avoidance behavior that is able to overcome the response to highly attractive odors.

RESULTS

Chemical informatics can be applied to describe activity of the ab1A neuron

Only a small number of odors have been tested on the ab1A associated receptor Or42b, however the testing of these odors represents a significant body of work. In the

largest analysis to date 47 odors were tested on the ab1A neuron in the *Drosophila* antenna using single unit electrophysiology (de Bruyne et al., 2001). In two additional analyses, 27 odors were tested on Or42b using single unit electrophysiology in the "empty neuron" expression system (Dobritsa et al., 2003; Kreher et al., 2008).

Volatile chemical space is vast (potentially >400,000) and can be exploited to find ligands of ab1A, however the cost of purchase and time required to test a large number of compounds to identify UP activators using traditional wet lab techniques could prove to be unfeasible. To circumvent these challenges, we have implemented a novel chemical informatics approach to identify UP activators of ab1A (Figure 5.1B). Our approach allowed us to sidestep these challenges by only purchasing and testing odors that share important structural features with known activating odors of ab1A, thus greatly reducing both the cost and time required for identification. While we did not directly predict UP activators, we greatly increased our efficiency at screening active compounds, some of which we hoped would show UP termination kinetics.

Since we wanted to train our pipeline from the largest and most diverse set of compounds that had been tested for activity on ab1A, we used the deBruyne analysis of 47 odors as our training set. To identify important structural features our chemical informatics pipeline selected molecular descriptors, which are mathematical values that describe the structural features of a chemical, that were highly correlated with the activity of a set of training odors (Figure 5.1B). Molecular descriptors can describe a great deal of features about a molecule, ranging from simple information such as molecular weight to much more complex characteristics such as the relationships between atoms in a 3D space. Using the commercially available Dragon suite we calculated 3,224 molecular descriptors for each of the 47 training odors, producing a broad range of information for

each odor. Calculated descriptors consist of 0, 1, 2, and 3 dimensional information, examples of which are vapor pressure, atom type counts, the number and type of atoms that connect each of the atoms in a molecule, and the 3D distances between atoms in energy minimized molecules, respectively. We then implemented a Sequential Forward Selection (SFS) approach that incrementally grows an optimal subset of molecular descriptors best describing the important structural features of activating odors (Whitney, 1971), which for descriptor optimization we considered to be only the most strongly activating odors (>113 spikes/sec) Figure 5.1B, See Supplemental Methods). The SFS approach began by selecting the single descriptor that is most highly correlated with the activity of the 47 training odors. The approach then iteratively built upon this single descriptor, identifying the second best descriptor when combined with the first. The process continued until a single iteration failed to identify a descriptor that further improved the correlation between activity and the growing descriptor set. The resulting list contained 13 optimized descriptors (Table 5.1).

Interestingly, the SFS approach selected four 1D descriptors, one 2D descriptor and eight 3D descriptors. The optimized descriptor set containing the 1D descriptors focusing on the number of aliphatic esters, number of acceptor atoms for H-bonds, number of hydroxyl groups, and the number of carbon atoms attached to heteroatom, as well as containing a large number of 3D descriptors, suggested that the most important structural features dependent for activity in ab1A are the absence of hydroxyl groups, the presence aliphatic esters, and their placement in the odor molecules 3D space. As expected from descriptor optimization, the ab1A activators cluster tightly together when the training is clustered using Euclidean distances calculated between odors using the optimized descriptor values (Figure 5.1C).

Next we used the optimized set of molecular descriptors to train a Support Vector Machine (SVM) to perform predictions, which is a well known and highly effective machine learning technique (Cortes and Vapnik, 1995) (Figure 5.1B, See Supplemental Methods). We then performed 100 independent 4-fold cross validations to computationally validate the predictive success, which is an established computational approach, followed by a Receiver-Operating-Characteristic (ROC) analysis (See Supplemental Methods). A mean ROC curve representing the prediction accuracy across all 100 iterations was generated and the Area UnderCurve (AUC) value was determined to be 0.999, indicating that our in silico approach to describe the activity of odors on the ab1A neuron using structural features of the odors themselves was almost perfect (Figure 5.1D).

Screening a large untested chemical space *in silico*

Since the Or-optimized descriptor sets are effective at grouping activating odors based on their structural features and we had successfully trained a SVM to perform predictions based on those descriptors, we next applied them to screen a large chemical space *in silico*. We assembled and calculated molecular descriptors for both an eMolecules library, which contains ~440,000 compounds with similar molecular weight and atom type constraints to known volatiles, and a natural odor library, which contains 3,197 natrually occurring odors (See Methods). We then applied the trained SVM to predict activity for each of the compounds in both libraries using the optimized molecular descriptor set, resulting in a ranked list of candidate agonists for Or42b (ab1A).

Identification of Or42b agonists through an electrophysiological validation of in silico screen

In order to identify agonists of Or42b, some of which we hoped would demonstrate prolonged termination kinetics, as well as to further validate our computational approach, we obtained 15 high ranking odors from our natural odor library predictions. We tested each odor for activity against ab1A and it's associated receptor Or42b using single unit electrophysiology (Figure 5.2A). We observed that all tested odors were agonists of Or42b. 13 Odors were highly active (>100 spikes/sec above spontaneous activity), 1 odor was active (>50 spikes/sec above spontaneous activity), and 1 odor validated as an inhibitor (>50% reduction in spontaneous activity). In order to further verify that the response was unequivocally from the Or42b receptor expressed in the ab1A neuron and not from neighboring neurons housed in the same ab1 sensilla, we performed single unit electrophysiology on Or42b -/- flies (Figure 5.2B). We observed a substantial reduction in response, indicating that Or42b is indeed the target receptor for these odors.

Identification of odors for ab1A with unusual ultra-prolonged termination kinetics

It has been proposed that not just frequency of activity, but the temporal properties of a response can potentially be a rich source of information encoded by sensory neurons. For a majority of odors, ORNs show rapid activation and termination kinetics, turning off abruptly shortly after the end of the odor stimulus (de Bruyne et al., 2001; Hallem and Carlson, 2006; Hallem et al., 2004). Now that we had identified several additional activating ligands, we could systematically inspect each odor for large differences in kinetics of response termination. Interestingly, 2 of the *in silico* predicted

odors that we tested using long-term electrophysiology recordings showed an ultra-prolonged response that lasted for several minutes beyond the 0.5 sec odor stimulus (Figure 5.3A). In both instances exposure of ab1A to a 0.5-second pulse of an ultra-prolonged (UP) activator elicits a strong initial response followed by decrease in the ORN firing rate to approximately half maximal frequency in about 8-10 seconds, after which the ORN has a relatively steady firing rate of ~30-70 spikes/sec above spontaneous activity for several minutes (Figure 5.3A -green line). This unusual pattern of activation is quite different from other equally strong activators that return close to baseline between 2-6 seconds after odor exposure (Figure 5.3A –grey line).

This form of a prolonged response elicited by a short 0.5 second stimulus appears even stronger than the activity pattern evoked by a continuous odor pulse of 30 secs, in which the adaptation of the initial strong response of the neuron occurs rapidly, followed by a low frequency of action potentials throughout the duration of the stimulus (de Bruyne et al., 2001). These observations suggest that the class of UP odors we have identified by cheminformatics have an atypical property to activate ORNs for periods >500 times the duration of the initial 0.5 sec stimulus.

A mechanism for short-term memory of odor exposure

Surprisingly, a brief 0.5-second pre-exposure to an UP activator renders the neuron unresponsive to changes in concentration of other activating odorants that the ORN normally responds to. The UP-activated ORNs are unable to respond to odorants normally for more than 100 seconds, in some cases up to the duration of the entire recording (300 seconds) (Figure 5.3B, green bars). This short-term memory effect is specific to the UP-activators and pre-exposure to regular odorants that activate to

comparable levels does not affect the ability of the ORN to detect subsequent repeated 0.5 second stimuli of an activating odorant (Figure 5.3B, grey bars). This form of change in coding capacity of a peripheral sensory level after ultra-prolonged activation is a novel phenomenon that has not been reported for neurons expressing Odor receptor proteins.

Although termination kinetics has been shown to be a property associated with the receptor (Hallem et al., 2004), it remains to be tested whether this novel type of short-term memory effect could be influenced by odor binding proteins (Wang et al., 2007) found in the sensillum lymph or other factors involved in signal transduction. In order to test directly whether the short-term memory effect depends upon the odor receptor protein, we expressed *Or42b* in the well-established "empty neuron" decoder system (Dobritsa et al., 2003). We find that the UP-activation effect is partially transferred to the Δhalo ab3a neuron upon ectopic expression of *Or42b* (Figure 5.4). Surprisingly, the short-term memory effect is transferred to the acceptor ORN simply by the ectopic misexpression of *Or42b* (Figure 5.4B, blue bars), suggesting that the integration of the two inputs and retention of the memory of prior exposure is maintained at the level of the receptor.

The ability of an animal to respond behaviorally to changes in concentration of a ligand along an odor gradient or navigate along an odor plume in the environment is presumed to depend on two important coding properties of the ORN: sensitivity and rapidity of detecting incremental changes in concentration. Since both these properties are severely compromised upon exposure to UP-odorants, we expect a long-term effect on behavior even after the animal is removed from the vicinity of the UP-odorant. The

Drosophila larvae are ideal for behavioral testing of the short-term memory effects to prior exposure to UP-activators since they have a simpler olfactory system and behavior assays are robust and quantitative (Fishilevich and Vosshall, 2005; Kreher et al., 2005; Kreher et al., 2008; Louis et al., 2008). An antennal receptor Or42b, for which we have identified UP-activators, is expressed in the larval system as well (Fishilevich and Vosshall, 2005; Kreher et al., 2005; Kreher et al., 2008). It has also been demonstrated that Or42b is the exclusive receptor for detection of ethyl acetate at 10^{-4} dilution, which causes attractive behavior (Kreher et al., 2008). We performed a two-choice assay where larvae were given 90 seconds to choose between an ethyl acetate (10^{-4}) side and a solvent side on a plate after they were pre-exposed to either the UP-activator (methyl proprionate), a regular activator of comparable initial strength with normal response termination (methyl isobutyrate), or solvent (paraffin oil) (Figure 5.5A). In solvent pre-exposed and activator pre-exposed larvae a robust attractive behavior towards ethyl acetate was observed, as expected (Figure 5.5B). However a 10-second pre-exposure to an UP-activator almost completely disrupted attraction. Even a brief 1-second pre-exposure to an UP-activator was sufficient to cause a significant reduction of attraction towards ethyl acetate (Figure 5.5B). Taken together these results demonstrate that pre-exposure to a super activator not only disrupts the ability of ab1A to detect odors, but also disrupts behavior elicited by the odorant.

In order to demonstrate that the change in behavior is caused directly by UP-activation of Or42b, we performed a similar pre-exposure experiment on larvae that were genetically manipulated to have only the *Or42b* expressing ORN pair functional. This was achieved by specifically rescuing expression of the obligate co-receptor *Orco* in only

the *Or42b* expressing neurons in an *Orco* mutant background. In both solvent pre-exposed and activator pre-exposed larvae robust attractive behavior towards ethyl acetate was observed, as expected (Figure 5.5C). However, pre-exposure to the UP-activator completely disrupted attraction towards ethyl acetate (Figure 5.5C). This shows that the olfactory system has the ability to integrate two odor stimuli, separated in time, directly at the level of the peripheral ORN, and compute two dramatically different outputs (change or no-change in firing frequency) to transmit to the CNS depending upon the identity of the first odorant.

Disruption of attraction towards apple cider vinegar in starved larvae

In the natural environment attractive cues are often more complex consisting of blends of odorants and insects. Moreover, the olfactory ability to find food odors may be enhanced by the NPF pathway-mediated starvation response (Root et al., 2011). For example, Or42b neurons were shown to be necessary for the starvation-induced food-search behavior in the adults (Root et al., 2011). In order to test whether exposure of Or42b, which was been demonstrated to be hardwired for behaviorally important appetitive choices, to an UP-activator was sufficient to interfere with the attraction of starved larvae towards complex food odor, we tested them in behavior assays. Surprisingly, a brief pre-exposure of 2-hour starved larvae (Koon et al., 2011) to an Or42b UP-activator was sufficient to dramatically reduce the ability to find an attractive source of apple cider vinegar (Figure 5.5D). This result suggests that The UP-activators may provide a powerful tool for behavior modification to protect against harmful pest species, many of which cause significant loss of agricultural produce and crops globally.

Inhibition of appetitive choice ORN causes strong avoidance response

While we have demonstrated that UP activators disrupt attraction and activators have been previously shown to cause a strong attractive response, the behavioral effect of inhibitors tuned to this receptor alone has not been previously reported. We performed a two-choice assay where larvae were given 90 seconds to choose between the inhibitor Isopentyl formate (10^{-2}), which was found in this study, and a solvent side (paraffin oil) on a plate. Interestingly, larvae behaviorally avoided the inhibitor, with a preference index of -0.2 (Figure 5.6). These results demonstrate that inhibitors are able to repel insects on their own.

Identification of odors that display Ultra Prolonged activation in additional ORNs

In order to test whether we could identify UP activators of additional Ors, we performed similar chemical informatics analysis for Or22a and Or85b. Here we trained our pipeline using the largest panel available for these receptors, which consists of 109 previously tested odors (Hallem and Carlson, 2006). In chapter 2 we predicted activating odors and validated several additional odors that displayed delayed termination kinetics reminiscent of UP activators (Figure 5.7A). We identified 5 UP activators for Or22a and 1 for 85b (Figures 5.7B, 5.8A), further suggesting that such ultra-prolonged termination kinetics are not properties specific only a select few receptors. The prolonged activation of Or85b by UP activator 2-heptanol elicits a strong initial response similar to the one caused by previously identified Or42b up activators, however the decrease in the ORN firing rate to approximately half maximal frequency requires a greatly extended time of 30 seconds, indicating that activation decay rates can vary significantly (Figure 5.7B).

Visual inspection of the 5 UP-activators of Or22a does not reveal any features that easily distinguish this class from that described for other activators of the same receptor (Figure 5.8B). To investigate whether we could identify an UP-optimized descriptor set that separates UP-activators from a large set of active compounds, we identified a subset of descriptors that cluster the UP-activators close together in a chemical space containing activators of Or22a (Figure 5.7C). We then ranked odors for UP-activation for Or22a, tested one using electrophysiology (Butyl propionate) and found that it had the hallmarks of an UP-activator (Figure 5.7D). These results suggest that the UP-optimized descriptors may be useful to distinguish odorants that cause prolonged activation even from structurally similar odorants.

We found instances where the UP-activator did not show prolonged activation of other ORNs it activates, suggesting that the prolonged response is not caused simply due to a property of the odor (Figure 5.8C). However, we also found that 2-heptanol can cause UP-activation of another ORN that expresses Or85a. These results suggest that although UP-activation is probably caused by specific interaction between odor and receptor, some odors may have properties that can cause prolonged activation in at least two different Or containing neurons.

Once again we find that a brief 0.5-second pre-exposure of these receptors to an UP activator renders the neuron unresponsive to changes in concentration of other activating odorants that the ORN normally responds to (Figure 5.9A). The UP-activated Or22a and Or85b expressing ORNs ab3A and ab3B are unable to respond to odorants normally for more 100 seconds for the case of ab3A and for the duration of the entire recording for ab3B (300 seconds) (Figure 5.9B, green bars). This short-term memory effect again contrasts from pre-exposure to regular odorants where activation to

comparable levels does not affect the ability of the ORN to detect subsequent repeated 0.5 second stimuli of an activating odorant (Figure 5.9B, grey bars).

DISCUSSION

We have designed a computational approach that allows us to identify UP activating odors and investigate their effects on odor receptors, including subsequent odor detection. Our computational pipeline builds an optimal set of molecular descriptors that explain odor activity and predicts Or42b activity to over 400,000 odors by applying them to a trained SVM. All experimentally tested odors either activate or inhibit Or42b, additionally, a number of validated odors exhibited prolonged termination kinetics. Through examination of their effects, we have identified that exposure to these UP odors disrupts the ability of *Drosophila* larvae to track towards highly attractive odors. We have also demonstrated the repellent properties of inhibitory odors of Or42b.

No analyses have been previously performed to computationally identify UP activating odors. Computational approaches have been made at predicting activating ligands, however if compared at the same activity threshold applied in our study (>=50 spikes/second), their success rates were far lower than ours at predicting ligands (25%)(Schmuker et al., 2007). We find that our computational approach was 100% efficient in predicting ligands of Or42b, as all odors tested with electrophysiology were either activators or inhibitors. Additionally, through rational prioritization of odors, the likelihood of identifying odors with unique properties becomes far higher. 13% (2/15) of validated odors displayed UP activation characteristics for Or42b. Without our initial computational prioritization to identify activating odors, this rate would reduce

significantly. For example, if we were screening for UP activators using the same ligand identification success rate of the hand selected training set at 19% (9/47) (de Bruyne et al., 2001), we could roughly expect our accuracy at identifying UP activators to be 13% of the 19%, or 2%. Clearly, a computational approach is advantageous.

The new ligands presented in this study include a small number of novel UP-activators whose discovery expands the olfactory code from activation and inhibition to now include ultra-prolonged activation. Most of these UP-odors, exposure to which is retained as a short-term memory trace by the odor-receptor, are present in natural substances suggesting that they may play a significant role in the natural olfactory behavior of animals. Functional insect odor receptors are thought to act as ion-channels (Sato et al., 2008) as well as GPCRs (Wicher et al., 2008) and have a unique inside-out 7-transmembrane structure (Benton et al., 2006) acting along with a heteromeric partner Orco (Benton et al., 2006; Larsson et al., 2004). Although outside the scope of this study, understanding mechanistically how these odors bind and cause prolonged activation will be of great interest.

We have demonstrated two very important characteristics of these UP activating odors: a brief exposure is sufficient to functionally silence the responses of Or42b, Or85b, and Or22a to normally excitatory odors for greater than 2 minutes and that a brief exposure is sufficient to disrupt robust attractive behavior. Olfactory signals are very important for many biological processes. Identification of subtle changes in odor plume identity and concentration allows for successful navigation by insects to odor sources. For example, casting, which involves identifying the edge of an odor plume and course correcting to towards it's center, is utilized by many species to travel up odor plumes. Our prolonged activators cause the insect to be functionally unable to determine these

concentration changes, leading to a loss of casting ability. By the time the effects of the UP activator wear off the insect could easily have flown far off course. In a case were they do reacquire the plume, they would also reacquire the UP activator, once again causing them to drift off course.

We have also demonstrated a very important characteristic of our inhibitory odor: inhibition of Or42b causes repellent behavior. Additionally, if the inhibitory odor is introduced next to a highly attractive odor for the same Or, it negates attraction. These results suggest that Or42b is not solely hardwired for innate attraction and is instead hardwired for behavior, with activating odors causing attraction and inhibitory odors causing avoidance. Further exploration on the effects of other inhibitory odors on this receptor as well as those for additional receptors should be an important focus.

An emerging area of research is the identification of odors that can modify host-seeking behavior in insect disease vectors, either by virtue of their ability to inhibit ORNs that detect host-seeking cues (Montague et al., 2011; Turner et al., 2011; Turner and Ray, 2009), by activating ORNs that cause avoidance behavior (Syed and Leal, 2008). UP-activators and inhibitors provide a novel paradigm for behavior modification strategies targeting important odor receptors (Figure 5.10). *In silico* screens can thus provide a rational foundation for identification of novel insect repellents and lures that are environmentally safe and can aid in the fight against insect-borne diseases.

