Title
Protein inhibition by targeted small molecule libraries

Author
Lu, Felice

Publication Date
2005

Peer reviewed|Thesis/dissertation

Protein Inhibition by Targeted Small Molecule Libraries

by

Felice Lu

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Preface

First and foremost, I would like to thank my thesis advisors Tack Kuntz and Kip Guy for their excellent mentorship.

I would also like to thank Robert Fletterick and Kevan Shokat for serving on my thesis committee and contributing helpful suggestions along the way.

Chapter 5 in this thesis was originally published in Chemistry and Biology, 2002. Large portions of Chapters 1 - 4 are being submitted for publication.

Abstract

The p53-MDM2 interaction serves to regulate cellular responses to DNA damage and the over expression of MDM2 is the cause of 7% of all cancers. We report a method of structure-based computational design that produces chemical libraries centered on a scaffold that projects side chain functionalities with distance and angular relationships equivalent to those seen in p53. One library of 173 compounds was synthesized using solution phase parallel chemistry. The *in vitro* competitive ability of the compounds to block p53 peptide binding to MDM2 was determined using a fluorescence polarization competition assay. The most active compound bound with K_d = 12μM and its binding was characterized by ^{1}H-^{15}N HSQC NMR.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Complex signal transduction networks involve series of protein-protein interactions, regulated by co-localization, co-expression, and control of the local physiochemical environment. These dynamic association events can be misregulated, leading to altered signaling responses often characteristic of cells in diseased states. The ability to manipulate these interactions holds the promise of a cure for diseases stemming from protein interactions gone awry. Antibodies, proteins, and peptides that inhibit protein-protein complexes exist, but many of these agents are plagued with poor bioavailability. Thus, we are faced with the challenge of identifying low-molecular weight ligands that disrupt protein-protein complexes.

In principle, protein binding interactions could be reproduced by small molecules in cases where small regions of a protein's binding surface account for the majority of the binding energy[1]. Efforts to mimic features of short peptides in extended or β-turn conformations have been quite successful[2]. Short peptides in extended conformations are recognized and processed by proteolytic enzymes, an important class of therapeutic targets. Mimicry of short extended peptides has mainly involved the use of molecular constraints to stabilize bioactive conformations. Numerous examples of these types of compounds have been developed as protease inhibitors, typically comprising of cyclic peptide-like macrocycles.[3] The strategies for the mimicry of β-turn conformations also involve the cyclization of naturally occurring bioactive peptides.[4] In addition, several monocyclic and polycyclic small molecule templates have been found to mimic various types of β-turns. Wong and coworkers provide many examples of strand and turn mimics in their review.[2]

Protein binding surfaces often do not comprise short linear sequences. The α-helix is a common structural element in proteins and is a key recognition element for receptor binding in many protein-protein interactions.[2] In the field of α-helix mimicry, helix stability is an important factor to consider. Because of the entropic penalty associated with conformational restraints of helix formation, most short peptides corresponding to helical regions of proteins are unordered in aqueous solution.[5,6] Helices can be stabilized by factors such as pH, ionic strength, temperature, and peptide composition. Some amino acids are considered to be helix breakers, while others, such as α-aminoisobutyric acid (Aib), can stabilize helical peptide conformations.[7] Helical turns can also be stabilized covalently by amide bonds between, for example, the *i* and *i* + 4 positions to form macrocyclic lactams.[8] The use of these various strategies have yielded increasing numbers of peptide-like α-helix mimics.

There are only a handful of examples in which non-peptidic small molecules mimic larger, discontinuous areas of protein surface such as one or more turns of an alpha helix.[9-11] Much of the progress towards alpha helix mimicry was reported by Hamilton and coworkers, who developed terphenyl derivatives as inhibitors of the interaction between CaM and smMLCK, Bak and Bcl-X_L, gp41 helical tertiary structures, and p53 and MDM2.[12-16] These inhibitors were developed as *i*, *i* + 3, *i* + 7 and *i*, *i* + 4, *i* + 7 mimics using a pharmacophore-based medicinal chemistry approach. Advanced screening technology has also enabled high throughput screens for inhibitors of all types, including those of protein-protein interactions. However, there is no general and reproducible method for the identification of α-helix mimics. As part of our interest in

helix surface recognition, we have chosen to use this outstanding problem as a test bed for the development of methods for scaffold and library design.

Although progress has been made in the design of enzyme inhibitors, it remains difficult to accurately predict the binding of small compounds to non-enzyme proteins, or to sites other than the active site of enzymes.[17] Some of the obstacles presented by protein-protein targets include flat nondescript binding interfaces and large bioactive surface areas. However, the binding energy is not usually distributed evenly over the large surface, leading to hot spots of binding composed of several residues in the protein interface.[18] Hot spots are enriched in tryptophan, tyrosine, and arginine, and may be surrounded by less important residues that likely serve to occlude solvent. Because hot spots serve as the optimal site for small molecule inhibition, we have chosen to target the p53-MDM2 complex, whose interface provides such a hot spot.[19]

The p53 gene encodes a tumor suppressor protein whose normal function is to induce apoptosis and cell-cycle arrest in mammalian cells in response to cellular stress such as DNA damage, hypoxia, certain cytokines, or other stimuli. [20] In the absence of such stressors, the activity of p53 is normally suppressed by the oncoprotein encoded by the mdm2 gene. MDM2 binds to the amino terminus of p53 and both inhibits the ability of p53 to activate downstream effectors and directly targets p53 for proteolytic degradation. Both inactivation of the p53 protein and the over-expression of MDM2 have been associated with increased tumor incidence in human patients. In particular, MDM2 is overexpressed in 20% of soft tissue tumors, 16% of osteosarcomas, 13% of esophageal carcinomas, and 8% of astrocytomas.[21] The roles of p53 and MDM2 in cellular signaling and cancer biology will be discussed further in Chapter 4.

The development of inhibitors of the p53-MDM2 interaction was the main focus of my thesis work. Unbeknownst to me at the time, this project began in the fall of 1999 with a short rotation in the Guy lab, in which I attempted to express and purify MDM2 without success. The overarching research theme in the lab was α-helix mimicry by peptides and small molecules. The largest project involved the pursuit of lactam-bridge peptides and small organic molecule inhibitors of thyroid hormone receptor. While Tim Geistlinger worked on the peptidic inhibitors, Jim Arnold and Tom Robertson worked on the *in silico* modeling and synthesis of small molecule organic inhibitors. Although I was interested in a synthetic chemistry project, this led me to a rotation in the Kuntz lab working with Jim Arnold on computational design.

Jim Arnold had used several modeling programs to design inhibitors of thyroid hormone receptor. He had proposed two inhibitors that were modified for synthesis reasons and were being synthesized in the Guy lab at that time. I used a similar approach in designing inhibitors of MDM2, with several modifications. First, I incorporated several filters aimed at increasing ease of synthesis. Second, I expanded the molecules screened by using more databases and increased conformational sampling within each database. Third, I increased the use of DOCKing to remove poorly scoring candidates early. Finally, I changed the descriptors used for the clustering of similar hits. The computational design of MDM2 inhibitors is reported in Chapter 2, along with an introduction covering the p53-MDM2 structure and known MDM2 inhibitors to date.

After the computational modeling, the decision to synthesize the compounds was relatively easy, due to my interest in synthetic chemistry. The proposed inhibitors were modular and structurally simple, unlike the proposed inhibitors for thyroid hormone

receptor. Although the compounds had not been previously reported in literature, we thought the synthesis would be straightforward. Chapter 3 describes the synthesis of the MDM2 inhibitor library. Initial proofing reactions were completed quickly, with the exception of a pinacolboronate intermediate that was difficult to synthesize. A reasonable amount of effort was also dedicated toward optimization of the two final and most frequently used transformations. After optimization, library synthesis proceeded smoothly, with the majority of the time spent on purification and quality control.

The evaluation and characterization of the library for biological activity is detailed in Chapter 4. A competitive fluorescence anisotropy assay was first used to test for MDM2 binding. This assay was chosen because it had been reported in the literature, the Guy lab had expertise in this type of binding assay, and the assay could be used to test all the compounds quickly using small amounts of MDM2 protein. The assay results showed that about 20 compounds of the 170 tested bound to MDM2. These compounds were then tested for cellular activity, but the results were negative. The binding of the best inhibitor, with K_i of 12 μM, was characterized by ^{1}H-^{15}N HSQC NMR by our collaborators in the lab of Kyou-Hoon Han. This confirmed the interaction of the inhibitor with MDM2 and also shows the MDM2 residues affected by inhibitor binding. While this information can be used to infer possible modes of binding, there is no proof that the actual binding mode is similar to the mode predicted computationally.

During my research in the Kuntz lab, I also worked on a collaborative project between the Bogyo and Kuntz groups investigating the classification of proteins into functionally distinct families based only on primary sequence information. Chapter 5 describes the production of a large data set of small molecule affinity fingerprints for a

group of closely related enzymes, the papain family of cysteine proteases. Experimental binding data was generated for a library of inhibitors based on the ability of each compound to block active site labeling of the target proteases by a covalent activity based probe (ABP). Experimental data were used to guide the development of a computational method that predicts small molecule inhibitors based on reported crystal structures. Clustering algorithms were used to classify proteases into subfamilies based on their experimental and computational small molecule affinity fingerprints.

References

(1) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 9997-10002.

(2) Fairlie, D. P.; West, M. L.; Wong, A. K. Towards protein surface mimetics. *Current Medicinal Chemistry* **1998**, *5*, 29-62.

(3) Leung, D.; Abbenante, G.; Fairlie, D. P. Protease inhibitors: Current status and future prospects. *Journal of Medicinal Chemistry* **2000**, *43*, 305-341.

(4) Fairlie, D. P.; Abbenante, G.; March, D. R. Macrocyclic Peptidomimetics - Forcing Peptides into Bioactive Conformations. *Current Medicinal Chemistry* **1995**, *2*, 654-686.

(5) Zimm, B. H.; Bragg, J. K. Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains. *The Journal of Chemical Physics* **1959**, *31*, 526-535.

(6) Scholtz, J. M.; Baldwin, R. L. The Mechanism of Alpha-Helix Formation by Peptides. *Annual Review of Biophysics and Biomolecular Structure* **1992**, *21*, 95-118.

(7) Marshall, G. R.; Hodgkin, E. E.; Langs, D. A.; Smith, G. D.; Zabrocki, J. et al. Factors Governing Helical Preference of Peptides Containing Multiple Alpha,Alpha-Dialkyl Amino-Acids. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87*, 487-491.

(8) Bracken, C.; Gulyas, J.; Taylor, J. W.; Baum, J. Synthesis and Nuclear-Magnetic-Resonance Structure Determination of an Alpha-Helical, Bicyclic, Lactam-Bridged Hexapeptide. *Journal of the American Chemical Society* **1994**, *116*, 6431-6432.

(9) Pagliaro, L.; Felding, J.; Audouze, K.; Nielsen, S. J.; Terry, R. B. et al. Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Current Opinion in Chemical Biology* **2004**, *8*, 442-449.

(10) Toogood, P. L. Inhibition of protein-protein association by small molecules: Approaches and progress. *Journal of Medicinal Chemistry* **2002**, *45*, 1543-1558.

(11) Berg, T. Modulation of protein-protein interactions with small organic molecules. *Angewandte Chemie-International Edition* **2003**, *42*, 2462-2481.

(12) Orner, B. P.; Ernst, J. T.; Hamilton, A. D. Toward proteomimetics: Terphenyl derivatives as structural and functional mimics of extended regions of an alpha-helix. *Journal of the American Chemical Society* **2001**, *123*, 5382-5383.

(13) Ernst, J. T.; Becerril, J.; Park, H. S.; Yin, H.; Hamilton, A. D. Design and application of an alpha-helix-mimetic scaffold based on an oligoamide-foldamer

strategy: Antagonism of the bak BH3/Bcl-xL complex. *Angewandte Chemie-International Edition* **2003**, *42*, 535-+.

(14) Ernst, J. T.; Kutzki, O.; Debnath, A. K.; Jiang, S.; Lu, H. et al. Design of a protein surface antagonist based on alpha-helix mimicry: Inhibition of gp41 assembly and viral fusion. *Angewandte Chemie-International Edition* **2001**, *41*, 278-+.

(15) Yin, H.; Hamilton, A. D. Terephthalamide derivatives as mimetics of the helical region of Bak peptide target Bcl-xL protein. *Bioorganic & Medicinal Chemistry Letters* **2004**, *14*, 1375-1379.

(16) Yin, H.; Lee, G. I.; Park, H. S.; Payne, G. A.; Rodriguez, J. M. et al. Terphenyl-based helical mimetics that disrupt the p53/HDM2 Interaction. *Angewandte Chemie-International Edition* **2005**, *44*, 2704-2707.

(17) Anderson, A. C. The process of structure-based drug design. *Chemistry & Biology* **2003**, *10*, 787-797.

(18) Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* **1998**, *280*, 1-9.

(19) Kortemme, T.; Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 14116-14121.

(20) Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **1997**, *88*, 323-331.

(21) Momand, J.; Jung, D.; Wilczynski, S.; Niland, J. The MDM2 gene amplification database. *Nucleic Acids Research* **1998**, *26*, 3453-3459.

Chapter 2

Library Design of Proteomimetic MDM2 Inhibitors

Introduction

Investigations into the structural basis for p53-MDM2 binding began with the mapping of p53 and MDM2 interaction domains, which determined that the interaction site was located at the N-terminal of p53.[1,2] Shortly thereafter, site-directed mutagenesis studies identified L14, F19, L22, and W23 residues as being critical for MDM2 binding, and L22 and W23 as being critical for transactivation.[3] Next, a study of the binding of phage-display library peptides to MDM2 revealed that the interaction affinity was between 0.3 and 0.5 μM and defined the critical binding points of p53 as being F19, D21, W23, and L26.[4] Later that year, the crystal structure of MDM2 bound to a peptide from the transactivation domain of p53 revealed a small amphipathic alpha helix bound to a relatively deep well-defined hydrophobic pocket of MDM2.[5] This pocket is filled primarily by the *i*, *i* + 4, and *i* + 7 side chains from the p53 sequence FSDLWKLLP. The p53 protein presents a hydrophobic face containing the F, W, and L residues to a hydrophobic pocket on the MDM2 surface. NMR studies show that this p53 region tends to from two β turns in solution, even though the MDM2-bound structure is an alpha helix.[6]

p53-based peptide libraries have produced compounds which inhibit the p53-MDM2 interaction *in vitro*.[4] This study found and optimized a small (12-20 residue) peptide that effectively competes against native p53 for binding to MDM2. Some inhibitors bound up to 100 times stronger than the p53-derived wild type peptide, illustrating the potential for high affinity binding in the MDM2 pocket. Furet and coworkers at Novartis continued to optimize peptide-based inhibitors.[7] First, they replaced residues that did not interact with MDM2 with α,α-disubstituted amino acids,

which are known to stabilize α- and 3_{10}-helices in short peptide motifs.[8] After a 4-fold increase in affinity, they focused on optimizing side chains. They ended up with a peptide, with a sequence of Ac-Phe[19]-Met-Aib-Pmp-6-Cl-Trp-Glu-Ac_3c-Leu[26]-NH_2, having an IC_{50} of 5 nM. The success in optimizing peptidic inhibitors set the stage for the small molecule inhibitors.

At the beginning of this project, three small molecule inhibitors of MDM2 had been reported – piperazine-4-phenyl derivatives,[9] chalcones,[10] and chlorofusin.[11] (Figure 2.1) Piperazine-4-phenyl derivatives, with K_i of 300 nM, were first discovered through screening by Luke and coworkers. Very little information has been reported, other than their general structure and binding affinity. Chalcones were then discovered through screening but had a low binding affinity with K_i of 5 μM. Lastly, fungal metabolite chlorofusin, with a binding affinity of 5 μM, was found in yet another screen.

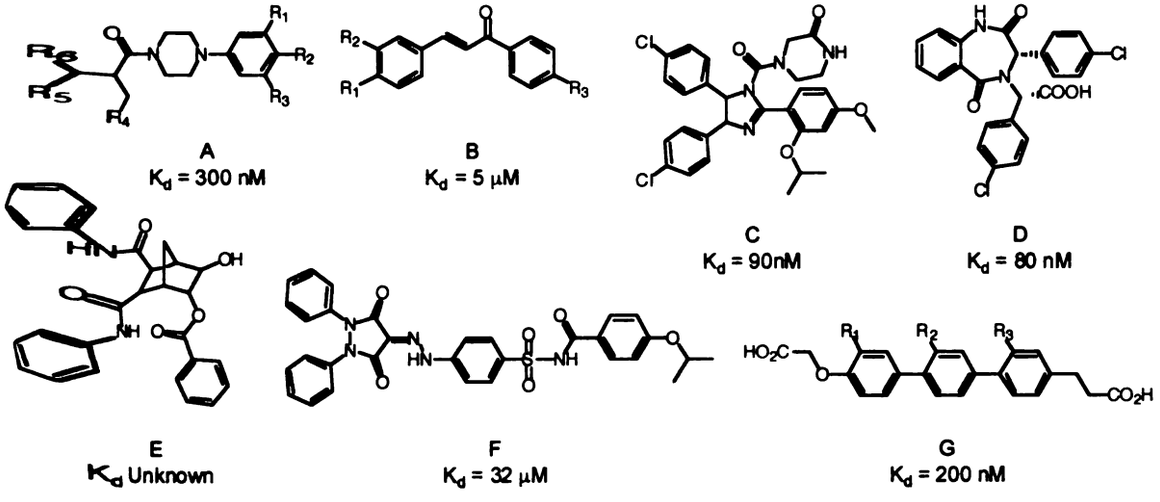


Figure 2.1 Reported MDM2 inhibitors A) piperazine-4-phenyl derivatives, B) chalcones, C) nutlins, D) benzodiazepinediones, E) norbornanes, F) sulfonamides, and G) terphenyls.

In recent years, many chemical inhibitors of MDM2 have been reported, attesting to the importance of the target as well as the suitability of the p53-MDM2 interface for small molecule binding. By far the most potent and well-characterized p53-MDM2 inhibitors are the nutlins, identified by high throughput screening. [12] The optimized inhibitors in this series bound to MDM2 with a K_d of 90 nM *in vitro*, and were shown to have anti-tumor activity in mice. X-ray crystallography and NMR structures revealed that nutlins bind to the p53-binding site of MDM2, mimicking the p53 peptide to a high degree. Another series of potent inhibitors discovered through screening are the benzodiazepinediones. [13] The activity of benzodiazepinediones is less well characterized but the compounds have potent *in vitro* activity, with the most potent compound binding at 80 nM. The crystal structure indicates that the inhibitor binding mode also mimics that of the p53 peptide.

Three of the nine classes of known inhibitors were discovered through design or computational screening – the norbornanes, [14] sulfonamides, [15] and terphenyls. [16] A sulfonamide, with IC_{50} of 32 μM, was discovered by Galatin and Abraham as a result of UNITY pharmacophore searches of the NCI chemical database. Not surprisingly, the three pharmacophores used were Phe19, Trp23, and Leu26 of the p53 peptide. Zao and coworkers also searched for the same pharmacophores using UNITY, and then docked the resulting compounds. They synthesized a series of norbornanes and found them active in several cell lines. The most potent MDM2 inhibitors discovered through design are the terphenyls, with a binding affinity of 200 nM. These inhibitors were originally designed in 2001 as *i*, *i* + 3, *i* + 7 alpha helix mimics binding to calmodulin. Since then,

terphenyl compounds have been shown to bind to several proteins that recognize and bind alpha helices: gp41,[17] Bcl-XL,[18] and MDM2.

Although the best drug candidates were discovered through screening, less potent inhibitors conceived through computational screening and design are highly significant because they embody a knowledge-based approach that has the potential to evolve into the lead discovery strategy of the future. This chapter, representing our efforts in advancing the field of computational design, describes a method of selecting scaffold candidates for proteomimetic libraries and applies this method in the small molecule friendly p53-MDM2 system.

Results

The structure of p53 and MDM2 revealed a deep hydrophobic binding interface in which the three amino acids of p53 at the protein interface are presented in a linear manner from the *i*, *i* + 4, and *i* + 7 positions of an alpha helix (Figure 2.2A). The amino acid side chains interact with MDM2, while the p53 peptide backbone makes few contacts. Thus, the stable folded alpha helical structure of a protein provides a scaffold from which side chains are delivered to the binding pockets of structure of a protein provides a scaffold from which side chains are delivered to the binding pockets of MDM2. We approached the same problem by designing a non-peptide molecule that can replace the peptide backbone while projecting side chain functionalities with distance and angular relationships equivalent to those seen in p53 (Figure 2.3). A semi-rigid scaffold that can pre-orient side chains should facilitate binding by lowering the entropic penalty of ordering the backbone into a helical structure. Since the amino acid functionalities

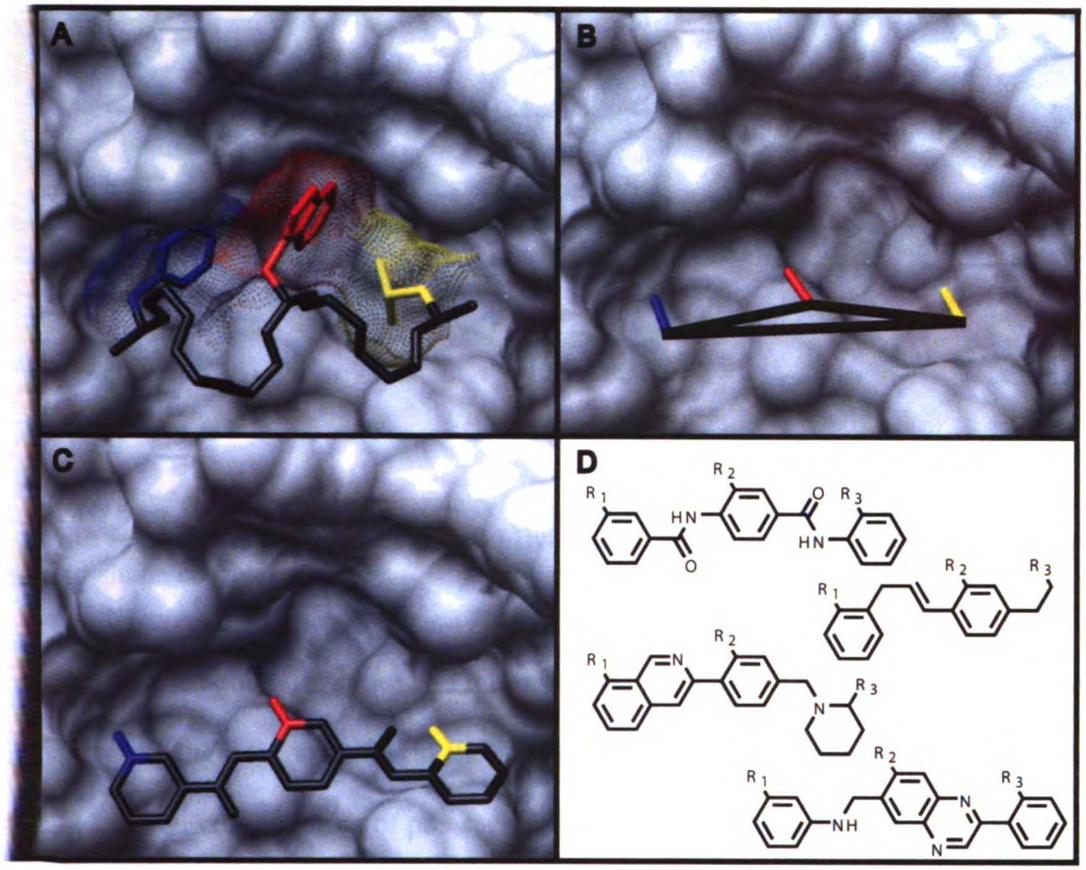


Figure 2.2 (A) Backbone trace of p53 peptide shown with MDM2. Residues 19F (blue), 23W (red), and 26L (yellow) occupy a deep hydrophobic pocket. (B) Cα-Cβ bonds of 19F, 23W, and 26L used in the CAVEAT search for scaffolds. (C) Library scaffold chosen for synthesis fulfills geometric requirements of CAVEAT search. (D) Structures of 4 compounds considered for synthesis.

extend from the Cα-Cβ bond, the relative positions and orientations of three bonds will need to be present in the scaffold. This general scaffold search strategy can be applied to other types of helices as well as other protein motifs, providing a general approach for proteomimetics (Figure 2.4).

To find scaffolds that fulfill the geometric properties necessary for correct side chain placement, our design method employs CAVEAT.[19] CAVEAT searches through 3-

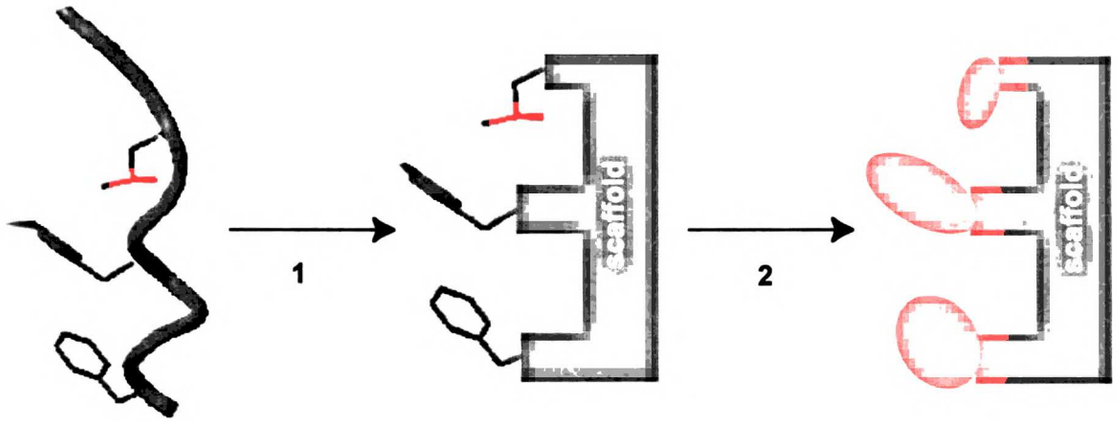


Figure 2.3 Inhibitor design strategy. (1) Replace flexible p53 peptide backbone with semi-rigid organic scaffold. (2) Evaluate diverse side chains to maximize size and shape complementarity.

dimensional structure databases and returns molecules containing bonds with the same distance and angle relationships as those deemed critical in the reference structure. The MDM2-bound conformation of p53 served as the reference structure, and the Cα-Cβ bonds of 19F, 23W, and 26L of p53 were used (Figure 2.2B). To allow for uncertainty in the crystal structure side chain binding conformation, tolerances of 11 degrees were used for bond angles and 0.24 Å were used for bond separations. A search of conformationally expanded versions of the ACD,[20] MDDR,[20] NCI,[21] CMC,[20] Iliad,[19] and Triad[19] databases yielded 40,000 structures that fulfilled the geometric requirements. Atoms extending beyond the Cβ equivalent positions were removed to allow side chains to be appended later in the design process.

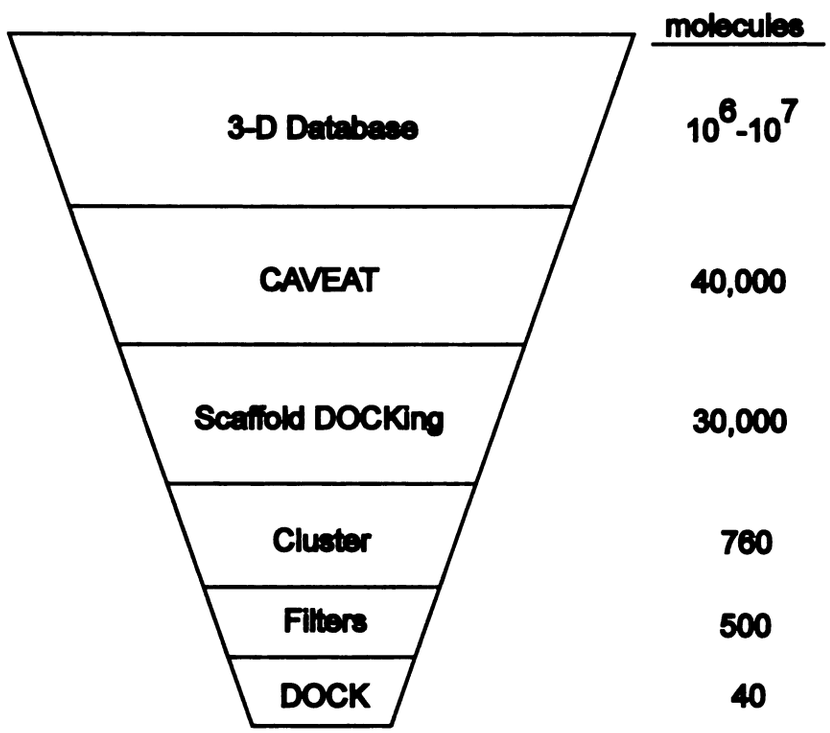


Figure 2.4 Overview of the multi-step approach to library design.