Figure 5.1: Activity of a behaviorally important neuron, Ab1A, can be described by an odor receptor neuron-optimized chemical inormatics approach

(A) Activation of Odor Receptor Neuron (ORN) ab1A, and thus glomerulus DM1, leads to a strong attractive response in *Drosophila melanogaster*. **(B)** Schematic of the cheminfomatics pipeline used to identify novel ligands from a larger chemical space. **(C)** Receiver-operating-characteristic curve (ROC) representing computational validation of ligand predictive ability of the ORN-optimization approach. The mean true-positive value from 100 independent 4-fold cross validation runs of the Support vector machine (SVM) approach is plotted. **(D)** Hierarchical cluster analysis of the 46 training set odorants using ORN optimized descriptor sets to calculate distances in chemical space for ab1A. Training set odorant activity is indicated using a color gradient scale. **(E)** Pharmacophore of odors that are highly active against ab1A.

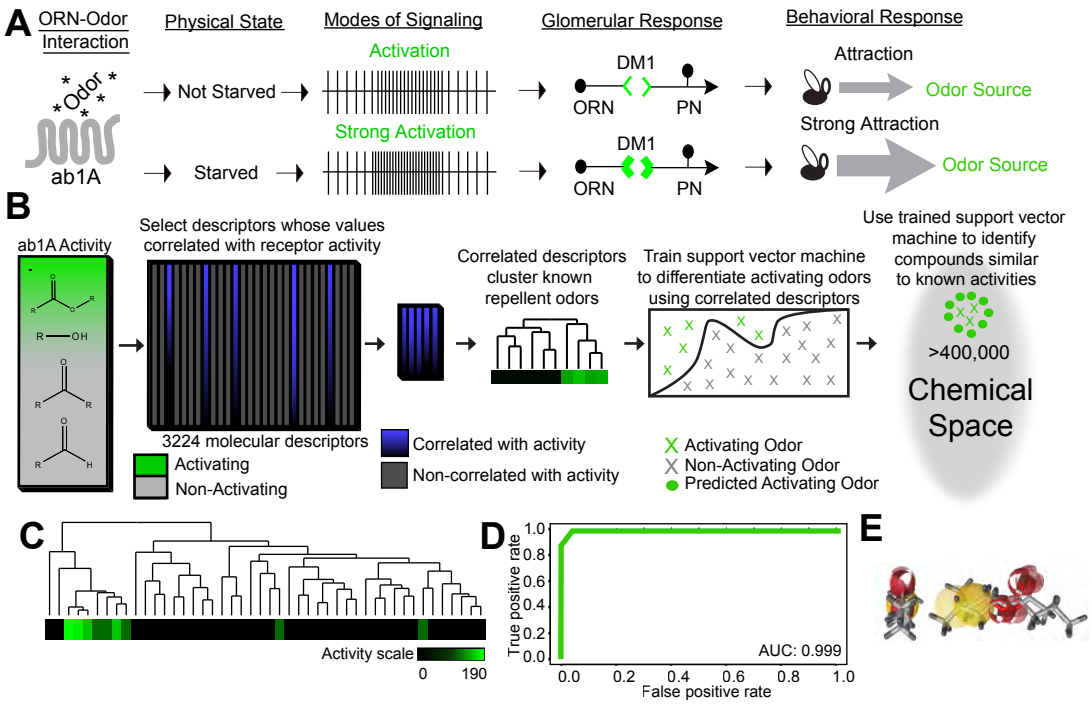

A
ORN-Odor Interaction
Physical State
Modes of Signaling
Glomerular Response
Behavioral Response
Odor
ab1A
Not Starved
Activation
DM1
ORN
PN
Attraction
Odor Source
Starved
Strong Activation
Strong Attraction
B
ab1A Activity
Select descriptors whose values correlated with receptor activity
3224 molecular descriptors
Activating
Non-Activating
Correlated descriptors cluster known repellent odors
Train support vector machine to differentiate activating odors using correlated descriptors
Correlated with activity
Non-correlated with activity
Activating Odor
Non-Activating Odor
Predicted Activating Odor
Use trained support vector machine to identify compounds similar to known activities
>400,000
Chemical Space
C
Activity scale
0
190
D
True positive rate
False positive rate
AUC: 0.999
E

Figure 5.1

Figure 5.2: Electrophysiology validates that ORN-optimized molecular descriptors can successfully identify new ligands for ab1A

Mean increase in response of neuron to 0.5-sec stimulus of indicated predicted odors in **(A)** ab1A (Or42b) ORN; **(B)** and ab1A (Or42b$^{-/-}$) (10^{-2} dilution). N=3, error bars=s.e.m.

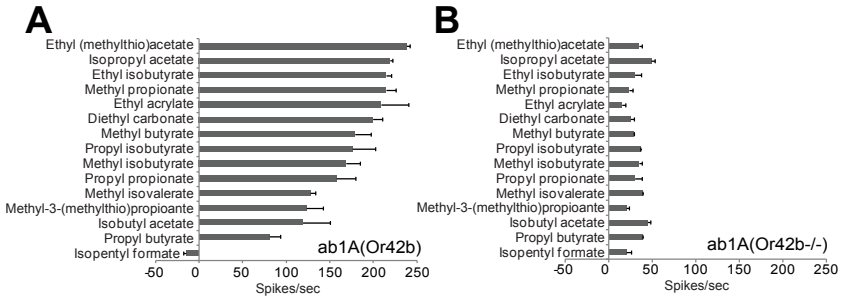

A
Ethyl (methylthio)acetate
Isopropyl acetate
Ethyl isobutyrate
Methyl propionate
Ethyl acrylate
Diethyl carbonate
Methyl butyrate
Propyl isobutyrate
Methyl isobutyrate
Propyl propionate
Methyl isovalerate
Methyl-3-(methylthio)propioante
Isobutyl acetate
Propyl butyrate
Isopentyl formate
ab1A(Or42b)
-50 0 50 100 150 200 250
Spikes/sec
B
Ethyl (methylthio)acetate
Isopropyl acetate
Ethyl isobutyrate
Methyl propionate
Ethyl acrylate
Diethyl carbonate
Methyl butyrate
Propyl isobutyrate
Methyl isobutyrate
Propyl propionate
Methyl isovalerate
Methyl-3-(methylthio)propioante
Isobutyl acetate
Propyl butyrate
Isopentyl formate
ab1A(Or42b-/-)
-50 0 50 100 150 200 250
Spikes/sec

Figure 5.2

Figure 5.3: Functional identification of Ultra-Prolonged activators and analysis of their long-term effects on ab1A using electrophysiology

(A) Mean long-term response of ab1A to a 0.5- sec stimulus of indicated odor at t=0. Each response curve is depicted in 3 separate graphs with different time windows, 10 sec, 90 sec and 300 sec. N=3, error bars=s.e.m **(B)** Mean increase in frequency of response of the ab1A to the indicated odor applied at indicated time points after pre-exposure to 0.5-sec odor stimulus indicated (grey=activator, green=UP-activator, blue=UP-activator response in UAS-Or42b expressing ab3A neurons in *ΔHalo; Or22a-G4,UAS-Or42b* flies). N=5, error bars= s.e.m.

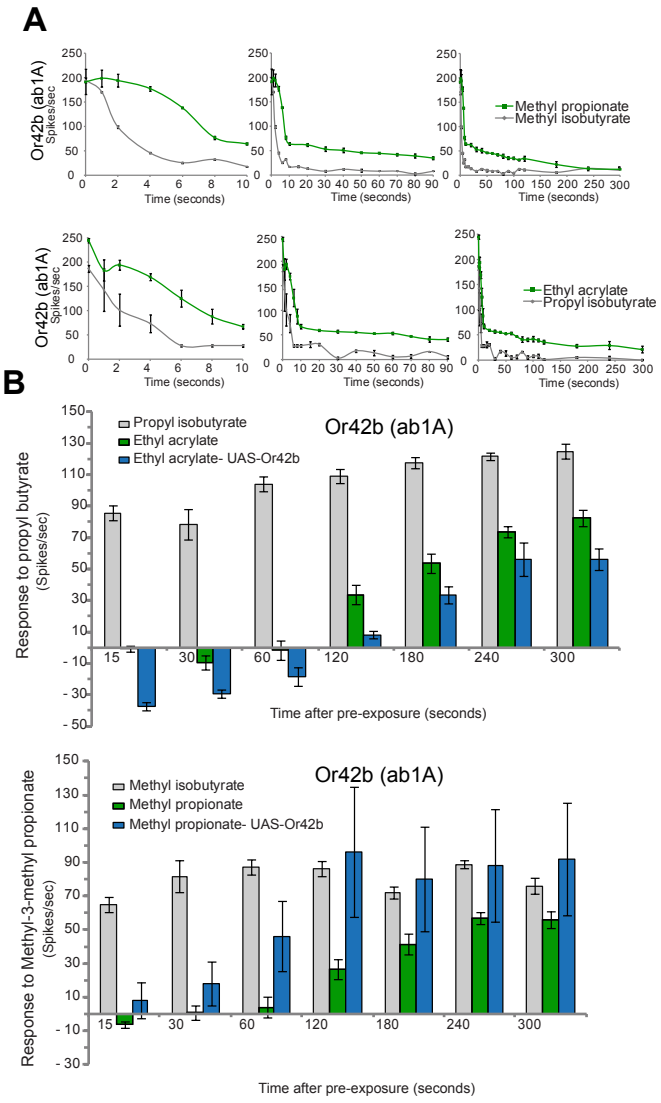

A
Or42b (ab1A)
Spikes/sec
Time (seconds)
Methyl propionate
Methyl isobutyrate
Ethyl acrylate
Propyl isobutyrate
B
Or42b (ab1A)
Propyl isobutyrate
Ethyl acrylate
Ethyl acrylate- UAS-Or42b
Response to propyl butyrate (Spikes/sec)
Time after pre-exposure (seconds)
Or42b (ab1A)
Methyl isobutyrate
Methyl propionate
Methyl propionate- UAS-Or42b
Response to Methyl-3-methyl propionate (Spikes/sec)
Time after pre-exposure (seconds)

Figure 5.3

Figure 5.4: Ectopic expression confers Or42b UP activation Mean long-term response of Or42b expressing neuron (ab1A) to a 0.5- sec stimulus of indicated odor. Each response curve is depicted in 2 separate graphs with different time windows, 10 sec and 90 sec. UAS-42b recordings were performed in ΔHalo; Or22a-G4,UAS-Or42b flies where ab1A neurons express Or42b. N=3, error bars=s.e.m

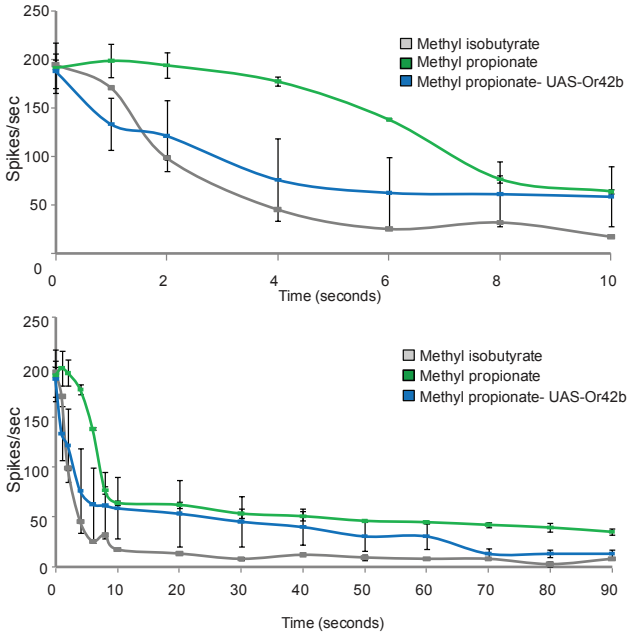

Methyl isobutyrate
Methyl propionate
Methyl propionate- UAS-Or42b
Spikes/sec
Time (seconds)
250
200
150
100
50
0
0
2
4
6
8
10
Methyl isobutyrate
Methyl propionate
Methyl propionate- UAS-Or42b
Spikes/sec
Time (seconds)
250
200
150
100
50
0
0
10
20
30
40
50
60
70
80
90

Figure 5.4

Figure 5.5: Behavioral effects of Ultra-Prolonged activators on odor detection in ab1A

(A) Overview of pre-exposure integration into larval two-choice preference assay. **(B)** Preference index of *Drosophila* larvae to Ethyl acetate 10^{-4}. Larvae were pre-exposed to the indicated odors for either 10-sec (left) or 1-sec (right) immediately prior to the preference assays. N=10 (~40 larvae/ trial), error bars=s.e.m. **(C)** Preference index of *Drosophila* larvae (*w;Or42b-GAL4/UAS-Orco; ΔOrco/ΔOrco*) to Ethyl acetate 10^{-4}. Larvae were pre-exposed to the indicated odors for either 10-sec immediately prior to the preference assays. N=10 (~40 larvae/ trial), error bars=s.e.m. **(D)** Similar experiment as (b) performed on 2-hr starved larvae given a choice between Apple Cider Vinegar (5%) and water. N=20 trials (~40 larvae/trial), error bars=s.e.m. P-values= *<0.05, **<0.001.

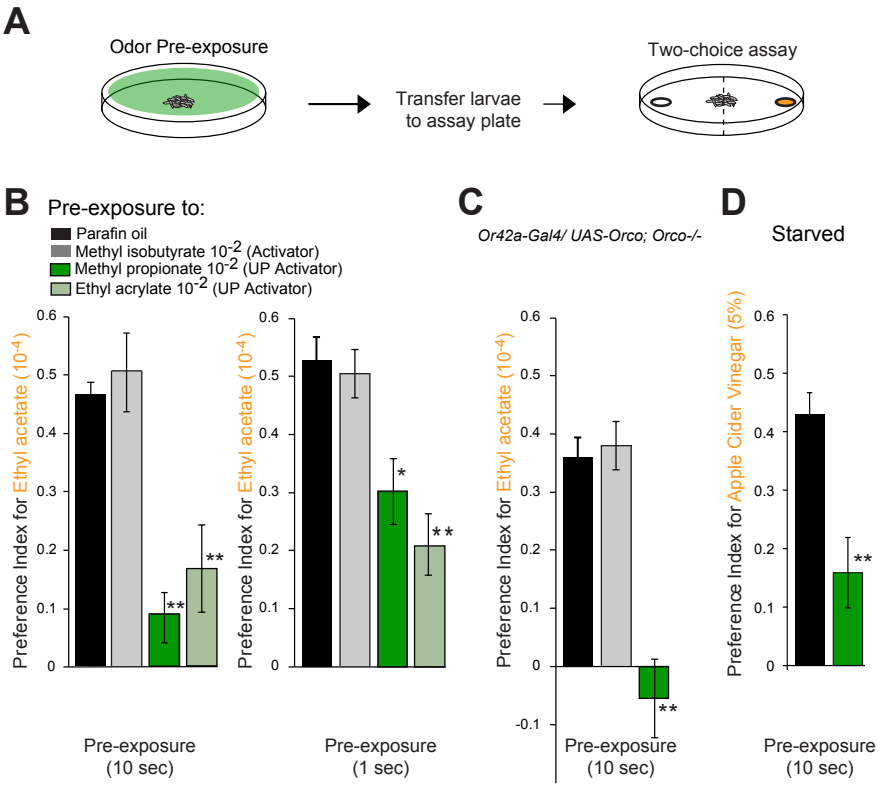

A
Odor Pre-exposure
Transfer larvae
to assay plate
Two-choice assay
B
Pre-exposure to:
Parafin oil
Methyl isobutyrate 10-2 (Activator)
Methyl propionate 10-2 (UP Activator)
Ethyl acrylate 10-2 (UP Activator)
Preference Index for Ethyl acetate (10-4)
Pre-exposure
(10 sec)
Pre-exposure
(1 sec)
C
Or42a-Gal4/ UAS-Orco; Orco-/-
Preference Index for Ethyl acetate (10-4)
Pre-exposure
(10 sec)
D
Starved
Preference Index for Apple Cider Vinegar (5%)
Pre-exposure
(10 sec)

Figure 5.5

Figure 5.6: Behavioral effects of inhibitory odor on odor detection in ab1A

(A) Overview of larval two-choice preference assay. Red circle denotes inhibitor (iso-amyl formate) placement and empty circle represents solvent (paraffin oil).

(B) Preference index of *Drosophila* larvae to iso-amyl formate 10^{-2}. Larvae were given a choice between solvent or iso-amyl formate 10^{-2} for the preference assay. N=8 (~40 larvae/ trial), error bars=s.e.m.

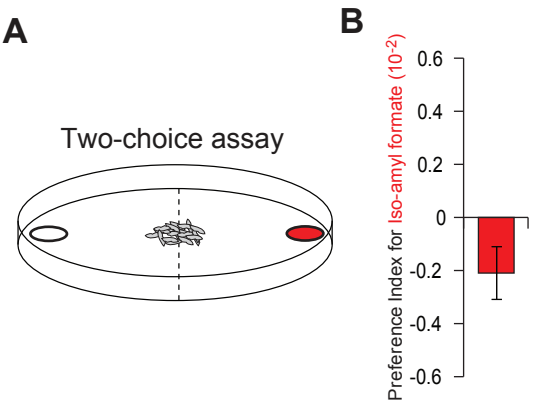

A
Two-choice assay
B
Preference Index for Iso-amyl formate (10-2)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6

Figure 5.6

Figure 5.7: Functional identification of Ultra-Prolonged activators of additional odor receptors using electrophysiology.

(A) Long-term response of a single Or22a expressing ab3A neuron to a brief 0.5-sec stimulus of ethyl valerate (indicated as a red bar). Action potentials from 0.25-sec windows shown from indicated regions of the response. **(B)** Mean long-term response of indicated receptor expressing neuron to a 0.5- sec stimulus of indicated odor at t=0. Each response curve is depicted in 3 separate graphs with different time windows, 10 sec, 90 sec and 300 sec. For ease of spike counting ab3B (Or85b) recordings were performed in a ΔHalo (ΔH) mutant background (Dobritsa et al., 2003) where ab3A neuron is unresponsive. N=3, error bars=s.e.m **(C)** Schematic representing the identification of prolonged activator-specific molecular descriptors that can cluster prolonged activators together in a tree containing all other Or22a activators. **(D)** Mean long-term response from an ab3A neuron expressing Or22a to a 0.5-sec stimulus of a predicted UP-activator depicted in 2 separate graphs with different time windows, 10 sec and 90 secs. N=3, error bars=s.e.m

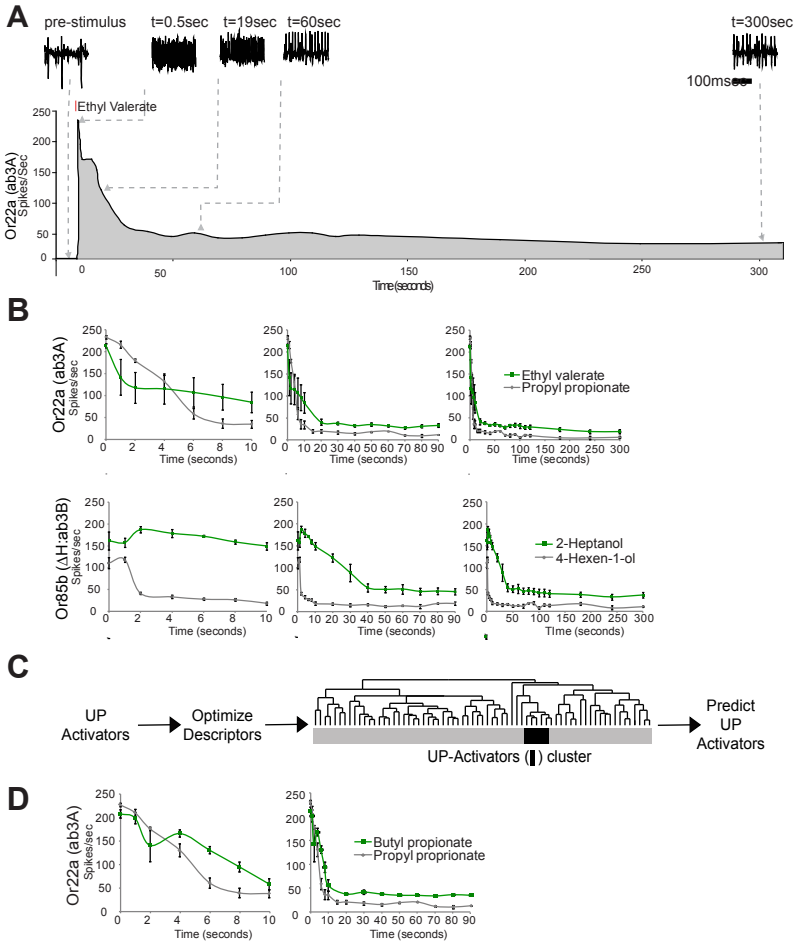

A
pre-stimulus
t=0.5sec
t=19sec
t=60sec
t=300sec
100msec
Ethyl Valerate
Or22a (ab3A)
Spikes/Sec
Time(seconds)
B
Or22a (ab3A)
Spikes/sec
Time (seconds)
Ethyl valerate
Propyl propionate
Or85b (ΔH:ab3B)
2-Heptanol
4-Hexen-1-ol
C
UP
Activators
Optimize
Descriptors
UP-Activators ( ) cluster
Predict
UP
Activators
D
Or22a (ab3A)
Butyl propionate
Propyl proprionate

Figure 5.7

Figure 5.8: Identification of ultra-prolonged activating odors for additional Ors

(A) Mean responses from ab3A neurons expressing Or22a across 10 seconds to a 0.5 second odor (10^{-2}) stimulus, red bar. **(B)** Pharmacophores of Or22a prolonged activators and Or22a activator pharmacophore. **(C)** Mean responses of indicated receptor expressing neurons to a 0.5 second stimulus of indicated odors (10^{-2}), red bar. For (A) and (C) N=3, error bars=s.e.m.

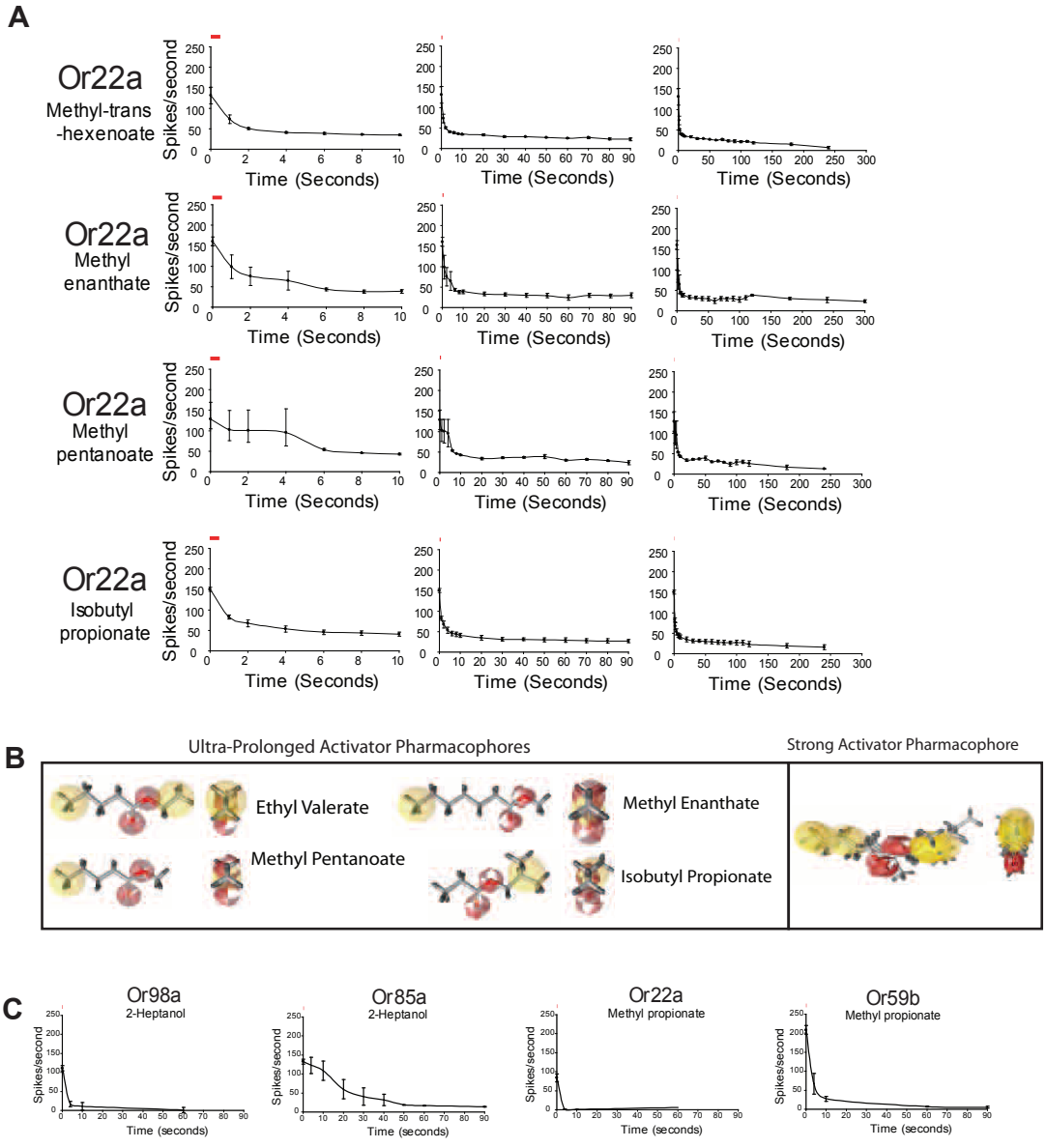

A
Or22a
Methyl-trans
-hexenoate
Or22a
Methyl
enanthate
Or22a
Methyl
pentanoate
Or22a
Isobutyl
propionate
Spikes/second
Time (Seconds)
B
Ultra-Prolonged Activator Pharmacophores
Strong Activator Pharmacophore
Ethyl Valerate
Methyl Pentanoate
Methyl Enanthate
Isobutyl Propionate
C
Or98a
2-Heptanol
Or85a
2-Heptanol
Or22a
Methyl propionate
Or59b
Methyl propionate
Spikes/second
Time (seconds)

Figure 5.8

Figure 5.9: Long-term effects of Ultra-Prolonged activators on odor detection

(A) Representative electrophysiology traces of Or85b expressing ab3B neuron starting ~13 sec after a 0.5-sec pre-exposure to an activator (4-hexen-1-ol) or an UP-activator (2- heptanol) with repeated exposures to another activator (5-hexen-1-ol). Boxed area denotes 2-sec regions at 15 sec and 30 sec after pre-exposure where 0.5-sec of 5-hexen-1-ol (black bars) was applied, which are magnified above and below. Ab3B (Or85b) recordings performed in a ΔHalo (ΔH) mutant background (Dobritsa et al., 2003). **(B)** Mean increase in frequency of response of the indicated neuron to the indicated odor applied at indicated time points after pre-exposure to 0.5-sec odor stimulus indicated (grey=activator, green=UP-activator). N=5, error bars= s.e.m.

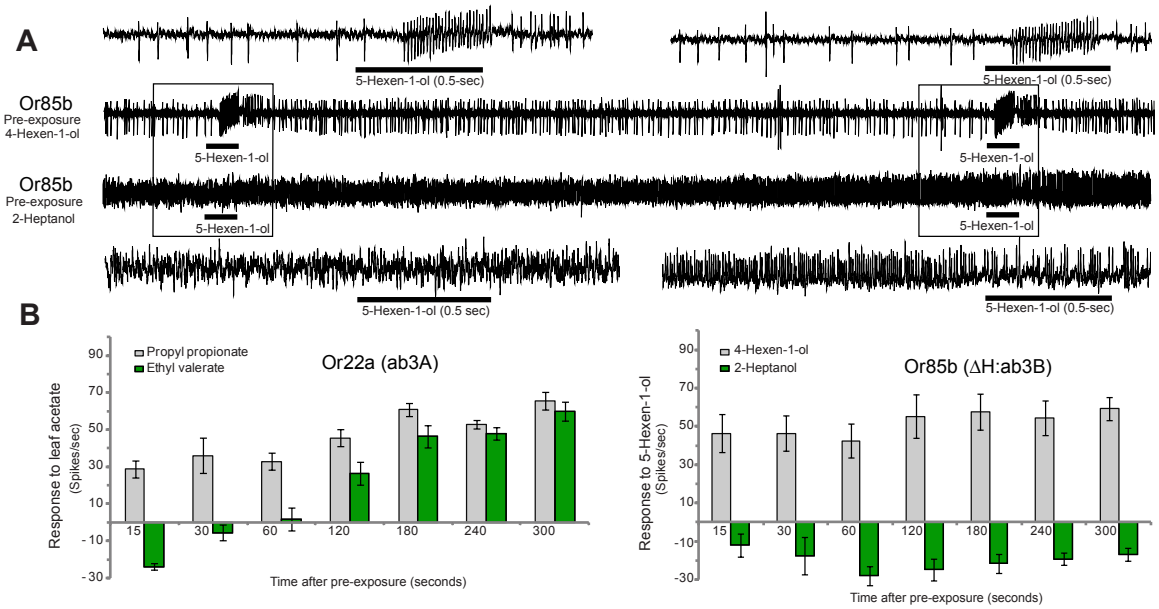

A
5-Hexen-1-ol (0.5-sec)
Or85b
Pre-exposure
4-Hexen-1-ol
5-Hexen-1-ol
Or85b
Pre-exposure
2-Heptanol
5-Hexen-1-ol
5-Hexen-1-ol (0.5 sec)
B
Propyl propionate
Ethyl valerate
Or22a (ab3A)
Response to leaf acetate
(Spikes/sec)
Time after pre-exposure (seconds)
4-Hexen-1-ol
2-Heptanol
Or85b (ΔH:ab3B)
Response to 5-Hexen-1-ol
(Spikes/sec)
Time after pre-exposure (seconds)

Figure 5.9

Figure 5.10: Modes of signaling and their behavioral responses

Expansion of the known behavioral responses from modes of signaling: inhibition of ab1A leads to repulsion and UP activation of ab1A leads to confusion. Flies are repulsed by inhibitory odors that specifically affect ab1A. Exposure to an UP activator causes flies to be unable to respond to normally attractive odors that activate ab1A.

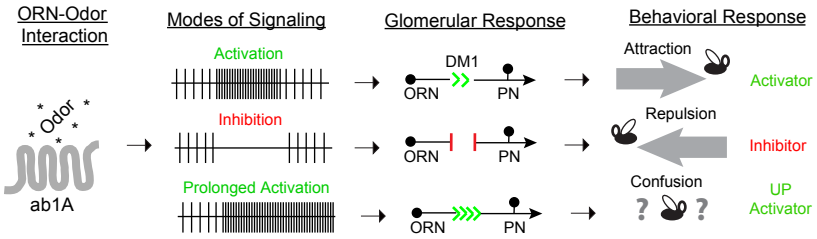

ORN-Odor Interaction
Modes of Signaling
Glomerular Response
Behavioral Response
Odor
ab1A
Activation
Inhibition
Prolonged Activation
DM1
ORN
PN
Attraction
Activator
Repulsion
Inhibitor
Confusion
UP Activator

Figure 5.10

Table 5.1: Optimized descriptor sets for ab1A (Or42b) Optimized descriptor symbol, brief description, class, and dimensionality are listed. Descriptors are listed in ascending order of when they were selected into the optimized set.

Optimized Descriptor List			
Symbol	**Description**	**Class**	**Dimensionality**
nRCOOR	number of esters (aliphatic)	Functional group counts	1
Mor10u	signal 10 / unweighted	3D-MoRSE descriptors	3
Mor04m	signal 04 / weighted by mass	3D-MoRSE descriptors	3
R1e+	R maximal autocorrelation of lag 1 / weighted by Sanderson electronegativity	GETAWAY descriptors	3
Mor27m	signal 27 / weighted by mass	3D-MoRSE descriptors	3
nHAcc	number of acceptor atoms for H-bonds (N,O,F)	Functional group counts	1
E1m	1st component accessibility directional WHIM index / weighted by mass	WHIM descriptors	3
GATS5m	Geary autocorrelation of lag 5 weighted by mass	2D autocorrelations	2
nROH	number of hydroxyl groups	Functional group counts	1
R5v	R autocorrelation of lag 5 / weighted by van der Waals volume	GETAWAY descriptors	3
Mor10p	signal 10 / weighted by polarizability	3D-MoRSE descriptors	3
C-006	CH2RX	Atom-centred fragments	1
Mor11e	signal 11 / weighted by Sanderson electronegativity	3D-MoRSE descriptors	3

Table 5.1

CHAPTER VI:

Applying Structure-Based Virtual Screening to Identify Potent Inhibitors of the Tuberculosis Target EthR

INTRODUCTION

While virtual screening is just now becoming an effective tool for modeling and screening odors in mammalian ORs, it is not yet an option for insect Ors, as the structure of receptors for insects have not yet been solved. However, as technologies advance the structures of insect and mammalian odor receptors will undoubtedly be determined in the near future. Once the structures are available, application of structure-based virtual screening approaches to predict ligand binding will be effective tools for predicting receptor-odor interactions. Additionally, application of these approaches will also provide valuable insight into how odors are functionally interacting with the receptors. Structure-based virtual screening can be applied as a standalone method or, as was done with the goldfish Or, it can be applied in combination with ligand-based virtual screening (Triballeau et al., 2008). Comparing the structures of activating or inhibiting compounds using ligand-based approaches, such as molecular descriptors, can quickly and effectively screen millions of compounds. A second screen can then be applied where potential ligands that share important structural features with known actives can be docked into the binding site of a target protein, which is a more time consuming process.

In order to learn the technique of structure-based virtual screening, I performed an external research project under the guidance of Professor Sir. Tom Blundell at the University of Cambridge, UK. Under the guidance of specialists in his laboratory, including Dr. Willaim Pitt, I successfully applied structure-based virtual screening to

identify potent inhibitors of the enzyme EthR, a promising second line drug target currently under investigation for Tuberculosis (TB) treatment.

TB is a major health challenge around the globe. Roughly 1/3 of the world population is infected with *Mycobacterium tuberculosis*, which is responsible to leading to the infectious disease Tuberculosis that is responsible for approximately 2 million deaths each year. The currently approved course of treatment, titled Directly Observed Therapy, Short-Course (DOTS), comprises 2 months of treatment using isoniazide, rifampicin, ethambutol, and pyrazinamide followed by a 4-6 month treatment with only 2 of the aforementioned drugs (Raviglione, 2003). Unfortunately as has been seen repeatedly in bacteria, a multidrug resistant strain has been emerging, with which it is estimated that roughly 50 million people are currently infected (Chan and Iseman, 2002; Mahmoudi and Iseman, 1993). As a response the efficacy of several "second line" drugs have been under investigation and considered increasingly important.

One of these second line drugs ethionamide (ETH) has shown good efficacy, but only at relatively high doses, leading to high rates of side effects (Frenois et al., 2004). ETH functions as a pro-drug, since it needs to be activated by a mycobacterial enzyme prior to being able to perform its antimicrobial effect (Baulard et al., 2000; DeBarber et al., 2000). The activator of ETH has been determined to be EthA protein, which is a monooxygenase that oxidizes ETH, thus activating the compound (Figure 6.1) (Baulard et al., 2000; Vannelli et al., 2002). A repressor gene EthR, which has been shown to transcriptionally regulate expression of EthA through a 55bp operator located in the ethA-R intergenic region, regulates EthA (Aramaki et al., 1995; Engohang-Ndong et al., 2004). Inhibition of EthR leads to increased ETH efficacy, as the prodrug is more readily activated.

Several small molecule inhibitors of EthR have been identified over the last decade. Initial crystal structures of the repressor identified a potential drugable cavity between helices 4 and 9, which contained two dioxane molecules (Dover et al., 2004) (Figure 6.2A). Shortly thereafter another group discovered a naturally bound ligand hexadecyl octanoate which was co-purified and co-crystallized with the repressor (Frenois et al., 2004)(Figure 6.2B). The same group also identified and tested several additional ketone based compounds, identifying benzlacetone as a likely inhibitor of the EthA repressor EthR. Over the years since, a number of groups have identified and co-crystalized additional EthR inhibitors, all of which function within the same long cylindrical cavity (Flipo et al., 2011; Flipo et al., 2012; Willand et al., 2010; Willand et al., 2009) (Figure 6.2C). The efficacy of these compounds ranges from micro to nanomolar.

In an attempt to identify more effective inhibitors of EthR we have applied a combination of both ligand-based and structure-based virtual screening to predict additional putative inhibitors of EthR. As we are able to use our approach to screen a vastly larger chemical space than has previously been inspected, we hoped to identify more potent EthR inhibitors.

RESULTS AND DISCUSSION

Analyzing the ligand binding pocket of EthR

As previously mentioned, a number of inhibitors have formerly been identified and their bound orientations within the EthR cavity had been determined. In addition to these published compounds, several additional molecules have been identified and co-crystallized by Dr. Sachin Surade, a member of the Blundel lab. Analysis of the EthR

structure displays a long narrow cavity that runs between helices 4, 5, 7, 8, and 9 (Figure 6.3). The cavity entrance is open and resides between helices 4, 5, and 7.

By Integrating all solved ligand binding locations and analyzing their orientations in EthR, it appears there are certainly two and possibly up to three distinct binding regions within the EthR cavity (Figures 6.4, 6.5). Very few ligands span the entire length of the cavity. The majority of compounds bind close to the cavity entrance at site 2 and a select few reside deep within the pocket in Site1. A few longer compounds extend into what could be considered site 3, which is an extension of site 2, and leave the binding pocket. From previous analysis it appears that site 2 is the most potent of these cavity positions, however it is important to note that the number of ligands identified for site 1 are limited, making a more thorough analysis required before a final decision can be made. This led us to focus our efforts on identifying and virtually docking ligands into sites 1 and 2 of the protein target, which we will refer to as small site 1 and small site 2. As a select few known ligands reside across both sites, we also performed a separate analysis where we attempted to dock ligands across the entire cavity, which we will call the combined site.

Preparing a Virtual Screening Library

We assembled a virtual library for inhibitory compound screening that contained only commercially available chemicals that were purchasable. Our library contained compounds from the Enamine (advanced library: 204,772 compounds), Maybridge (screening library: 56,000 compounds), and Asinex (merged library: 439,946 compounds) libraries containing 700,930 unique structures. Upon removal of duplicate

molecules shared across libraries (9,661 removed) and molecules containing salts or unappealing atom types, we remained with a sizable chemical library of 588,723 compounds (See Methods).

Compounds identified to bind within small site 1 were duplicates of those bound within small site 2 (Sachin Unpublished). Interestingly, binding site 1 was only occupied when site 2 was filled, raising the possibility that this site may be more challenging to fill. As a result, a single library was docked within both site 1 and site 2 and a separate and larger compound set was created for docking into the combined site.