Another important consideration in the scaffold design was creating chemical complementarity to MDM2. This factor was assessed using several rounds of scoring with DOCK 4.0.[22] (Figure 2.4) The computational time and design effort increased with each round, so it was important to remove poor scaffolds as early as possible using increasingly stringent scoring schemes. The DOCK ligand orientation function was not needed because CAVEAT had already oriented the ligand to match the corresponding bonds in the p53 reference structure. In the first round, no energy minimization was allowed and the score was based solely on Van der Waals interactions. Scaffolds scoring greater than 1000 were deemed to have an irreparable clash with MDM2 and discarded from further consideration. The surviving structures were hierarachically clustered using

2-dimensional Daylight fingerprint descriptors, a closest linkage algorithm, and a 0.85 Tanimoto coefficient.[23] Because fingerprint descriptors recognize chemical diversity, all heteroatoms were changed to carbon to effect geometric, rather than chemical clustering. This resulted in 761 clusters.

Between the first and second rounds of DOCK scoring, several filters were used to remove poor scaffolds. Because we preferred a conformationally rigid scaffold, structures with no rings or greater than 5 consecutive rotatable bonds were removed. Because strained or complex molecules would involve difficult syntheses, molecules with 4-membered rings or more than 3 fused rings were also discarded. The application of these filters left 500 remaining clusterheads, which were minimized using DOCK 4.0 scored with a Van der Waals scoring function. In this step, the ligand geometry was allowed to move with respect to the MDM2 target. Conformational variances and heteroatoms were restored to the molecules scoring in the top half, yielding 1369 three dimensional structures.

At this point, the scaffolds alone could not be further distinguished by DOCK because the most important aspect of the interaction, the side chains, was missing. Hence, phenylalanine, phenylalanine, and leucine side chains were added to the R_1, R_2, and R_3 positions on the scaffold, corresponding to the 19F, 23W, and 26L positions of p53. The structures were evaluated with a Van der Waals and electrostatics score using DOCK 4.0. Thus, the side chains aided in scaffold evaluation. The top structures were visually inspected for synthetic accessibility (Figure 2.5).

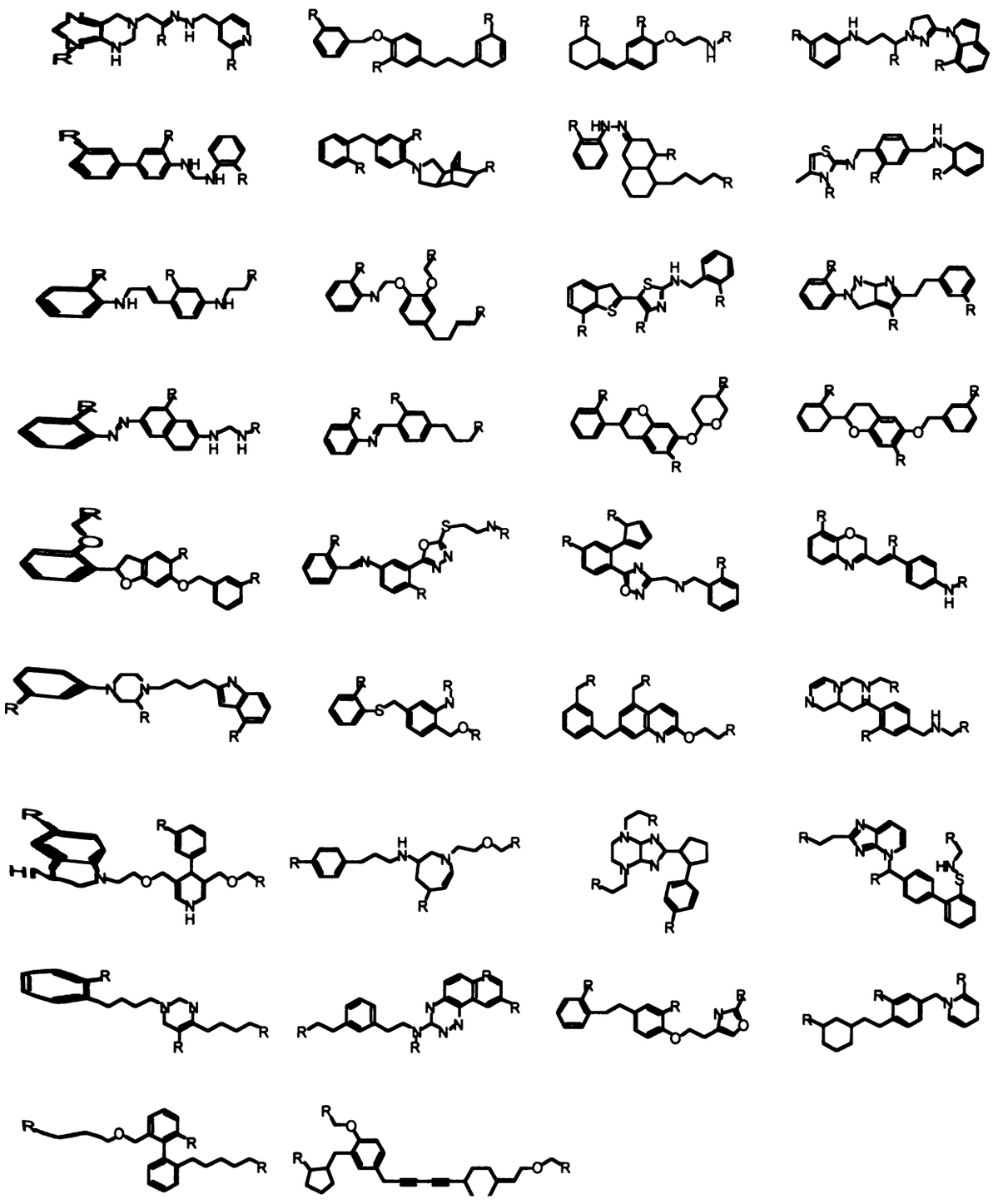


Figure 2.5 Best-scoring scaffolds considered for synthesis.

Of the four best scaffolds (Figure 2.2D), the main advantage of the selected scaffold (Figure 2.2C) was that it could be synthesized in a modular fashion with simple,

seemingly straightforward chemistry. The only disadvantage was its high hydrophobicity, which can be addressed by replacing scaffold carbons with heteroatoms and by attaching hydrophilic functionalities to the solvent-facing side of the scaffold.

Theoretically, the selected scaffold with phenylalanine, tryptophan, and leucine side chains should bind to MDM2 with similar or greater affinity than the p53 peptide. There was also potential for increasing the binding affinity by optimizing side chain contacts (Figure 2.3). Preliminary evaluations of a limited set of side chains using the DOCK scoring function yielded very similar scores, with the exception of saturated rings, which scored poorly. Because the project reported in Chapter 5 demonstrated the difficulty and time required to accurately predict the binding of minor variations in side chains, we decided to optimize side chains empirically rather than theoretically.[24]

Discussion

We have constructed inhibitors of the p53-MDM2 interaction using a computational design strategy that can be applied to any protein-protein interaction for which a co-crystal structure exists. These inhibitors have a novel structure and represent a complementary approach to the screening methods used in the discovery of previous inhibitors. Because the inhibitor scaffolds were designed as alpha helix mimics, the compound libraries may have activity in other protein-protein interactions in which an *i*, *i* + 4, *i* + 7 alpha helix plays a role. Some of these proteins include but are not limited to Bak, NF-κB, and VP16. The success of the libraries in these other systems will speak toward the extent of their alpha helix mimicry as will the synthesis and testing of the remaining scaffolds.

As the first step in the library design process, CAVEAT and its parameters played a large role in determining the scaffold structure. One such parameter is the angle and distance tolerances that describe the allowed deviation from the input structure. These tolerances can be increased or decreased based on the quality of the crystal structure, receptor site flexibility, and desired number of hits. Pilot studies showed a dramatic increase in hits as the tolerance is relaxed. While 11 degrees and 0.24 angstroms is a conservative margin, larger tolerances would have resulted in too many hits. The bonds chosen for the CAVEAT search are also important. The multitude of options includes other peptide bonds such as the Cβ-Cγ bond and a combination of bonds from a non-peptidic inhibitor. Libraries derived from Cα-Cβ bonds of protein secondary structures have added value because the recurrence of the secondary structures increases the relevance and applicability of the library in other protein-protein interactions. Ultimately, the numerous paths for entering a binding site are equally valid until further investigation and will result in highly varied libraries and inhibitors. Of the 4 scaffolds we examined closely, only one resembled a known MDM2 inhibitor: the chalcone (Figure 2.2D). The overlap of hits resulting from a screening approach and our computational approach was a significant positive benchmark.

CAVEAT can be classified as a pharmacophore-based modelling program, but it differs from UNITY, the pharmacophore package used in discovery of the norbornane and sulfonamide classes of MDM2 inhibitors. In Galatin and Abraham's search for MDM2 inhibitors, UNITY identifies molecules that contain three pharmacophores described by 19F, 23W, and 26W. Although both programs emphasize the same side chains, UNITY hits incorporate the side chains, while CAVEAT hits only contain the

scaffold. As a result, the CAVEAT-based inhibitors are larger, its side chains are easily modified, and the full inhibitor may differ greatly from the original database molecule. Most importantly, the number of hits is dramatically increased with CAVEAT. Searching through the NCI database using a 20% tolerance, UNITY found 7 hits. Searching through the NCI database using a much smaller tolerance, CAVEAT found 4,000 unique scaffolds.

Because pharmacophore modelling does not account for a molecule's receptor complementarity, it should be used in conjunction with a docking method when a receptor structure is available. Abraham and Galatin checked for receptor clashes in UNITY, Zhao and coworkers used DOCK, and our method also uses several rounds of DOCK scoring. While DOCK was extremely useful in removing molecules with receptor clashes, it would not perform well as the sole tool in a search for inhibitors of a protein-protein interaction. DOCK has successfully identified inhibitors in the past, but these examples have generally used enzyme targets with a small and well-defined binding site. With such a large hydrophobic MDM2 surface, many molecules scoring well in DOCK may not fill the three binding pockets of MDM2. Using CAVEAT compensates for this shortcoming by finding scaffolds that place side chains into the binding pocket.

After scaffold selection, the next step in inhibitor design was side chain selection. Preliminary DOCK calculations yielded side chains that scored similarly, suggesting that more in-depth studies were necessary. Due to the high similarity amongst the modelled side chains, this result was expected. In Chapter 5, theoretical predictions of side chain preferences for members of the cysteine protease family were successful only after molecular dynamics techniques were employed. The original approach using a

DOCKing strategy did not correctly predict the difference between dramatically different side chains such as phenylalanine, isoleucine, and glutamine. Thus, we chose side chains that were derivatives of phenylalanine because they were commercially available, easy to handle and incorporate into the synthesis, and could reveal shape and electrostatic SAR. Because DOCK disfavored saturated rings, cyclohexane was also included to test the accuracy of the prediction. In the future, experimental results could be used as a training set for developing DOCKing methods for predicting library side chains.

Experimental Methods

The structure-based design process began with the coordinates for MDM2 bound to a short segment of p53 (PDB code 1ycr). The Cα-Cβ atoms and bonds of p53 residues19, 23, and 26 were used as vectors in a CAVEAT search with geometric tolerances of 11 degrees for bond angles and 0.24Å. The CAVEAT search database containing an average of 10 low-energy conformations of molecules from the combination of the Available Chemicals Directory (ACD), the MDL Drug Data Repository (MDDR), the National Cancer Institute (NCI), comprehensive medicinal chemistry (CMC), Iliad, and Triad was generated using OMEGA (OpenEye). The 40,000 structures identified by CAVEAT were scored with the DOCK 4.0 Van der Waals scoring function. The crystal structure was prepared for docking in a standard manner by removing the p53 peptide and assigning charges by the method of Cornell, et al.[25] A 0.15Å spacing energy grid comprised of a Lennard-Jone 12-6 potential was used to score the rigidly DOCKed molecules. Approximately 10,000 structures scored greater than 1,000 and were discarded.

The structures were characterized by 2-dimensional Daylight fingerprint descriptors and hierarchically clustered with a closest linkage algorithm using a Tanimoto coefficient of 0.85. Because fingerprint descriptors recognize chemical diversity, all heteroatoms were changed to carbon to effect geometric, rather than chemical clustering. Of the remaining 761 structures, those with more than 6 consecutive rotatable bonds or zero rings were removed in a screen for conformational rigidity. Structures with 4-membered rings or more than 4 fused rings were removed due to synthetic difficulty. The application of these filters left 500 remaining clusterheads, which were minimized using DOCK 4.0 scored with a Van der Waals scoring function and a narrower cutoff. Conformational variances and heteroatoms were restored to the molecules scoring in the top half, yielding 1369 three dimensional structures.

Phenylalanine, phenylalanine, and leucine side chains were added to the R_1, R_2, and R_3 positions on the scaffold, corresponding to the 19F, 23W, and 26L positions of p53. These molecules were charged with Gasteiger charges and DOCKed using Van der Waals (previously described) and a distance dependent dielectric of 4.[26] The top 250 structures contained approximately 40 unique scaffolds that were considered for library synthesis.

Appendix 2 contains OMEGA, DOCK, and CAVEAT parameters used.

References

(1) Oliner, J. D.; Pietenpol, J. A.; Thiagalingam, S.; Gvuris, J.; Kinzler, K. W. et al. Oncoprotein Mdm2 Conceals the Activation Domain of Tumor Suppressor-P53. *Nature* **1993**, *362*, 857-860.

(2) Chen, J. D.; Marechal, V.; Levine, A. J. Mapping of the P53 and Mdm-2 Interaction Domains. *Molecular and Cellular Biology* **1993**, *13*, 4107-4114.

(3) Lin, J. Y.; Chen, J. D.; Elenbaas, B.; Levine, A. J. Several Hydrophobic Amino-Acids in the P53 Amino-Terminal Domain Are Required for Transcriptional Activation, Binding to Mdm-2 and the Adenovirus-5 E1b 55-Kd Protein. *Genes & Development* **1994**, *8*, 1235-1246.

(4) Bottger, V.; Bottger, A.; Howard, S. F.; Picksley, S. M.; Chene, P. et al. Identification of novel mdm2 binding peptides by phage display. *Oncogene* **1996**, *13*, 2141-2147.

(5) Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J., Pavletich, N.P. Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain. *Science* **1996**, *274*, 948-953.

(6) Botuyan, M. V. E.; Momand, J.; Chen, Y. Solution conformation of an essential region of the p53 transactivation domain. *Folding & Design* **1997**, *2*, 331-342.

(7) Garcia-Echeverria, C.; Chene, P.; Blommers, M. J. J.; Furet, P. Discovery of potent antagonists of the interaction between human double minute 2 and tumor suppressor p53. *Journal of Medicinal Chemistry* **2000**, *43*, 3205-3208.

(8) Marshall, G. R.; Hodgkin, E. E.; Langs, D. A.; Smith, G. D.; Zabrocki, J. et al. Factors Governing Helical Preference of Peptides Containing Multiple

Alpha,Alpha-Dialkyl Amino-Acids. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87*, 487-491.

(9) Luke, R. W. A., Cotton, R., Jewsbury, P.J. Piperizine-4-phenyl derivatives as inhibitors of the interaction between MDM2 and p53: Great Britain, 2000.

(10) Stoll, R., Renner, C., Hansen, S., Plame, S., Klein, C., Belling, A., Zeslawski, W., Kamionka, M., Rehm, T. Muhlhahn, P., Schumacher, R., Hesse, F., Kaluza, B.,Voelter, W., Engh, R., Holak, T. Chalcone Derivatives Antagonize Interactions between the Human Oncoprotein MDM2 and p53. *Biochemistry* **2001**, *40*, 336-344.

(11) Duncan, S. J.; Gruschow, S.; Williams, D. H.; McNicholas, C.; Purewal, R. et al. Isolation and structure elucidation of chlorofusin, a novel p53-MDM2 antagonist from a Fusarium sp. *Journal of the American Chemical Society* **2001**, *123*, 554-560.

(12) Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F. et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **2004**, *303*, 844-848.

(13) Grasberger, B. L.; Lu, T. B.; Schubert, C.; Parks, D. J.; Carver, T. E. et al. Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *Journal of Medicinal Chemistry* **2005**, *48*, 909-912.

(14) Zhao, J.; Wang, M.; Chen, J.; Luo, A.; Wang, X. et al. The initial evaluation of non-peptidic small-molecule HDM2 inhibitors based on p53-MDM2 complex structure. *Cancer Letters* **2002**, *183*, 69-77.

(15) Galatin, P. S.; Abraham, D. J. A nonpeptidic sulfonamide inhibits the p53-mdm2 interaction and activates p53-dependent transcription in mdm2-overexpressing cells. *Journal of Medicinal Chemistry* **2004**, *47*, 4163-4165.

(16) Yin, H.; Lee, G. I.; Park, H. S.; Payne, G. A.; Rodriguez, J. M. et al. Terphenyl-based helical mimetics that disrupt the p53/HDM2 Interaction. *Angewandte Chemie-International Edition* **2005**, *44*, 2704-2707.

(17) Ernst, J. T.; Kutzki, O.; Debnath, A. K.; Jiang, S.; Lu, H. et al. Design of a protein surface antagonist based on alpha-helix mimicry: Inhibition of gp41 assembly and viral fusion. *Angewandte Chemie-International Edition* **2001**, *41*, 278-+.

(18) Kutzki, O.; Park, H. S.; Ernst, J. T.; Orner, B. P.; Yin, H. et al. Development of a potent Bcl-x(L) antagonist based on alpha-helix mimicry. *Journal of the American Chemical Society* **2002**, *124*, 11838-11839.

(19) Lauri, G., Bartlett, P.A. CAVEAT: A program to facilitate the design of organic molecules. *J. Comput-Aided Mol. Design* **1994**, *8*, 51-66.

(20) MDL Information Systems. *MDL Information Systems*, www.mdli.com.

(21) National Cancer Institute. *National Cancer Institute*, cactus.nci.nih.gov/ncidb/download.html.

(22) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design* **2001**, *15*, 411-428.

(23) Willett, P. *Similarity and Clustering in Chemical Information Systems*; John Wiley & Sons, 1987; 266.

(24) Greenbaum, D. C.; Arnold, W. D.; Lu, F.; Hayrapetian, L.; Baruch, A. et al. Small molecule affinity fingerprinting: a tool for enzyme family subclassification, target identification, and inhibitor design. *Chemistry & Biology* **2002**, *9*, 1085-1094.

(25) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. et al. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(26) Gasteiger, J., Marsili, M. Iterative Partial Equalization of Orbital Electronegativity-Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219-3288.

Acknowledgements

The following people contributed scientifically and/or socially to my time down in the dungeon: Bill Arnold, Jim Arnold, Alex Aronov, Tiba Aynechi, Pradipta Bandyopadthy, Natasja Brooijmans, Jim Caldwell, Jim Frazine, Jose Haresco, Peter Kollman, Jennifer Krumrine, Fernando Martin, Terry Lang, Matt Lee, Kevin Masukawa, Demetri Moustakas, Anang Shelat, Scott Pegg, Geoff Skillman, Stepan Sklenak, David Sullivan, Sam Toba, and Andy Verras.

Tack Kuntz, for all of his wisdom, funding, encouragement, and group meeting sandwiches, gets his own run-on sentence, in which I'd like to thank him profusely.

Chapter 3

Synthesis of Proteomimetic Combinatorial Libraries

Introduction

Amongst the few dozen computationally superior scaffolds, one scaffold stood out as being the most synthetically tractable. The scaffold consisted of three phenyl rings connected to eachother with two atom linkers. This scaffold was found repeatedly in the top scoring DOCKed structures, although the linkers varied as combinations of amine, azo, ether, amide, and sulfide groups. The amide linker was chosen for several reasons. First, amide bond formations were well-studied and multiple reaction conditions had been published. Second, the amide linker was less rigid and more stable than the azo linker, but more rigid than the amine, ether, and sulfide linkers. Third, the amide group could be reduced to an amine group, yielding two libraries.

The library members were composed of six parts – three scaffold aryl rings connected by amide bonds, and three side chains each connected to a scaffold ring by carbon-carbon bonds. We envisaged synthesizing the library in solution. We considered assembling the scaffold first, but thought there would be great difficulty in attaching three different side chains selectively. Therefore we decided to add side chains to each of the three aryl rings, and then use amidations to connect the individual subunits (Figure 3.1).

An advantage of this strategy was the utilization of a common chemistry for production of diversity elements – Suzuki couplings were used to add side chains to each monomer at the diversity points. Carboxylic acid and amine groups were protected as methyl ester, formyl, and nitro groups during the production of diversity reagents.

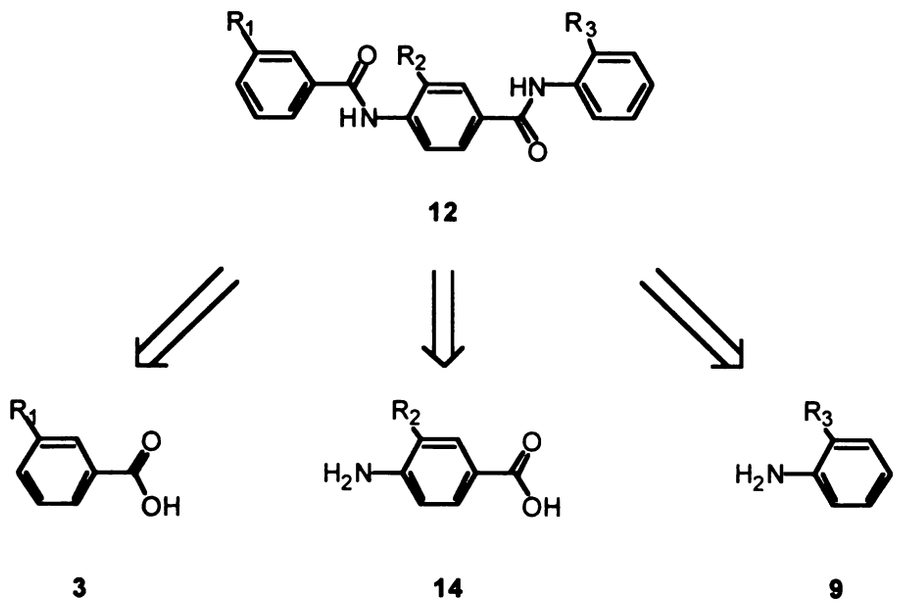


Figure 3.1 Retrosynthetic analysis of proposed library.

After careful consideration and planning, we believed the library synthesis would be straightforward and simple. This chapter describes the successes and failures that led to a successful synthetic route and the production of 173 compounds (Figure 3.2).

Results and Discussion

The first objective was to synthesize **3** and **9** using benzyl side chains. These reactions were successful during the first attempt and the production of these diversity elements proceeded without difficulty. **2** and **8** were easily synthesized with 30-85% yield from commercially available reagents **1** and **7**, respectively, under conditions reported by Klarner et al.[1] The reactions proceeded cleanly, as the purity of the crude reaction was greater than 80% when analyzed by thin layer chromatography (TLC). Reduction of **8** proceeded quantitatively by catalytic hydrogenation with palladium on

carbon. The oxidation of **2** was carried out using sodium chlorite as a mild oxidant and hydrogen peroxide as a scavenging agent for the hypochlorite byproduct.[2]

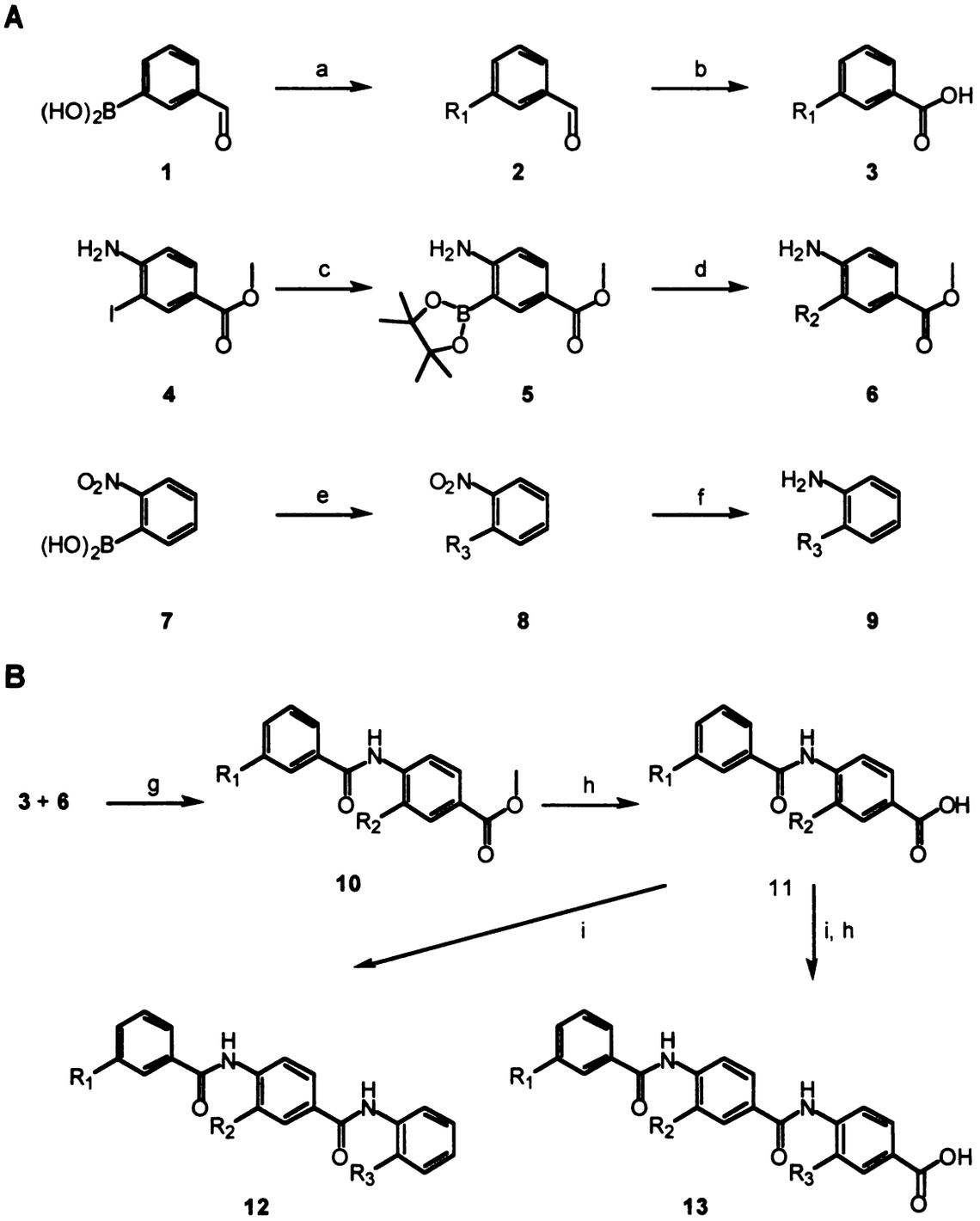


Figure 3.2 (A) Synthesis of diversity elements. a) R_1Br, $Pd(PPh_3)_4$, THF, K_2CO_3, 80°C, 16 h, 28-84%; b) $NaClO_2$, NaH_2PO_4, H_2O_2, CH_3CN, 4°C to room

temperature, 2 h, 56-99%; c) pinacolborane, 1,4-dioxane, Et_3N, $PdCl_2(dppf)$, 80°C, 8 h, 20%; d) R_2Br, $Pd(PPh_3)_4$, THF, K_2CO_3, 80°C, 16 h, 14-67%; e) R_3Br, $Pd(PPh_3)_4$, THF, K_2CO_3, 80°C, 16 h, 44-100 ; f) Pd/C, H_2, 1 h, 80-100% (B) Library synthesis. g) PS-TsCl, DMAP, CH_2Cl_2, 40°C, 48 h, 71-90% ; h) NaOH, 1,4-dioxane, 72 h, 43-91%; i) PS-TsCl, DMAP, DMF, 40°C, 48 h, 9-60%

Synthesis of **6** was highly problematic. The initial strategy, based on the work of Seaman and Johnson on the nitration of phenylboronic acid, involved the synthesis of **17** via nitration of boronic acid **15**.[3] (Figure 3.3) These attempts yielded mostly proteodeboronylated nitrobenzaldehydes and large amounts of starting material. Nitration of a more activated substrate **18** yielded an inseparable mixture of nitroboronic acid regioisomers. Because of this regioselectivity problem, we developed a new approach starting from commercially available trisubstituted benzene **4**.