As it would take a great deal of time to computationally dock >550,000 chemicals into each of our binding sites, we analyzed previously identified ligands to reduce the size of our screening library. We determined structural features that may be important for fitting within the cavity and for binding to EthR. By analyzing known ligands, it appears that the inhibitory cavity of EthR is long and fairly thin, thus all known compounds bound to either site are also relatively thin and unbranched. We also noticed that compounds that bound only into sites 1 or 2 were small, contained at least one aromatic ring, and often contained a ketone within the structure that formed a hydrogen bond within each of the cavities (Asparagine 179 – Site 2, Glutamine 125 - Site 1). Additionally, we identified 2 potential hydrogen bonding sites that are rarely satisfied by any of the known ligands (Asparagine 176 – Site2, Threonine 121 – Site 1). Considering these criteria, we assembled our two separate libraries, one to individually dock into binding sites 1 and 2 and another for the cavity spanning combined site.

We selected compounds for our small ligand library from the full set of >550,000 compounds using the following criteria: <= 3 H-bond acceptors/donors, at least 1 aromatic ring, a molecular weight of less than 250, and had no rings of size greater than

7. As the compounds in this library were relatively small, we were not concerned with removing compounds containing side chain branches that would be to wide for the cavity. These criteria resulted in selection of 38,893 compounds for a small site library that could be docked into binding sites 1 and 2.

We selected our large ligand library by application of both molecular descriptors and machine learning. While the combined site library was also selected from the full set of >550,000 compounds, it needed to contain much larger structures. As there were many compounds with sizes that would span the entire cavity in the full library, a more meticulous criterion was used in its selection. We first attempted to identify a single molecular descriptor from either the commercially available Dragon descriptor suite or pipeline pilot that defined the length and narrowness of a compound. Unfortunately, we were unable to identify such a descriptor. As a result, we compared three different sets of general shape descriptors (Balaban, Chi, and Kappa) for their ability to separate out long thin compounds. At the same time we also compared several machine learning approaches including random forests, support vector machines, and linear discriminate analysis for their ability to separate out the ideal compound shapes using each of the three descriptors. Each of these methods requires both a positive and a negative training set, which in this case would be a set of compounds that inhibit EthR and another set that do not. We assembled a training compound set containing ~20 previously know ligands (positive training set) and, as no negative ligands were known, 100 randomly select ligands from the Asinex library (negative training set) in order to compare each method and descriptor for the ability to separate known binding from non-binding chemicals. Application of the random forest approach trained on balaban descriptor values was the most effective at differentiating the training set with the largest Area

Under the Curve (AUC) of the Reciever Operating Characteristic (ROC) of 0.95. We next applied the random forest technique to discriminate between similar and dissimilar compounds based on their balaban descriptor values, reducing the full original library from 588,723 to 112,692 compounds. As it requires a great deal longer to dock compounds into the long combined binding site, the library was still far too large. We further reduced the size of the large ligand library by randomly selecting 6,000 chemicals meeting the following requirements that were determined based upon the structures of previously identified compounds that bound into the combined cavity: MW > 400, >= 1 aromatic ring, and no ring sizes greater than 6 atoms.

Selecting Docking Structures of EthR

We next selected the crystal structures and binding pockets that would be used to perform the virtually screen of our final compound libraries. In total, 32 previously determined EthR crystal structures were available, consisting of 24 unpublished (Surade Unpublished) and 8 previously published (3G1M, 1T56, 1U9N, 1U9O, 3G1L, 3G1O, 3O8H, and 3O8G) structures. While initial attempts were made to use ensemble docking, which uses multiple crystal structures of the same protein to allow for protein backbone and R group movements, the computational docking time greatly exceeded our time constraints. As a result, we selected a single structure that best defined each binding site for docking.

We identified the single structure that underwent the largest conformational shift in the DNA binding region, which in theory would also cause the largest reduction in repression of EthA, and contained the widest ligand cavity. All protein structures were aligned to their largest alpha helix using Pymol. Each structure was compared for R

group shifts leading to increase or decrease in binding cavity shape. After careful inspection, it appeared that MB1047 and NT06 had the most open configurations in site 2 and site 1, respectively, and were selected for docking (Surade Unpublished). NT06 was also selected for combined site docking as this structure had the largest cavity spanning both binding sites.

Selecting Hydrogen Bond Locations Important for Protein-Ligand Interface

The hydrogen bonding between a ligand and its protein-binding site is critically important for describing the mechanism of attachment. As such, it is essential that we identify which binding locations appear to be important for EthR. Visual analysis of all previously identified ligands, their binding sites, and the protein cavity itself uncovered Asparagine 179 for site 1 and Glutamine 125 for site 2, as being highly important to hydrogen bonding at the protein-ligand interface. Asparagine 176 and Threonine 149 in site 1 and Threonine 121 in site 2 may also play a less substantial role, as evidence for their interactions with known ligands is present, but less pronounced.

Self-Docking and Cross-Docking Validations

We next compared how well the Gold software suite self-docked, which involves docking a ligand back into its own crystal structure, and cross-docked, which involves docking other solved ligands for the same receptor into a crystal structure, for the structures NT06 and MB1047. As we know exactly where each of these compounds should bind, they provide good indicators as to how well the docking software and protein system will work.

For self-Docking, ligands for both NT06 and MB1047 were re-docked into their own solved structures using the GOLD suite (Verdonk et al., 2003). H-Bond weighted constraints for Asparagine 179 in site 1 and Glutamine 125 in site 2 were compared to docking runs where no constraints were implemented. Self-docking runs were often able to successfully reinsert their own ligands back into the correct position and orientation in the receptor (Figures 6.6, 6.7). For our cross-docking analysis, we docked a variety of ligands known to bind to EthR. Cross-docking is a more stringent test of binding as a structure that is used for docking was not solved with a ligand that is attempted to be docked creating a test to see how "accepting" the binding site is compounds with different shapes that are known to bind as well. The analyses yielded promising results, as many of the ligands were positioned very similarly to the natural binding positions of their own solved structures.

We next compared the effectiveness of both the presence and absence of H-bond constraints. Interestingly, docking with and without H-bond constraints provided similar results, indicating that even without constraints this interaction is likely energetically preferred. However, as a large number of known ligands contain H-bond interactions, we feel they will be highly important in discriminating between the tens of thousands of compounds in our virtual screen, many of which may not be able to form this H-bond. As such, we have decided to include constraints for all docking.

Performing structure-based virtual screening on EthR

We performed 3 independent docking experiments. Each experiment attempted to dock a single library into a binding site using the Gold software suite. While the many of the same settings were applied across experiments, several were unique to a binding

site, such as H-bond constraints. Each compound was docked multiple times into a binding site to increase the likelihood of identifying the correct location and orientation in the binding pocket. The structure of the EthR protein was kept rigid for all 3 docking scenarios.

We first docked our small library into the EthR binding site 1 of the solved structure NTO6. Each of the 38,893 compounds from the small site library were independently docked into binding site 1, 10 times. H-bond constraints were applied for both Asparagine 179 and Asparagine 176.

We next docked our small library into the EthR binding site 2 of the solved structure MB1047. Each of the 38,893 compounds from the small site library were independently docked into binding site 2 10 times. An H-bond constraint was applied for Glutamine 125.

Lastly, we docked our combined library into the full EthR cavity of the solved structure NT06, which spans nearly the entire cavity including from deep within the pocket at biding site 1 to binding site 2 and beyond to the opening of the tunnel. Each of the 6,000 structures was independently docked into the cavity 3 times. We had reduced the number of docking iterations to 3 due to computational constraints. As the size of both the ligands and docking cavity was much larger for the combined site, a far longer time was required for each individual docking. All H-bond constraints applied in both sites 1 and 2 were implemented for combined site docking (Asparagine 179, Asparagine 176, Glutamine 125).

Selecting docked ligands for experimental validation

We analyzed Gold docking results using Goldmine, which is an integrated portion of the Gold software suite. The Goldmine package allows for visual inspection and thresholding of docking results based on H-bond constraints and fitness values. Fitness values and H-bond constratins were individually applied to each docking set, resulting in a reduction to 708 compounds for site 2, 619 compounds for site 1, and 69 compounds for the combined site. We then sorted the remaining compounds by H-bond score and visually inspected each docking, selecting the most promising leads. We ultimately decided on 14 compounds for purchase and experimental validation. 6 of the selected compounds originated from binding site 1, 3 from binding site 2, and 5 from the combined site.

Experimental validation of predicted ligands

All experimental validation was performed by Dr Sachin Surade. The 14 compounds were assayed for their inhibition of EhtR binding affinity to the promoter sequence of EthA. Surface Plasmon Resonance (SPR) assays were performed to determine the inhibition caused by each compound at a 50 micro molar concentration. SPR assay activity determined for the compounds ranged from 95% inhibition to 159% activation (Table 6.1). We believe activation values may be a product of interference with the assay itself.

Analysis of predicted ligand binding and experimentally validated activity

We next inspected each of the docking conformations and compared their validated inhibitory values to orientations within their associated binding pocket. It is important to note that predicted binding orientations may not be exactly as they are

when truly bound. Our screening approach only considered the rigid structure of a single enzyme. Since protein structures in cellular conditions are able to shift and flex and many amino acid side chains are able to rotate, the true binding conditions will likely allow for more ligand movement and a better orientation in the binding pocket.

Compound BAS118136, which was docked into site 1, had the highest validated inhibition of 95% (Figure 6.8 (Top)). From analysis of the predicted binding location it appears that two hydrogen bonds may be strengthening the interaction. It appears that the ligand may actually be shifted forward and rotated forming interactions with Glutamine 125 and Glutamine 180, however this is only a hypothesis would need to be determined with a crystal sctructure. The wide shape of the structure placed into binding site 1 should cause a large shift in DNA binding region of EthR, thus reducing the repression of EthA and increasing ETH effectiveness.

Compound T6977356, which is docked into site 2 had the second highest validated inhibiton of 92% (Figure 6.8 (Middle)). From analysis of the predicted binding location it appears that a hydrogen bond with Asparagine 179 is formed. The ligand shape is long and flat, which fits very well into the cavity and spans from the edge of binding site 1 to the cavity opening. This long addition in the cavity should cause a large shift to the backbone of EthR.

Compound T6779777, which is also docked into site 2 had the third highest validated inhibition of 86% (Figure 6.8 (Bottom)). This compound has a very similar orientation in the binding cavity to T6977356. There appears to be a hydrogen bond forming with Asparagine 179, however this molecule contains two additional H-bond accepting oxygen atoms that do not appear to be forming specific interactions. It is possible that the ligand has a slightly different orientation in its naturally bound state.

The SPR analysis determined that 4 additional compounds have inhibitory values that are between 25% and 75% (Figure 6.9). These moderate inhibitors range in shape and size. One was predicted to interact with binding site 1, one is predicted to interact at binding site 2, and two are predicted to interact across the whole combined cavity. All 4 molecules appear to form H-bonds with at least one amino acid. The two compounds from the combined docking set may form multiple H-bond interactions. 3 of the molecules have bi or tri-cyclic rings that are likely effective in expanding the cavity and thus effectively reducing the ability of EthR to bind to the DNA promoter region upstream of EthA.

Finally, 6 of the compounds had validated inhibitory values lower than 25% and were poor inhibitors of EthR (Figure 6.10). Poor inhibitory compounds originated from all three binding sets. Several compounds contained very large side groups including an adamantane group and a bicyclo[2.2.2]octante. Interestingly, several of the compounds validated as activators of EthR. While it is possible these compounds increased the binding affinity of EthR to DNA, it is probably more likely they have some off target effect on the assay itself.

Future directions in validation

Preliminary results presented here show great promise and we will continue validation of these compounds in several ways. Firstly, It will be important to determine IC50 values so that they can be compared to the current best in class inhibitors of EthR. Each of these compounds has been tested for activity at a 50 micro molar concentration. Since two of our compounds caused very high inhibition (>90%), it is possible that these compounds could be as effective as the best in class compounds. Secondly, it will be

important to test for inhibition across multiple assays. The responses of EthR to a compound may change in a different assay. Finally, it will be very important to determine the true binding orientations of the best compounds within the binding pocket. In order to achieve this we will need to co-crystallize compounds into EthR. As EthR has been co-crystallized with a number of other compounds, this should be an achievable goal. It will be very interesting to compare the predicted with the actual ligand orientations in EthR.

The high success rate in identifying highly inhibitory ligands is a testament to the utility that structure-based virtual screening can have on ligand identification. In a period of 6 weeks we assembled a large untested library of >550,000 compounds, applied ligand-based virtual screening to analyze structural features that are important for ligand binding to limit the library sizes for individual binding sites, and performed structure-based virtual screens for 3 independent binding sites on EthR. From our very small experimental validation we find an accuracy of 36% at identifying inhibitors (>50% inhibition), with 14% being very strong inhibitors (>90% inhibition). The ligand specificity of the EthR receptors is almost certainly vastly more selective than for insect odor receptors, as many insect odor receptors are activated or inhibited by an array of structural features and sizes. While the accuracy of this analysis was lower than what we have observed by applying ligand-based virtual screening in insect species (~72% in Drosophila and ~65% in Anopheles), we would expect a far higher success rate in odor receptors by application of a two stage (ligand-based and then structure-based) screen. As the first structures of insect odor receptors and additional mammalian odor receptors become available, application of structure-based virtual screening should be of high priority.

Figure 6.1: Schematics of Ethionamide Activation Pathway

The Ethionamide activation pathway. EthR represses the expression of EthA. EthA is responsible for activating Ethionamide causing the antibacterial effect. By inhibiting our target EthR, a larger amount of Ethionamide is converted into active form, thus increasing the effectiveness of the drug.

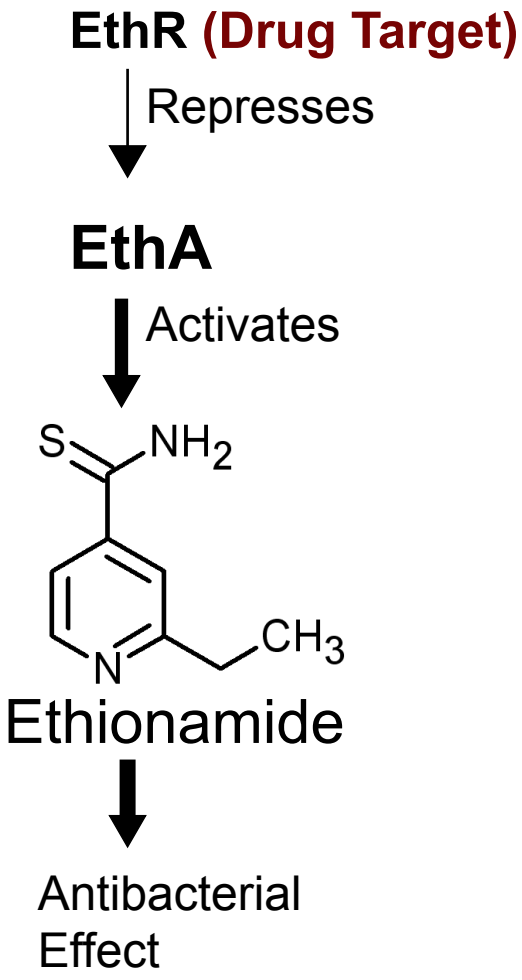

EthR (Drug Target)
Represses
EthA
Activates
S
NH2
CH3
N
Ethionamide
Antibacterial
Effect

Figure 6.1

Figure 6.2: Solved EthR structures

(A) The first solved structure of EthR. Fortuitously, the receptor was co-crystallized with two dioxane molecules, allowing for identification of a cavity within the protein. This structure was previously determined, obtained from PDB, and visualized in PyMol (Dover et al., 2004). **(B)** The first identified structure of EthR bound to a naturally occurring ligand. EthR is bound to hexadecyl octanoate, which spans a very large cavity in EthR. This structure was previously determined, obtained from PDB, and visualized in PyMol (Frenois et al., 2004). **(C)** An overlay of several bound EthR ligands. Individually co-crystallized structures have been aligned by their largest alpha helix, producing a clear representation of the biding cavity. All structures were previously determined, obtained from PDB, and visualized by PyMol (Dover et al., 2004; Flipo et al., 2011; Frenois et al., 2004; Willand et al., 2010; Willand et al., 2009).

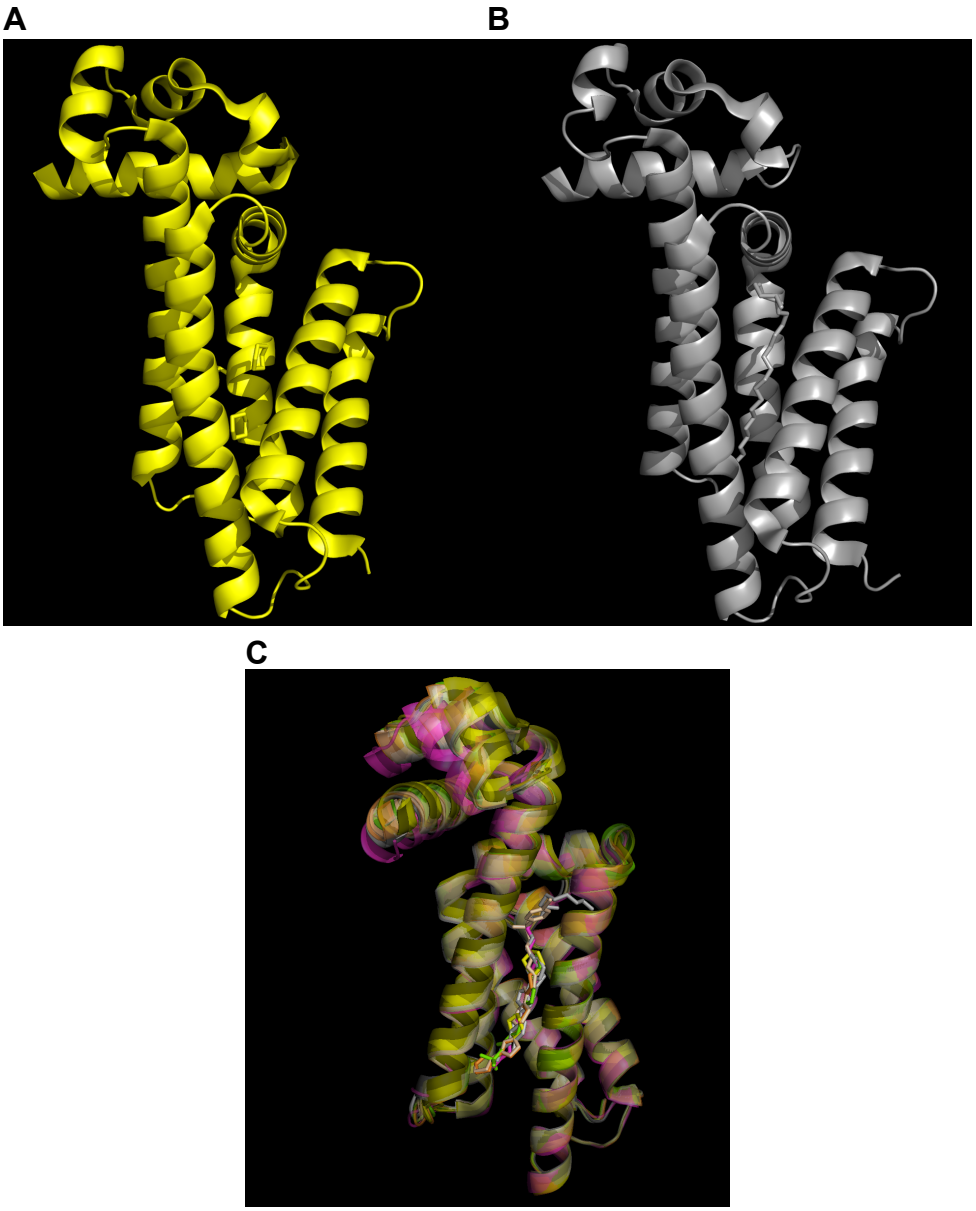


Figure 6.2

Figure 6.3: The shape of the EthR cavity

A structure of EthR (NT06), which was recently solved by members of the Blundell lab, was used along with PyMol to visualize a long and unbranged cylindrical cavity that traverses nearly the entire length of EthR .