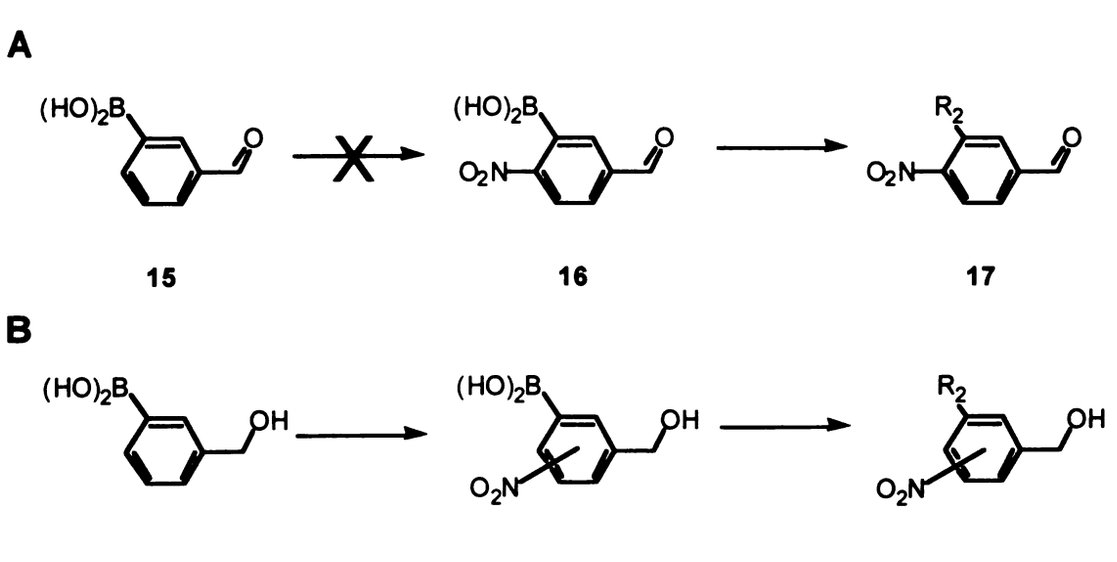


Figure 3.3 Initial approach toward synthesis of the R_2 diversity element.

Literature searches revealed that the formation of boronic acids required a palladium catalyst, a base, a boron source, and a suitable solvent. The abundance of

reagents and reaction conditions was problematic in this case because there was no known rationale for deducing optimal conditions. After trying several different combinations of conditions, none were successful. Finally, the publication of the synthesis of (2-aminophenyl)boronate guided us toward favorable reaction conditions.

The synthesis of pinacolboronate **5** was based on conditions reported by Baudoin et al. for the synthesis of pinacol (2-aminophenyl)boronate from 2-bromoaniline.[4] Because their substrate was highly similar to **4**, their report contained the most promising published conditions for the synthesis of **5**. Their optimal conditions employing palladium acetate as a catalyst and a biphenylphosphine ligand worked poorly for our substrate. I continued to test other conditions reported by Baudoin and coworkers and found that conditions employing palladium chloride diphenylphosphinoferrocene, suboptimal in their case, resulted in 20% yield of **5**. The low yield can be partially attributed to product loss during silica chromatography, as evidenced by a long product tail. The subsequent Suzuki coupling to form **6**{1} proceeded smoothly under the conditions of Klarner, et al. To prevent the significant product loss from silica purification of **5**, the boronate synthesis and Suzuki coupling were performed consecutively in one pot, increasing the yield for the two reactions to 35%.

The original approach for forming the amide bond linking diversity elements involved a catch and release strategy using tetrafluorophenol resin. The carboxylic acid substrate reacts with the resin to form an activated resin-bound ester. After the beads are washed and the reagents removed, the amine nucleophile is added and the amide bond formed. This approach failed because the activated resin-bound ester was never formed. The focus then shifted toward a stronger activating reagent, oxalyl chloride. The amide

bond was formed using oxalyl chloride, but the yield was 10% and TLC revealed 6 products. Most importantly, the use of oxalyl chloride required anhydrous conditions, creating difficulty in adapting the reaction for library synthesis. Further experiments with other coupling agents, such as DCC or TFFH resulted in low yields or low purity.

Amide bond formation between each of the diversity elements was easily carried out under the conditions optimized by Soloshonok and coworkers for sterically hindered poorly nucleophilic amines[5]. Treatment of the acid component with tosyl chloride followed by addition of the aniline gave a clean reaction with good yield. Furthermore, the reaction conditions were easily adapted for parallel synthesis in 48-well FlexChem reaction blocks. The FlexChem's fritted reaction well facilitated the use of resin-bound tosyl chloride reducing the number of species in the crude product.

Proofing reactions for the synthesis of **3**, **6**, and **9** were carried out using side chain 1. Purified yields were greater than 75% for all reactions except c and d. The synthesis of chemsets **3**, **6**, and **9** followed with similar results. The exception was the nitro side chain, whose presence reduced the yield in Suzuki couplings by about half. Half gram amounts of chemsets **3**, **6**, and **9** were synthesized and used as needed for library synthesis. The library synthesis reactions g-i were tested with methyl side chains. While test reactions proceeded quantitatively, the conversion time was longer for the actual side chains.

The size of the first production library of chemset **12** was limited to roughly 100 total members to ease handling. The side chains chosen were based on commercial availability and structural similarity to the benzyl side chain used in the proofing

reactions (Figure 3.4). These side chains are highly similar to the benzyl side chains but could potentially reveal shape and electrostatic structure-activity relationships.

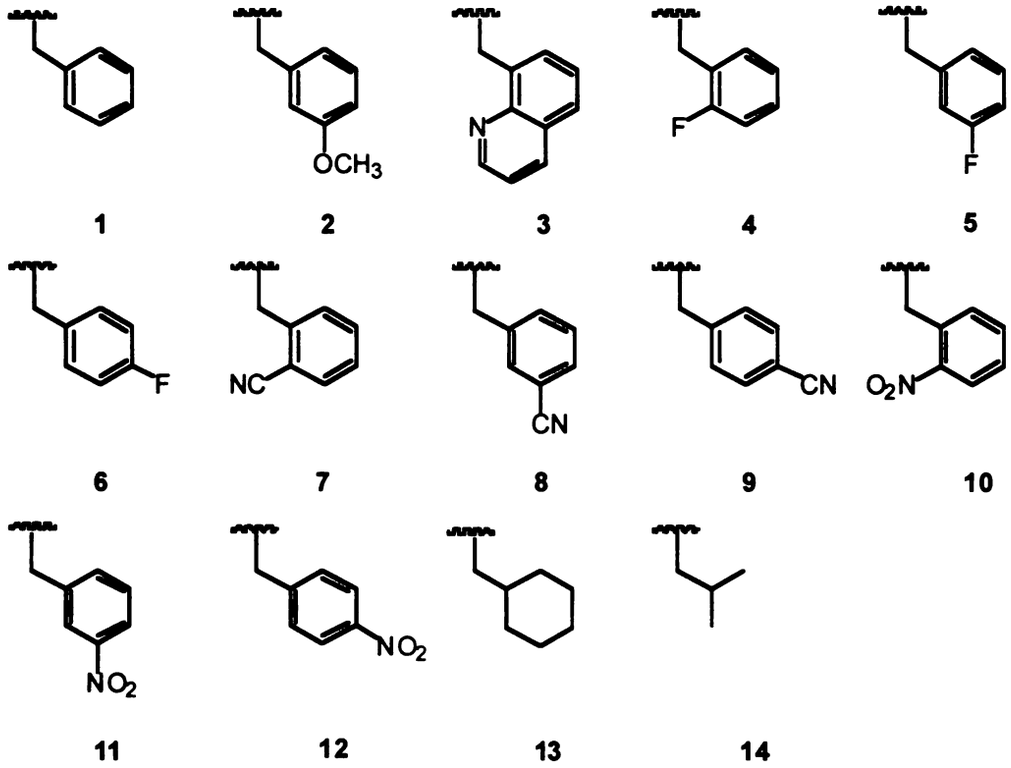


Figure 3.4 Side chains chosen for synthesis.

Preliminary calculations with the DOCK scoring function did not clearly favor one aromatic side chain over another, with one exception – saturated side chains such as cyclohexane scored poorly. Thus, side chain 13 was used as a negative control for the DOCKing predictions. The tryptophan side chain was not included because its reactivity rendered it synthetically difficult. With three diversity elements and roughly 10 possible side chains for each diversity element, an all-by-all by all library would contain over a thousand compounds. Instead, 13 random combinations of chemset **10** were synthesized and converted to chemset **11**. Chemset **11** was coupled to nine R_3 elements to form

chemset **12** for a total of 117 compounds. Members of chemset **11** were also tested for activity. Due to positive assay results (Chapter 4), I synthesized a second library focused on chemset **11**. The all by all library of chemset **11** was aimed at probing SAR and increasing potency. I also synthesized a few members of chemset **13**{7,6,X} to test the importance of the carboxylic acid functionality.

Library synthesis progressed smoothly after conditions were optimized. It was important to optimize each reaction until high yield and high purity was reached. During library synthesis, it is laborious and impractical to isolate products from unknown mixtures, a task that is common in conventional syntheses. The key reaction in the synthesis, which enabled the production of over 100 library members, was the amide bond formation. Resin-bound TsCl and DMAP effected the transformation quantitatively and did not require anhydrous conditions. As with the liquid form of tosyl chloride, reagent deterioration occurs over a time period of weeks and it is best to use unopened bottles. This reaction proved to be robust with the majority of missing or failed compounds attributed to tosyl chloride deterioration, human or machine error. It would be feasible to make a library of many hundreds of compounds using these reaction conditions, with the rate-limiting step being the post-purification fraction sorting and characterization.

Experimental

General Synthetic Methods: All reagents and starting materials were purchased from commercial sources and used without further purification; solvents were HPLC grade and degassed and dried with activated alumina. Proofing reactions were carried out in

standard glassware, while the production of library intermediates was carried out using Radleys 6 and 12-place reactors. Analytical reverse phase HPLC was performed using an Xterra RPC18 column (3.5 uM, 4.6x50 μm, Waters) on an Alliance 2695 HPLC. Preparative reverse phase HPLC was performed using an YMC ODS-AQ column (20x50 mm, particle size S-5) on a Parallex Flex HPLC System. MALDI-TOF was carried out with the Voyager-EE STR instrument from Applied Biosystems. Mass spectra were also obtained using a Waters ZQ4000 mass spectrometer with an electrospray probe and single quadrupole detector. ^{1}H NMR were recorded using a Varian 400-Mhz spectometer. Chemical shifts were measured in parts per million (δ) relative to tetramethylsilane as the internal standard. Coupling constants were measured in hertz.

General procedures for Preparation of Chemset 12.

3-Benzyl-benzaldehyde 2{1}: Under argon atmosphere: To a mixture of THF (12.5 mL) and aqueous K_2CO_3 (2 M, 5 mL, 10 mmol) were added 3-formylphenylboronic acid (0.50 g, 3.3 mmol, 1.1 eq.), benzyl bromide (0.36 mL, 3 mmol, 1 eq.), and $Pd(PPh_3)_4$ (0.087 g, 0.075 mmol, 0.025 eq.). The reaction was heated to 80°C and monitored by periodic thin layer chromatography (silica, 12:1 hexanes/ethyl acetate, $R_f = 0.3$). Full conversion was reached after 16 h. The reaction was quenched with aqueous HCl (1 M, 50 mL) and the aqueous phase was extracted with ethyl acetate (3 × 30 mL). The combined organic layers were dried using $MgSO_4$ and solvent was removed *in vacuo* giving the crude product. The crude material was purified by flash chromatography [silica gel, hexanes/ethyl acetate (12:1)] to give 0.5 g (80%) of the product. ^{1}H NMR

(400 MHz, $CDCl_3$): δ = 9.98 (s, 1H), δ = 7.723 (m, 1H), δ = 7.459 (d, J = 5.6 Hz, 1H), δ = 7.26 (m, 7H), δ = 4.062 (s, 2H).

3-Benzyl-benzoic acid 3{1}: An aqueous solution of $NaClO_2$ (3.5 M, 4 mL, 14 mmol, 7 eq.) was added dropwise in 1 h to a stirred mixture of 3-benzyl-benzaldehyde (0.39 g, 2.0 mmol, 1 eq.), aqueous NaH_2PO_4 (0.7 M, 7 mL, 4.9 mmol, 2.5 eq.), and 35% H_2O_2 (1 mL, 10 mmol, 5 eq.) in acetonitrile (15 mL), keeping the temperature below 10°C using an ice bath. After the addition was complete, the ice bath was removed and the reaction was monitored by periodic thin layer chromatography following the disappearance of starting material (silica, 12:1 hexanes/ethyl acetate, $R_f = 0$). In two hours, the reaction had proceeded to completion and sodium sulfite (1.8 g, 14 mmol, 7 eq.) was added to quench the reaction. The solution was acidified with aqueous HCl to pH 3 as indicated using pH paper. The organic phase was separated and dried *in vacuo* to afford 0.54 g (75%) of the product. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.95 (m, 2H), δ = 7.41 (m, 2H), δ = 7.25 (m, 5H), δ = 6.65 (bs, 1H), δ = 4.046 (s, 2H); ^{13}C NMR (400 Mhz, d_6-DMSO): 167.4, 141.9, 140.9, 133.5, 133.0, 131.0, 129.6, 129.3, 129.0, 128.8, 128.6, 128.5, 128.4, 127.2, 126.9, 126.0, 40.8; MS calcd for $C_{14}H_{12}O_2$ 212.08, found 213.40; λ_{max} at 316 nm.

2-Benzyl-nitrobenzene 8{1}: Under argon atmosphere: To a mixture of THF (12.5 mL) and aqueous K_2CO_3 (2 M, 5 mL, 10 mmol) were added 2-nitrophenylboronic acid (0.55 g, 3.3 mmol, 1.1 eq.), benzyl bromide (0.36 mL, 3 mmol, 1 eq.), and $Pd(PPh_3)_4$ (0.087 g, 0.075 mmol, 0.025 eq.). The reaction was heated to 80 °C and monitored by periodic thin layer chromatography (silica, 12:1 hexanes/ethyl acetate, $R_f = 0.4$). Full conversion was reached after 16 h. The reaction was quenched with HCl (1 M, 50 mL) and the aqueous phase was extracted with ethyl acetate (3 × 30mL). Solvent was removed *in*

vacuo giving the crude product. The crude material was purified by flash chromatography [silica gel, hexanes/ethyl acetate (12:1)] to give 0.22 g (33%) of the product. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.932 (dd, J = 8,1.4 Hz, 1H), δ = 7.512 (td, J = 7.6, 1.2 Hz, 1H), δ = 7.375 (td, J = 7.6,1.2 Hz, 1H), δ = 7.4 (m, 6H), δ = 4.312 (s, 2H).

2-Benzyl-phenylamine 9{1}: Under hydrogen atmosphere: 10% palladium on carbon (20 mg, 50% wet) was added to a solution of 2-benzyl-nitrobenzene (0.22 g, 1 mmol) in MeOH (15 mL). The reaction was conducted using a Parr apparatus under 30 psi of H_2. Full conversion was reached after 1 h, as indicated by thin layer chromatography (silica, 12:1 hexanes/ethyl acetate, R_f = 0.3). The mixture was filtered and the solvent was removed *in vacuo* to afford 0.16 g (87%) of the product. ^{1}H NMR (400 MHz, $CDCl_3$): δ =7.4 (m, 6H), δ = 6.768 (td, J = 7.6, 1.2 Hz, 1H), δ = 6.678 (d, J = 8 Hz), δ = 3.908 (s, 2H), δ = 3.5 (bs, 2H); ^{13}C NMR (400 Mhz, d_6-DMSO): 146.0, 140.3, 130.1, 128.8, 128.3, 127.0, 125.9, 124.4, 116.6, 115.0, 37.0; MS calcd for $C_{13}H_{13}N$ 183.10, found 184.47; λ_{max} at 212 nm.

4-Amino-3-(4,4,5,5-tetramethyl-[1,3,2]dioxaborolan-2-yl)-benzoic acid methyl ester (5): Under argon atmosphere: To a mixture of methyl 4-amino-3-iodo-benzoate (2.3 g, 8.2 mmol, 1 eq.) in 1,4-dioxane (20 ml), triethylamine (4.6 ml, 33 mmol, 4 eq.) and $PdCl_2$(dppf) (0.30 g, 0.4 mmol, 0.005 eq.), pinacolborane (3.6 ml, 25 mmol, 3 eq.) was added dropwise at rt. The reaction was heated to 80°C and monitored by thin layer chromatography (silica, dichloromethane, R_f = 0.1 - 0.5). Full conversion was reached after 8 h. The reaction was slowly quenched with aqueous saturated NH_4Cl (30 mL) and the aqueous phase was extracted with diethyl ether (7 × 25 mL). After drying over $MgSO_4$, the solution was filtered over a patch of silica. Subsequently the silica was

washed with methylene chloride (1 L). Concentration of the solution in vacuo gave 0.49 g of a mixture of the product and methyl-4-aminobenzoate. ^{1}H NMR (400MHz, $CDCl_3$): δ = 8.310 (d, J = 2 Hz, 1H), δ = 7.888 (d, J = 2.4 Hz, 1H), δ = 6.551 (d, J = 8.8 Hz, 1H), δ = 3.844 (s, 3H), δ = 5.184 (bs, 2H), δ = 1.346 (s, 12H); ^{13}C NMR (400 Mhz, d_6-DMSO): 139.1, 133.7, 130.8, 115.3, 113.9, 113.2, 83.8, 51.2, 24.5; MS calcd for $C_{14}H_{20}BNO_4$ 277.15, found 278.50; λ_{max} at 309 nm.

Methyl 4-amino-3-benzyl-benzoate 6{1}: Under argon atmosphere: To a mixture of THF (8 mL) and aqueous K_2CO_3 (2 M, 1.6 mL, 3.2 mmol) were added crude 4-amino-3-(4,4,5,5-tetramethyl-[1,3,2]dioxaborolan-2-yl)-benzoic acid methyl ester (0.49 g,1.8 mmol, 1 eq.), benzyl bromide (0.40 mL, 3.6 mmol, 2 eq.), and $Pd(PPh_3)_4$ (0.050 g, 0.043 mmol, 0.025 eq.). The reaction was heated to 80°C and monitored by TLC (silica, dichloromethane, R_f = 0.4). The reaction was quenched with aqueous HCl (1 M, 50 mL) and the aqueous phase was extracted with ether (3 × 30 mL). Solvent was removed *in vacuo* from the combined organic layers. The crude material was purified by flash chromatography [silica gel, dichloromethane/hexanes (5:1)] to give 0.1 g (20%) of the product. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.81 (m, 2H), δ = 7.2 (m, 5H), δ = 6.634 (d, J = 8.4 Hz, 1H), δ =3.930 (s, 2H), δ = 3.898 (bs, 2H), δ =3.860 (s, 3H); ^{13}C NMR (400 Mhz, d_6-DMSO): 149.3, 138.5, 132.9, 130.0, 128.8, 128.3, 126.7, 123.7, 119.9, 114,8, 51.6, 38.1; MS calcd for $C_{15}H_{15}NO_2$ 241.11, found 242.50; λ_{max} at 294 nm.

Methyl 4-amino-3-benzyl-benzoate 6{1}: Under argon atmosphere: To a mixture of methyl 4-amino-3-iodo-benzoate (2.27 g, 8.19 mmol, 1 eq.) in 1,4-dioxane (20 mL), triethylamine (4.6 ml, 33 mmol, 4 eq.) and $PdCl_2$(dppf) (0.30 g, 0.4 mmol, 0.005 eq.), pinacolborane (3.6 ml, 25 mmol, 3 eq.) was added dropwise at rt. After 16 hours at 80°C,

benzyl bromide (0.9 mL, 8 mmol, 1 eq.), aqueous K_2CO_3 (2 M, 6.3 mL, 13 mmol), and $Pd(PPh_3)_4$ (0.22 g, 0.19 mmol, 0.02 eq.). After 24 hours, the reaction was quenched with saturated NH_4Cl and the aqueous phase was extracted with ethyl acetate. Solvent was removed *in vacuo* from the combined organic layers. The crude material was purified by flash chromatography [silica gel, dichloromethane/hexanes (5:1)] to give 0.7 g (35%) of the product.

3-Benzyl-4-(3-benzyl-benzoylamino)-benzoic acid methyl ester 10{1,1}: Under argon atmosphere: To a mixture of 3-benzyl-benzoic acid (0.21 g,1 mmol, 1 eq.), DMAP (0.49 g, 4 mmol, 4 eq.), and PS-TsCl (1.0 g, 1.5 mmol, 1.5 eq.) was added methylene chloride (15 mL) and methyl 4-amino-3-benzyl-benzoate (0.24 g, 1 mmol, 1 eq.). The reaction was heated to 40°C and monitored by thin layer chromatography (silica, 10:1 hexanes/ethyl acetate , $R_f = 0.3$). The crude material was filtered and purified by flash chromatography [silica gel, hexanes/ethyl acetate (10:1)] to give 0.39 g (90%) of the product. ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 8.371$ (d, J = 9.2, 1H), $\delta = 8.04$ (m, 2H), $\delta =$ 7.745 (s, 1H), $\delta = 7.4$ (m, 1H), $\delta = 7.2$ (m, 11H), $\delta = 6.994$ (d, J = 6.8, 1H), $\delta = 3.954$ (s, 3H), $\delta = 3.948$ (s, 2H), $\delta = 3.843$ (s, 2H); ^{13}C NMR (400 Mhz, d_6-DMSO): 173.2, 166.7, 166.2, 165.2, 143.0, 141.8, 141.5, 140.7, 140.2, 139.9, 138.6, 138.4, 137.7, 134.9, 134.5, 133.1, 132.7, 132.6, 130.0, 129.7, 129.5, 129.4, 129.3, 129.1, 129.0, 128.84, 128.80, 128.63, 128.59, 128.52, 128.3, 127.4, 127.3, 127.0, 126.6, 126.4, 126.3, 125.8, 124.7, 121.6, 52.2, 52.1, 41.6, 41.4, 38.9, 37.0; MS calcd for $C_{31}H_{31}NO_3$ 465.23, found 466.50; λ_{max} at 238 nm.

3-Benzyl-4-(3-benzyl-benzoylamino)-benzoic acid 11{1,1}: To 3-benzyl-4-(3-benzyl-benzoylamino)-benzoic acid methyl ester (0.35 g, 0.8 mmol, 1 eq.), a 4:1 solution

of THF/MeOH was added until the ester became soluble (30 - 50 mL). Aqueous NaOH (12 mL, 50% w/v) was added to the solution and the reaction was monitored by thin layer chromatography (silica, 20:1 dichloromethane/acetic acid, R_f = 0.2). Full conversion was reached after 72 hours and the reaction was quenched with 6 M HCl to pH 3, as indicated by pH paper. The quenched reaction was extracted with dichloromethane (3 × 30 mL) and the solvent was removed *in vacuo* from the combined organic layers. The crude material was purified by flash chromatography [silica gel, dichloromethane/acetic acid (20:1)] to give 0.30 g (90%) of the product. ^{1}H NMR (400 MHz, d_6-DMSO): δ = 9.884 (s, 1H), 7.774 (d, J = 8.4, 1H), δ = 7.714 (s, 1H), δ = 7.58 (m, 4 H), δ = 7.38 (m, 2 H), δ = 7.2 (m, 8 H), δ = 4.049 (s, 2H), δ = 3.958 (s, 2H); MS calcd for $C_{28}H_{23}NO_3$ 421.17, found 422.48; λ_{max} at 275 nm.

12{1,1,1}: These reactions were carried out in parallel in polypropylene fritted FlexChem 48-well reaction blocks rotating at 700rpm. To a mixture of 3-benzyl-4-(3-benzyl-benzoylamino)-benzoic acid (17 mg, 0.04 mmol, 1 eq.), DMAP (20 mg, 0.16 mmol, 4 eq.), and PS-TsCl (40 mg, 0.06 mmol, 1.5 eq.) was added DMF (2 mL) and 2-benzyl-phenylamine (7.3 mg, 0.04 mmol, 1 eq.). After 72 hours at 40°C the reaction was filtered and purified by reverse phase HPLC (see purification procedure below) to give 14 mg (60%) of the product. ^{1}H NMR (400 MHz, d_6-DMSO): δ = 9.927 (s, 1H), δ = 9.858 (s, 1H), δ = 7.750 (m, 2H), δ = 7.689 (s, 1H), δ = 7.658 (d, J = 6.8, 1H), δ = 7.559 (d, J = 8.0, 1H), δ = 7.3 (m, 21H), δ = 4.099 (s, 1H), δ = 4.011 (s, 2H); ^{13}C NMR (400 Mhz, d_6-DMSO): 165.6, 141.6, 140.9, 140.2, 139.7, 139.1, 136.8, 136.1, 136.0, 134.6, 132.0, 131.9, 130.2, 130.0, 128.73, 128.70, 128.6, 128.5, 128.33, 128.27, 128.0, 127.1,

126.6, 126.2, 126.1, 126.07, 125.9, 125.3, 40.9, 37.1, 37.2; HRMS calcd for $C_{41}H_{34}N_2O_2$ 586.2620, found 587.2708; λ_{max} at 277 nm.

General Procedure for Purification of Chemset 12.

The crude compounds in a solution of DMF were purified with a preparative YMC ODS-AQ column (20 × 50 mm, particle size S-5) running a 5 - 95% gradient of acetonitrile/0.05% trifluoracetic acid with a 20 mL/min flow rate on a Parallex Flex HPLC System. Chromatographs were monitored with a dual wavelength UV detector at 220 and 254 nm. Fraction collection was automatically triggered by UV absorption above 0.05 AU at either wavelength. All fractions eluted with 50% acetonitrile or greater were analyzed with an Xterra RPC18 column (3.5 uM, 4.6 × 50 μm, Waters) running a 0-100% gradient of acetonitrile/0.05% trifluoroacetic acid with a 1 mL/min flow rate on a Alliance 2695 HPLC (Waters). Peaks were integrated at 254 nm using Millenium software (Waters). Samples with 95% purity or greater were further characterized by MALDI-TOF (Voyager-EE STR, Applied Biosystems). Fractions containing the correct product of 95% purity or greater were pooled. Pooled fractions were dried down using a GeneVac Mega 980 solvent evaporator.

General Procedures for Preparation and Purification of Chemsets 11 and 13.

3-Benzyl-4-(3-benzyl-benzoylamino)-benzoic acid methyl ester 10{1,1}: These reactions were carried out in parallel in polypropylene fritted FlexChem 48-well reaction blocks rotating at 700 rpm. To a mixture of 3-benzyl-benzoic acid (0.02 g, 0.1 mmol, 1 eq.), DMAP (50 mg, 0.4 mmol, 4 eq.), and PS-TsCl (100 mg, 0.15 mmol, 1.5 eq.) was

added DMF (2 mL) and methyl 4-amino-3-benzyl-benzoate (0.02 g, 0.1 mmol, 1 eq.). After 48 hours at 40°C, the crude reaction was filtered and purified on a Parallex Flex HPLC System as described above. All fractions eluted with 30% acetonitrile or greater were characterized by LC/MS using an Alliance 2695 HPLC and Waters ZQ4000 mass spectrometer with an electrospray probe and single quadrupole detector operating in positive ion mode. Fractions containing the correct product of 30% purity or greater were pooled and dried down using a GeneVac Mega 980 solvent evaporator.

3-Benzyl-4-(3-benzyl-benzoylamino)-benzoic acid 11{1,1}: These reactions were carried out in parallel in glass test tubes rested in FlexChem 48-well reaction blocks rotating at 300 rpm. To the pooled fractions from the previous reaction, 2 mL of 15:1 solution of dioxane/50% NaOH was added. Full conversion was reached after 72 hours and the reaction was quenched with HCl (6 M, 2 mL, 12 mmol). The quenched reaction was extracted with dichloromethane (2 × 2 mL). The crude material was purified on a Parallex Flex HPLC System as described above. All fractions eluted with 50% acetonitrile or less were characterized by LC/MS using an Alliance 2695 HPLC and Waters ZQ4000 mass spectrometer with an electrospray probe and single quadrupole detector operating in positive ion mode. Fractions containing the correct product of 95% purity or greater were pooled and dried using a GeneVac Mega 980 solvent evaporator.

Appendix 1 contains tables of yield and purity data for all compounds, ^{1}H-NMR, HPLC, and MS for 20 library members.

References

(1) Klarner, C., Greiner, A. Synthesis of polybenzyls by Suzuki Pd-catalyzed crosscoupling of boronic acids and benzyl bromides: Model reactions and polyreactions. *Macromolecular Rapid Communications* **1998**, *19*, 605.

(2) Dalcanale, E., Montanari, F. Selective Oxidation of Aldehydes to Carboxylic Acids with Sodium Chlorite-Hydrogen Peroxide. *Journal of Organic Chemistry* **1986**, *51*, 567.

(3) Seaman, W., Johnson, J.R. Derivatives of phenyboric acid, their preparation and action upon bacteria. *Journal of the American Chemical Society* **1931**, *53*, 711.

(4) Baudoin, O.; Guenard, D.; Gueritte, F. Palladium-catalyzed borylation of ortho-substituted phenyl halides and application to the one-pot synthesis of 2,2 '-disubstituted biphenyls. *Journal of Organic Chemistry* **2000**, *65*, 9268.

(5) Ueki, H.; Ellis, T. K.; Martin, C. H.; Boettiger, T. U.; Bolene, S. B. et al. Improved synthesis of proline-derived Ni(II) complexes of glycine: Versatile chiral equivalents of nucleophilic glycine for general asymmetric synthesis of alpha-amino acids. *Journal of Organic Chemistry* **2003**, *68*, 7104.

Acknowledgements

I would like to thank: Marc Anderson, Leggy Arnold, Mike Cohen, Samantha Cooper, Naoaki Fujii, Sarah Galicia, Tim Geistlinger, Sabina Gerber, Jeremy Mallari, Peter Madrid, Andrea McReynolds, Nick Mills, Jamie Moore, Kathleen Novak, Anang Shelat, John Sherrill, Martina Sigal, Wes Straub, and Hong Yu.

I would also like to thank Kip Guy for his patience, advice, support, and humor.

Chapter 4

Evaluation and Characterization of MDM2 Binding

Felice Lu, Seung-Wook Chi, Do-Hyoung Kim, Kyou-Hoon Han, Irwin D. Kuntz, R. Kiplin Guy

Introduction

p53 is a tumor suppressor protein at the center of a complex network that regulates responses to genotoxic stress. The signaling pathway downstream of p53 involves many gene products which lead to apoptosis or G1/S cell cycle arrest. The route to apoptosis proceeds through p53 binding of Bcl-2, Bcl-XL, Bax, and Bak as well as p53 transcriptional activation of genes such as Bax, Puma, Noxa, Fas, PERP, and DR4/5 (Figure 4.1).[1-4] The primary p53 target in cell cycle arrest is p21, which inhibits transcription of E2F-regulated genes that are required for DNA replication.[5] Investigation of p53 and its downstream targets is ongoing, as components of the p53 pathway are yet unidentified or uncharacterized.

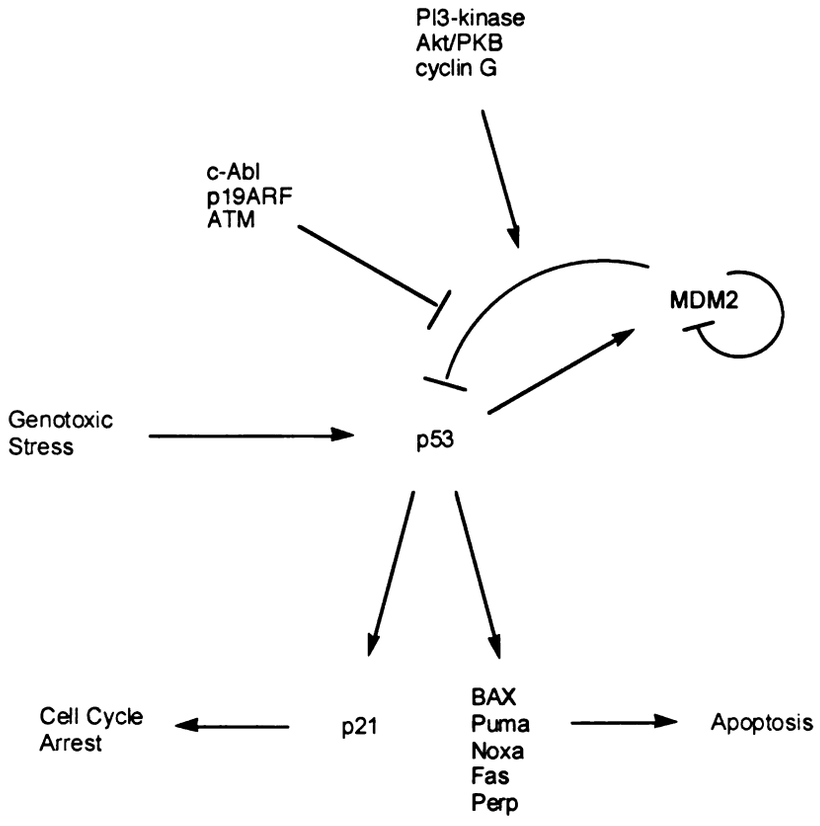


Figure 4.1 A simplified model of the p53 pathway.

The first negative regulator of p53 to be discovered was MDM2.[6] p53 and MDM2 are involved in a regulatory feedback loop which involves p53's transcriptional activation of MDM2 and MDM2's inhibition of p53 activity, resulting in decreased MDM2. MDM2 also acts as an E3 ubiquitin ligase that binds and ubiquitinates p53 promoting its degradation. In addition to self-regulation through self-ubiquitination, MDM2 is also regulated by many other proteins.[7,8] p19ARF interacts with MDM2 and prevents MDM2 shuttling of p53 between the nucleus and the cytoplasm, required for efficient degradation.[9,10] PI3-kinase and Akt phosphorylate MDM2, allowing MDM2 shuttling of p53.[11,12] ATM kinase and c-Abl phosphorylate MDM2 and impairs degradation and nuclear export of p53, while cyclin G dephosphorylates and thus activates MDM2.[13-15]

Many mouse knockouts have been generated to investigate the role of p53 and MDM2, emphasizing the importance of the inhibition of p53 by MDM2. p53-null mice are viable but highly susceptible to tumors,[16] MDM2-null mice are early embryonic lethal, and the MDM2-null/p53-null mice appear normal.[17,18] These experiments suggest that embryo lethality was due to active p53 and that MDM2 is a critical inhibitor of p53 activity *in vivo*. Mice expressing 30% of wild type MDM2 levels are small with low numbers of haematopoietic cells, a p53-dependent phenotype suggesting that MDM2 may have other growth-promoting functions.[19] Hence, a chemical genetics approach to disabling MDM2 later in life when cell differentiation and growth requirements have subsided is important for understanding MDM2's other functions.

In conjunction with the mouse models, *in vitro* experiments have investigated the roles of p53 and MDM2 in tumorigenesis. The addition of p53 cDNA to tumor cell lines

has reduced its tumorigenic potential.[20] In addition, antisense inhibition of MDM2 in tumor cells is associated with decreased p53-MDM2 complex formation, nuclear p53 accumulation, and apoptosis.[21] These two experiments show that decreasing p53 or increasing MDM2 levels promote tumors. A chemical inhibitor of p53-MDM2 complex formation has the potential to allow normal activation of p53 in tumors and thus promote the death and clearance of cancerous cells. In recent years, this has been considered an important therapeutic strategy and MDM2 has become a target at many pharmaceutical companies.

Our interest in MDM2 inhibitors stemmed from the attractiveness of its binding site for the structure-based design of alpha helical protein mimics. After the design and synthesis of compounds libraries described in Chapters 2 and 3, this chapter details the evaluation and characterization of MDM2 binding.

Results

Biochemical Assay

The library was initially tested using a biochemical model for the interaction of p53 and MDM2 based on fluorescence polarization, previously reported by Kay and coworkers.[22,23] This was implemented as a competition experiment using a fluorescently labeled p53 peptide of 19 amino acids in length and a recombinant $(His)_6$-tagged MDM2 protein expressed in *E. coli*. Protein expression proceeded under standard conditions and the purification was moderately problematic, due to large amounts of insoluble MDM2 as well as several impurities requiring purification by anion exchange chromatography. Both factors reduced the protein yield to 0.1 mg/L, which was not problematic for testing

one or two hundred compounds, but create a large obstacle in high-throughput screens. Samples of MDM2 that were not purified by anion exchange chromatography were unsuccessful in tests of direct binding to p53 peptide. The p53 peptide was synthesized on a peptide synthesizer using standard Fmoc chemistry. The peptide was labeled using fluorescein isothiocyanate (Molecular Probes) using low pH conditions provided by the vendor. The low pH conditions prevented the labeling of lysine and allowed selective labeling of the N-terminal amine.

Initial studies were carried out to optimize assay conditions. I first noticed that a small amount of detergent was necessary to form a consistent meniscus in the 384-well plate. Second, the most consistent readings resulted from adding the smallest volume component first, and the largest volume component last. The addition of assay components produces turbulence in the assay wells, a built-in mechanism for mixing. Addition of the largest volume component last results in the best mixing. Finally, I tested two different pH conditions: 7.5 and 8.0. Because these conditions gave similar results, I used pH 8.0 because it was the most convenient.

Binding studies of the fluorescently labeled peptide with purified his-MDM2 (1-222) showed saturable binding with a K_d of 1.6 μM, which is in agreement with literature values (Figure 4.2).[22,23] The validity of the assay was established with several controls: a positive control consisting of nutlin-3, and a negative control consisting of a peptide similar to the p53 probe containing alanine substitutions at 19F, 23W, and 26L. The difference in anisotropy between the negative controls and nutlin-3 was 25-30 units. Samples incubated less than an hour showed a smaller difference in anisotropy. Using

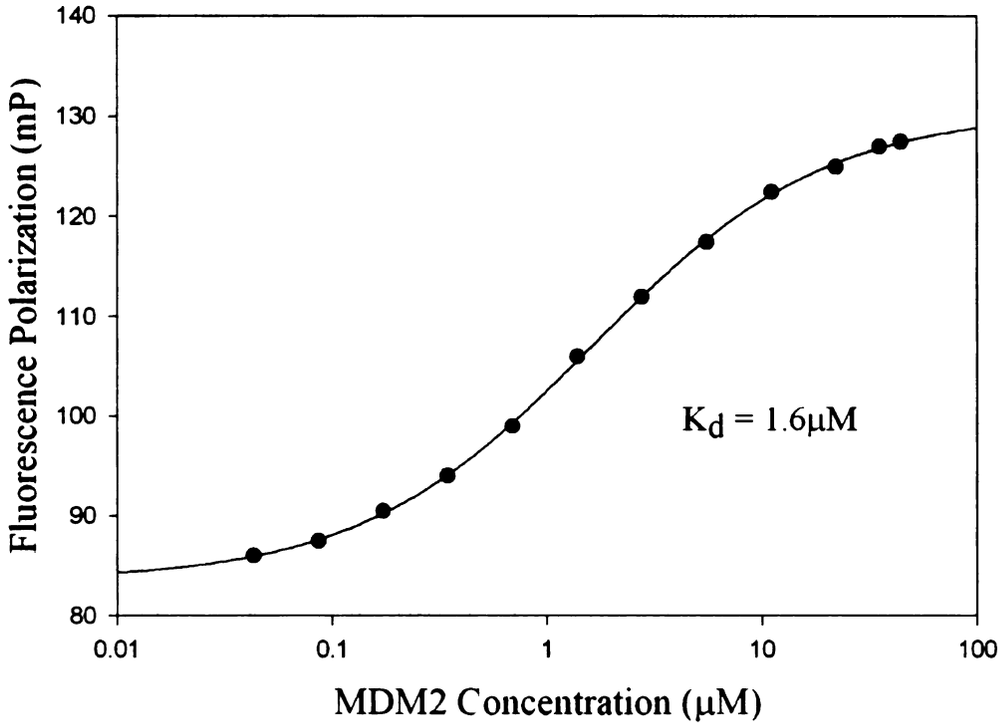


Figure 4.2 Direct binding of p53-FITC to MDM2 (1-222) gives K_d of 1.6 μM and Hill coefficient of 1.12.

this assay, the entire library was screened at a fixed concentration of 30 μM. Compounds showing an inhibitory ability were then subjected to dose response analysis. The binding curves for compounds with a binding constant of 30 μM or less are shown in Figure 4.3. Because three of the thirteen compounds in chemset **11** were active, we synthesized a second library focused on chemset **11**, as well as a few members of chemset **13**{7,6,X}. This second library yielded a plethora of weakly binding compounds with K_d greater than 30 μM (Table 4.1).

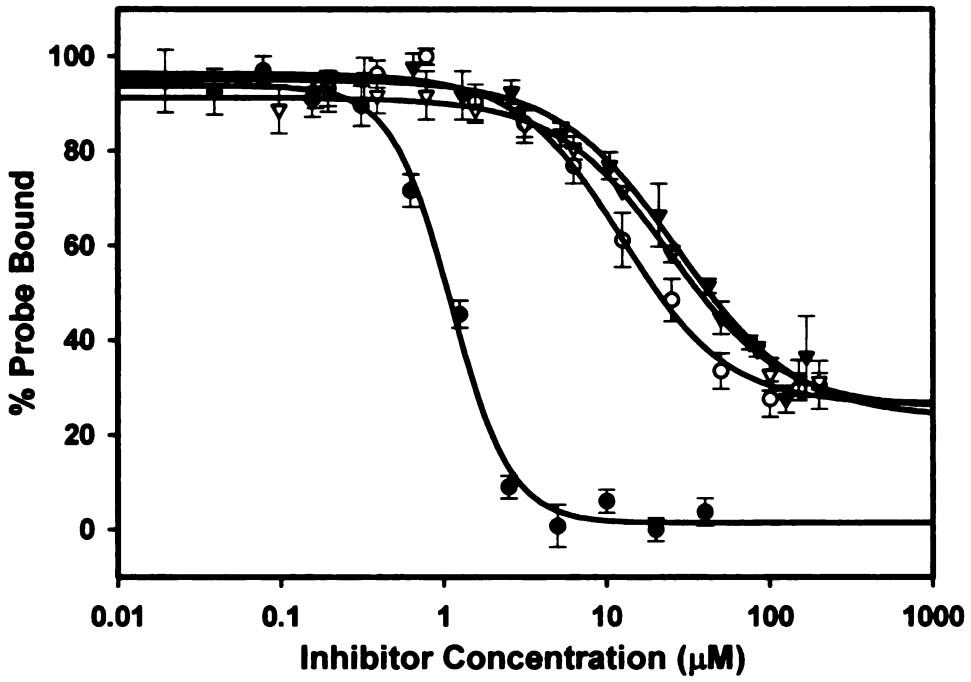


Figure 4.3 Competition of p53-FITC by nutlin-3 (∇, K_d = 1.1 μM), **11**{7,6} (●, K_d = 12 μM), **11**{8,9} (O, K_d = 24 μM), **11**{5,9} (▼, K_d = 27 μM).

Cellular Assay

With positive biochemical results, the next logical step was to test the compounds in cellular assays for activation of the p53 pathway. These experiments were conducted in the laboratory of Gerard Evan with the help of two postdoctoral fellows, Andrew Finch and Abigail Hunt. Finch and Hunt developed a cell line in which MRC-5 human embryonic primary lung fibroblasts immortalized by expression of hTERT were engineered to express $MycER^{TAM}$, generating MRC-5/Myc/TERT (MMT) cells (manuscript in preparation). While robust apoptosis could be induced in control MMT cells, MMT cells overexpressing either MDM2 or the dominant negative C-terminal fragment of p53 (DN-p53) failed to arrest or upregulate $p21^{cip1}$.[24] These 3 cell lines were

well-suited for my studies because the p53 pathway was inhibited by MDM2 and the cells were shown to undergo apoptosis via the p53 pathway.

Compound	K_i (μM)
11{7,6}	12
11{8,9}	24
11{5,9}	27
11{4,4}	>30
11{6,9}	>30
11{6,6}	>30
11{1,6}	>30
13{7,6,7}	>30
11{4,9}	>30
11{8,8}	>30
11{5,11}	>30
11{1,11}	>30
13{7,6,8}	>30
13{7,6,5}	>30
13{7,6,2}	>30
11{5,6}	>30
11{5,7}	>30
11{1,9}	>30
11{5,8}	>30
11{1,8}	>30
11{7,4}	>30

Table 4.1 Active compounds tested by competitive fluorescence polarization.

Initial experiments were conducted with nutlin-3 to determine assay conditions. MMT, MMT/MDM2, and MMT/DN-p53 cells were treated with DMSO, 2 μM nutlin-3, or 10 μM nutlin-3. One hour after treatment, cells were treated with ionizing radiation or doxorubicin while a control set of cells were left undisturbed. All cells were incubated for 4 hours and then harvested, lysed, and probed for p53 and p21 by Western blot. The results showed that the neither the radiation nor the concentration of nutlin-3 affected p53 or p21 levels. Both the MMT and MMT/MDM2 cells treated with nutlin-3 showed

increased p53 and p21 , while the MMT/DN-p53 cells treated with nutlin-3 did not. We concluded that a compound screen would be conducted with MMT control cells that would not be irradiated. Any active compounds would then be tested in all three cell lines at various concentrations with and without ionizing radiation.

In the compound screen, MMT cells were treated with each of the compounds from Table 4.1 (50μM) as well as **10**{7,6} and incubated for 5 hours. After harvest, cell lysates were probed for p53 and p21. The results showed that none of the compounds increased p53 or p21 levels, with the exception of positive control nutlin-3. Compounds **10**{7,6}, **11**{7,6}, and **11**{8,9} were further studied using 10 hour and 24 hour treatment times. However, the p53 and p21 levels remained low and cell morphology did not change. In contrast, immediate cell death was observed under the microscope after the addition of 50 μM nutlin-3. Finally, various lipofectamine reagents were used due to the possibility that the compounds were unable to cross the cellular membrane. These studies also failed to produce increased p53 and p21 levels.

Characterization of Binding by NMR

In order to characterize the binding of **11**{7,6} on mdm2, we performed NMR titration study on ^{15}N-labeled mdm2 (3-109) with the compound. Figure 4.4 shows the chemical shift perturbation in mdm2 residues upon binding of **11**{7,6} and the location of the significantly perturbed residues on mdm2 surface. The most significantly perturbed mdm2 residues are Glu25, Phe55, His73, and Val93 (Figure 4.4A). All of these residues are located within the p53 helix-binding pocket of mdm2, indicating that **11**{7,6} binds to the p53-binding pocket in mdm2 (Figure 4.4B). Therefore, **11**{7,6} is expected to competitively block p53 binding to mdm2. Interestingly, binding of **11**{7,6} and the p53

helix differentially influences the mdm2 residues within the same binding pocket. The **11**{7,6} affects amide proton resonances in the following order, His73 > Phe55 > Glu25 ≅ Val93, whereas in the p53 helix, His73 > Val93 > Tyr100 > Ser22. Therefore, the exact binding modes of the compound and the p53 helix might be slightly different from each other, although both bind to the same helix-binding pocket.

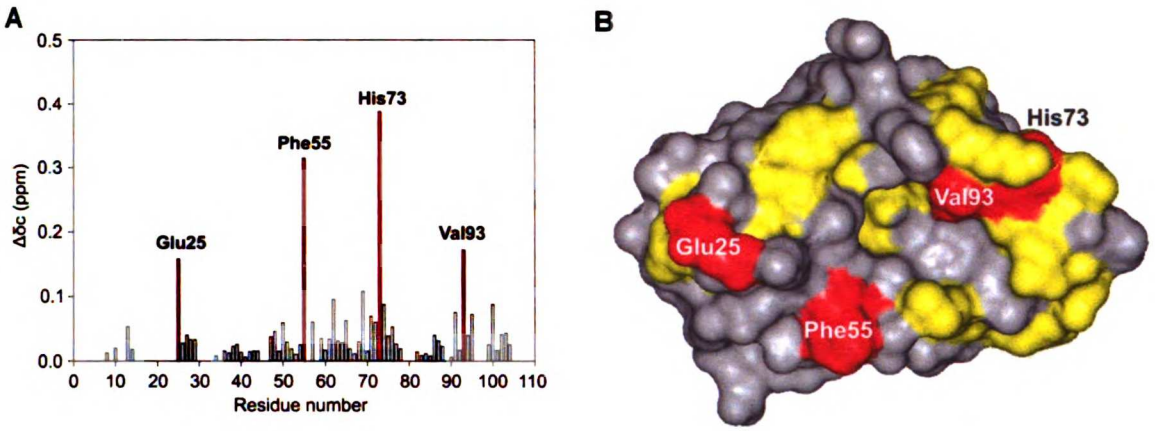


Figure 4.4 (A) Chemical shift perturbation in mdm2 upon binding to **11**{7,6}. The $\Delta\delta_c(^1H, ^{15}N)$ value was calculated as described earlier[25] when the molar ratio of mdm2 to p53 is 1:0.6. (B) Color-coded structures of mdm2 (3-109) showing the sites of major chemical shift perturbation. The residues showing the major chemical shift perturbation are colored on the surface of mdm2 from the crystal structure of mdm2 bound with the p53 peptide[26]. Color-coding is based upon the degree of chemical shift perturbation: gray, $\Delta\delta_c < 0.04$ ppm; yellow, 0.04 ppm < $\Delta\delta_c < 0.15$ ppm; red, $\Delta\delta_c > 0.15$ ppm.

Discussion

This project culminates in assays for binding, cellular activity, and characterization of binding. The first test, involving the ability of inhibitors to complete with p53 peptide in MDM2 binding, was successful. Active compounds were members of chemsets **11** and **13**, suggesting that the carboxylic acid functionality serves in an

important interaction. To further support this hypothesis, the methyl ester analogue, **10**{7,6}, did not have detectable activity. Previous studies have suggested that acidic groups can disturb the salt bridge between Lys51 and Glu25 of MDM2.[25-27] The chemical shift perturbation of Glu25 from the NMR studies indicates that the carboxylic acid group is affecting this salt bridge.

The data also suggests that electron-withdrawing groups are preferred at the para position of the R_2 benzyl ring. Since the R_2 binding pocket is large enough to accommodate a tryptophan group, a meta/para disubstituted benzyl ring in this position may improve binding. The data is insufficient to draw other conclusions regarding the SAR.

The second test, involving the ability of the inhibitors to upregulate the p53 pathway in cells, was unsuccessful. While there are multiple causes that could explain their inactivity, the most probable cause is low cellular uptake. The addition of a 50 μM dose of a foreign chemical, such as nutlin-3 or other inhibitor, often causes cytotoxicity. The additon of nutlin-3 at this concentration resulted in visibly obvious cell death within several minutes, while no major changes occurred after the addition of any of the inhibitors studied. This suggests that the cell is unaffected by the inhibitors and with a logP of approximately 10, there is little chance the inhibitor crossed the cellular membrane.

A high logP was a major drawback of the inhibitor library, whose hydrophobicity posed some solubility problems in addition to the cell permeability issues discussed previously. In the conditions used for the cellular assay, 50 μM with 1% DMSO, crystals of **10**{7,6} were found under the microscope. The high concentrations used in

the competitive inhibition assays (Figure 4.3) is another example where the assay was limited by inhibitor solubility. The solubility of **11**{7,6} was determined to be 89 μM in 1% DMSO and 499 μM in 5% DMSO while the solubility of **10**{7,6} was found to be greater than 68 μM in 1%. Two possibilities exist for reducing hydrophobicity and increasing aqueous solubility: the addition of polar side chains to the solvent-exposed face of the inhibitor, and the substitution of scaffold carbons with heteroatoms. Because cellular permeability is crucial for further biological studies of the inhibitor, increasing hydrophilicity would be my first priority in any future work on this project.

Another important goal is to fully characterize the binding of **11**{7,6} with MDM2. NMR studies indicate that **11**{7,6} binds to MDM2 in the same pocket in which the p53 helix binds, a result that is consistent with the results of the competitive fluorescence polarization experiments. The most significantly shifted MDM2 residues are Glu25, Phe55, His73, and Val93. These four residues have also shifted during the binding of p53 to MDM2.[25] More specifically, the p53 peptide's Phe19 interacts with His73 of MDM2, and Trp23 of p53 interacts with Val93 of MDM2. Phe55 is a solvent-exposed residue engaging in aromatic interactions with chalcones, and most likely with **11**{7,6} as well.[25] These results suggest that **11**{7,6} binds to MDM2 in the Phe19 and Trp23 binding pockets, as the inhibitor design process intended. However, we did not anticipate that a carboxylic acid would play a role in binding, or that the leucine pocket was less important for binding. While these structural studies reveal important information, further structural studies are necessary to compare the predicted *vs* actual binding modes.

Experimental

Recombinant MDM2: Plasmid encoding His-hDM2 (1-222) was kindly donated by R. Tjian.[28] Protein was expressed in BL21(DE3) RAI cells (Stratagene) grown at 37°C to OD_{600} = 0.6 and induced for 3 hours under 1 mM β-D-thiogalactopyranoside. The cells were harvested, sonicated, and centrifuged at 35,000 × g for 15 min. A large fraction of the MDM2 protein was insoluble and in the pellet. The supernatant was loaded onto Ni-NTA resin and eluted with 250 mM imidazole. The yield was 2 mg/L. An SDS-PAGE gel showed a large band near 35 kD, even though the expected molecular weight was 27.5 kD. Coomassie stain and western blot analysis revealed several bands of 2-5 kD lower molecular weight, though to be products of proteolytic degradation. These bands remained despite the use of protease inhibitors aprotinin, leupeptin, pepstatin, and Pefabloc (Roche). When tested by direct binding of the fluorescein-labeled p53 peptide, the K_d was greater than 10 μM. The protein was further purified by ion exchange chromatography using a Source 15Q column (20 mM Tris-HCl, pH 8.0,1 mM DTT, 110 mM NaCl). The use of a smaller 1 mL anion exchange column caused the protein to elute at 1M NaCl. Fractions were characterized by SDS-PAGE and MALDI-TOF mass spectrometry (Voyager-EE STR, Applied Biosystems). The dominant band near 35 kD was identified as His-MDM2 (1-222). The pure protein was concentrated and quantified by Coomassie protein assay. The yield was 0.1 mg/L.

Peptides: The following peptides were synthesized on an Applied Biosystems Model 433A peptide synthesizer using Fmoc chemistry: GSGSSQETFSDLWKLLPEN, FSFSSQETASDLAKLAPEN. TFA cleavage was performed using Reagent K, as

described by Method 3-18, "General TFA Cleavage", in the 2004/5 Novabiochem catalog. The peptides were isolated according to Method 3-29, "Post-Cleavage Work-up", in the 2004/5 Novabiochem catalog. Peptides were purified using the Parallex Flex HPLC System as described in Chapter 3.

p53-FITC: To aqueous $NaHCO_3$ (0.2 M, 0.5 mL, 0.1 mmol, pH 7.0), p53 peptide (2 mg, 1 μmol, 1 eq.) and FITC (Molecular Probes) (50 mM in DMF, 100 μL, 5 μmol, 5 eq.) were added. These reaction conditions are also described in "Amine-Reactive Probes", product information distributed by Molecular Probes. The reaction was monitored on an Alliance 2695 HPLC (Waters) running a 0-100% gradient of acetonitrile/0.05% trifluoroacetic acid with a 1 mL/min flow rate. Because the p53 peptide contains lysine, the low pH and reaction time were important to ensure labeling at a single site. Full conversion was achieved after 4 hours and the crude material was purified using the Parallex Flex HPLC System as described in Chapter 4.

Fluorescence Polarization Assays: Measurements were made with an LJL Biosystems Analyst AD plate reader using a 485 nM excitation filter and a 535 nM emission filter. Assays were performed in Corning 384-well black plates. Pilot experiments demonstrated that the binding of p53-FITC was saturable and our observed K_d agreed well with the reported value of 2 μM. Nutlin-3 (Cayman Chemicals) was used as a positive control, while both DMSO and p53 peptide with alanine substitutions at 19F, 23W, and 26L were used as negative controls. Assays were performed in duplicate and repeated at least twice on separate days with different batches of protein. Competition experiments were carried

out in a total volume of 20 μL 40 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM DTT, 5% DMSO and 0.05% Tween 20. Probe peptide was present at a final concentration of 10 nM and MDM2 was present at a final concentration of 2 μM. Plates were allowed to incubate at room temperature for 1 hour prior to measurement. Data was analyzed with SigmaPlot.