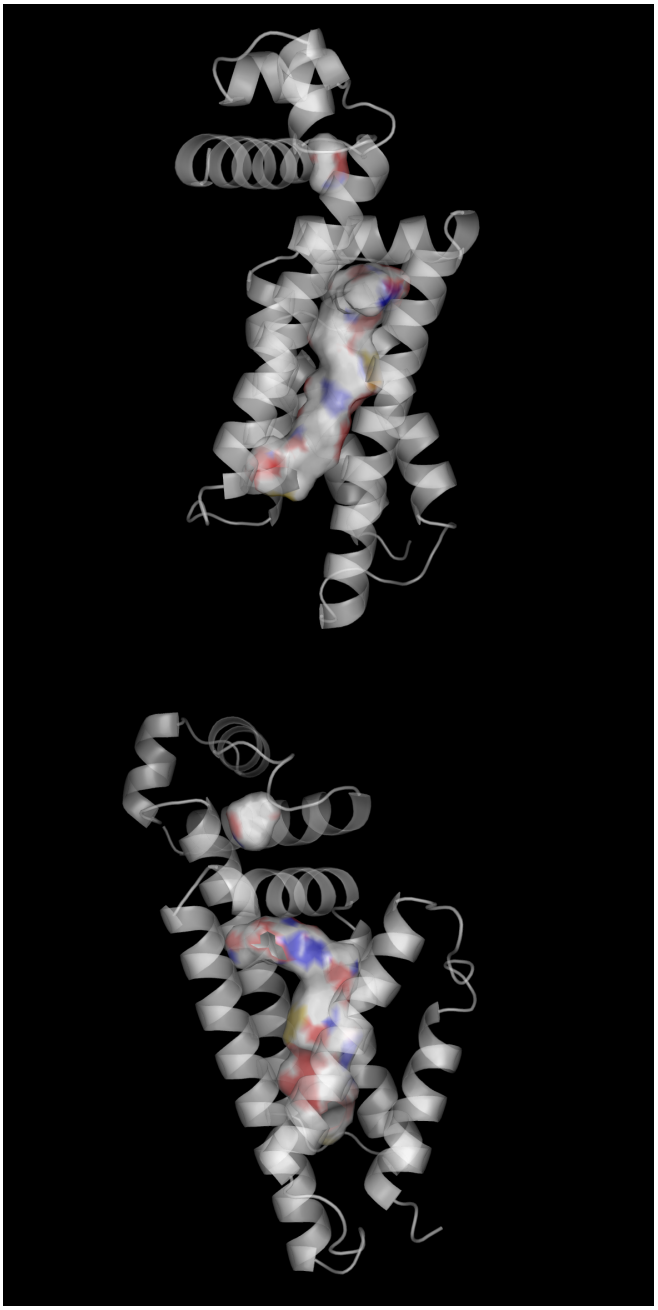


Figure 6.3

Figure 6.4: Previously identified EthR ligands bind within the proposed cavity

Previously determined ligands were oriented into the EthR binding cavity. Three distinct binding regions can be observed. Similar compounds bind into sites 1 and 2. Compounds that extend from the site 2 occupy a hypothetical site 3.

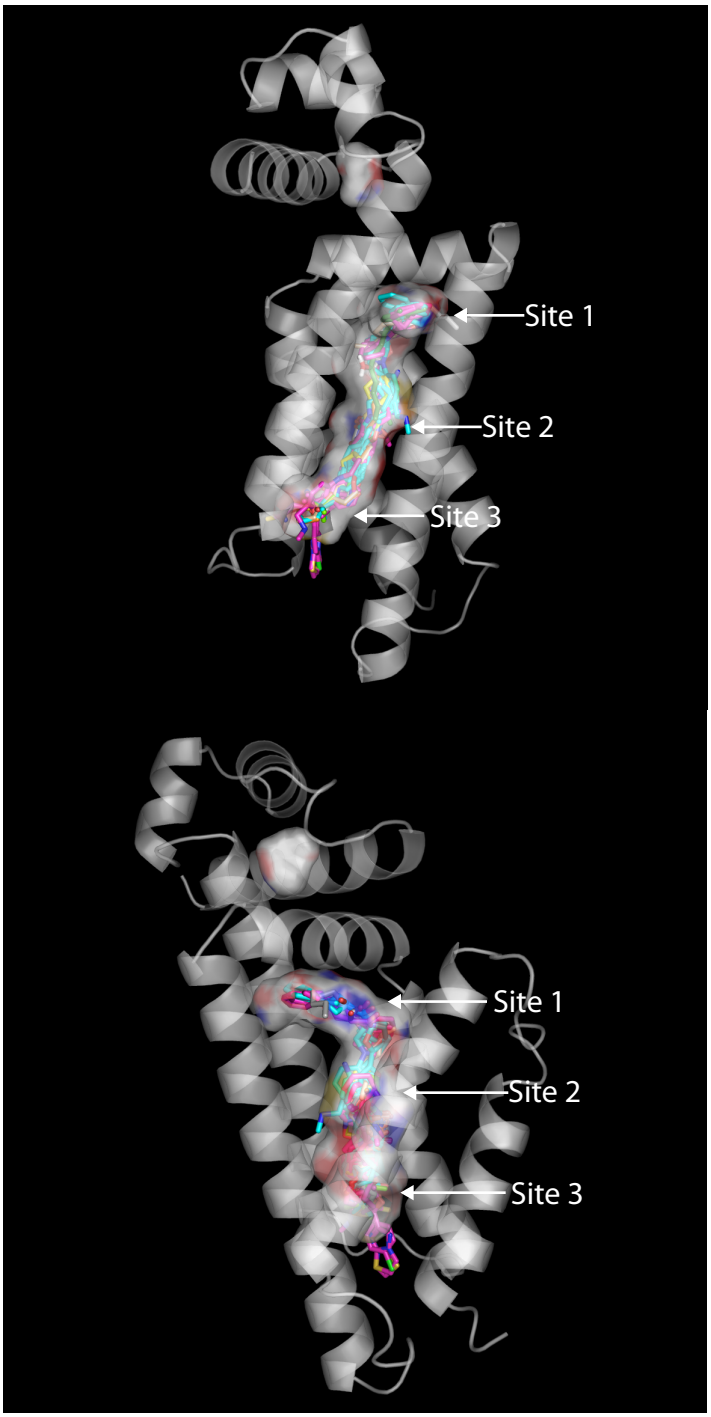


Figure 6.4

Figure 6.5: Visualizing previously identified EthR ligands

Three different views of previously identified EthR ligands residing within the cavity.

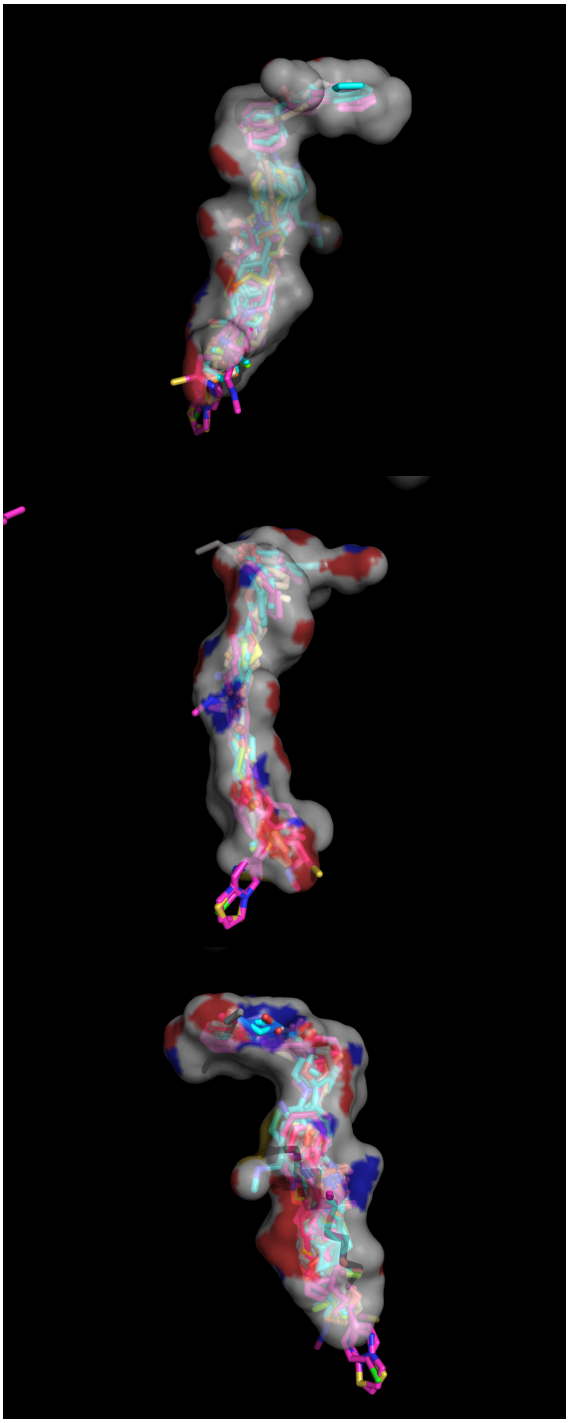


Figure 6.5

Figure 6.6: Single example of successful EthR self-docking

The EthR ligand (brown ligand) has been successfully self-docked back into the correct orientation (yellow ligand) in the EthR binding site 1.

Figure 6.6

Figure 6.7: The results of many EthR site 1 self-docking runs

The Gold software suite is highly successful at self-docking ligands back into EthR site 1. Each overlayed ligand represents an individual self-docking into binding site 1. The majority of ligands reside in the correct orientation.

Figure 6.7

Figure 6.8: The predicted orientation of our strongest identified inhibitors in EthR

The resulted docking of our top three strongest validated inhibitors bound into EthR. The percent inhibition represents the inhibition observed by SPR experimental analysis. An individual image is provided for each inhibitor.

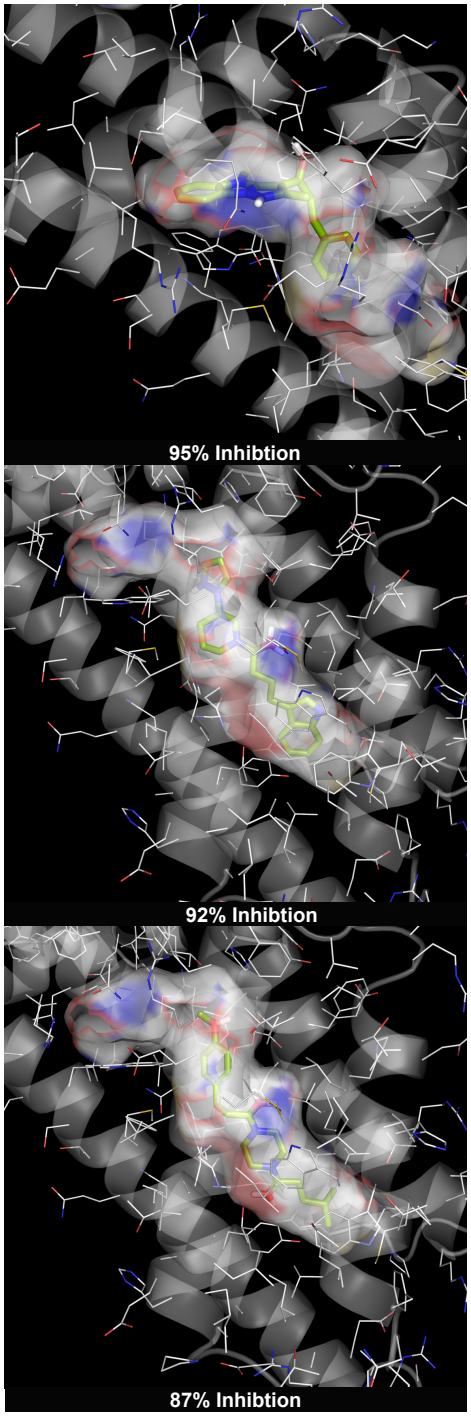


Figure 6.8

Figure 6.9: The predicted orientations of 4 modest inhibitors of EthR

The resulted dockings of our modest validated inhibitors bound into EthR. The percent inhibition represents the inhibition observed by SPR experimental analysis. An individual image is provided for each inhibitor.

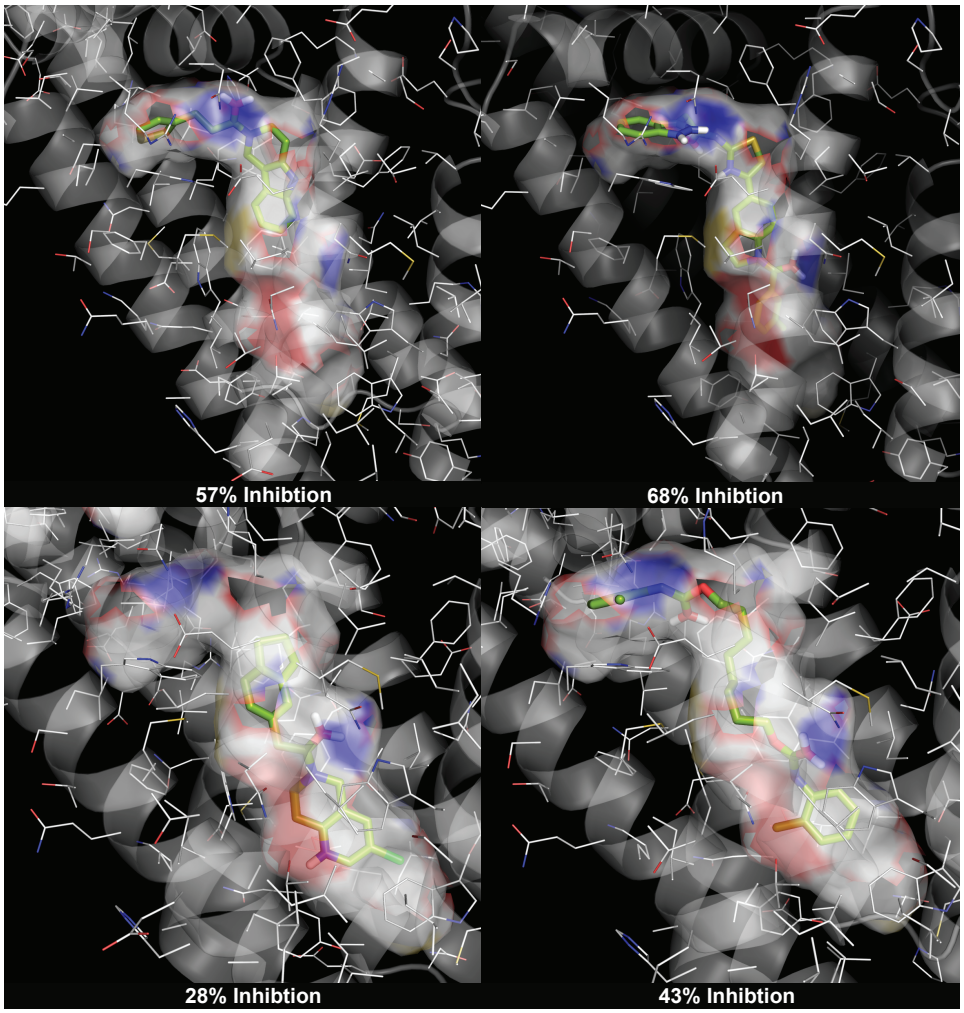


Figure 6.9

Figure 6.10: The predicted orientations of the poor inhibitors of EthR

The resulted dockings of our poor inhibitors bound into EthR. The percent inhibition or activation represents either the inhibition or activation observed by SPR experimental analysis. An individual image is provided for each compound.

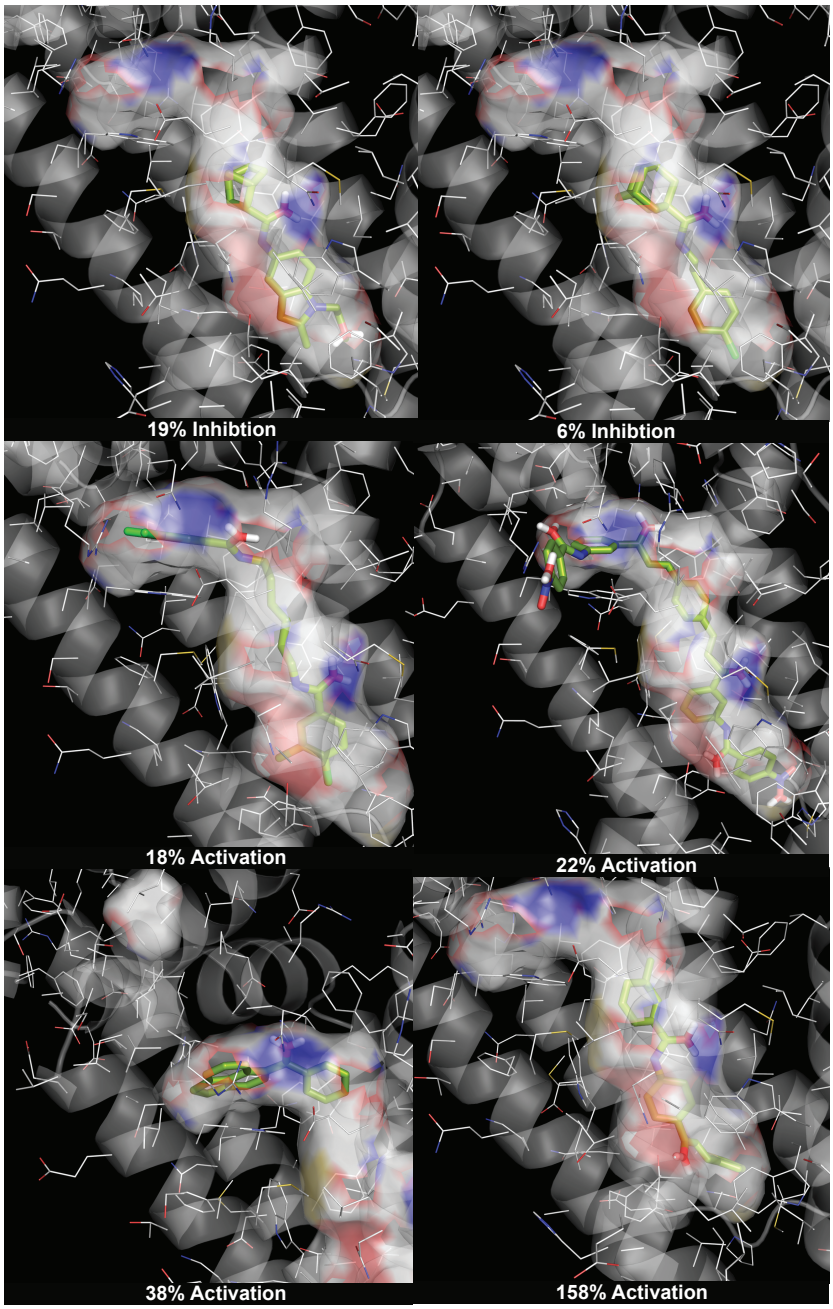

19% Inhibtion
6% Inhibtion
18% Activation
22% Activation
38% Activation
158% Activation

Figure 6.10

Table 6.1: A breakdown of our structure-based virtual screening accuracy for EthR

The accuracy of our virtual screening approach was validated by SPR. The number of compounds that were determined to inhibit EthR are listed by percent inhibition.

Inhibition Criteria	Number of Compounds
> 90% Inhibition	2/14
> 75% Inhibition	3/14
> 50% Inhibition	5/14
> 25% Inhibition	8/14
< 25% Inhibition	6/14

Table 6.1

CHAPTER VII:

Methods

Chemical informatics

Energy minimized 3-D structures were generated using Omega2 software (OpenEye)(Hawkins et al., 2010). Optimized descriptor subsets were iteratively identified from the large Dragon (Talete) and Cerius2 (Accelrys) molecular descriptor sets using a Sequential Forward Selection (SFS) method (Whitney, 1971). The SFS approach functioned by selecting descriptors that increased the correlation between the training set odor distances calculated by activity (spikes/sec) and the growing optimized descriptor set. A 5-fold cross-validation and a receiver-operating-characteristics (ROC) analysis was applied to analyse the performance of the receptor-optimized descriptor sets. Each receptor-optimized descriptor set was then used to rank the combined Natural Odor and Pubchem Libraries (>240,000 compounds) based on each odors distance to a known activator for the corresponding receptor.

Electrophysiology

Extracellular single-sensillum electrophysiology was performed as before (de Bruyne et al., 2001b; Dobritsa et al., 2003; Hallem and Carlson, 2006) with a few modifications. Diagnostic odorants were used to distinguish individual classes of ORNs in sensilla (ab1-ab7) and therefore unequivocally identify the target ORN for testing (de Bruyne et al., 2001b; Hallem et al., 2004). 50ml odor at 10^{-2} dilution in paraffin oil was applied to cotton wool in odor cartridge. Odor stimulus flow = 12ml/second. Due to variability in temporal kinetics of response across various odors, the counting window was shortened

to 250 milliseconds from the start of odor stimulus. A diagnostic panel of odorants were used to distinguish individual classes of sensilla (ab1-ab7) and therefore unequivocally identified the target ORN(de Bruyne et al., 2001a; Hallem et al., 2004).

Natural odor compound library

We assembled a subset of 3197 volatile compounds from annotated origins including plants(Knudsen et al., 2006), insects(El-Sayed, 2009), humans, and a fragrance collection(Sigma-Aldrich, 2007) that may have additional fruit and floral volatiles(Cork and Park, 1996; Curran et al., 2005; Gallagher et al., 2008; Knudsen et al., 2006; Logan et al., 2008; Meijerink et al., 2000; Zeng et al., 1991; Zeng et al., 1996).

Pubchem compound library

We assembled a subset of 241,150 odors from Pubchem, which have similar characteristics to known odor molecules. Compounds met a criteria of MW < 200 and only being composed of the following atoms (C, O, N, H, I, Cl, S, F).

eMolecules compound library

We assembled a subset of >440,000 odors from eMolecules(eMolecules), which have similar characteristics to known odor molecules. Compounds met a criteria of MW < 350 and only being composed of the following atoms (C, O, N, H, I, Cl, S, F).

Calculation of 3D conformations

The 3-Dimensional structures were predicted for compounds through use of the Omega2 software package(Bostrom et al., 2003; Hawkins et al., 2010). The Omega2 software

package functions in three major stages: assembly of an initial 3D structure from a library of fragments; exhaustive enumeration of all rotatable torsions using values drawn from a knowledge-based list of angles, creating a large number of conformations; and sampling of this set by geometric and energy criteria(Hawkins et al., 2010). The lowest energy 3D conformer for each compound in our Pubchem and Natural compound libraries were stored for use in molecular descriptor calculation.

Calculation of molecular descriptors

Commercially available software packages Cerius2 (200 individual descriptors) and Dragon (3224 individual descriptors) from Accelrys and Talete were used to calculate molecular descriptors from 3D molecular structures. Descriptor values were normalized across compounds to standard scores by subtracting the mean value for each descriptor type and dividing by the standard deviation. Molecular descriptors that did not show variation across compounds were removed. Maximum Common Substructures were determined using an existing algorithm(Cao et al., 2008b). Atom Pairs were computed from the version implemented in ChemmineR(Cao et al., 2008a).

Classification of active compounds

Since we were interested in identifying descriptors which best described active compounds, we needed to first determine which compounds to classify as “active” based on their electrophysiology activity for the receptor being studied. All of the training odors were clustered using hierarchical clustering by activity individually for each Or. The resulting tree can then be pruned such that the branch containing the majority of activating odors (>50 spikes/second) are selected. The activity threshold therefore was

set as the lowest spike/second activity of any odor present in the selected branch.

Calculation of Accumulative Percentage of Actives (APoA)

The accumulative percentage of actives is calculated for each descriptor set individually as previously described(Chen and Reynolds, 2002). Compounds are ranked according to their distance from each known active using the optimized descriptor values as distance, resulting in one set of ranked compound distances from each activating odor. Moving down the list for each of these rankings, ratios are calculated for the number of active compounds observed divided by the total number of compounds inspected, or the APoA. APoA values are averaged across all active compound rankings, creating a single set of mean values representing the APoA for a single Or and descriptor set. Using this approach, ApoA mean values are calculated for each of the 24 Ors separately for each descriptor set used, including optimized sets, all Dragon, all Cerius2, Atom Pair, and Maximum Common Substructure. The Area Under the Curve (AUC) scores were calculated by approximation of the integral under each plotted APoA line.

Determination of optimized descriptor subsets

A compound-by-compound activity distance matrix was calculated using training odor activity data for each of the Ors(Hallem and Carlson, 2006). A separate compound-by-compound descriptor distance matrix was calculated using the 3,424 descriptor values for training odors calculated by Dragon and Cerius2. Active compounds for each Or were identified individually through activity thresholds, as described above. The

correlation between the compound-by-compound activity and compound-by-compound descriptor distance matrices were compared for each actively classified compound, considering their distances to all other compounds. The goal was to identify descriptors that correlate most closely with activity. Using a Sequential Forward Selection (SFS) approach, which involves incrementally adding a single best choice item to growing list in an attempt to produced an optimal final set of items(Whitney, 1971), all descriptors are individually compared for their ability to increase this correlation. The descriptor that correlates best is retained and the process is iteratively used to search for additional descriptors. Each iteration aims to further increases in correlation values. In this manner, the size of the optimized descriptor set increases by one in each iteration, as the best descriptor set from the previous step is combined with all possible descriptors to find the next best descriptor. This process is halted when all possible descriptor additions in an iteration fails to improve the correlation value from the previous step. This whole process is run independently for each Or resulting in unique descriptor sets that are optimized for each Or.

Clustering Ors by most common descriptors

The first 20 descriptors selected by our optimized descriptor selection algorithm for each Or were used to create an identity matrix. Each row representing an Or and column a specific descriptor. Ors that share common descriptors contain 1s in the same column. This matrix was then converted into an Or-by-Or Euclidean distance matrix and clustered using hierarchical clustering and complete linkage.

Clustering compounds by activity of Or

The responses of each of the Ors that had previously been tested against a panel of compounds were converted into an Or-by-Or Euclidean distance matrix (Hallem and Carlson, 2006). Ors were clustered using hierarchical clustering and complete linkage. Specifically, this was achieved by creating a compound-by-compound distance matrix using the differences in activity between compounds tested on a singe Or. Hierarchical clustering using each Or distance matrix and then identifying the sub cluster which contained the most compounds.

Clustering Ors by predicted ligand space

Percentages of overlapping predictions within the top 500 predicted compounds were calculated pair-wise for all Ors. Euclidean distances were calculated from the similarity between Ors.

Calculation of Or prediction distribution frequencies

Initially, all extreme outliers were removed from the dataset for each Or. On average 5.82 compounds were removed for each Or, resulting in a mean dataset reduction of 0.0024%. Next, all compounds whose distance was greater than 3 standard deviations from the strongest activating compound were removed to reduce outliers. Distribution frequencies were produced for each Or. All compound distances were converted into a percentage of the most distant compound for each Or. Frequencies of compounds in the top 15% were plotted.

Or-ligand interaction map

The Or-ligand interaction map was developed using Cytoscape(Shannon et al., 2003).

Each predicted Or-ligand interaction from the top 500 predicted ligands for all of the Ors listed were used to calculate the map (Figure 5A). All predicted interactions are labelled in purple. In addition all interactions identified in this study and the previous study(Hallem and Carlson, 2006) were included and labeled in gray. All compounds are represented as small black circles and Ors are represented as large colored circles. Or names are provided on the upper right corner of each Or.

Computational validation of ligand-based virtual screening approach (non- SVMs)

We performed 5 independent 5-fold cross-validations. For each independent validation the dataset was divided into 5 equal sized partitions containing roughly 22 compounds each. During each run, one of the partitions is selected for testing, and the remaining 4 sets are used for training. The training process is repeated 5 times with each unique odorant set being used as the test set exactly once. For every training iteration, a unique set of descriptors was calculated from the training compound set. These descriptors were then used to calculate distances of the test set compounds to the closest active compound, exactly as we use to predict ligands in our ligand discovery pipeline. Once test set compounds have been ranked by distance from closest to furthest to a known active in the training set, a receiver operating characteristics (ROC) analysis is used to analyze the performance of our computational ligand prediction approach. Using ROC we were able to determine our predictive ability for the 12 receptors. We decided to perform this validation only on receptors for which sufficient training odors had previously been identified. We consider this to consist of at least one very strongly activating known odor (>150 spikes/sec) and at least five strongly activating odors (>100 spikes/sec), thus allowing for at least one activating odor for each of the 5 test sets in the

cross-validation (DmOr7a, DmOr9a, DmOr10a, DmOr22a, DmOr35a, DmOr43b, DmOr12, DmOr59b, DmOr67a, DmOr67c, DmOr85b, DmOr98a). Test set validations for all 12 Ors were combined and a single ROC curve representing an average across all Ors was plotted (Fig. 2A).

Calculation of LogP and vapor pressure values

SMILES structures of the predicted odors were used with EPI Suite (http://www.epa.gov/oppt/exposure/pubs/episuite.htm) to calculate predicted LogP and Vapor Pressure values.

Repellency behavior testing

Repellency was tested using *Drosophila melanogaster* 2-choice trap assay as described previously (Reeder et al., 2001; Syed et al., 2011).

Preference Index= number of flies in treated trap/(number of flies in treated + control traps).

The *Drosophila* T-maze assay was conducted as described previously (Turner et al., 2011; Turner and Ray, 2009).

Preference index = (number of flies in test arm - number of flies in control arm)/(number of flies in test arm + number of flies in control arm).

Repellency was tested in mated and starved *Ae. aegypti* females using a hand-in-glove assay. Briefly, a gloved hand with an opening exposing skin odorants protected by 2 layers of netting was presented to mosquitoes for 5 min inside a cage and video taped for landing and avoidance responses. Mosquitoes were unable to bite due to the

outer protective layer of netting and the inner layer of netting was treated with either test compound (10%) or solvent, such that mosquitoes were able to respond to volatiles but unable to make physical contact. The number of mosquitoes present for more than 5 seconds, and the numbers departing during the same period were counted from the videos at minutes 2,3,4, and 5 mins and repellency percentage and escape index calculated by comparing with similar numbers in solvent treated controls.

Percentage repellency = 100 x [1 - (mean cumulative number of mosquitoes on the window of treatment for 5 seconds at time points 2,3,4,5 min/ mean cumulative number of mosquitoes that remained on window of solvent treatment for 5 seconds at time points 2,3,4,5 min)].

Percentage present = average number of mosquitoes on window for 5 seconds at a given time-point across trials. All values were normalized to percentage of the highest value for the comparison, which was assigned a 100 percent present.

Mean Escape Index = (Average Number of mosquitoes in treatment that landed yet left the mesh during a five second window over the following time points: 2 minutes, 3 minutes, 4 minutes, 5 minutes) / (Average Number of mosquitoes that landed yet left the mesh during a five second window over the same time points in (treatment + control)). Each time point has N=5 trials, 40 mosquitoes per trial, Except for EA, where N=4.

The humidity and warmth attraction assay *A. aegypti* were placed in a cylindrical clear acrylic cage covered with insect screen mesh on one end. At 0.5 cm above the cage a moist filter paper and heated source was presented to cause attraction. In between the moist filter paper and the cage either DEET or acetone treated nets were placed far enough above the top of the cage so there could be no contact. For 5 second windows at 30 second intervals throughout the 5 minutes assay the number of

mosquitoes present at the top of the mesh cage were measured and an average was calculated for each condition.

Number of mosquitoes probing = Average number of mosquitoes present at top mesh for 5-sec or more, across the 10 time-point measurements/trial, for 3 trials.

Classification of repellent compounds

Training odors were clustered by their repellency values using Euclidean distance and hierarchical clustering. The resulting cluster containing the majority of strongest repellents was then selected and the repellency threshold for identification of descriptors determined as the lowest protection time within the cluster.

Support Vector Machine (SVM) predictions

The R Package e1071 interface with libsvm, a well established program, was used in the analysis(Chang and Lin, 2001; Karatzoglou et al., 2006). The Tune.SVM function was used to determine optimal gamma (0.01) and cost (100) values. The SVM was trained with the determined optimized descriptors for the training compound set using regression and a radial basis function kernel. The trained SVM then ranked both the eMolecules and natural odor libraries by repellency using their optimized descriptors.

Computational validation of ligand-based virtual screening (SVM)

We performed 20 independent 5-fold cross-validations by dividing the dataset into 5 equal sized partitions containing roughly 40 compounds each. During each run, one of the partitions is selected for testing, and the remaining 4 sets are used for training. The optimized descriptor values for the training set were used in training the SVM to

recognize repellent chemicals. The trained SVM is then used to predict repellent activity for the withheld test set. This process is repeated five times, each trial excluding a different subset of compounds as the training set and assigning the remainder as the test set. The whole process is repeated 20 times to improve consistency. A receiver operating characteristics (ROC) analysis is then used to analyze the performance of our computational repellency prediction approach.

Vinyl solubility assay

One 3 x 3 mm square of 4 gauge vinyl was submerged in 1mL of each test compound in a glass container and stirred at a constant rate on a shaker and checked every 30 minutes until the vinyl square in DEET was completely dissolved (6 Hrs). The vinyl pieces in each of the other compounds was removed, rinsed in ethanol and weighed. The process was repeated at 30 Hrs (24 Hrs after the vinyl square completely disappeared in DEET).

Olfactory avoidance assay trap assay for *Drosophila*

Trap Assay experiments were preformed as described previously(Reeder et al., 2001), (Syed et al., 2011) with minor modifications. Briefly, traps were made with two 1.5 ml microcentrifuge tubes (USA Scientific) and 200 ul pipette tips (USA Scientific), each cap contained standard cornmeal medium. T-shape piece of filter paper (Whatman #1) was impregnated with 5 ul of acetone (control) or 5 ul of 10%, 1%, 0.10% test odor, diluted in acetone. Traps were placed within a petri dish (100 x 15mm, Fisher) containing 10ml of 1% agarose to provide moisture. Ten flies wCs 4-7days old were used per trial which lasted 48 hours by which time point nearly all flies in the assays had made a choice. For

the 24 hour time point data was considered only if 30% of flies had made a choice, at 48 hours the majority of flies had made choices.

Preference Index= number of flies in treated trap/(number of flies in treated + control traps).

T-maze Assay Methods

Carbon Dioxide and DEET trials were conducted as described previously (Turner and Ray, 2009). Sawyer Jungle Juice 100 Insect Repellent (DEET) was dissolved at 10% in dimethyl sulfoxide (DMSO) and impregnated in a filter paper disc and placed at the bottom of test tube. An equal volume of DMSO impregnated on filter paper disc was used in control arm.

Preference index = (number of flies in test arm - number of flies in control arm)/(number of flies in test arm + number of flies in control arm). For trials with DEET and carbon dioxide, tubes were set up with 10%DEET and DMSO in respective arms. Carbon dioxide was injected into DMSO control tube.

Modified hand-in-glove olfactory repellency assay for mosquitoes

Ae. aegypti mosquitoes (eggs obtained from Benzon Research Inc.) were maintained at ~27 ^{o}C and 70% RH on 14h: 10h L: D cycle. Behavioral tests were done with 40 mated, non-blood fed, ~24 hour starved, 4-10 day old females in 30 x 30 x 30 cm cages with a glass top to allow for video recording (Figure 1C, Figure S1). Each test compound solution (500µl) of 10% concentration in acetone solvent was applied evenly to a white rectangular 7 x 6 cm polyester netting (mesh size 26 x 22 holes per square inch) in a glass petri-dish and suspended in the air for 30 minutes to allow solvent evaporation.

The more volatile 2,3-dimethyl-5-isobutyl pyrizine was dissolved in paraffin oil. Acetone or paraffin oil (500µl) served as control. A nitrile glove (Sol-vex) was modified as described in Supplementary Figure 1 such that a 5.8 x 5cm window was present for skin odor exposure. A set of magnetic window frames were designed to secure the treated net ~1.5 mm above skin, and a second untreated netting ~4.5 mm above the treated net in a manner so that mosquitoes were attracted to skin emanations in the open window but unable to contact treated nets with tarsi, or contact and pierce skin. Additionally the test compound had minimal contact with skin. A clean set of glove and magnets were used for every trial. Care was taken that experimenter did not use cosmetics, soap etc on arms. For each trial the arm was first inserted for 5 min and the number landing or escaping test window recorded on video for 5 min period. Solvent controls were always tested prior to treatment, No cage was tested more than once within 1 hour of a testing session and not more than twice on any single day. Videos were analyzed offline on computers and numbers of mosquitoes present for a 5-sec continuous duration were counted every minute. Mosquitoes reliably started accumulating in controls at the 2 min point, and data from this time point was considered for analysis.

Percentage repellency = 100 x [1 - (mean cumulative number of mosquitoes on the window of treatment for 5 seconds at time points 2,3,4,5 min/ mean cumulative number of mosquitoes that remained on window of solvent treatment for 5 seconds at time points 2,3,4,5 min)].

Percentage present = average number of mosquitoes on window for 5 seconds at a given time-point across trials. All values were normalized to percentage of the highest value for the comparison, which was assigned a 100 percent present.

Mean Escape Index = (Average Number of mosquitoes in treatment that landed yet left the mesh during a five second window over the following time points: 2 minutes, 3 minutes, 4 minutes, 5 minutes) / (Average Number of mosquitoes that landed yet left the mesh during a five second window over the same time points in (treatment + control)) Each time point has N=5 trials, 40 mosquitoes per trial, Except for EA, where N=4.

Humidity and warmth attraction assay

Experiments with female *Aedes aegypti* for DEET repellency were conducted at 27±1^{o}C and 50-55% under fluorescent lighting from 1400-1700hrs. Each assay lasted 5 minutes. In the humidity-and-heat-attraction assay, 10 females (24hr starved mosquitoes) were held in cylindrical clear acrylic cages (7cm diameter x 5 cm high) covered with insect screen mesh on one end and sealed with clean manila paper on the open end. Mosquitoes were exposed to DEET-treated or acetone(solvent)-treated mesh placed 0.5cm above the screened end of the cage. A 55mm Whatman #1 filter paper disc moistened with 400µL of water was placed over the net to provide humidity. Heat was provided by placing a 148mL dram plastic vial (Thornton Plastics, Utah USA) containing 75mL of water at 37^{o}C over the filter paper. Mosquito behavior around the treated net was recorded on HD video and the number of probing mosquitoes sitting on the top screen mesh for >5-sec were counted every 30-sec interval of the 5-min duration of the assay.

Number of mosquitoes probing = Average number of mosquitoes present at top mesh for 5-sec or more, across the 10 time-point measurements/trial, for 3 trials.