Cell Culture and Immunoblotting: MMT, MMT/MDM2, and MMT/DN-p53 were generously donated by the Evan lab (REF). Cells were grown in DMEM with 10% fetal bovine serum and split when they reached 95% confluence. Twenty-four hours prior to treatment, cells were plated at a density of 7.5×10^4 cells/mL. Compounds were diluted to 50% DMSO and 50% growth medium and added at afinal concentration of 50 μM. After incubation for 5, 10, or 24 hours, media was removed and the cells were frozen. Cells were defrosted on ice and lysed in RIPA buffer (20 mM HEPES pH 7.4, 0.5 mM EDTA, 0.5 mM EGTA, 0.2 M NaCL, 1% IGEPAL, 0.5% sodium deoxycholate, 0.1% SDS) containing 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl chloride, 10 μg/mL aprotinin, and 10 μM E64. Lysates were clarified by centrifugation prior to fractionation by SDS-polyacrylamide gel electrophoresis, electroblot transfer to PVDF membrane (Millipore) and probing with antibodies. Primary antibodies were purchased from Santa Cruz Biotech (p21) and Oncogene (p53). Horseradish peroxidase-conjugated secondary antibodies were purchased from Amersham. Lipofectamine, Oligofectamine, Superfect, DMRIE-C, Lipofectamine with Plus Reagent, and Lipofectamine 2000 were used according to manufacturer's instructions.

NMR Spectroscopy: NMR spectra were acquired using a Varian Unity INOVA 600 spectrometer equipped with a cold probe. NMR sample contained 0.1 mM ^{15}N-labeled MDM2 (amino acid residues 3-109) in 90% H_2O/10% ^2H_2O, 25 mM TrisHCl (pH 7.5), 150 mM NaCl, 2 mM DTT, 0.1 mM PMSF, 0.1 mM EDTA, 0.1 mM benzamidine, and 0.02% NaN_3. Aliquots of **11**{7,6} were added in a stepwise fashion to the ^{15}N-labeled mdm2 (3-109) during titration. The ^{15}N-^{1}H HSQC spectra were collected for the unbound mdm2 (3-109) alone or with **11**{7,6} at 25 °C. The final molar ratio of MDM2 to **11**{7,6} was 1:1. The backbone assignment of the MDM2 (3-109) was previously obtained (accession number 2410, BioMagResBank)[29].

References

(1) Mihara, M.; Erster, S.; Zaika, A.; Petrenko, O.; Chittenden, T. et al. p53 has a direct apoptogenic role at the mitochondria. *Molecular Cell* **2003**, *11*, 577-590.

(2) Chipuk, J. E.; Kuwana, T.; Bouchier-Hayes, L.; Droin, N. M.; Newmeyer, D. et al. Direct activation of Bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. *Science* **2004**, *303*, 1010-1014.

(3) Leu, J. I. J.; Dumont, P.; Hafey, M.; Murphy, M. E.; George, D. L. Mitochondrial p53 activates Bak and causes disruption of a Bak-Mcl1 complex. *Nature Cell Biology* **2004**, *6*, 443-450.

(4) Yu, J.; Zhang, L. The transcriptional targets of p53 in apoptosis control. *Biochemical and Biophysical Research Communications* **2005**, *331*, 851-858.

(5) Eldeiry, W. S.; Tokino, T.; Velculescu, V. E.; Levy, D. B.; Parsons, R. et al. Waf1, a Potential Mediator of P53 Tumor Suppression. *Cell* **1993**, *75*, 817-825.

(6) Momand, J.; Zambetti, G. P.; Olson, D. C.; George, D.; Levine, A. J. The Mdm-2 Oncogene Product Forms a Complex with the P53 Protein and Inhibits P53-Mediated Transactivation. *Cell* **1992**, *69*, 1237-1245.

(7) Fang, S. Y.; Jensen, J. P.; Ludwig, R. L.; Vousden, K. H.; Weissman, A. M. Mdm2 is a RING finger-dependent ubiquitin protein ligase for itself and p53. *Journal of Biological Chemistry* **2000**, *275*, 8945-8951.

(8) Honda, R.; Yasuda, H. Activity of MDM2, a ubiquitin Ligase, toward p53 or itself is dependent on the RING finger domain of the ligase. *Oncogene* **2000**, *19*, 1473-1476.

(9) Tao, W. K.; Levine, A. J. P19(ARF) stabilizes p53 by blocking nucleo-cytoplasmic shuttling of Mdm2. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 6937-6941.

(10) Roth, J.; Dobbelstein, M.; Freedman, D.; Shenk, T.; Levine, A. Nucleo-cytoplasmic shuttling of the hdm2 oncoprotein regulates the levels of p53 protein via a pathway used by the human immunodeficiency virus rev protein. *Gastroenterology* **1998**, *114*, A669-A669.

(11) Ogawara, Y.; Kishishita, S.; Obata, T.; Isazawa, Y.; Suzuki, T. et al. Akt enhances Mdm2-mediated ubiquitination and degradation of p53. *Journal of Biological Chemistry* **2002**, *277*, 21843-21850.

(12) Mayo, L. D.; Donner, D. B. A phosphatidylinositol 3-kinase/Akt pathway promotes translocation of Mdm2 from the cytoplasm to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 11598-11603.

(13) Sionov, R. V.; Moallem, E.; Berger, M.; Kazaz, A.; Gerlitz, O. et al. c-Abl neutralizes the inhibitory effect of Mdm2 on p53. *Journal of Biological Chemistry* **1999**, *274*, 8371-8374.

(14) Khosravi, R.; Maya, R.; Gottlieb, T.; Oren, M.; Shiloh, Y. et al. Rapid ATM-dependent phosphorylation of MDM2 precedes p53 accumulation in response to DNA damage. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 14973-14977.

(15) Okamoto, K.; Li, H. Y.; Jensen, M. R.; Zhang, T. T.; Taya, Y. et al. Cyclin G recruits PP2A to dephosphorylate Mdm2. *Molecular Cell* **2002**, *9*, 761-771.

(16) Donehower, L. A.; Harvey, M.; Slagle, B. L.; McArthur, M. J.; Montgomery, C. A. et al. Mice Deficient for P53 Are Developmentally Normal but Susceptible to Spontaneous Tumors. *Nature* **1992**, *356*, 215-221.

(17) Jones, S. N.; Roe, A. E.; Donehower, L. A.; Bradley, A. Rescue of Embryonic Lethality in Mdm2-Deficient Mice by Absence of P53. *Nature* **1995**, *378*, 206-208.

(18) Luna, R. M. D.; Wagner, D. S.; Lozano, G. Rescue of Early Embryonic Lethality in Mdm2-Deficient Mice by Deletion of P53. *Nature* **1995**, *378*, 203-206.

(19) Mendrysa, S. M.; McElwee, M. K.; Michalowski, J.; O'Leary, K. A.; Young, K. M. et al. mdm2 is critical for inhibition of p53 during lymphopoiesis and the response to ionizing irradiation. *Molecular and Cellular Biology* **2003**, *23*, 462-473.

(20) Casey, G.; Lohsueh, M.; Lopez, M. E.; Vogelstein, B.; Stanbridge, E. J. Growth Suppression of Human Breast-Cancer Cells by the Introduction of a Wild-Type P53 Gene. *Oncogene* **1991**, *6*, 1791-1797.

(21) Chen, L. H.; Agrawal, S.; Zhou, W. Q.; Zhang, R. W.; Chen, J. D. Synergistic activation of p53 by inhibition of MDM2 expression and DNA damage. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 195-200.

(22) Zhang, R. M.; Mayhood, T.; Lipari, P.; Wang, Y. L.; Durkin, J. et al. Fluorescence polarization assay and inhibitor design for MDM2/p53 interaction. *Analytical Biochemistry* **2004**, *331*, 138-146.

(23) Knight, S. M. G.; Umezawa, N.; Lee, H. S.; Gellman, S. H.; Kay, B. K. A fluorescence polarization assay for the identification of inhibitors of the p53-DM2 protein-protein interaction. *Analytical Biochemistry* **2002**, *300*, 230-236.

(24) Shaulian, E.; Zauberman, A.; Ginsberg, D.; Oren, M. Identification of a Minimal Transforming Domain of P53 - Negative Dominance through Abrogation of Sequence-Specific DNA-Binding. *Molecular and Cellular Biology* **1992**, *12*, 5581-5592.

(25) Stoll, R., Renner, C., Hansen, S., Plame, S., Klein, C., Belling, A., Zeslawski, W., Kamionka, M., Rehm, T. Muhlhahn, P., Schumacher, R., Hesse, F., Kaluza, B.,Voelter, W., Engh, R., Holak, T. Chalcone Derivatives Antagonize Interactions between the Human Oncoprotein MDM2 and p53. *Biochemistry* **2001**, *40*, 336-344.

(26) Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J., Pavletich, N.P. Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain. *Science* **1996**, *274*, 948-953.

(27) Kumar, S. K.; Hager, E.; Pettit, C.; Gurulingappa, H.; Davidson, N. E. et al. Design, synthesis, and evaluation of novel boronic-chalcone derivatives as antitumor agents. *Journal of Medicinal Chemistry* **2003**, *46*, 2813-2815.

(28) Thut, C. J.; Goodrich, J. A.; Tjian, R. Repression of p53-mediated transcription by MDM2: a dual mechanism. *Genes & Development* **1997**, *11*, 1974-1986.

(29) Stoll, R.; Renner, C.; Muhlhahn, P.; Hansen, S.; Schumacher, R. et al. Sequence-specific H-1, N-15, and C-13 assignment of the N-terminal domain of the human oncoprotein MDM2 that binds to p53. *Journal of Biomolecular Nmr* **2000**, *17*, 91-92.

Acknowledgements

I would like to thank Nathan Sallee, Abigail Hunt, and Andy Finch, who generously shared their lab bench and time with me so that I could learn and perform the assays.

Chapter 5

Small Molecule Affinity Fingerprinting: A Tool for Enzyme Family Subclassification, Target Identification, and Inhibitor Design

Doron C. Greenbaum, William D. Arnold, Felice Lu, Linda Hayrapetian, Amos Baruch, Jennifer Krumrine, Samuel Toba, Kareem Chehade, Dieter Brömme, Irwin D. Kuntz and Matthew Bogyo

Introduction

The recent genomics revolution has provided us with the first low-resolution roadmap of the human genome. However, the true challenge lies in using this raw sequence information to create a better understanding of the role of specific gene products in both normal and disease processes. Functional genomics efforts have begun to address this challenge using sequence-alignment algorithms and transcriptional profiling as a way to link biological functions to specific genes and gene products [1]. Indeed, this process has lead to the annotation of a substantial number of enzyme and protein families. In many cases, these families will serve as a starting point in the process of target selection for the development of preclinical drug candidates. However, many protein families are populated with dozens of closely related members. For example, the protease family alone comprises 1%–2% of the human genome and represents over 500 enzymes grouped within only a few distinct subfamilies. Therefore, potential drug targets such as these must be viewed not as single entities but as members of closely related protein networks. Therapeutic design must focus not only on issues of potency toward a single target but also, and often more importantly, on selectivity within the context of a target's nearest functional relatives.

Traditionally, the problem of specificity has been addressed using medicinal chemistry to generate compounds that have been optimized for a single protein target. Correlation of structural elements of small molecule leads with their inhibition potencies is used to generate structure activity relationships (SARs). These data can be used to rank individual compounds and ultimately to sort out the best candidates for further development. To aid in this process, several groups have developed complementary *in*

silico methods to define molecular similarity among a class of protein targets [2 and 3]. Additionally, computational methods have been developed that allow small molecule binding to be addressed by virtual docking to a protein active site [4, 5 and 6]. From these computational SAR studies, a set of physicochemical descriptors can be generated that define the binding properties of many related small molecule inhibitors. Ultimately, such computational approaches allow a large number of theoretical compounds to be virtually assayed prior to embarking on costly and time consuming medicinal chemistry efforts.

In addition to providing a starting point for lead optimization, SAR data also provide information that can be used to generally define the topology of the small molecule binding pocket of a target protein. Furthermore, compilation of SAR data obtained from chemical library screening against a set of proteins provides affinity fingerprints for each target. As an increasing number of diverse compounds are assayed against these targets, the fingerprints that are generated become more refined. If these fingerprints become sufficiently unique, they can be used to establish subtle differences among members of a large protein family with a high degree of sequence homology.

Several methods for protein classification based on affinity fingerprints have been proposed. One such method relies upon a training set of inhibitors that is screened against a panel of disparate proteins to predict affinity fingerprints for other nonrelated proteins. Ultimately, this method could be used to allow chemists to quickly predict pharmacophores within a chemical library that will serve as lead compounds for further development [7 and 8]. Yet another classification method has introduced structure activity relationship homologies (SARAH) as a means to cluster proteins within a family. The kinase family of enzymes was used to highlight the utility of inhibitor fingerprinting

as a rapid classification method for members of this large family of highly related proteins [9]. Once a functional classification is established based on SARAH, it becomes possible to group newly sequenced kinases into chemical subgroups to optimize the drug-screening process. Furthermore, this method of classification provides critical information concerning the "nearest neighbors" in the family that are likely to be of concern when trying to design a selective small molecule drug.

Here, we outline a combined chemical- and computational-based approach to generate and analyze affinity fingerprints for the papain family of cysteine proteases. An affinity labeling methodology has been employed to assess the inhibitory characteristics of a set of small molecule libraries toward this panel of closely related protease targets. This resulting inhibition data set is a compilation of affinity fingerprints for the set of purified targets and was used as a method to classify individual family members. In addition, the identity of proteases from crude cellular lysates could be determined by clustering affinity fingerprints of "unknown" targets with the data set of purified targets. A computational protocol was then developed and used to generate predictions for cysteine proteases based on experimentally determined crystal structures. Ultimately, this method could aid the process of development of small molecule inhibitors for families of related targets when only limited structural and functional information is available.

Results and Discussion

We have previously described a set of positional scanning libraries (PSLs) based on the epoxide electrophile scaffold found in the natural product E-64 [10 and 11]. This scaffold can be used to generate compounds that are mechanism-based irreversible

inhibitors of the papain family of cysteine proteases [12]. The compounds in these libraries are made up of a primary tripeptide backbone linked to a reactive epoxide electrophile. The amino acids found adjacent to the epoxide moiety are expected to occupy the S2–S4 binding pockets of the protease targets (termed the P2, P3, and P4 amino acids; Figure 5.1). The S2 pocket has been shown to be the primary site of substrate discrimination for this family of proteases [13].

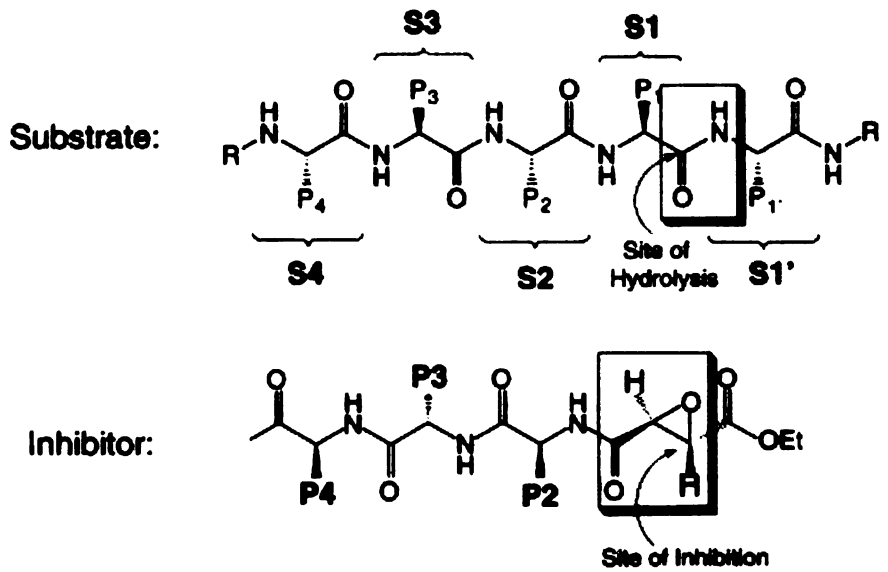


Figure 5.1 Comparison of binding mode of peptidyl epoxide inhibitors and peptide substrates. Peptidyl epoxides bind to cysteine protease active sites in a manner analogous to a peptide substrate. The three amino acid side chains adjacent to the epoxide, termed the P2, P3, and P4 residues, align in the active site such that they occupy the S2, S3, and S4 binding pockets. Note that no side chain fills the S1 pocket due to the structure of the epoxide building block.

Initially, three sets of PSLs were synthesized by fixing each of the P2, P3, and P4 positions with each of the 20 possible natural amino acids (minus cysteine and methionine, plus norleucine as a mimetic of methionine). A mixture of the same natural amino acids was used in the remaining two amino acid positions, resulting in 19 P2, P3, and P4 sublibraries, with each made up of a mixture of 361 compounds.

The three sets of PSLs were assayed against purified protease targets by competition with the radiolabeled active-site-directed probe ^{125}I-DCG-04. Samples were analyzed by SDS-PAGE followed by phosphorimaging to determine the intensity of labeled bands using a commercial software package (Figure 5.2). Competition (i.e., loss of labeling) was indicative of inhibition by the unlabeled library member. Competition assays are performed by preincubation of protease targets with inhibitor libraries followed by labeling with the general probe. Since the extent of inhibition by the inhibitor libraries is a function of preincubation and labeling times, these parameters had to be carefully controlled, and assays were performed in triplicate to confirm the run-to-run reproducibility of the assay. Furthermore, for this method to provide a valid readout, final concentration of inhibitors (10–50 μM) must be held in excess over concentrations of the target protease (100–300 nM) throughout the assay. Using this method it was possible to determine a percent competition for each fixed position library by determining the ratio of intensity of labeled bands in the treated samples to the intensity of the untreated control. These data were subsequently used to generate affinity fingerprints.

Covalent irreversible inhibitors such as the peptide epoxides function mechanistically through a two-step process involving an initial reversible binding event (measured as an equilibrium constant, K_i) followed by an irreversible alkylation step (measured as rate constant kinact). Potency values for this class of inhibitors are expressed as a ratio of the K_i/k_{inact}. Detailed kinetic studies of the peptide epoxides have shown that the rates of inactivation (k_{inact}) remain relatively constant across structurally diverse inhibitor scaffolds [14]. As a result, competition data obtained for libraries of peptide epoxides provide mainly information that relates to the relative K_i values of an

inhibitor for a given target. Furthermore, any small molecule that binds in the active site of a target will block the reversible binding step of the probe and will lead to loss of labeling (competition). Therefore, this method is suitable for screening of both reversible and irreversible inhibitors. In fact, similar screens with libraries of reversible cysteine protease inhibitors have been carried out for the parasitic protease target cruzain. These competition results were found to closely correlate with kinetic inhibition values obtained by standard substrate-based methods (D.C.G., M.B., and J. Ellman, unpublished results).

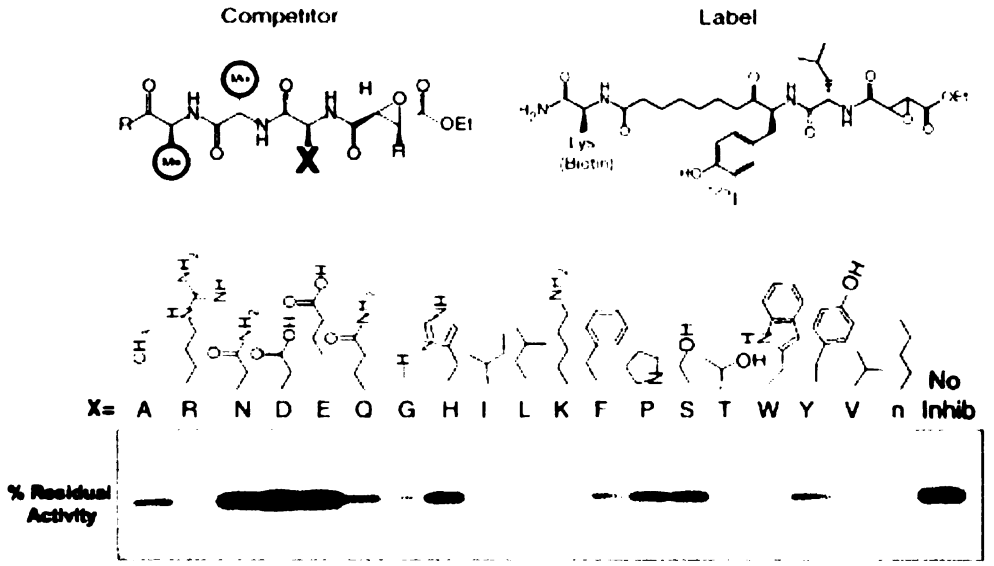


Figure 5.2 Methods for generating affinity fingerprints. Example of an affinity fingerprint generated by screening of a P2 diverse peptide epoxide library. Purified cathepsin K was pretreated with individual constant P2 sublibraries (X position on competitor) followed by labeling with ^{125}I-DCG-04 (label). Samples were separated on a 12.5% SDS-PAGE gel and visualized by PhosphoImaging (Molecular Dynamics). Labeling intensity of each target relative to the control untreated sample was used to generate percent competition values. This method was used to generate competition values for multiple enzymes and for libraries with diversity at the P2, P3, and P4 positions on the inhibitor scaffold.

While substrate-based kinetic assays provide for high-throughput screening of targets, the competition-based method can be multiplexed to accommodate multiple targets in a single gel-based assay. Additionally, this screening method allows for rapid

analysis of multiple related targets without the need to optimize substrate and kinetic conditions for each enzyme. Finally, the competition screen allows separation of the target from the substrates and small molecules in the screen, thereby eliminating problems of insoluble and intrinsically fluorescent compounds that can hinder an absorbance-based detection method. To increase the assay throughput, we have also designed a dot-blot-based readout for competition. In the case where a single protein target is screened, filtering of samples through a PVDF membrane provides a method to isolate and measure the amount of labeled target protein. This assay method circumvents the need for SDS-PAGE gels and allows the assay to be performed in a 96-well plate format (data not shown).

This affinity-probe-based method of screening of PSLs has been validated by our laboratories in a representative crude proteome [11] and for a specific protease target [15]. These studies show that it is possible to use this screening method to rapidly identify selective inhibitors of protease targets. It was therefore of interest to apply the same set of PSLs to profiling the specificity of an expanded set of papain family enzymes. While this family of proteolytic enzymes has been extensively studied, most inhibitor SAR studies have been focused on a limited number of compounds screened against a small set of family members. It was therefore of interest to determine if a large data set could be used to classify this set of proteases into distinct subfamilies based on substrate/inhibitor binding.

PSLs were screened against a set of purified and recombinant papain family cysteine proteases that were obtained from commercial and public sources. To aid in the analysis of the data, numerical competition values were visualized by conversion to a

color format using software developed by Eisen and coworkers designed for data generated from microarray analysis [16]. This software assigns colors based on the numerical competition values in the range from 0%–100%. Compounds that were potent inhibitors (i.e., 100% competition) were assigned a red (hot) color, while compounds that were weak inhibitors, showing little or no competition, were assigned a blue (cold) color. Compounds with intermediate activities were assigned lighter shades of red and blue, with white assigned to compounds with 50% inhibition. Furthermore, hierarchical clustering software was used to group the data based on similarities among profiles of enzymes (y axis) or small molecules (x axis).

Cluster analysis of inhibition data from each of the P2, P3, and P4 library sets against 12 papain family proteases revealed patterns of specificity for each of the three primary substrate binding pockets (Figure 5.3). The resulting specificity data agreed with previously reported findings identifying the P2 position as the primary site for enzyme-substrate interactions [13]. Furthermore, the S2 pocket of the papain family enzymes preferred many of the hydrophobic and aromatic amino acids, suggesting the need for a more diverse set of hydrophobic P2 residues in order to obtain distinct binding profiles for this class of enzymes.

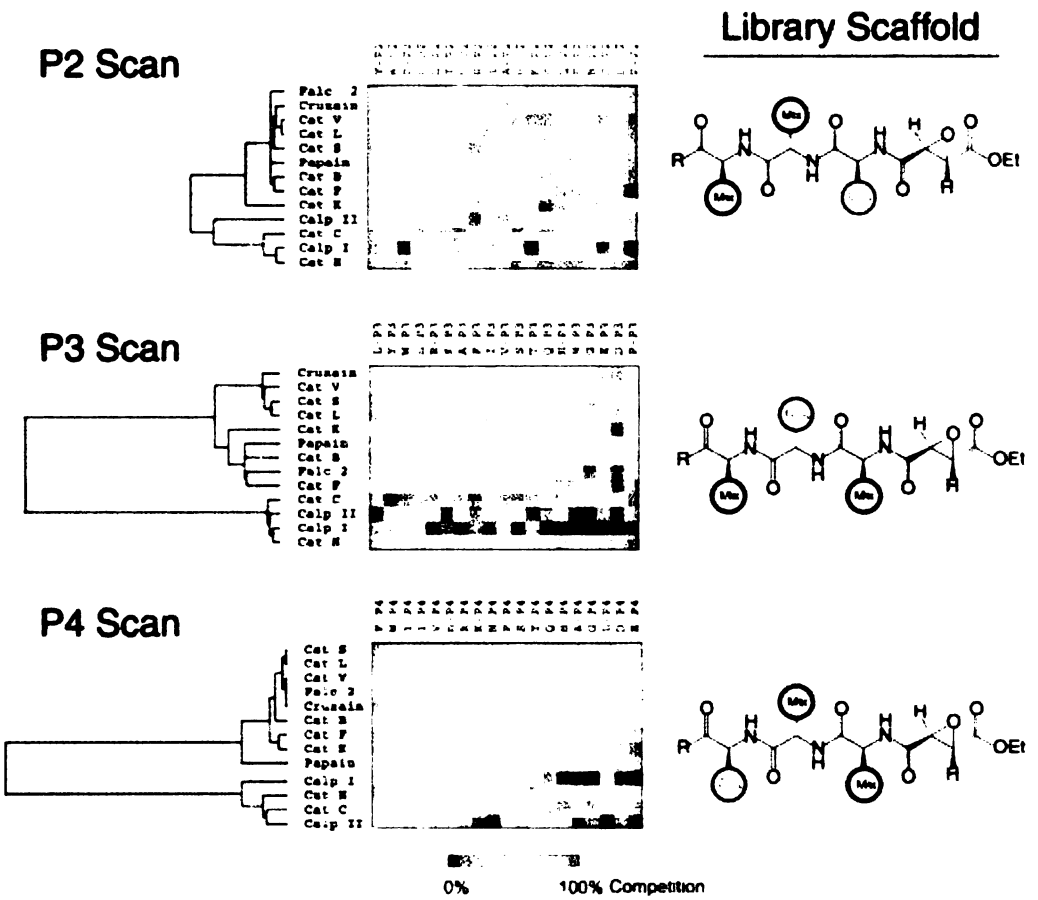


Figure 5.3 Cluster analysis of affinity fingerprints for a set of papain family proteases: subsite specificities within the active sites. Inhibition data from screening of P2, P3, and P4 diverse inhibitor libraries (scaffold structures indicated on right of data panels). Sublibraries were composed of a single constant amino acid position that was varied through all natural amino acids (C_{1-19}) and two variable positions composed of a mixture of all 19 amino acids (mix). Competition data were obtained as describe in Figure 5.2 and were clustered and visualized using programs designed for analysis of microarray data (see Experimental Procedures). Colors indicate the potency of a sublibrary with the indicated fixed amino acid for a designated target protease. Potent (hot) inhibitors are assigned a red color, and weak or ineffective (cold) inhibitors are assigned a blue color. Target enzymes are arrayed along the *y* axis, and each of the constant amino acids is arrayed along the *x* axis. The tree structures at the left of the diagrams were obtained by hierarchical clustering and indicate the degree of similarity of enzymes as a function of the height of the lines connecting profiles. The color key is shown at the bottom. Amino acids are indicated by their single-letter code, with n used for norleucine.

A set of 41 hydrophobic nonnatural amino acids was selected and used to generate a nonnatural P2 library. For this extended P2 library, each of the 41 nonnatural amino acids was held constant in the P2 position, while the P3 and P4 positions were composed of a mix of all possible natural amino acids. The mixture method was chosen rather than using general favorable binding P2 and P3 amino acids because this resulted in sublibraries that had greater overall utility for screening. These libraries were not biased in the P3 and P4 positions and therefore could be used to assay the contribution of the P2 element for virtually any cysteine protease target. In order to further increase the diversity of compounds for affinity fingerprinting, a second set of libraries was synthesized using the complete set of natural amino acid building blocks in the P2 position attached to the enantiomeric form of the epoxide electrophile (2R, 3R, versus 2S, 3S; Figure 5.4). Previous work has shown that this change in stereochemistry is likely to favor binding of the inhibitors in the prime side of the active site, thus increasing the potential for finding binding pockets unique to each papain family protease [17].