Larval behavior assays

Behavior assays were conducted as in (Kreher et al., 2008) with some modifications. Odorant solution and paraffin oil solvent were presented in small plastic containers fashioned out of inverted eppendorf tube caps. For pre-exposure experiments ~50 larvae per trial were placed on a 1% agarose base in a 90 mm disposable petri dish. 750 ml of odor solution was evenly spread on the lid and it was placed for indicated times (1 sec or 10 sec) on dish. Larvae were gently removed using a paint brush and placed on a separate petridish and assayed for odor preference in the dark for 1 min 30 sec as described above. The mutant *Orco-/-* (previously called *Or83b*^{1}) (Larsson et al., 2004) was obtained from Bloomington stock center, and the *Or42b-Gal4* flies were a kind gift from Dr. John Carlson.

Assembling of EthR screening library

Three large libraries were assembled consisting of the advanced library from Enamine (204,772 compound), the screening library from Maybridge (56,000 compounds) and the merged library from Asinex (439,946 compounds). All chemical from each library were combined and duplicates or any compound that contained an atom other than C, O, H, N, S, Br, and F were removed. The resulting library contained 588,723 unique structures.

Calculating molecular descriptors for EthR analysis

Molecular descriptors were calculated independently using both the Dragon software (Talete) suite and Pipeline pilot (Accelrys). The 3D conformations for each compound were calculated using the Omega application (OpenEye). Only a single best conformation was saved for each compound resulting in 588,723 unique 3D structures.

These structures were then fed into Dragon, which calculated 3,224 unique molecular descriptor values for each of the compounds. These same structures were also fed into pipeline pilot for descriptor calculation (Balaban, Chi, and Kappa values) and machine learning.

Applying machine learning with Pipeline Pilot

The Pipeline pilot suite was used to compare three different machine learning approaches (Random Forests, Support Vector Machines, and linear discriminate analysis) for their ability to identify what shape characteristics separate compounds that bind into EthR from those that do not. As we only knew compounds that have been effectively docked into the binding site and machine learning techniques require both a positive and negative training set, we had to estimate a set of compounds that would not. We randomly selected 100 compounds from the Asinex library, which as EthR is selective in it's binding, are unlikely to bind into EthR by chance. Each of the three machine learning methods was then individually trained to separate out inhibitors from our hypothetical non-inhibitors based on molecular descriptor values. Each method had a built in built in validation approach, such as cross fold validation or out of bag, that was applied along with a Receiver Operating Characteristic (ROC) analysis to compare between methods.

Performing structure-based virtual screening with Gold

The Gold suite was used for all protein-ligand docking and Goldmine was applied for analysis of docking results. An individual docking run was performed for each of the three binding sites. Other than for the following exceptions, default settings were applied

for the gold algorithm. Protein H-bond constraints were individually applied for each binding site with the following values: Asparagine 179 and Asparagine 176 for binding site 1, Glutamine 125 for binding site 2, and Asparagine 179, Asparagine 176, Glutamine 125 for the combined site. We performed 10 independent docking iterations for both binding site 1 and 2 and 3 independent iterations for the combined binding site. Early termination was allowed if 3 independent iterations failed to improve the rms_tolerance value by better than a value of 1.5. The binding cavity for each of the sites was set to a 10 angstrom radius from the natural ligands bound into site 1 or 2 of structures NT06 and MB1047, respectively. The binding cavity was set to be a 10 angstrom radius from either of the natural bound ligands in both sites 1 and 2 in structure NT06 for the combined site, which created a single site spanning nearly the entire cavity. None of the amino acid side chains were allowed to rotate in any of the docking sets. Both goldscore and chemscore scoring functions were calculated for each binding site.

REFERNCES:

Abuin, L., Bargeton, B., Ulbrich, M.H., Isacoff, E.Y., Kellenberger, S., and Benton, R. (2011). Functional architecture of olfactory ionotropic glutamate receptors. Neuron *69*, 44-60.

Aceves-Pina, E.O., and Quinn, W.G. (1979). Learning in normal and mutant Drosophila larvae. Science *206*, 93-96.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.*, et al.* (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185-2195.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Aramaki, H., Yagi, N., and Suzuki, M. (1995). Residues important for the function of a multihelical DNA binding domain in the new transcription factor family of Cam and Tet repressors. Protein Eng *8*, 1259-1266.

Araneda, R.C., Kini, A.D., and Firestein, S. (2000). The molecular receptive range of an odorant receptor. Nat Neurosci *3*, 1248-1255.

Ayyub, C., Paranjape, J., Rodrigues, V., and Siddiqi, O. (1990). Genetics of Olfactory Behavior in Drosophila-Melanogaster. J Neurogenet *6*, 243-262.

Baldwin, E., Plotto, A., Manthey, J., McCollum, G., Bai, J.H., Irey, M., Cameron, R., and Luzio, G. (2010). Effect of Liberibacter Infection (Huanglongbing Disease) of Citrus on Orange Fruit Physiology and Fruit/Fruit Juice Quality: Chemical and Physical Analyses. J Agr Food Chem *58*, 1247-1262.

Baulard, A.R., Betts, J.C., Engohang-Ndong, J., Quan, S., McAdam, R.A., Brennan, P.J., Locht, C., and Besra, G.S. (2000). Activation of the pro-drug ethionamide is regulated in mycobaeteria. Journal of Biological Chemistry *275*, 28326-28331.

Belluscio, L., Gold, G.H., Nemes, A., and Axel, R. (1998). Mice deficient in G(olf) are anosmic. Neuron *20*, 69-81.

Benton, R., Sachse, S., Michnick, S.W., and Vosshall, L.B. (2006). Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. PLoS Biol *4*, e20.

Benton, R., Vannice, K.S., Gomez-Diaz, C., and Vosshall, L.B. (2009). Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in Drosophila. Cell *136*, 149-162.

Bohacek, R.S., McMartin, C., and Guida, W.C. (1996). The art and practice of structure-based drug design: A molecular modeling perspective. Med Res Rev *16*, 3-50.

Bohbot, J., Pitts, R.J., Kwon, H.W., Rutzler, M., Robertson, H.M., and Zwiebel, L.J. (2007). Molecular characterization of the Aedes aegypti odorant receptor gene family. Insect Mol Biol *16*, 525-537.

Bolton, E.E., Wang, Y., Thiessen, P.A., and S.H., B. (2008). PubChem: Integrated Platform of Small Molecules and Biological Activities In Annual Reports in Computational Chemistry (Washington DC, American Chemical Society).

Bostrom, J., Greenwood, J.R., and Gottfries, J. (2003). Assessing the performance of OMEGA with respect to retrieving bioactive conformations. J Mol Graph Model *21*, 449-462.

Bove, J.M. (2006). Huanglongbing: A destructive, newly-emerging, century-old disease of citrus. J Plant Pathol *88*, 7-37.

Brockhoff, A., Behrens, M., Niv, M.Y., and Meyerhof, W. (2010). Structural requirements of bitter taste receptor activation. P Natl Acad Sci USA *107*, 11110-11115.

Buck, L., and Axel, R. (1991). A Novel Multigene Family May Encode Odorant Receptors - a Molecular-Basis for Odor Recognition. Cell *65*, 175-187.

Butler, D. (2011). Mosquitoes score in chemical war. Nature *475*, 19.

Cao, Y., Charisi, A., Cheng, L.C., Jiang, T., and Girke, T. (2008a). ChemmineR: a compound mining framework for R. Bioinformatics *24*, 1733-1734.

Cao, Y., Jiang, T., and Girke, T. (2008b). A maximum common substructure-based algorithm for searching and predicting drug-like compounds. Bioinformatics *24*, i366-374.

Carde, R.T., and Gibson, G. (2010). Long Distance orientation of mosquitoes to host odours and their clues. in Ecology of Vector-Borne Diseases. Vol. 2. Olfaction in Vector-Host Interactions (Wageningen, The Netherlands, Wageningen Academic Publishers).

Carey, A.F., Wang, G.R., Su, C.Y., Zwiebel, L.J., and Carlson, J.R. (2010). Odorant reception in the malaria mosquito Anopheles gambiae. Nature *464*, 66-U77.

Carhart, R.E., Smith, D.H., and Venkataraghavan, R. (1985). Atom Pairs as Molecular-Features in Structure Activity Studies - Definition and Applications. J Chem Inf Comp Sci *25*, 64-73.

Chan, E.D., and Iseman, M.D. (2002). Current medical treatment for tuberculosis. Brit Med J *325*, 1282-1286.

Chang, C., and Lin, C. (2001). Libsvm: A Library for Support Vector Machines.

Chen, X., and Reynolds, C.H. (2002). Performance of similarity measures in 2D fragment-based similarity searching: comparison of structural descriptors and similarity coefficients. J Chem Inf Comput Sci *42*, 1407-1414.

Clyne, P.J., Warr, C.G., and Carlson, J.R. (2000). Candidate taste receptors in Drosophila. Science *287*, 1830-1834.

Clyne, P.J., Warr, C.G., Freeman, M.R., Lessing, D., Kim, J., and Carlson, J.R. (1999). A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in Drosophila. Neuron *22*, 327-338.

Corbel, V., Stankiewicz, M., Pennetier, C., Fournier, D., Stojan, J., Girard, E., Dimitrov, M., Molgo, J., Hougard, J.M., and Lapied, B. (2009). Evidence for inhibition of cholinesterases in insect and mammalian nervous systems by the insect repellent deet. Bmc Biol *7*, -.

Cork, A., and Park, K.C. (1996). Identification of electrophysiologically-active compounds for the malaria mosquito, Anopheles gambiae, in human sweat extracts. Medical and Veterinary Entomology *10*, 269-276.

Cortes, C., and Vapnik, V. (1995). Support-Vector Networks. Mach Learn *20*, 273-297.

Couto, A., Alenius, M., and Dickson, B.J. (2005). Molecular, anatomical, and functional organization of the Drosophila olfactory system. Curr Biol *15*, 1535-1547.

Curran, A.M., Rabin, S.I., Prada, P.A., and Furton, K.G. (2005). Comparison of the volatile organic compounds present in human odor using SPME-GC/MS. Journal of Chemical Ecology *31*, 1607-1619.

Dagraca, J.V. (1991). Citrus Greening Disease. Annu Rev Phytopathol *29*, 109-136.

Dahanukar, A., Foster, K., van Naters, W.M.V.D.G., and Carlson, J.R. (2001). A Gr receptor is required for response to the sugar trehalose in taste neurons of Drosophila. Nat Neurosci *4*, 1182-1186.

Dahanukar, A., Hallem, E.A., and Carlson, J.R. (2005). Insect chemoreception. Curr Opin Neurobiol *15*, 423-430.

de Bruyne, M., and Baker, T.C. (2008). Odor detection in insects: volatile codes. J Chem Ecol *34*, 882-897.

de Bruyne, M., Clyne, P.J., and Carlson, J.R. (1999). Odor coding in a model olfactory organ: the Drosophila maxillary palp. J Neurosci *19*, 4520-4532.

de Bruyne, M., Foster, K., and Carlson, J. (2001). Odor coding in the Drosophila antenna. Neuron *30*, 537-552.

DeBarber, A.E., Mdluli, K., Bosman, M., Bekker, L.G., and Barry, C.E. (2000). Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis. P Natl Acad Sci USA *97*, 9677-9682.

Ditzen, M., Pellegrino, M., and Vosshall, L.B. (2008). Insect odorant receptors are molecular targets of the insect repellent DEET. Science *319*, 1838-1842.

Dobritsa, A.A., van der Goes van Naters, W., Warr, C.G., Steinbrecht, R.A., and Carlson, J.R. (2003). Integrating the molecular and cellular basis of odor coding in the Drosophila antenna. Neuron *37*, 827-841.

Dobson, C.M. (2004). Chemical space and biology. Nature *432*, 824-828.

Dover, L.G., Corsino, P.E., Daniels, I.R., Cocklin, S.L., Tatituri, V., Besra, G.S., and Futterer, K. (2004). Crystal structure of the TetR/CamR family repressor Mycobacterium tuberculosis EthR implicated in ethionamide resistance. Journal of Molecular Biology *340*, 1095-1105.

EFSA (2008). Consideration of Anthranilate derivatives evaluated by JECFA (65th meeting) Opinion of the scientific Panel on Food Additives, Flavourings, Processing Aids and Materials in Contact wtih Food. The EFSA Journal, 1-24.

El-Sayed, A. (2009). The Pherobase: Database of Insect Pheromones and Semiochemicals.

Engohang-Ndong, J., Baillat, D., Aumercier, M., Bellefontaine, F., Besra, G.S., Locht, C., and Baulard, A.R. (2004). EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. Mol Microbiol *51*, 175-188.

Fishilevich, E., and Vosshall, L.B. (2005). Genetic and functional subdivision of the Drosophila antennal lobe. Curr Biol *15*, 1548-1553.

Flipo, M., Desroses, M., Lecat-Guillet, N., Dirie, B., Carette, X., Leroux, F., Piveteau, C., Demirkaya, F., Lens, Z., Rucktooa, P.*, et al.* (2011). Ethionamide Boosters: Synthesis, Biological Activity, and Structure-Activity Relationships of a Series of 1,2,4-Oxadiazole EthR Inhibitors. Journal of Medicinal Chemistry *54*, 2994-3010.

Flipo, M., Desroses, M., Lecat-Guillet, N., Villemagne, B., Blondiaux, N., Leroux, F., Piveteau, C., Mathys, V., Flament, M.P., Siepmann, J.*, et al.* (2012). Ethionamide Boosters. 2. Combining Bioisosteric Replacement and Structure-Based Drug Design To Solve Pharmacokinetic Issues in a Series of Potent 1,2,4-Oxadiazole EthR Inhibitors. Journal of Medicinal Chemistry *55*, 68-83.

Floriano, W.B., Vaidehi, N., Goddard, W.A., Singer, M.S., and Shepherd, G.M. (2000). Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. P Natl Acad Sci USA *97*, 10712-10716.

Fox, A.N., Pitts, R.J., Robertson, H.M., Carlson, J.R., and Zwiebel, L.J. (2001). Candidate odorant receptors from the malaria vector mosquito Anopheles gambiae and evidence of down-regulation in response to blood feeding. P Natl Acad Sci USA *98*, 14693-14697.

Frenois, F., Engohang-Ndong, J., Locht, C., Baulard, A.R., and Villeret, V. (2004). Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis. Mol Cell *16*, 301-307.

Galizia, C.G., Munch, D., Strauch, M., Nissler, A., and Ma, S.W. (2010). Integrating Heterogeneous Odor Response Data into a Common Response Model: A DoOR to the Complete Olfactome. Chemical Senses *35*, 551-563.

Gallagher, M., Wysocki, J., Leyden, J.J., Spielman, A.I., Sun, X., and Preti, G. (2008). Analyses of volatile organic compounds from human skin. Brit J Dermatol *159*, 780-791.

Ghaninia, M., Ignell, R., and Hansson, B.S. (2007). Functional classification and central nervous projections of olfactory receptor neurons housed in antennal trichoid sensilla of female yellow fever mosquitoes, Aedes aegypti. Eur J Neurosci *26*, 1611-1623.

Ghose, A.K., Viswanadhan, V.N., and Wendoloski, J.J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. J Comb Chem *1*, 55-68.

Gillies, M.T. (1980). The Role of Carbon-Dioxide in Host-Finding by Mosquitos (Diptera, Culicidae) - a Review. B Entomol Res *70*, 525-532.

Goldman, A.L., van Naters, W.V., Lessing, D., Warr, C.G., and Carlson, J.R. (2005). Coexpression of two functional odor receptors in one neuron. Neuron *45*, 661-666.

Goto, S., Okuno, Y., Hattori, M., Nishioka, T., and Kanehisa, M. (2002). LIGAND: database of chemical compounds and reactions in biological pathways. Nucleic Acids Research *30*, 402-404.

Grant, A.J., and Oconnell, R.J. (1996). Electrophysiological responses from receptor neurons in mosquito maxillary palp sensilla. Ciba F Symp *200*, 233-253.

Grosse-Wilde, E., Kuebler, L.S., Bucks, S., Vogel, H., Wicher, D., and Hansson, B.S. (2011). Antennal transcriptome of Manduca sexta. P Natl Acad Sci USA *108*, 7449-7454.

Guo, S., and Kim, J. (2007). Molecular evolution of Drosophila odorant receptor genes. Mol Biol Evol *24*, 1198-1207.

Guo, S., and Kim, J. (2010). Dissecting the molecular mechanism of drosophila odorant receptors through activity modeling and comparative analysis. Proteins *78*, 381-399.

Gupta, R.K.a.B., A.K. (2007). Discovery and Design of New Arthropod/Insect Repellents by Computer-Aided Molecular Modeling. In Insect Repellents: principles, methods, and uses, M. Debboun, Frances, S.P., Strickman, D., ed. (Boca Raton, Taylor & Francis Group), pp. 195-228.

Haddad, R., Khan, R., Takahashi, Y.K., Mori, K., Harel, D., and Sobel, N. (2008). A metric for odorant comparison. Nat Methods *5*, 425-429.

Halbert, S.E., and Manjunath, K.L. (2004). Asian citrus psyllids (Sternorrhyncha : Psyllidae) and greening disease of citrus: A literature review and assessment of risk in Florida. Fla Entomol *87*, 330-353.

Hallem, E.A., and Carlson, J.R. (2006). Coding of odors by a receptor repertoire. Cell *125*, 143-160.

Hallem, E.A., Ho, M.G., and Carlson, J.R. (2004). The molecular basis of odor coding in the Drosophila antenna. Cell *117*, 965-979.

Hastie, T., Tibshirani, R., and Friedman, J.H. (2001). The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations (New York, Springer).

Hawkins, P.C.D., Skillman, A.G., Warren, G.L., Ellingson, B.A., and Stahl, M.T. (2010). Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. Journal of Chemical Information and Modeling *50*, 572-584.

Hendrickson, J.B. (1991). Concepts and Applications of Molecular Similarity - Johnson,Ma, Maggiora,Gm. Science *252*, 1189-1189.

Hill, S.R., Hansson, B.S., and Ignell, R. (2009). Characterization of antennal trichoid sensilla from female southern house mosquito, Culex quinquefasciatus Say. Chem Senses *34*, 231-252.

Ishimoto, H., Takahashi, K., Ueda, R., and Tanimura, T. (2005). G-protein gamma subunit 1 is required for sugar reception in Drosophila. Embo J *24*, 3259-3265.

JECF (2007). Safety evaluation of ertain food additives and contaminants. Sixty-fifth meeting of the joint FAO/WHO Expert Committee of Food Additives. WHO Food Additives Series: 56 IPCS, WHO Geneva.

Jefferis, G.S., Potter, C.J., Chan, A.M., Marin, E.C., Rohlfing, T., Maurer, C.R., Jr., and Luo, L. (2007). Comprehensive maps of Drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. Cell *128*, 1187-1203.

Jones, W.D., Cayirlioglu, P., Kadow, I.G., and Vosshall, L.B. (2007). Two chemosensory receptors together mediate carbon dioxide detection in Drosophila. Nature *445*, 86-90.

Kain, P., Chakraborty, T.S., Sundaram, S., Siddiqi, O., Rodrigues, V., and Hasan, G. (2008). Reduced odor responses from antennal neurons of G(q)alpha, phospholipase C beta, and rdgA mutants in Drosophila support a role for a phospholipid intermediate in insect olfactory transduction. Journal of Neuroscience *28*, 4745-4755.

Kaluza, J.F., and Breer, H. (2000). Responsiveness of olfactory neurons to distinct aliphatic aldehydes. J Exp Biol *203*, 927-933.

Karatzoglou, A., Meyer, D., and Hornik, K. (2006). Support Vector Machines in R. J Stat Softw *15*.

Katada, S., Hirokawa, T., Oka, Y., Suwa, M., and Touhara, K. (2005). Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: Mapping the odorant-binding site. Journal of Neuroscience *25*, 1806-1815.

Katritzky, A.R., Wang, Z., Slavov, S., Tsikolia, M., Dobchev, D., Akhmedov, N.G., Hall, C.D., Bernier, U.R., Clark, G.G., and Linthicum, K.J. (2008). Synthesis and bioassay of improved mosquito repellents predicted from chemical structure. Proc Natl Acad Sci U S A *105*, 7359-7364.

Keiser, M.J., Setola, V., Irwin, J.J., Laggner, C., Abbas, A.I., Hufeisen, S.J., Jensen, N.H., Kuijer, M.B., Matos, R.C., Tran, T.B.*, et al.* (2009). Predicting new molecular targets for known drugs. Nature *462*, 175-U148.

Khafizov, K., Anselmi, C., Menini, A., and Carloni, P. (2007). Ligand specificity of odorant receptors. J Mol Model *13*, 401-409.

Kim, J., and Carlson, J.R. (2002). Gene discovery by e-genetics: Drosophila odor and taste receptors. J Cell Sci *115*, 1107-1112.

Kitchen, D.B., Decornez, H., Furr, J.R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. Nat Rev Drug Discov *3*, 935-949.

Kline, D.L., Bernier, U.R., Posey, K.H., and Barnard, D.R. (2003). Olfactometric evaluation of spatial repellents for Aedes aegypti. Journal of Medical Entomology *40*, 463-467.

Klocke, J.A., Darlington, M.V., and Balandrin, M.F. (1987). Biologically-Active Constituents of North-American Plants .3. 1,8-Cineole (Eucalyptol), a Mosquito Feeding and Ovipositional Repellent from Volatile Oil of Hemizonia-Fitchii (Asteraceae). J Chem Ecol *13*, 2131-2141.

Klun, J.A., Strickman, D., Rowton, E., Williams, J., Kramer, M., Roberts, D., and Debboun, M. (2004). Comparative resistance of Anopheles albimanus and Aedes aegypti to N,N-diethyl-3-methylbenzamide (Deet) and 2-methylpiperidinyl-3-cyclohexen-1-carboxamide (AI3-37220) in laboratory human-volunteer repellent assays. J Med Entomol *41*, 418-422.

Knudsen, J.T., Eriksson, R., Gershenzon, J., and Stahl, B. (2006). Diversity and Distribution of Floral Scent. The Botanical Reveiw *72*, 1-120.

Koon, A.C., Ashley, J., Barria, R., DasGupta, S., Brain, R., Waddell, S., Alkema, M.J., and Budnik, V. (2011). Autoregulatory and paracrine control of synaptic and behavioral plasticity by octopaminergic signaling. Nat Neurosci *14*, 190-U275.

Krajick, K. (2006). Medical entomology - Keeping the bugs at bay. Science *313*, 36-38.

Kreher, S.A., Kwon, J.Y., and Carlson, J.R. (2005). The molecular basis of odor coding in the Drosophila larva. Neuron *46*, 445-456.

Kreher, S.A., Mathew, D., Kim, J., and Carlson, J.R. (2008). Translation of sensory input into behavioral output via an olfactory system. Neuron *59*, 110-124.

Kurland, M.D., Newcomer, M.B., Peterlin, Z., Ryan, K., Firestein, S., and Batista, V.S. (2010). Discrimination of Saturated Aldehydes by the Rat I7 Olfactory Receptor. Biochemistry *49*, 6302-6304.

Kwon, J.Y., Dahanukar, A., Weiss, L.A., and Carlson, J.R. (2007). The molecular basis of CO2 reception in Drosophila. P Natl Acad Sci USA *104*, 3574-3578.

Lai, P.C., Bahl, G., Gremigni, M., Matarazzo, V., Clot-Faybesse, O., Ronin, C., and Crasto, C.J. (2008). An olfactory receptor pseudogene whose function emerged in humans: a case study in the evolution of structure-function in GPCRs. J Struct Funct Genomics *9*, 29-40.

Larsson, M.C., Domingos, A.I., Jones, W.D., Chiappe, M.E., Amrein, H., and Vosshall, L.B. (2004). Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. Neuron *43*, 703-714.

Lee, Y., Kim, S.H., and Montell, C. (2010). Avoiding DEET through Insect Gustatory Receptors. Neuron *67*, 555-561.

Liu, C., Pitts, R.J., Bohbot, J.D., Jones, P.L., Wang, G., and Zwiebel, L.J. (2010). Distinct olfactory signaling mechanisms in the malaria vector mosquito Anopheles gambiae. PLoS Biol *8*.

Logan, J.G., Birkett, M.A., Clark, S.J., Powers, S., Seal, N.J., Wadhams, L.J., Mordue, A.J., and Pickett, J.A. (2008). Identification of human-derived volatile chemicals that interfere with attraction of Aedes aegypti mosquitoes. Journal of Chemical Ecology *34*, 308-322.

Louis, M., Huber, T., Benton, R., Sakmar, T.P., and Vosshall, L.B. (2008). Bilateral olfactory sensory input enhances chemotaxis behavior. Nat Neurosci *11*, 187-199.

Lu, T., Qiu, Y.T., Wang, G., Kwon, J.Y., Rutzler, M., Kwon, H.W., Pitts, R.J., van Loon, J.J.A., Takken, W., Carlson, J.R.*, et al.* (2007). Odor coding in the maxillary palp of the malaria vector mosquito Anopheles gambiae. Curr Biol *17*, 1533-1544.

Mahmoudi, A., and Iseman, M.D. (1993). Pitfalls in the Care of Patients with Tuberculosis - Common Errors and Their Association with the Acquisition of Drug-Resistance. Jama-J Am Med Assoc *270*, 65-68.

Maldonado, A.G., Doucet, J.P., Petitjean, M., and Fan, B.T. (2006). Molecular similarity and diversity in chemoinformatics: From theory to applications. Mol Divers *10*, 39-79.

Martin, Y.C., Kofron, J.L., and Traphagen, L.M. (2002). Do structurally similar molecules have similar biological activity? Journal of Medicinal Chemistry *45*, 4350-4358.

Meijerink, J., Braks, M.A.H., Brack, A.A., Adam, W., Dekker, T., Posthumus, M.A., Van Beek, T.A., and Van Loon, J.J.A. (2000). Identification of olfactory stimulants for Anopheles gambiae from human sweat samples. Journal of Chemical Ecology *26*, 1367-1382.

Montague, S.A., Mathew, D., and Carlson, J.R. (2011). Similar Odorants Elicit Different Behavioral and Physiological Responses, Some Supersustained. Journal of Neuroscience *31*, 7891-7899.

Monte, P., Woodard, C., Ayer, R., Lilly, M., Sun, H., and Carlson, J. (1989). Characterization of the Larval Olfactory Response in Drosophila and Its Genetic-Basis. Behav Genet *19*, 267-283.

Nikolova, N., and Jaworska, J. (2004). Approaches to measure chemical similarity - A review. Qsar Comb Sci *22*, 1006-1026.

Olsen, S.R., Bhandawat, V., and Wilson, R.I. (2007). Excitatory interactions between olfactory processing channels in the Drosophila antennal lobe. Neuron *54*, 89-103.

Olsen, S.R., and Wilson, R.I. (2008). Lateral presynaptic inhibition mediates gain control in an olfactory circuit. Nature *452*, 956-U953.

Onagbola, E.O., Rouseff, R.L., Smoot, J.M., and Stelinski, L.L. (2011). Guava leaf volatiles and dimethyl disulphide inhibit response of Diaphorina citri Kuwayama to host plant volatiles. J Appl Entomol *135*, 404-414.

Pellegrino, M., Steinbach, N., Stensmyr, M.C., Hansson, B.S., and Vosshall, L.B. (2011). A natural polymorphism alters odour and DEET sensitivity in an insect odorant receptor. Nature.

Pelz, D., Roeske, T., Syed, Z., de Bruyne, M., and Galizia, C.G. (2006). The molecular receptive range of an olfactory receptor in vivo (Drosophila melanogaster Or22a). J Neurobiol *66*, 1544-1563.

Pitts, R.J., Rinker, D.C., Jones, P.L., Rokas, A., and Zwiebel, L.J. (2011). Transcriptome profiling of chemosensory appendages in the malaria vector Anopheles gambiae reveals tissue- and sex-specific signatures of odor coding. Bmc Genomics *12*.

Raviglione, M.C. (2003). The TB epidemic from 1992 to 2002. Tuberculosis *83*, 4-14.

Reeder, N.L., Ganz, P.J., Carlson, J.R., and Saunders, C.W. (2001). Isolation of a DEET-insensitive mutant of Drosophila melanogaster (Diptera: Drosophilidae). J Econ Entomol *94*, 1584-1588.

Robertson, H.M., and Kent, L.B. (2009). Evolution of the gene lineage encoding the carbon dioxide receptor in insects. Journal of Insect Science *9*.

Robertson, H.M., Warr, C.G., and Carlson, J.R. (2003). Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. Proc Natl Acad Sci U S A *100 Suppl 2*, 14537-14542.

Rodriguez-Gil, D.J., Treloar, H.B., Zhang, X.H., Miller, A.M., Two, A., Iwema, C., Firestein, S.J., and Greer, C.A. (2010). Chromosomal Location-Dependent Nonstochastic Onset of Odor Receptor Expression. Journal of Neuroscience *30*, 10067-10075.

Root, C.M., Ko, K.I., Jafari, A., and Wang, J.W. (2011). Presynaptic Facilitation by Neuropeptide Signaling Mediates Odor-Driven Food Search. Cell *145*, 133-144.

Root, C.M., Semmelhack, J.L., Wong, A.M., Flores, J., and Wang, J.W. (2007). Propagation of olfactory information in Drosophila. P Natl Acad Sci USA *104*, 11826-11831.

Ruta, V., Datta, S.R., Vasconcelos, M.L., Freeland, J., Looger, L.L., and Axel, R. (2010). A dimorphic pheromone circuit in Drosophila from sensory input to descending output. Nature *468*, 686-U106.

Saito, H., Chi, Q., Zhuang, H., Matsunami, H., and Mainland, J.D. (2009). Odor coding by a Mammalian receptor repertoire. Sci Signal *2*, ra9.

Saito, H., Kubota, M., Roberts, R.W., Chi, Q., and Matsunami, H. (2004). RTP family members induce functional expression of mammalian odorant receptors. Cell *119*, 679-691.

Sanger, F., and Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol *94*, 441-448.

Sato, K., Pellegrino, M., Nakagawa, T., Vosshall, L.B., and Touhara, K. (2008). Insect olfactory receptors are heteromeric ligand-gated ion channels. Nature *452*, 1002-1006.

Sato, K., Tanaka, K., and Touhara, K. (2011). Sugar-regulated cation channel formed by an insect gustatory receptor. P Natl Acad Sci USA *108*, 11680-11685.

Schmuker, M., de Bruyne, M., Hahnel, M., and Schneider, G. (2007). Predicting olfactory receptor neuron responses from odorant structure. Chem Cent J *1*, -.

Schneider, G., Schneider, P., and Renner, S. (2006). Scaffold-hopping: How far can you jump? Qsar Comb Sci *25*, 1162-1171.

Schwarz, D., Robertson, H.M., Feder, J.L., Varala, K., Hudson, M.E., Ragland, G.J., Hahn, D.A., and Berlocher, S.H. (2009). Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot Rhagoletis pomonella. Bmc Genomics *10*.

Semmelhack, J.L., and Wang, J.W. (2009). Select Drosophila glomeruli mediate innate olfactory attraction and aversion. Nature *459*, 218-U100.

Shang, Y.H., Claridge-Chang, A., Sjulson, L., Pypaert, M., and Miesenbock, G. (2007). Excitatory local circuits and their implications for olfactory processing in the fly antennal lobe. Cell *128*, 601-612.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

Sigma-Aldrich (2007). Flavors and fragarances 2007-2008 catalog. Sigma-Aldrich Fine Chemicals Company, Milquakee, WI.

Singer, M.S. (2000). Analysis of the molecular basis for octanal interactions in the expressed rat I7 olfactory receptor. Chemical Senses *25*, 155-165.

Singer, M.S., and Shepherd, G.M. (1994). Molecular modeling of ligand-receptor interactions in the OR5 olfactory receptor. Neuroreport *5*, 1297-1300.

Smadja, C., Shi, P., Butlin, R.K., and Robertson, H.M. (2009). Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, Acyrthosiphon pisum. Mol Biol Evol *26*, 2073-2086.

Stanczyk, N.M., Brookfield, J.F., Ignell, R., Logan, J.G., and Field, L.M. (2010). Behavioral insensitivity to DEET in Aedes aegypti is a genetically determined trait residing in changes in sensillum function. P Natl Acad Sci USA *107*, 8575-8580.

Stensmyr, M.C., Giordano, E., Balloi, A., Angioy, A.M., and Hansson, B.S. (2003). Novel natural ligands for Drosophila olfactory receptor neurones. The Journal of experimental biology *206*, 715-724.

Syed, Z., and Leal, W.S. (2008). Mosquitoes smell and avoid the insect repellent DEET. P Natl Acad Sci USA *105*, 13598-13603.

Syed, Z., Pelletier, J., Flounders, E., Chitolina, R.F., and Leal, W.S. (2011). Generic insect repellent detector from the fruit fly Drosophila melanogaster. PLoS ONE *6*, e17705.

Takken, W. (1996). Synthesis and future challenges: The response of mosquitoes to host odours. Ciba F Symp *200*, 302-320.

Tan, P.-N., Steinbach, M., and Kumar, V. (2006). Introduction to data mining, 1st edn (Boston, Pearson Addison Wesley).

Tentschert, J., Bestmann, H.J., Holldobler, B., and Heinze, J. (2000). 2,3-dimethyl-5-(2-methylpropyl)pyrazine, a trail pheromone component of Eutetramorium mocquerysi Emery (1899) (Hymenoptera: Formicidae). Naturwissenschaften *87*, 377-380.

Triballeau, N., Van Name, E., Laslier, G., Cai, D., Paillard, G., Sorensen, P.W., Hoffmann, R., Bertrand, H.O., Ngai, J., and Acher, F.C. (2008). High-Potency Olfactory Receptor Agonists Discovered by Virtual High-Throughput Screening: Molecular Probes for Receptor Structure and Olfactory Function. Neuron *60*, 767-774.

Turner, S.L., Li, N., Guda, T., Githure, J., Carde, R.T., and Ray, A. (2011). Ultra-prolonged activation of CO2-sensing neurons disorients mosquitoes. Nature *474*, 87-91.

Turner, S.L., and Ray, A. (2009). Modification of CO2 avoidance behaviour in Drosophila by inhibitory odorants. Nature *461*, 277-281.

van der Goes van Naters, W., and Carlson, J.R. (2006). Insects as chemosensors of humans and crops. Nature *444*, 302-307.

van Naters, W.V.G., and Carlson, J.R. (2007). Receptors and neurons for fly odors in Drosophila. Curr Biol *17*, 606-612.

Vannelli, T.A., Dykman, A., and de Montellano, P.R.O. (2002). The antituberculosis drug ethionamide is activated by a flavoprotein monooxygenase. Journal of Biological Chemistry *277*, 12824-12829.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., and Taylor, R.D. (2003). Improved protein-ligand docking using GOLD. Proteins-Structure Function and Genetics *52*, 609-623.

Vosshall, L.B., Amrein, H., Morozov, P.S., Rzhetsky, A., and Axel, R. (1999). A spatial map of olfactory receptor expression in the Drosophila antenna. Cell *96*, 725-736.

Vosshall, L.B., and Hansson, B.S. (2011). A unified nomenclature system for the insect olfactory coreceptor. Chemical senses *36*, 497-498.

Walker, J.D., Rodford, R., and Patlewicz, G. (2003). Quantitative structure-activity relationships for predicting percutaneous absorption rates. Environ Toxicol Chem *22*, 1870-1884.

Wang, J.W., Wong, A.M., Flores, J., Vosshall, L.B., and Axel, R. (2003). Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. Cell *112*, 271-282.

Wang, P., Lyman, R.F., Shabalina, S.A., Mackay, T.F.C., and Anholt, R.R.H. (2007). Association of polymorphisms in odorant-binding protein genes with variation in olfactory response to benzaldehyde in Drosophila. Genetics *177*, 1655-1665.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.

Weiss, L.A., Dahanukar, A., Kwon, J.Y., Banerjee, D., and Carlson, J.R. (2011). The molecular and cellular basis of bitter taste in Drosophila. Neuron *69*, 258-272.

Whitney, A.W. (1971). Direct Method of Nonparametric Measurement Selection. Ieee T Comput *C 20*, 1100-&.

Wicher, D., Schafer, R., Bauernfeind, R., Stensmyr, M.C., Heller, R., Heinemann, S.H., and Hansson, B.S. (2008). Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. Nature *452*, 1007-1011.

Willand, N., Desroses, M., Toto, P., Dirie, B., Lens, Z., Villeret, V., Rucktooa, P., Locht, C., Baulard, A., and Deprez, B. (2010). Exploring Drug Target Flexibility Using in Situ Click Chemistry: Application to a Mycobacterial Transcriptional Regulator. Acs Chem Biol *5*, 1007-1013.

Willand, N., Dirie, B., Carette, X., Bifani, P., Singhal, A., Desroses, M., Leroux, F., Willery, E., Mathys, V., Deprez-Poulain, R.*, et al.* (2009). Synthetic EthR inhibitors boost antituberculous activity of ethionamide. Nat Med *15*, 537-544.

Wilson, R.I., and Laurent, G. (2005). Role of GABAergic inhibition in shaping odor-evoked spatiotemporal patterns in the Drosophila antennal lobe. Journal of Neuroscience *25*, 9069-9079.

Wilson, R.I., Turner, G.C., and Laurent, G. (2004). Transformation of olfactory representations in the Drosophila antennal lobe. Science *303*, 366-370.

Wishart, D.S., Knox, C., Guo, A.C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008). DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Research *36*, D901-D906.

World Health Organization. (2011). World malaria report (Geneva, Switzerland, World Health Organization), pp. v.

Xia, Y., Wang, G., Buscariollo, D., Pitts, R.J., Wenger, H., and Zwiebel, L.J. (2008). The molecular and cellular basis of olfactory-driven behavior in Anopheles gambiae larvae. Proc Natl Acad Sci U S A *105*, 6433-6438.

Yao, C.A., and Carlson, J.R. (2010). Role of G-proteins in odor-sensing and CO2-sensing neurons in Drosophila. J Neurosci *30*, 4562-4572.

Yao, C.A., Ignell, R., and Carlson, J.R. (2005). Chemosensory coding by neurons in the coeloconic sensilla of the Drosophila antenna. Journal of Neuroscience *25*, 8359-8367.

Zeng, X.N., Leyden, J.J., Lawley, H.J., Sawano, K., Nohara, I., and Preti, G. (1991). Analysis of Characteristic Odors from Human Male Axillae. Journal of Chemical Ecology *17*, 1469-1492.

Zeng, X.N., Leyden, J.J., Spielman, A.I., and Preti, G. (1996). Analysis of characteristic human female axillary odors: Qualitative comparison to males. Journal of Chemical Ecology *22*, 237-257.

Zhan, S., Merlin, C., Boore, J.L., and Reppert, S.M. (2011). The Monarch Butterfly Genome Yields Insights into Long-Distance Migration. Cell *147*, 1171-1185.

Zhang, X., and Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. Nat Neurosci *5*, 124-133.

Zwiebel, L.J., and Takken, W. (2004). Olfactory regulation of mosquito-host interactions. Insect Biochem Molec *34*, 645-652.