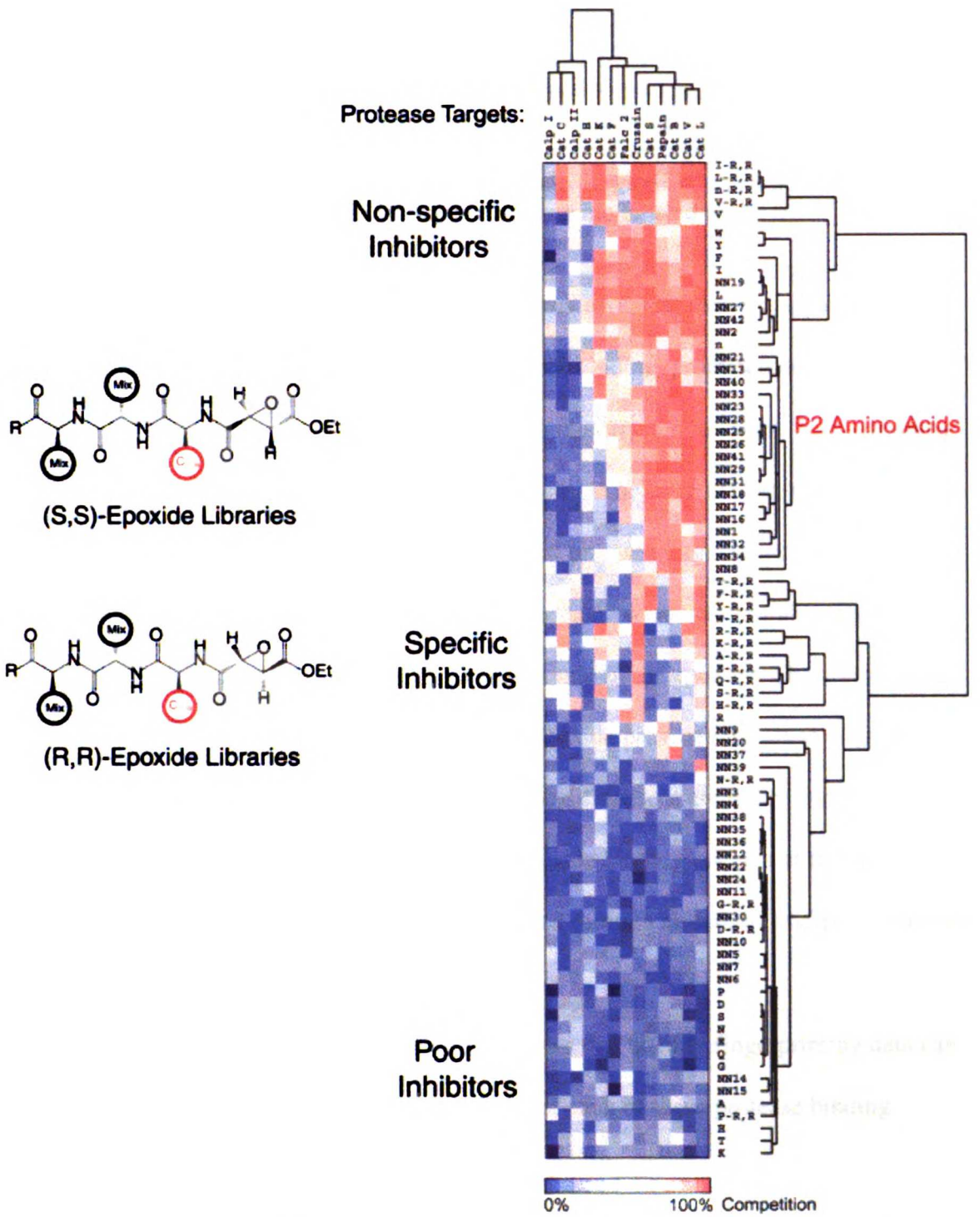


Figure 5.4 Cluster analysis of an extended P2 diversity library. A large set of P2 amino acids including the 19 natural amino acids and 41 nonnatural hydrophobic amino acids was selected and used to generate an extended P2 inhibitor library. In addition to the set of 60 natural and nonnatural amino acids coupled to the epoxide moiety containing the (S,S) stereochemistry, the natural 19 amino acids were coupled to the enantiomeric form of the epoxide (R,R isomer; see structures at left).

The resulting 79 sublibraries were assayed against the reference set of 12 papain family protease as described in Figure 5.2 and Figure 5.3. Single-letter codes were used for natural amino acids, with n being assigned to norleucine. The 41 nonnatural amino acids were assigned arbitrary numbers (1–41) and listed with the NN prefix. Libraries containing the R,R enantiomer of the epoxide are listed with "R,R" following the single-letter amino acid code. Regions of weak binding, nonselective strong binding, and selective binding are labeled at the left.

The clustering of the extended P2 library data revealed underlying patterns of inhibition by grouping compounds with overall poor binding, promiscuous binding, or selective binding together (see annotation at left of clustergram in Figure 5.4). Grouping the data in this manner immediately identified P2 amino acids in the central region of the clustergram that conferred specificity for individual protease targets. Interestingly, the bulk of the amino acids found in this "specificity region" were nonnatural amino acids and natural amino acids linked to the (R,R) enantiomer of the epoxide. These results suggest that changing the stereochemistry of the epoxide provided access to different binding sites in the protease active-site cleft. These differences are likely due to interactions of the R,R compounds with the prime-side binding pockets of the papain family proteases. This hypothesis will be confirmed through structural studies of inhibitor binding and will be the focus of future work.

This clustering methodology therefore shows that affinity fingerprinting data can be used to reveal information about the topology of each of these protease binding pockets. Ultimately a screen of a larger, more structurally diverse small molecule library is likely to provide a higher-resolution image of these inhibitor/enzyme interactions.

Another powerful application of this affinity-fingerprinting methodology is its ability to classify an unknown protease activity from a crude cell or tissue lysate by clustering its affinity fingerprint within a database of standard protease fingerprints. We

have previously demonstrated the utility of activity-based probes as a means to profile cysteine protease activities within intact cells or crude cell lysates. This technology therefore allowed the extended P2 inhibitor library to be screened against several cysteine proteases in a crude cell extract [11].

The rat liver proteome was chosen for initial studies due to its high content of proteolytic enzymes and because the major protease activities in this sample were previously identified by purification and sequencing [11]. Total protein extracts were probed for cysteine protease activity using ^{125}I-DCG-04 (Figure 5A). Four major protease activities were observed by affinity labeling and SDS-PAGE analysis (Figure 5B). This profile exactly matched the results reported by our laboratory in an earlier publication [11], indicating that the labeling method is highly reproducible.

Affinity fingerprints were generated for each protease activity by pretreatment of extracts with inhibitor PSL sublibraries followed by affinity labeling. The resulting data sets were clustered with the database of extended P2 cysteine protease inhibition fingerprints (Figure 5.5C, black boxes). Protease band 2 clustered into a small subgroup of cathepsin proteases, with the greatest similarity to cathepsin B. The identity of this band was confirmed to be cathepsin B by isolation and sequencing by mass spectrometry [11]. Protease bands 3 and 4 had identical fingerprints and clustered together in the cluster tree as a distinct branch, which included cathepsin H. Again, this cluster-based assignment of bands 3 and 4 was confirmed by purification, sequencing, and identification of these two bands as differentially processed forms of cathepsin H [11]. Protease band 1, unlike the other proteases, clustered into its own branch and had no direct counterpart in the database. This protease activity was identified as cathepsin Z

[11], an enzyme that was not fingerprinted and therefore had no reference points in the database. Thus, the clustering method was able to predict the identity of enzyme activities within a crude tissue lysate by virtue of their unique affinity fingerprints.

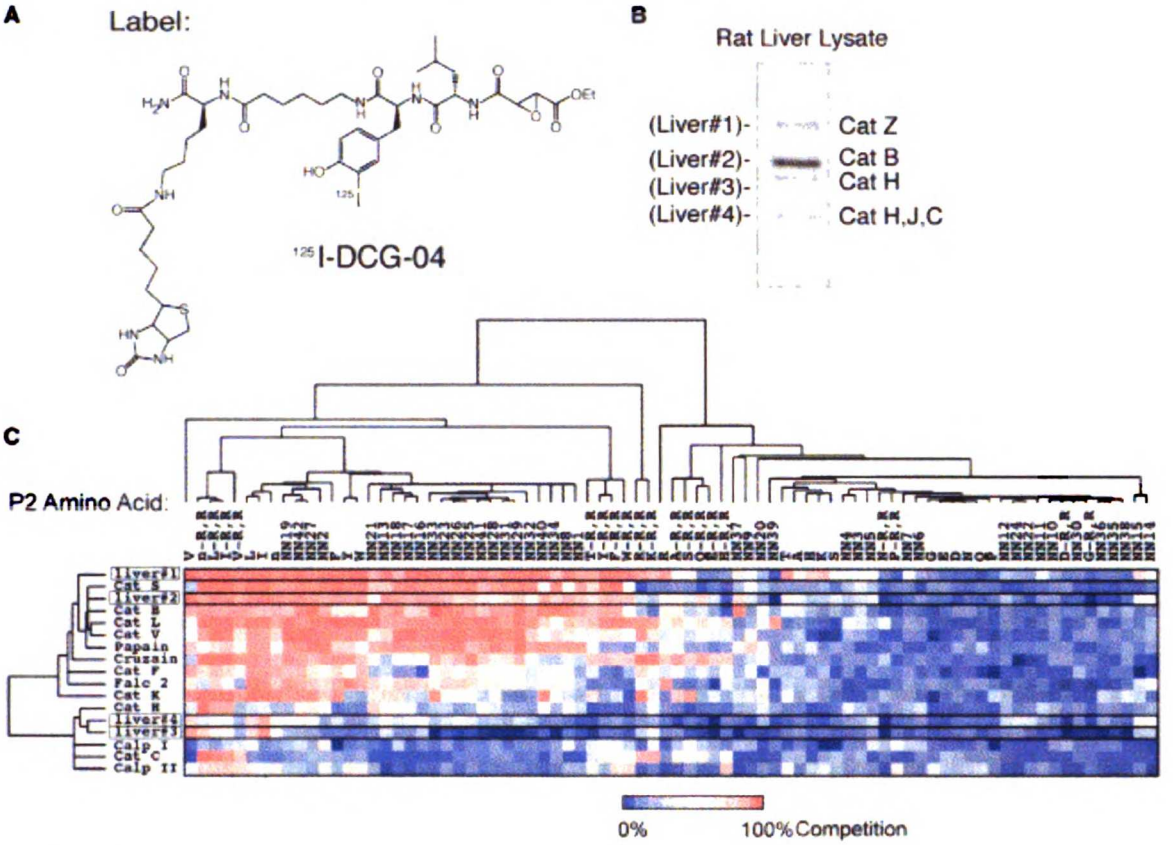


Figure 5.5 Identifying unknown proteases' targets using fingerprint clustering. (A) Structure of the activity based probe, ^{125}I-DCG-04. (B) Profile of active papain family cysteine proteases in crude rat liver homogenates. "Unknown" proteases are labeled 1–4 (liver #1–liver #4 at left). The true identity of each protease was determined by mass-spectromety-based sequencing and is listed for reference at right. (C) Competition data obtained by treatment of crude homogenates with the extended P2 sublibraries described in Figure 5.4 followed by labeling with the probe. Competition data for each unknown (see black boxes) were added to the reference protease data, and the complete data set was clustered as described in Figure 5.3 and Figure 5.4. Identity of the unknown proteases could be inferred by inspection of the closest neighbors in the vertical dendrogram shown on the left.

The results from this experiment highlight several strengths of combined inhibitor screening and clustering technology. First, the inhibitor libraries allow screening against cysteine proteases present in a crude cell and tissue proteome. The ability to use crude protein extracts, rather than recombinant or purified protein, greatly reduces the effort required to screen large inhibitor libraries and allows rapid lead identification for endogenously expressed enzymes. Second, the tight clustering of endogenous cathepsins with their recombinant counterparts suggests that this methodology could be used for rapid, crude characterization of unknown enzymes from complex protein samples without absolute knowledge of their identity.

In addition to being useful for optimization of small molecule inhibitors, clustergrams of affinity fingerprints also yield functional information about the topology of the active site of the protein. The dendrogram that results from clustering of the library data using the programs Cluster and TreeView [16] pictorially describes the relationships amongst individual proteases. This dendrogram is analogous to homology trees that are generated through sequence alignments. However, it provides inhibitor-generated functional alignments, in contrast to traditional sequence alignments based on linear amino acid relationships.

For comparison, a dendrogram of proteases was generated using the sequence alignment program Clustal W and compared against the affinity-fingerprint alignment. The two dendrograms have overall similarities but upon closer inspection reveal significant differences (Figure 5.6). For example, cathepsin B and cathepsin C cluster together based on primary sequence alignments. Although these are both exoproteases, cathepsin B is a carboxypeptidase while cathepsin H is an aminopeptidase, and their true

functions are highly divergent. The fingerprint clustering yields a more satisfying picture of the large functional difference between cathepsin B and C (Figure 5.6, red labels). On the other hand, sequence alignment of cathepsin K clusters it within a subfamily with cathepsins S, V, and L. However, affinity-fingerprint clustering identified cathepsin F as its closest neighbor and, therefore, the major concern for efforts to design cathepsin K selective inhibitors (Figure 5.6, green labels). Furthermore, the fingerprint clustering identified cathepsins K, F, and H as the best candidates in this family of proteases for design of selective inhibitors due to the uniqueness of their specificity profiles (i.e., distinct branches in the clustering tree). Such information may also help to prioritize targets in large protein families based on the chances for successful development of selective inhibitors.

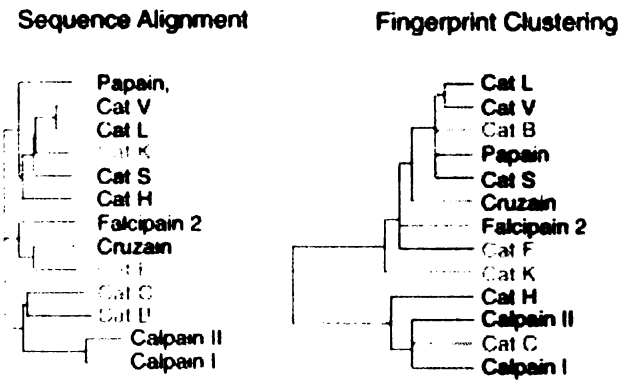


Figure 5.6 Comparison of fingerprint clustering and sequence alignment-based clustering. Hierarchical clustering of affinity fingerprints for the 12 reference cysteine proteases produced dendrogram trees that indicate the degree of functional similarity between enzymes as a function of the height of the lines connecting profiles. A dendrogram tree generated using affinity fingerprints was compared to a tree generated by primary sequence alignment using Clustal W, as described in the Experimental Procedures section. Examples of enzymes with divergent clustering based on sequence alignment but with similarities in affinity fingerprints are shown in green, while enzymes that show similar sequence alignment but dramatic differences in classification based on affinity fingerprinting are shown in red.

The affinity fingerprints generated for a control set of cysteine proteases was also used to tailor the design of a computational protocol for generating in silico fingerprints based on structural data. A molecular docking scheme [18], which had proven successful for the design of both peptidic and nonpeptidic inhibitors in a series of serine proteases, was unable to distinguish specificity in the lysosomal cysteine proteases. We found that the covalent linkage between the inhibitor and the enzyme necessitated a complete molecular mechanical forcefield for proper inhibitor placement. The DOCK program, however, employs only the intermolecular van der Waals and coulombic terms as an energy scoring function. We therefore combined docking with molecular dynamics (MD) to develop a new strategy in the spirit of the MMPBSA (molecular mechanics Poisson-Bolzmann surface area) approach [19]. Relative binding free energies can be derived from MD trajectories using the theories of statistical thermodynamics. In this case, however, a simulation of each inhibitor for each enzyme would require over a hundred individual MD runs. In order to make the problem computationally tractable, we performed MD just once for each enzyme, using only the common portion of each inhibitor. Benzyl groups served as "dummy" side chains at the P2-P4 scaffold positions during these simulations and acted as placeholders in the enzyme pockets. Following the dynamics runs, full side chains at the P2 position were added in an incremental fashion and rank ordered according to the DOCK energy score [20]. The top 20 conformations of each side chain were then minimized in AMBER [21] and rescored using a PBSA solvation model [19]. Since the scaffold and enzyme conformational degrees of freedom were sampled during the dynamics runs, the resulting coordinates were preserved in subsequent steps. The side chain degrees of freedom were sampled using the less

expensive incremental growth and energy minimization routines. Because we did not carry forth the thermodynamic ensemble of structures derived from the MD simulation, the results cannot be considered as time averaged free energies of binding. Although there is no physically rigorous way to isolate individual members of an MD ensemble for docking, we chose the member closest to a corresponding X-ray structure [22, 23, 24, 25, 26 and 27], which itself is part of a larger, physical ensemble.

The predictions derived from the six enzymes considered are in good qualitative agreement with the experimental data (Figure 5.7). Overall, the computational results accurately predict the general nature of favorable S2 sidechains for each enzyme. The computational results also agree with some of the fine discrimination seen between enzymes experimentally. Tryptophan, for example, is predicted to be a poor P2 sidechain for cathepsin K, and arginine is predicted to be poor for both cathepsin K and cathepsin S. These results demonstrate that qualitatively accurate results can be derived by DOCKing sidechains onto one member of an MD ensemble. It is reasonable to expect that individual predictions would improve as we averaged the docking results of more members of the scaffold-enzyme MD ensemble.

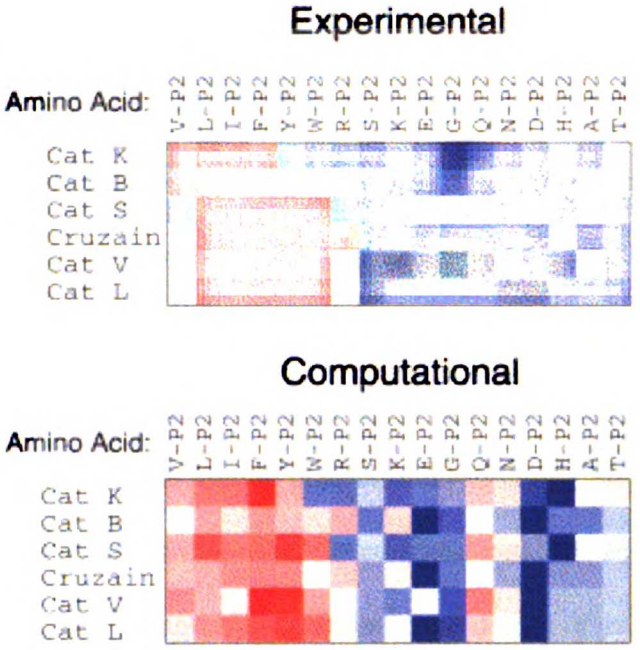


Figure 5.7 Comparison of *in silico* affinity fingerprints with experimental fingerprints. The affinity-fingerprint inhibition data generated using a subset of the PSL P2 data were compared to data generated using a combination of molecular dynamics and DOCKing algorithms (see text). Computationally derived values for relative free energies were converted to color format similarly to experimentally obtained competition data. Cluster analysis highlights similarities between the two sets of data.

The largest differences between the in silico predictions and the experimental results are seen with the lysine, glutamine, and arginine residues (Figure 5.7). There are several differences between the conditions of the experiment and the assumptions of the models that could account for this. First, the experiment represents a measurement of relative residual enzymatic activity following treatment with each inhibitor sublibrary rather than a K_i. The calculations attempt to rank order the relative binding affinities of each P2 side chain. Second, the modeled inhibitors were constructed with alanine at the P3 and P4 sites, while the positional scanning libraries have equimolar mixtures of all amino acids at these sites. Third, the protonation states of the modeled acidic and basic residues were estimated based upon the experimental pH; the actual protonation states

depend upon the local environments of each amino acid. Fourth, the inhibitor could adopt secondary structure in solution, thereby affecting its binding surface in a manner not considered during the simulations. Given these factors, it is reasonable that the theoretical predictions do not agree perfectly with the experimental results.

Ultimately, the computational protocol generated affinity fingerprints that can be used to predict most of the critical elements that control substrate specificity. Therefore, this method has the potential to be used to predict small molecule binding properties for other papain family proteases. Furthermore, the computational strategy allows for the screening of a virtual library of inhibitors to assist in the design of selective compounds for targets within a family of highly related enzymes.

Experimental Procedures

Ethyl (2S,3S)-Oxirane-2,3-Dicarboxylate and Ethyl (2R,3R)-Oxirane-2,3-Dicarboxylate and DCG-04: The synthesis of (2R,3R)-oxirane-2,3-dicarboxylate is identical to that reported for the (2S,3S) isomer [28]. The synthesis of DCG-04 is reported elsewhere [10].

Synthesis of Positional Scanning Libraries: Synthesis of the PSL libraries was reported elsewhere [11].

Gel Electrophoresis: One-dimensional SDS-PAGE and two-dimensional IEF was performed as described [29].

Competition Labeling and Analysis of Data: Rat liver lysates (100 μg total protein in 100 μL buffer A: 50 mM Tris [pH 5.5], 5 mM MgCl2, 2 mM DTT) or purified cathepsins (1 μg protein in 100 μL buffer A) were preincubated with 10 μM of each library member (diluted from 10 mM DMSO stocks) for 30 min at room temperature. Samples were then labeled by addition of ^{125}I-DCG-04 to each sample followed by further incubation at room temperature for 1 hr. Samples were quenched by the addition of 4× sample buffer, resolved by SDS-PAGE, and analyzed by PhosporImaging (Molecular Dynamics). Bands corresponding to each labeled protease were quantified. Intensities of inhibitor-treated samples were divided by the intensity of an untreated control sample to obtain a percent competition value. Numerical values for percent competition were analyzed as described previously [11 and 15] using the programs Tree View and Cluster written by Eisen and coworkers [16]. These programs can be obtained from www.microarrays.org.

Cluster Analysis Based on Sequence Alignment: Amino acid sequences for all proteins were obtained from GenBank. All sequences used were human with the exception of falcipain 2 and cruzain. Sequence alignments were performed according to their primary structure similarity using the default settings for CLUSTAL W, version 1.5 (EMBL-EBI; www.ebi.ac.uk/clustalw/).

Computational Strategy: Initial geometries for the ligand-receptor structures were constructed by analogy to the E64-cathepsin K complex (Protein Data Bank ID code 1atk) [22 and 30]. The structures of cathepsin L, V, B, K, S, and Cruzain were each aligned to the 1atk structure by matching the C atoms of four residues in the active site:

Q19, C25, Y67, and W184 (papain numbering) [22, 23, 24, 25, 26, 27 and 30]. A minimal scaffold was then built into each receptor structure by analogy to the atomic coordinates of E-64. The resulting complexes were energy minimized for 500 steps using the AMBER program suite [21]. A 28 Å cap of TIP3P water [31] centered at the scaffold center-of-mass was added to each complex, and a subset of atoms 15 Å from the ligand as well as all water atoms were selected to be mobile during subsequent molecular dynamics simulations. Each complex was equilibrated to 300 K over 80 ps, and "snapshots" were then acquired every 4 ps over a 400 ps production MD run. From the resulting 100 scaffold-receptor poses, one was selected based upon minimum root-mean-square deviations from the original crystal structure and a minimum $C\alpha_{scaffold}$-$C\alpha_{E64}$ distance.

For each scaffold-receptor pose selected from the MD runs, side chains were incrementally grown away from the P2 scaffold position according to a previously reported methodology [18]. The resulting conformations of each side chain were then rank ordered by DOCK score [20], and the top twenty conformations of each added side chain on each scaffold-receptor pose were energy minimized using the AMBER program suite [21]. Cartesian restraints were applied to the scaffold and receptor atoms during the minimization. A 1 kcal/mol restraint was imposed upon the backbone atoms of the P2 residue, while a 500 kcal/mol restraint was imposed upon all other scaffold atoms and all receptor atoms. Only the P2 side chain atoms were allowed to move freely during 500 steps of minimization. Following minimization, each of the twenty conformations of each P2 side chain in each pose was rescored using a previously reported Poison-Bolzmann continuum solvation scheme [19]. Here, the free energy of binding is approximated by

decomposition into molecular mechanical, solvation, and conformational entropy (ignored in this work) contributions.

References

1. D. Eisenberg, E.M. Marcotte, I. Xenarios and T.O. Yeates, Protein function in the post-genomic era. *Nature* **405** (2000), pp. 823–826.

2. J.S. Fetrow and J. Skolnick, Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and T1 ribonucleases. *J. Mol. Biol.* **281** (1998), pp. 949–968.

3. R.D. Hull, S.B. Singh, R.B. Nachbar, R.P. Sheridan, S.K. Kearsley and E.M. Fluder, Latent semantic structure indexing (LaSSI) for defining chemical similarity. *J. Med. Chem.* **44** (2001), pp. 1177–1184.

4. E.K. Kick, D.C. Roe, A.G. Skillman, G. Liu, T.J. Ewing, Y. Sun, I.D. Kuntz and J.A. Ellman, Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin D. *Chem. Biol.* **4** (1997), pp. 297–307.

5. S.C. Hopkins, R.D. Vale and I.D. Kuntz, Inhibitors of kinesin activity from structure-based computer screening. *Biochemistry* **39** (2000), pp. 2805–2814.

6. S. Huo, J. Wang, P. Cieplak, P.A. Kollman and I.D. Kuntz, Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. *J. Med. Chem.* **45** (2002), pp. 1412–1419.

7. L.M. Kauvar, Affinity fingerprinting. *Biotechnology (N Y)* **13** (1995), pp. 965–966.

8. L.M. Kauvar, D.L. Higgins, H.O. Villar, J.R. Sportsman, A. Engqvist-Goldstein, R. Bukar, K.E. Bauer, H. Dilley and D.M. Rocke, Predicting ligand binding to proteins by affinity fingerprinting. *Chem. Biol.* **2** (1995), pp. 107–118.

9. S.V. Frye, Structure-activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chem. Biol.* **6** (1999), pp. R3–R7.

10. D. Greenbaum, K.F. Medzihradszky, A. Burlingame and M. Bogyo, Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7** (2000), pp. 569–581.

11. D. Greenbaum, A. Baruch, L. Hayrapetian, Z. Darula, A. Burlingame, K. Medzihradszky and M. Bogyo, Chemical approaches for functionally probing the proteome. *Mol. Cell Proteomics* **1** (2002), pp. 60–68.

12. A.J. Barrett, A.A. Kembhavi, M.A. Brown, H. Kirschke, C.G. Knight, M. Tamai and K. Hanada, L-trans-Epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. *Biochem. J.* **201** (1982), pp. 189–198.

13. D. Turk, G. Guncar, M. Podobnik and B. Turk, Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol. Chem.* **379** (1998), pp. 137–147. 14. J.P. Meara and D.H. Rich, Mechanistic studies on the inactivation of papain by epoxysuccinyl inhibitors. *J. Med. Chem.* **39** (1996), pp. 3357–3366.

15. T. Nazif and M. Bogyo, Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors. *Proc. Natl. Acad. Sci. USA* **98** (2001), pp. 2967–2972.

16. M.B. Eisen, P.T. Spellman, P.O. Brown and D. Botstein, Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95** (1998), pp. 14863–14868.

17. N. Schaschke, I. Assfalg-Machleidt, W. Machleidt, D. Turk and L. Moroder, E-64 analogues as inhibitors of cathepsin B. On the role of the absolute configuration of the epoxysuccinyl group. *Bioorg. Med. Chem.* **5** (1997), pp. 1789–1797.

18. M.L. Lamb, K.W. Burdick, S. Toba, M.M. Young, A.G. Skillman, X. Zou, J.R. Arnold and I.D. Kuntz, Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins* **42** (2001), pp. 296–318.

19. P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang *et al.*, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33** (2000), pp. 889–897.

20. T.J. Ewing, S. Makino, A.G. Skillman and I.D. Kuntz, Dock 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* **15** (2001), pp. 411–428.

21. D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, III, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energenic properties of molecules. *Comp. Phys. Commun.* **91** (1995), pp. 1–41.

22. B. Zhao, C.A. Janson, B.Y. Amegadzie, K. D'Alessio, C. Griffin, C.R. Hanning, C. Jones, J. Kurdyla, M. McQueney, X. Qiu *et al.*, Crystal structure of human osteoclast cathepsin K complex with E-64. *Nat. Struct. Biol.* **4** (1997), pp. 109–111.

23. D. Turk, M. Podobnik, T. Popovic, N. Katunuma, W. Bode, R. Huber and V. Turk, Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors. *Biochemistry* **34** (1995), pp. 4791–4797.

24. J.R. Somoza, H. Zhan, K.K. Bowman, L. Yu, K.D. Mortara, J.T. Palmer, J.M. Clark and M.E. McGrath, Crystal structure of human cathepsin V. *Biochemistry* **39** (2000), pp. 12543–12551.

25. G. Guncar, G. Pungercic, I. Klemencic, V. Turk and D. Turk, Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO J.* **18** (1999), pp. 793–803.

26. L.S. Brinen, E. Hansell, J. Cheng, W.R. Roush, J.H. McKerrow and R.J. Fletterick, A target within the target: probing cruzain's P1' site to define structural determinants for the Chagas' disease protease. *Struct. Fold. Des.* **8** (2000), pp. 831–840.

27. M.E. McGrath, J.T. Palmer, D. Bromme and J.R. Somoza, Crystal structure of human cathepsin S. *Protein Sci.* **7** (1998), pp. 1294–1302.

28. M. Bogyo, S. Verhelst, V. Bellingard-Dubouchaud, S. Toba and D. Greenbaum, Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs. *Chem. Biol.* **7** (2000), pp. 27–38.

29. M. Bogyo, S. Shin, J.S. McMaster and H.L. Ploegh, Substrate binding and sequence preference of the proteasome revealed by active-site-directed affinity probes. *Chem. Biol.* **5** (1998), pp. 307–320.

30. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.E. Bourne, The protein data bank. *Nucleic Acids Res.* **28** (2000), pp. 235–242.

31. W.L. Jorgensen, J. Chandrasekhar, J. Madura and M.L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79** (1983), pp. 926–935.

Acknowledgements

We would like to thank Mark Rice and Paul Sprengeler (Celera, South San Francisco) for helpful discussion of the data and manuscript. We thank Dr. Vito Turk (Jozef Stefan Institute, Ljubljana, Slovenia) for the generous gift of purified recombinant human cathepsin L. This work was supported by funding from the Sandler Program in Basic Science (M.B., D.C.G., L.H., and K.C.) and the National Institutes of Health (GM31497 to I.D.K., GM64097 to W.D.A., GM56531 to P. Ortiz de Montellano, Principal Investigator, and CA72006 to M. Shuman, Principal Investigator).

Chapter 6

Future Directions

We have constructed a library of p53-MDM2 inhibitors through structure-based design. One of the most pressing opportunities for future experiments is the optimization of binding, structural characterization of binding mode, and enhancement of drug-like properties. The optimization of binding may require the synthesis and testing of a new set of side chains. Side chains can be chosen empirically from structure-activity relationships seen in the current data or computationally through modeling of the binding interaction. More detailed structural information, which may improve side chain selection, will require further NMR or x-ray crystallography experiments. The incorporation of heteroatoms and hydrophilic groups will bring the inhibitors one step closer to obeying Lipinski's Rule of 5.

Another promising research experiment lies in the synthesis of libraries based on other scaffolds. The synthetic route for the libraries of the three scaffolds not discussed here has been proposed (Figure 2.2D). The synthesis and evaluation of these remaining libraries may shed some light on the success of the library design method. Comparing and contrasting the characteristics of successful and unsuccessful scaffolds will help guide the development of this relatively new and untested design method.

These inhibitors have a high potential for activity in other protein-protein interactions involving an alpha helix. One potential direction for future research includes testing the library in other biological systems associated with an i, $i + 4$, $i + 7$ alpha helix. If the central scaffold is successful in mimicking the helical backbone, the addition of side chains tailored for each protein could give multiple potent inhibitors of protein-protein interactions. The common scaffold used will ease synthesis and give insight into

other ADMET and pharmacokinetic properties once one representative inhibitor has been characterized.

This library represents one of several libraries of compounds targeting an *i*, *i* + 4, *i* + 7 alpha helical system, and the method used for *in silico* design can be applied to other helical motifs as well as other classes of protein substructures. After creating libraries populating each substructure class, one could screen new protein-protein targets whose binding features have been classified. Thus, recurring protein motifs provide an added advantage by facilitating lead discovery.

We have also generated an affinity fingerprinting method to functionally characterize a family of cysteine proteases both chemically and computationally. This method allows for the rapid visual analysis of inhibitor specificity and enzyme active site topology. Enzymes can then be subclassified based on functional relationships rather than simply by linear amino acid sequences. Furthermore, this method provides a direct readout of the overall inhibitory characteristics of compounds under a variety of assay conditions. This method will ultimately aid in the process of target selection, prioritization, and inhibitor design.

Appendix 1

Yield, Purity, and Spectral Data

Table A.1 Yields and purities of chemset 12 precursors.

Entry	Product	% Yield (% purity)
1	2{1}	84 (95)
2	2{2}	81 (95)
3	2{3}	22 (60)
4	2{4}	74 (95)
5	2{5}	74 (95)
6	2{6}	80 (95)
7	2{7}	68 (95)
8	2{8}	71 (95)
9	2{9}	52 (95)
10	2{10}	36 (75)
11	2{11}	62 (95)
12	2{12}	31 (95)
13	2{13}	28 (90)
14	3{1}	75 (95)
15	3{2}	98 (95)
16	3{3}	56 (95)
17	3{4}	93 (95)
18	3{5}	96 (95)
19	3{6}	92 (95)
20	3{7}	95 (95)
21	3{8}	92 (95)
22	3{9}	85 (95)
23	3{10}	75 (95)
24	3{11}	60 (95)
25	3{12}	99 (95)
26	3{13}	56 (95)
27	6{1}	35 (95)
28	6{2}	58 (95)
29	6{3}	14 (95)
30	6{4}	24 (95)
31	6{5}	28 (95)
32	6{6}	64 (95)
33	6{7}	67 (95)
34	6{8}	36 (95)
35	6{9}	47 (95)
36	6{11}	20 (60)
37	8{1}	100 (95)
38	8{2}	87 (95)
39	8{4}	74 (95)
40	8{5}	79 (95)
41	8{6}	64 (95)
42	8{7}	65 (95)
43	8{8}	88 (95)
44	8{9}	44 (95)
45	8{14}	70 (95)
46	9{1}	87 (95)
47	10{1,1}	90 (95)
48	10{1,7}	87 (95)
49	10{2,4}	77 (95)
50	10{4,2}	83 (95)

51	10{5, 9}	88 (95)
52	10{6,11}	84 (95)
53	10{7,6}	71 (95)
54	10{8,9}	86 (95)
55	10{9,8}	84 (95)
56	10{11,6}	80 (95)
57	10{13,1}	82 (95)
58	10{3,4}	Failed
59	10{1,3}	Failed
60	11{1,1}	90 (95)
61	11{1,7}	72 (95)
62	11{2,4}	43 (95)
63	11{4,2}	87 (95)
64	11{5, 9}	84 (95)
65	11{6,11}	43 (95)
66	11{7,6}	69 (95)
67	11{8,9}	72 (95)
68	11{9,8}	70 (95)
69	11{11,6}	91 (95)
70	11{13,1}	81 (95)
71	12{1,1,1}	60 (95)
72	12{1,7,1}	33 (95)
73	12{2,4,1}	22 (95)
74	12{4,2,1}	37 (95)
75	12{5, 9,1}	27 (95)
76	12{6,11,1}	16 (95)
77	12{7,6,1}	9 (95)
78	12{8,9,1}	40 (95)
79	12{9,8,1}	15 (95)
80	12{13,1,1}	37 (95)

Table A.2 Representative analytical data of chemset 12.

Entry	Compound	1H-NMR (400 MHz, DMSO-d6)	HRMS Calc.	HRMS Found
1	12{13,1,1}	δ = 9.883 (s, 1H), δ = 9.861 (s, 1H), δ = 7.760 (m, 2H), δ = 7.610 (d, J = 7.6, 1H), δ = 7.552 (d, J = 7.6, 1H), δ = 7.516 (s, 1H), δ = 7.3 (m, 3H), δ = 7.2 (m, 13H), δ = 4.120 (s, 2H), δ = 4.015 (s, 2H), δ = 2.5 (m, overlap w/solvent), δ = 1.17 (m, 4H), δ = 0.94 (m, 2H)	592.3090	593.3175
2	12{2,4,1}	δ = 10.048 (s, 1H), δ = 9.863 (s, 1H), δ = 7.74 (m, 3H), δ = 7.667 (s, 1H), δ = 7.520 (d, J = 6, 1H), δ = 7.44 (m, 2H), δ = 7.361 (d, J = 6, 1H), δ = 7.1 (m, 13H), δ = 7.83 (m, 2H), δ = 6.778 (dd, J = 8.2, J = 1.2, 1H), δ = 4.108 (s, 2H), δ = 3.996 (s, 2H), δ = 3.982 (s, 2H), δ = 3.724 (s, 3H)	634.2632	635.2708
3	12{1,7,1}	δ = 10.112 (s, 1H), δ = 9.872 (s, 1H), δ = 7.800 (dd, J = 8, J = 1.6, 1H), δ = 7.752 (d, J = 8, 1H), δ = 7.67 (m, 2H), δ = 7.615 (s, 1H), δ = 7.552 (d, J = 8, 2H), δ = 7.40 (m, 2H), δ = 7.359 (d, J = 8, 2H), δ = 7.300 (d, J = 8, 2H), δ = 7.2 (m, 12H), δ = 4.327 (s, 2H), δ = 4.001 (s, 2H), δ = 3.990 (s, 2H)	611.2573	612.2635
4	12{9,8,1}	δ = 10.033 (s, 1H), δ = 9.881 (s, 1H), δ = 7.78 (m, 3H), δ = 7.739 (s, 1H), δ = 7.68 (m, 2H), δ = 7.616 (d, J = 6.8, 1H), δ = 7.518 (s, 1H), δ = 7.4 (m, 9H), δ = 7.20 (m, 3H), δ = 7.12 (m, 4H), δ = 4.145 (s, 2H), δ = 4.124 (s, 2H), δ = 4.017 (s, 2H)	636.2525	637.2602
5	12{8,9,1}	δ = 10.021 (s, 1H), δ = 9.883 (s, 1H), δ = 7.778 (m, 2H), δ = 7.731 (s, 1H), δ = 7.67 (m, 6H), δ = 7.5 (m, 4H), δ = 7.372 (d, J = 8, 1H), δ = 7.2 (m, 7H), δ = 7.1 (m, 3H), δ = 4.186 (s, 2H), δ = 4.090 (s, 2H), δ = 4.011 (s, 2H)	636.2525	637.2592
6	12{7,6,1}	δ = 9.995 (s, 1H), δ = 9.863 (s, 1H), δ = 7.848 (d, J = 7.6, 1H), δ = 7.747 (d, J = 8, 1H), δ = 7.70 (m, 5H), δ = 7.513 (t, J = 8,4, 2H), δ = 7.459 (m, 3H), δ = 7.369 (d, J = 8, 1H), δ = 7.19 (m, 4H), δ = 7.11 (m, 5H), δ = 7.02 (m, 2H), δ = 4.261 (s, 2H), δ = 4.066 (s, 2H), δ = 4.007 (s, 2H)	629.2479	630.2559
7	12{6,11,1}	δ = 10.023 (s, 1H), δ = 9.894 (s, 1H), δ = 8.004 (d, J = 8, 1H), δ = 7.962 (s, 1H), δ = 7.78 (m, 2H), δ = 7.639 (m , 2H), δ = 7.532 (d, J = 8, 1H), δ = 7.510 (s, 1H), δ = 7.486 (t, J = 7.6, 1H), δ = 7.40 (m, 3H), δ = 7.27 (m, 3H), δ = 7.20 (m, 4H), δ = 7.12 (m, 5H), δ = 4.244 (s, 2H), δ = 4.015 (s, 2H), δ = 3.991 (s, 2H)	649.2377	650.2465
8	12{5,9,1}	δ = 10.018 (s, 1H), δ = 9.882 (s, 1H), δ = 7.778 (d, J = 8, 1H), δ = 7.732 (s, 1H), δ = 7.66 (m, 3H), δ = 7.518 (d, J = 8, 1H), δ = 7.464 (s, 1H), δ = 7.440 (d, J = 7.2, 1H), δ = 7.36 (m, 2H), δ = 7.28 (m, 3H), δ = 7.18 (m, 3H), δ = 7.143 (d, J = 7.2, 1H), δ = 7.1 (m, 6H), δ = 7.03 (m, 1H), δ = 4.187 (s, 2H), δ = 4.038 (s, 2H), δ = 4.011 (s, 2H)	629.2479	630.2565
9	12{2,4,1}	δ = 9.951 (s, 1H), δ = 9.862 (s, 1H), δ = 7.75 (m, 2H), δ = 7.70 (m, 2H), δ = 7.547 (d, J = 8.8, 1H), δ = 7.42 (m, 2H), δ = 7.2 (m, 14H), δ = 6.70 (m, 3H), δ = 4.060 (s, 2H), δ = 4.051 (s, 2H), δ = 4.010 (s, 2H), δ = 3.585 (s, 3H)	634.2632	635.2712

Table A.3 Representative analytical data of chemset 11.

Entry	Compound	1H-NMR (400 MHz, DMSO-d6)	LCMS Calc [M+1]	LCMS Found [M+1]
1	11{11.6}	δ = 9.986 (s, 1H), δ = 8.151 (m, 1H), δ = 8.089 (d, J = 7.2, 1H), δ = 7.831 (dd, J = 8.8, J = 2, 1H), δ = 7.75 (m, 3H), δ = 7.704 (d, J = 7.6, 1H), δ = 7.620 (t, J = 8, 1H), δ = 7.577 (d, J = 8.4, 1H), δ = 7.523 (d, J = 7.6, 1H), δ = 7.469 (d, J = 8.0, 1H), δ = 4.196 (s, 2H), δ = 4.081 (s, 2H)	485.1435*	485.1508*
2	11{2.4}	($CDCl_3$) δ = 8.152 (d, J = 8, 1H), δ = 8.02 (m, 2H), δ = 7.479 (m, 2H), δ = 7.368 (m, 2H), δ = 7.331 (s, 1H), δ = 7.221 (m, 2H), δ = 7.07 (m, 2H), δ = 6.75 (m, 3H), δ = 4.090 (s, 2H), δ = 3.977 (s, 2H), δ = 3.773 (s, 3H)	470.17	470.54
3	11{13.1}	δ = 9.847 (s, 1H), δ = 7.844 (d, J = 8, 1H), δ = 7.791 (s, 1H), δ = 7.645 (d, J = 8, 2H), δ = 7.546 (s, 1H), δ = 7.36 (m, 2H), δ = 7.235 (d, J = 6.4, 2H), δ = 7.180 (d, J = 6.4, 1H), δ = 7.135 (d, J = 7.6, 2H), δ = 4.127 (s, 2H), δ = 2.5 (m, overlap w/solvent), δ = 1.63 (m, 5H), δ = 1.17 (m, 4H), δ = 0.94 (m, 2H)	428.21	428.60
4	11{1.7}	δ =10.148 (s, 1H), δ = 7.888 (dd, J = 8, J = 1.6, 1H), δ = 7.780 (d, J = 8, 1H), δ = 7.785 (m, 3H), δ = 7.58 (m, 2H), δ = 7.42 (m, 3H), δ = 7.30 (m, 5H), δ = 4.339 (s, 2H), δ = 4.013 (s, 2H)	447.50	447.50
5	11{9.8}	δ = 10.021 (s, 1H), δ = 9.883 (s, 1H), δ = 7.778 (m, 2H), δ = 7.731 (s, 1H), δ = 7.67 (m, 6H), δ = 7.5 (m, 4H), δ = 7.372 (d, J = 8, 1H), δ = 7.2 (m, 7H), δ = 7.1 (m, 3H), δ = 4.186 (s, 2H), δ = 4.090 (s, 2H), δ = 4.011 (s, 2H)	472.16	472.58
6	11{8.9}	δ =10.045 (s, 1H), δ = 7.873 (d, J = 8, 1H), δ = 7.815 (s, 1H), δ = 7.779 (s, 1H), δ = 7.68 (m, 6H), δ = 7.558 (m, 2H), δ = 7.465 (m, 2H), δ = 7.290 (d, J = 8, 2H), δ = 4.211 (s, 2H), δ = 4.094 (s, 2H)	472.16	472.51
7	11{5.9}	δ =10.045 (s, 1H), δ = 7.878 (d, J = 8, 1H), δ = 7.825 (s, 1H), δ = 7.67 (m, 4H), δ = 7.578 (d, J = 8, 1H), δ = 7.46 (m, 2H), δ = 7.35 (m, 1H), δ = 7.293 (d, J = 8, 2H), δ = 7.112 (d, J = 8, 2H), δ = 7.041 (m, 1H), δ = 4.219 (s, 2H), δ = 4.043 (s, 2H)	465.15	465.51
8	11{4.2}	δ = 9.980 (s, 1H), δ = 7.851 (d, J = 8, 1H), δ = 7.809 (s, 1H), δ = 7.720 (m, 2H), δ = 7.615 (d, J = 8, 1H), δ = 7.433 (m, 2H), δ = 7.321 (m, 2H), δ = 7.17 (m, 3H), δ = 6.714 (m, 3H), δ = 4.085 (s, 2H), δ = 4.058 (s, 2H), δ = 3.612 (s, 3H)	470.17	470.54
9	11{7.6}	δ = 9.995 (s, 1H), δ = 7.81 (m, 2H), δ = 7.69 (m, 4H), δ = 7.48 (m, 5H), δ = 7.12 (m, 2H), δ = 7.02 (m, 2H), δ = 4.243 (s, 2H), δ = 4.061 (s, 2H)	465.15	465.51

*HRMS

Table A.4 Compounds tested showing no activity.

11{5,4}	10{1,1}	11{12,11}	12{4,11,1}	12{7,6,1}	12{9,8,9}
11{5,5}	10{1,7}	11{12,2}	12{4,11,14}	12{7,6,14}	12{11,6,1}
11{6,1}	10{11,6}	11{12,3}	12{4,11,2}	12{7,6,2}	12{11,6,14}
11{6,2}	10{2,4}	11{12,4}	12{4,11,4}	12{7,6,4}	12{11,6,2}
11{6,4}	10{3,5}	11{12,5}	12{4,11,5}	12{7,6,5}	12{11,6,4}
11{6,5}	10{4,11}	11{13,1}	12{4,11,6}	12{7,6,6}	12{11,6,5}
11{6,7}	10{5,1}	12{1,1,1}	12{4,11,7}	12{7,6,7}	12{11,6,6}
11{6,8}	10{5,9}	12{1,1,5}	12{4,11,8}	12{7,6,8}	12{11,6,7}
11{7,5}	10{7,6}	12{1,7,1}	12{4,11,9}	12{7,6,9}	12{11,6,8}
11{7,7}	10{8,7}	12{1,7,14}	12{4,2,1}	12{8,9,1}	12{11,6,9}
11{7,8}	10{8,9}	12{1,7,2}	12{4,2,14}	12{8,9,14}	12{13,1,1}
11{8,1}	11{1,1}	12{1,7,4}	12{4,2,2}	12{8,9,2}	12{13,1,14}
11{8,2}	11{1,4}	12{1,7,5}	12{4,2,4}	12{8,9,4}	12{13,1,2}
11{8,4}	11{1,5}	12{1,7,6}	12{4,2,5}	12{8,9,5}	12{13,1,4}
11{8,5}	11{1,7}	12{1,7,7}	12{4,2,6}	12{8,9,6}	12{13,1,5}
11{9,1}	11{2,1}	12{1,7,8}	12{4,2,7}	12{8,9,7}	12{13,1,6}
11{9,8}	11{2,2}	12{1,7,9}	12{4,2,8}	12{8,9,8}	12{13,1,7}
11{10,1}	11{2,4}	12{2,4,1}	12{4,2,9}	12{8,9,9}	12{13,1,7}
11{11,1}	11{4,1}	12{2,4,14}	12{5,9,1}	12{9,8,1}	12{13,1,8}
11{11,1}	11{4,11}	12{2,4,2}	12{5,9,2}	12{9,8,14}	12{13,1,9}
11{11,11}	11{4,2}	12{2,4,4}	12{5,9,4}	12{9,8,2}	13{7,6,4}
11{11,2}	11{4,5}	12{2,4,5}	12{5,9,5}	12{9,8,4}	13{7,6,6}
11{11,5}	11{4,6}	12{2,4,6}	12{5,9,6}	12{9,8,5}	
11{11,6}	11{4,7}	12{2,4,7}	12{5,9,7}	12{9,8,6}	
11{11,7}	11{4,8}	12{2,4,8}	12{5,9,8}	12{9,8,7}	
11{12,1}	11{5,2}	12{2,4,9}	12{5,9,9}	12{9,8,8}	

Figure A.1 ^{1}HNMR spectra for **12**{13,1,1}

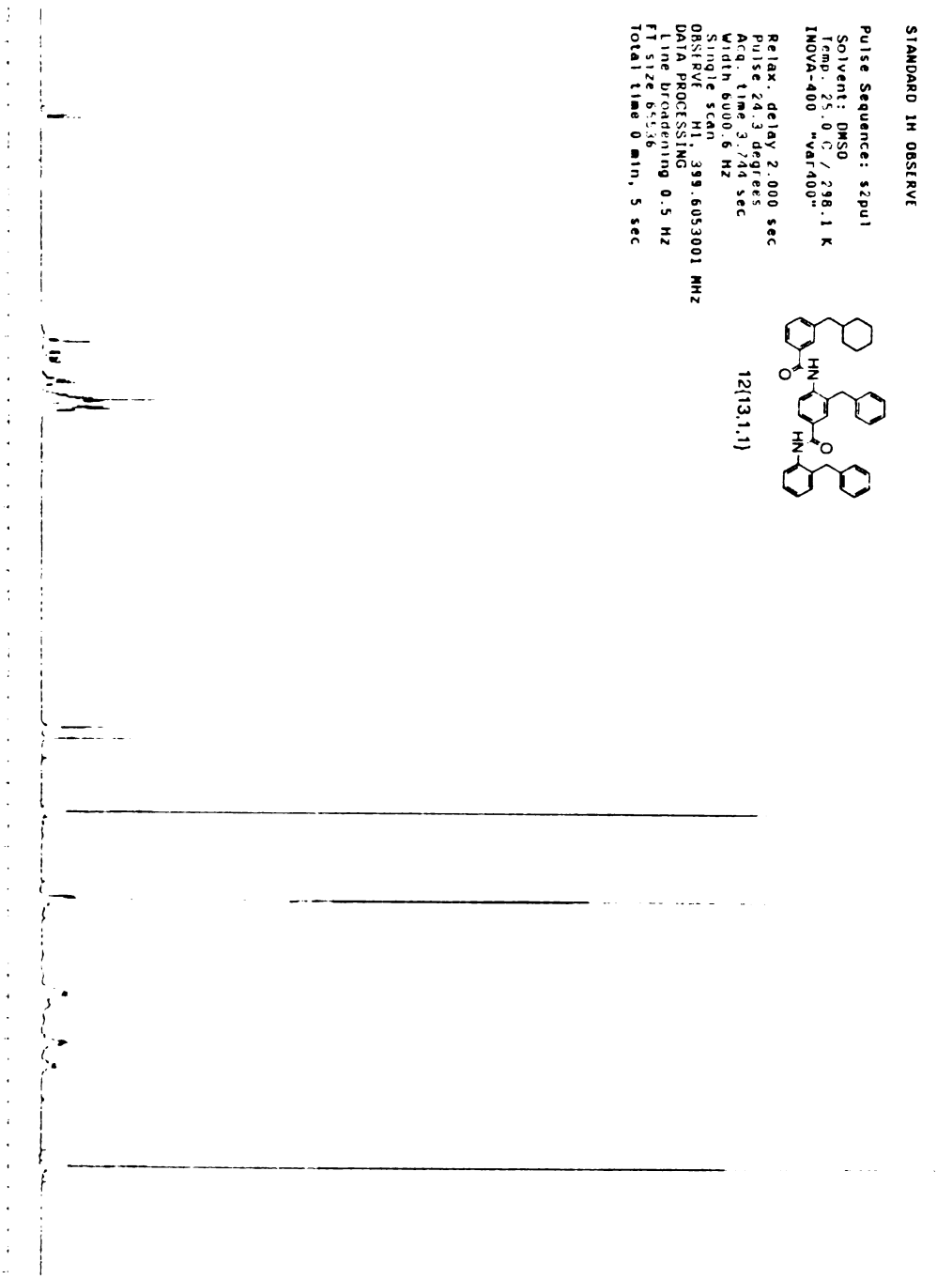
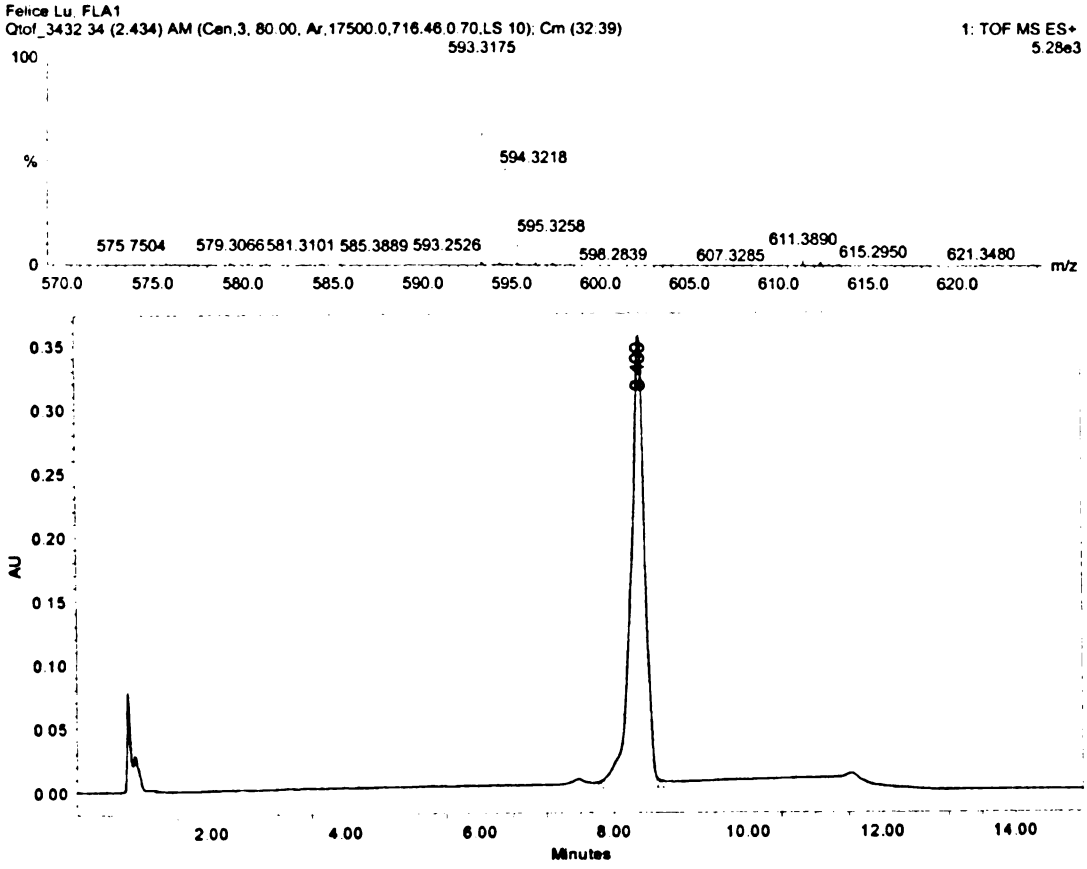


Figure A.2 HPLC and HRMS spectra for **12**{13,1,1}

Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
122 formula(e) evaluated with 69 results within limits (up to 50 closest results for each mass)



	RT	Area	% Area	Height
1	8.400	4842537	100.00	348376

Figure A.3 ^{1}HNMR spectra for **12**{2,4,1}

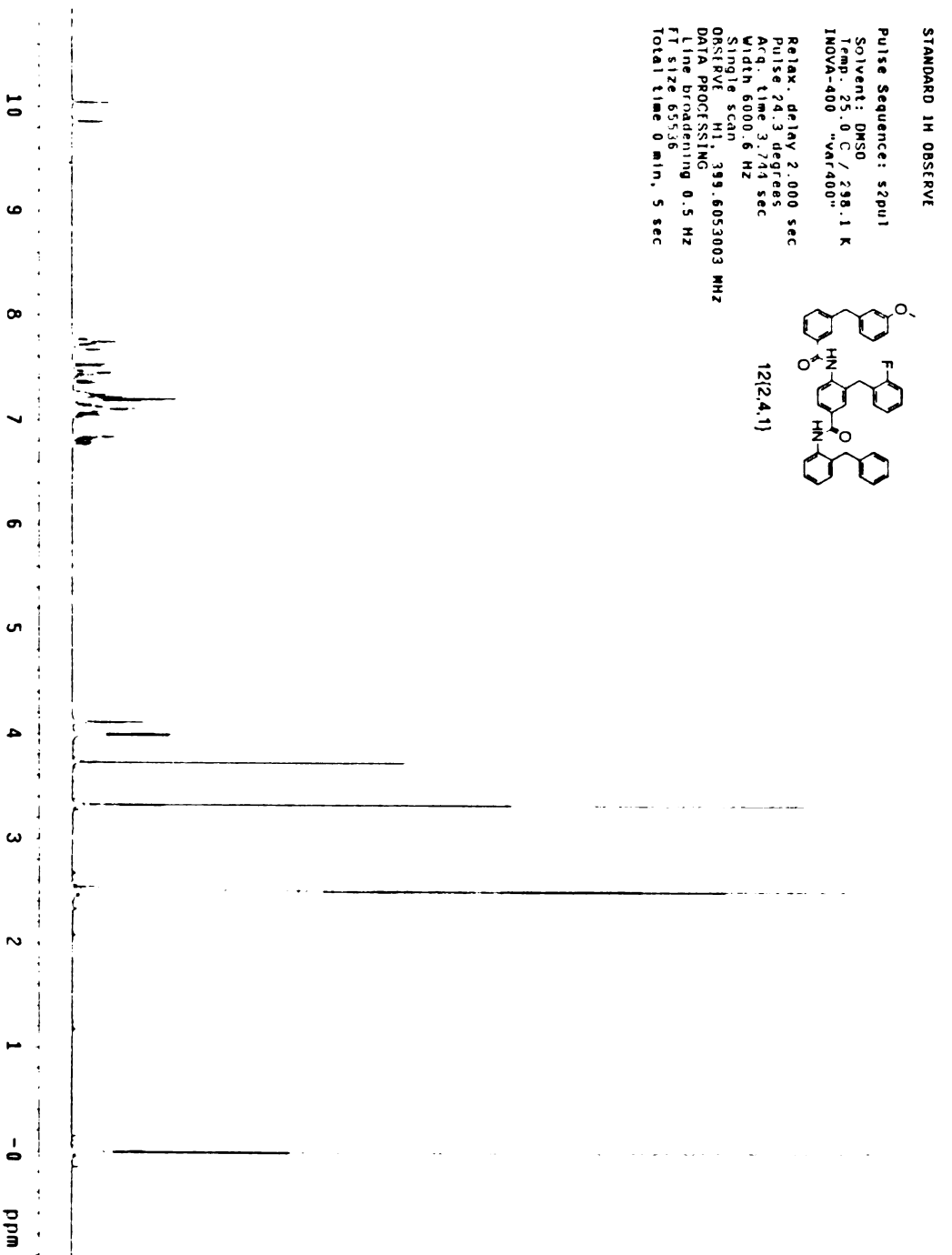


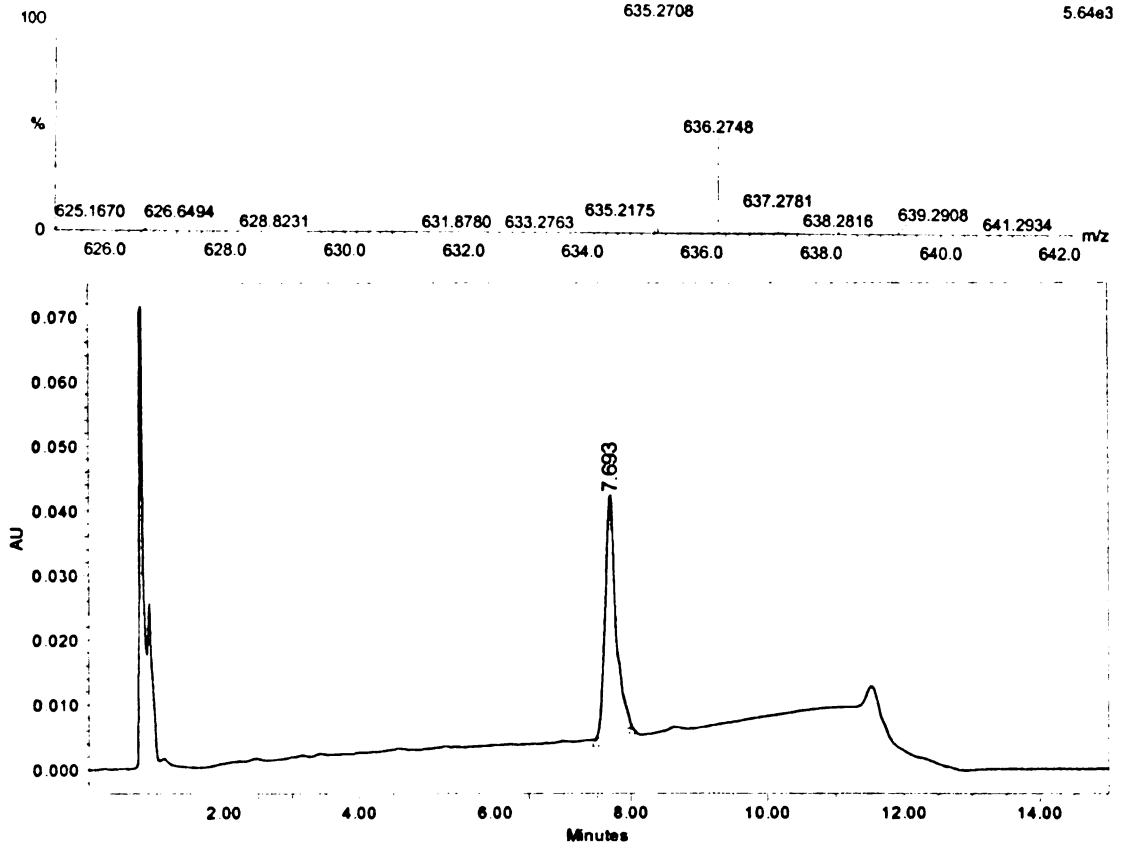
Figure A.4 HPLC and HRMS spectra for **12**{2,4,1}

Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
250 formula(e) evaluated with 134 results within limits (up to 50 closest results for each mass)

Felice Lu, FLA2
Qtof_3433 34 (2.434) AM (Cen,3, 80.00, Ar,17500.0,716.46,0.70,LS 10); Cm (33:39) 1: TOF MS ES+ 5.64e3



	RT	Area	% Area	Height
1	7.693	407029	100.00	37080

Figure A.5 ^{1}HNMR spectra for **12**{1,7,1}

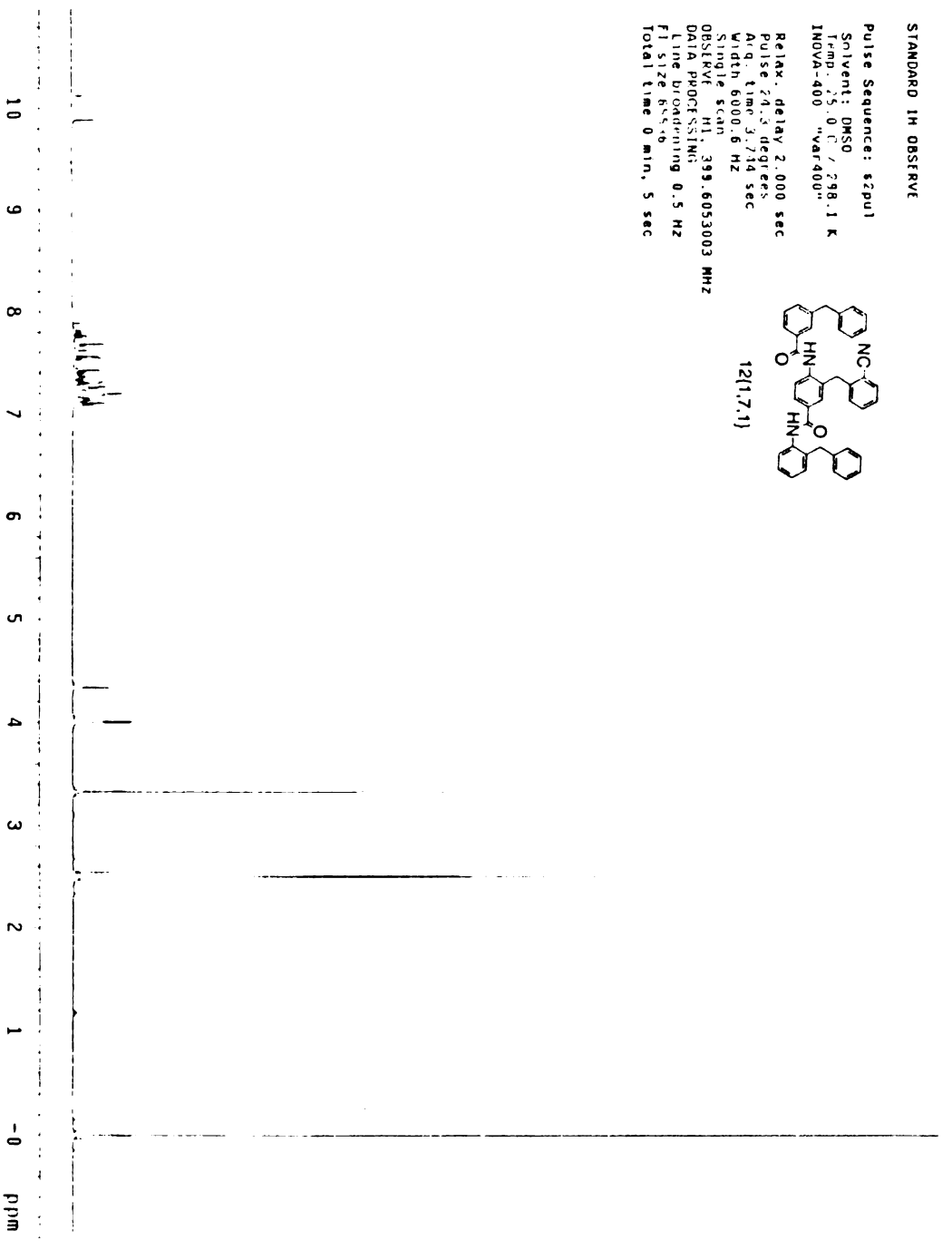
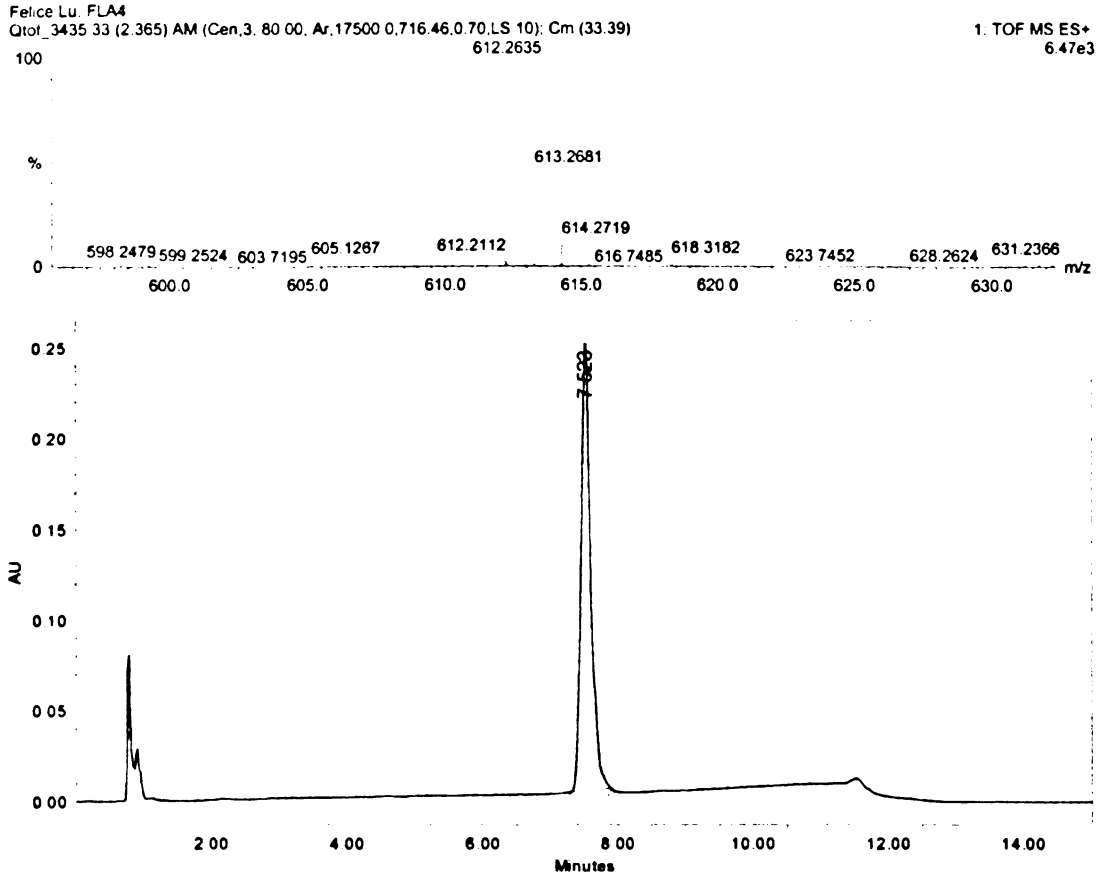


Figure A.6 HPLC and HRMS spectra for **12**{1,7,1}

Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
210 formula(e) evaluated with 119 results within limits (up to 50 closest results for each mass)



	RT	Area	% Area	Height
1	7 523	2451160	100 00	248076

Figure A.7 ^{1}HNMR spectra for **12**{9,8,1}

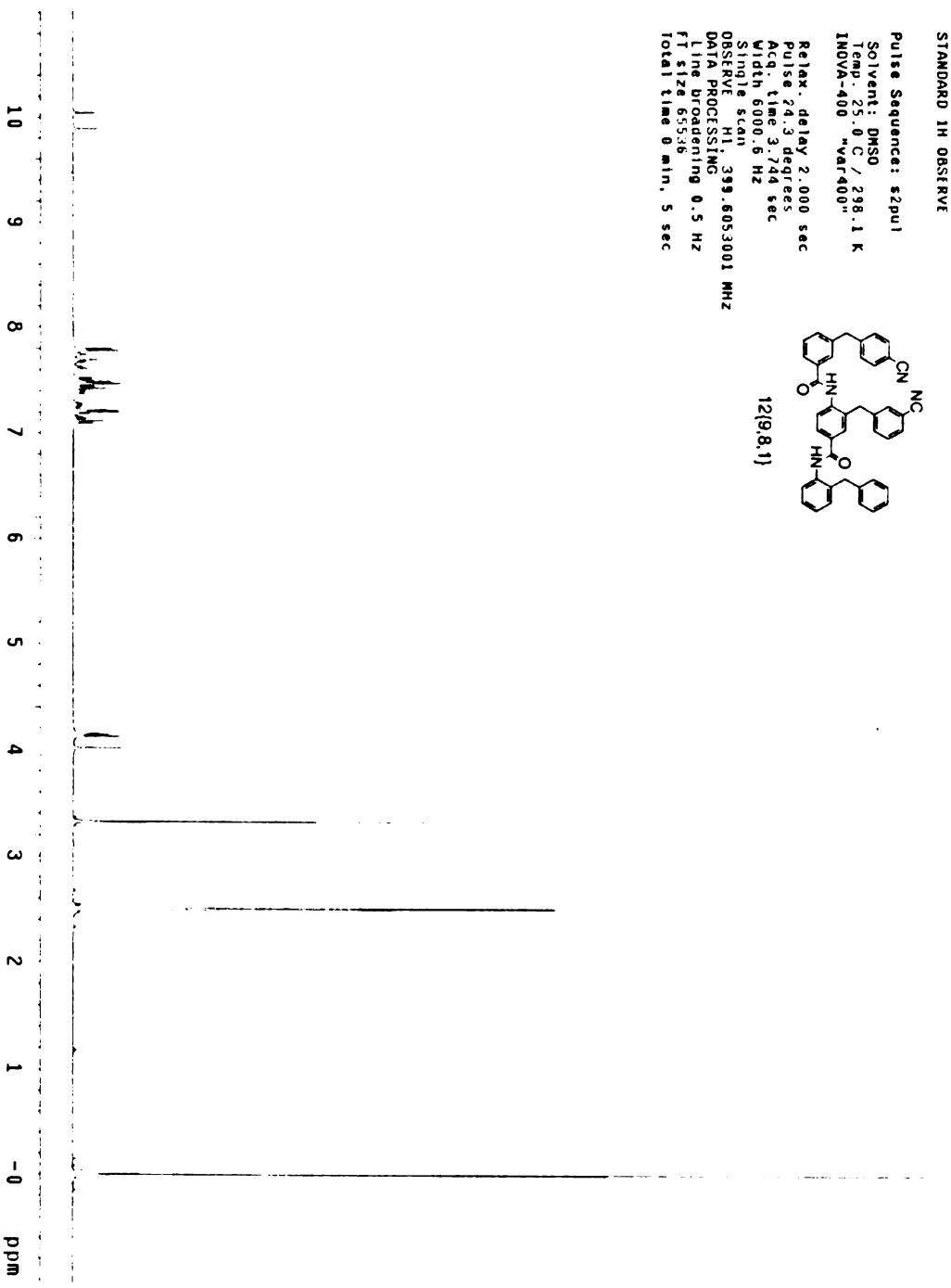


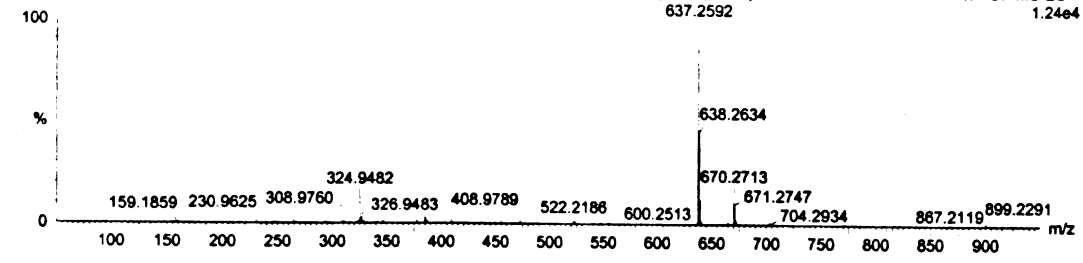
Figure A.8 HPLC and HRMS spectra for **12**{9,8,1}

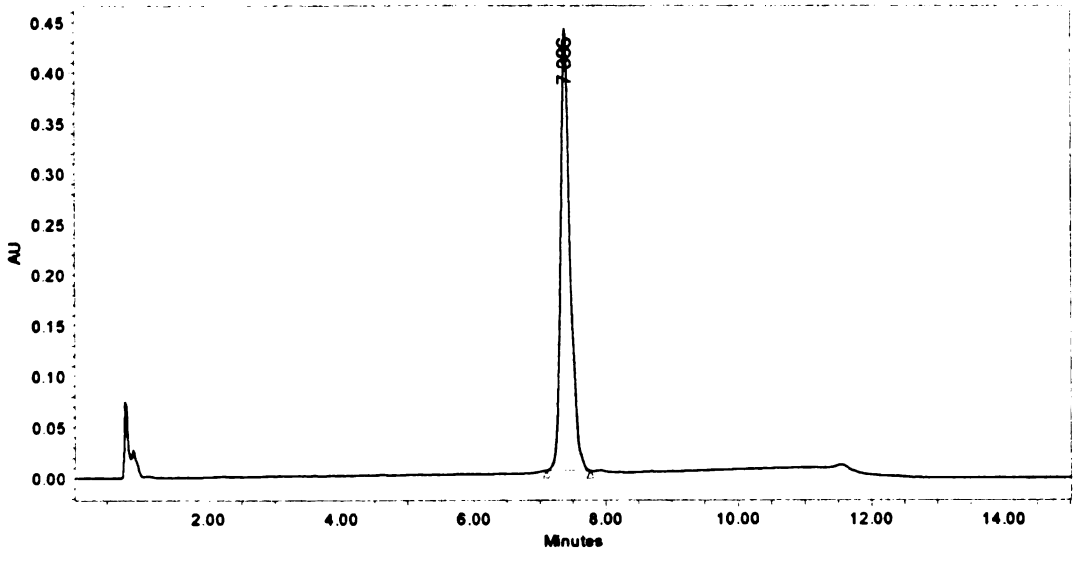
Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
115 formula(e) evaluated with 60 results within limits (up to 50 closest results for each mass)

FLA6, Lu
Qtof_3545 34 (2.440) AM (Cen,3, 80.00, Ar,17500.0,716.46,0.70,LS 10); Sm (SG, 2x3.00); Cm (32:46)
1: TOF MS ES+
1.24e4





	RT	Area	% Area	Height
1	7.386	4307452	100.00	436813

Figure A.9 [1]HNMR spectra for **12**{8,9,1}

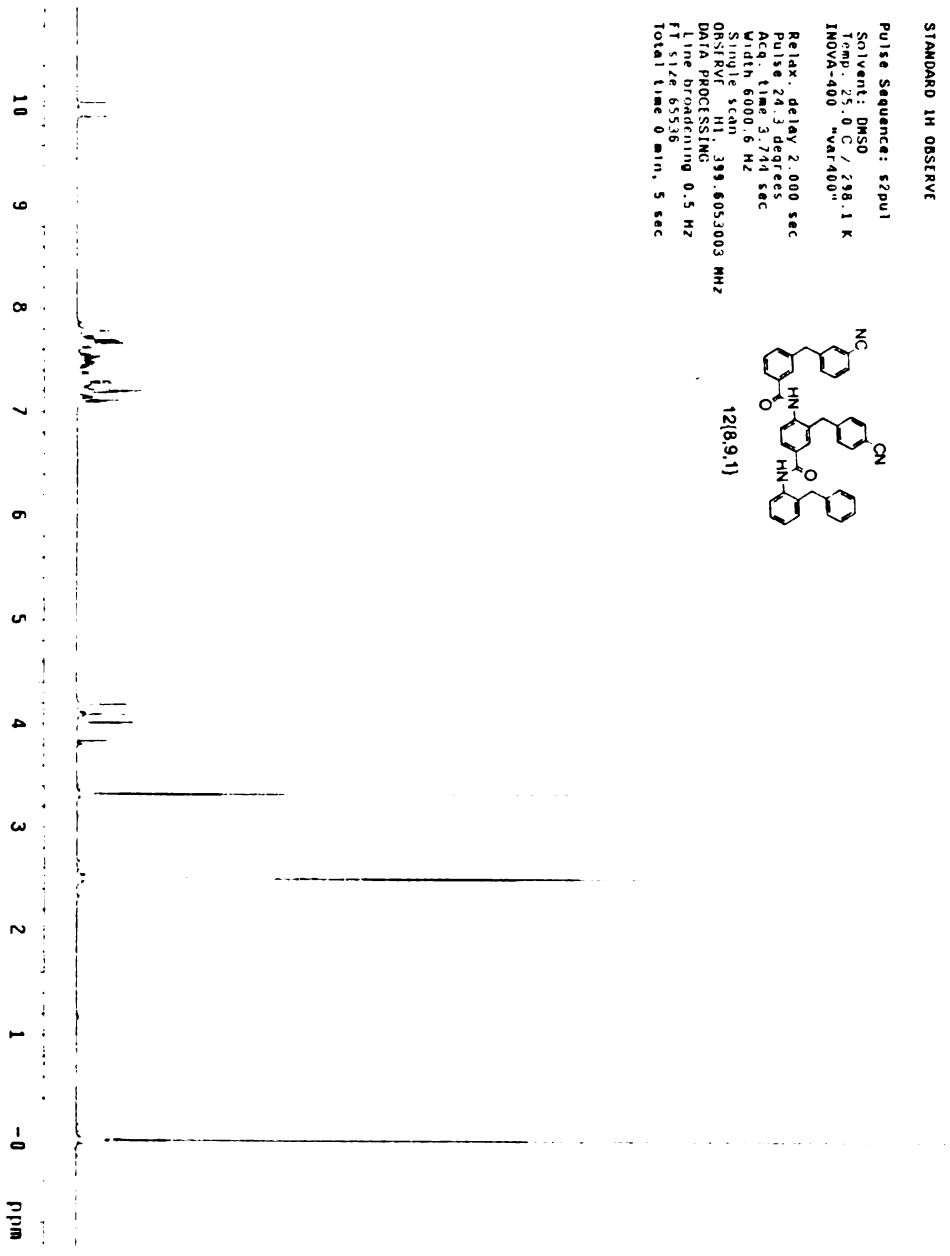


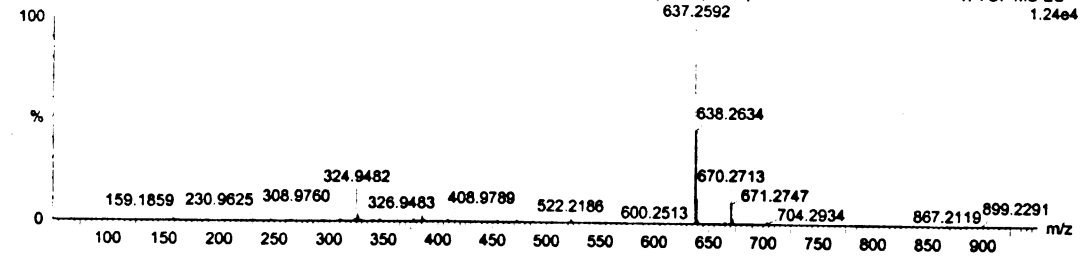
Figure A.10 HPLC and HRMS spectra for **12**{8,9,1}

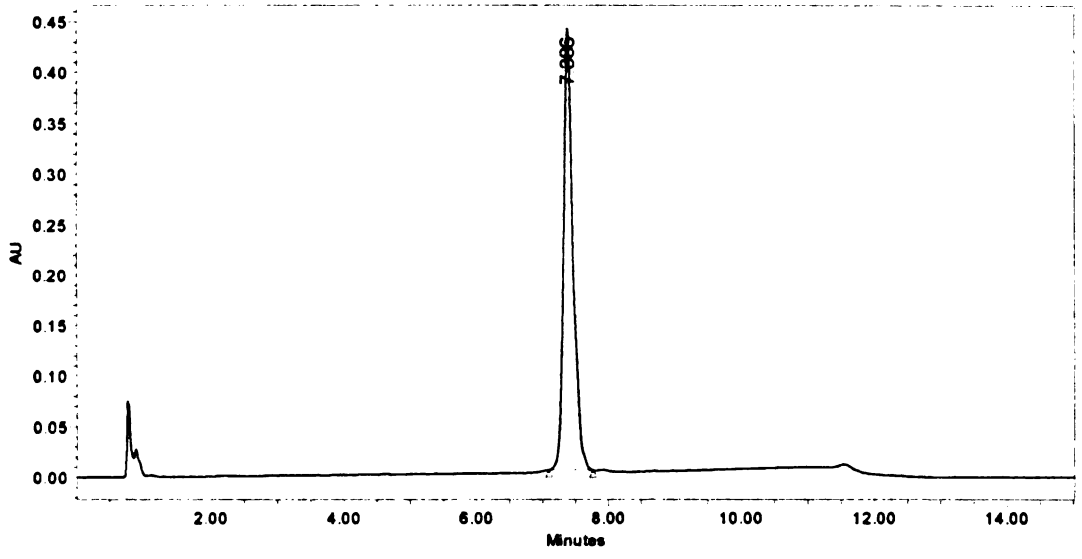
Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
115 formula(e) evaluated with 60 results within limits (up to 50 closest results for each mass)

FLA6, Lu
Qtof_3545 34 (2.440) AM (Cen,3, 80.00, Ar,17500.0,716.46,0.70,LS 10); Sm (SG, 2x3.00); Cm (32:46)





	RT	Area	% Area	Height
1	7.386	4307452	100.00	436813

Figure A.11 ^{1}HNMR spectra for **12**{7,6,1}

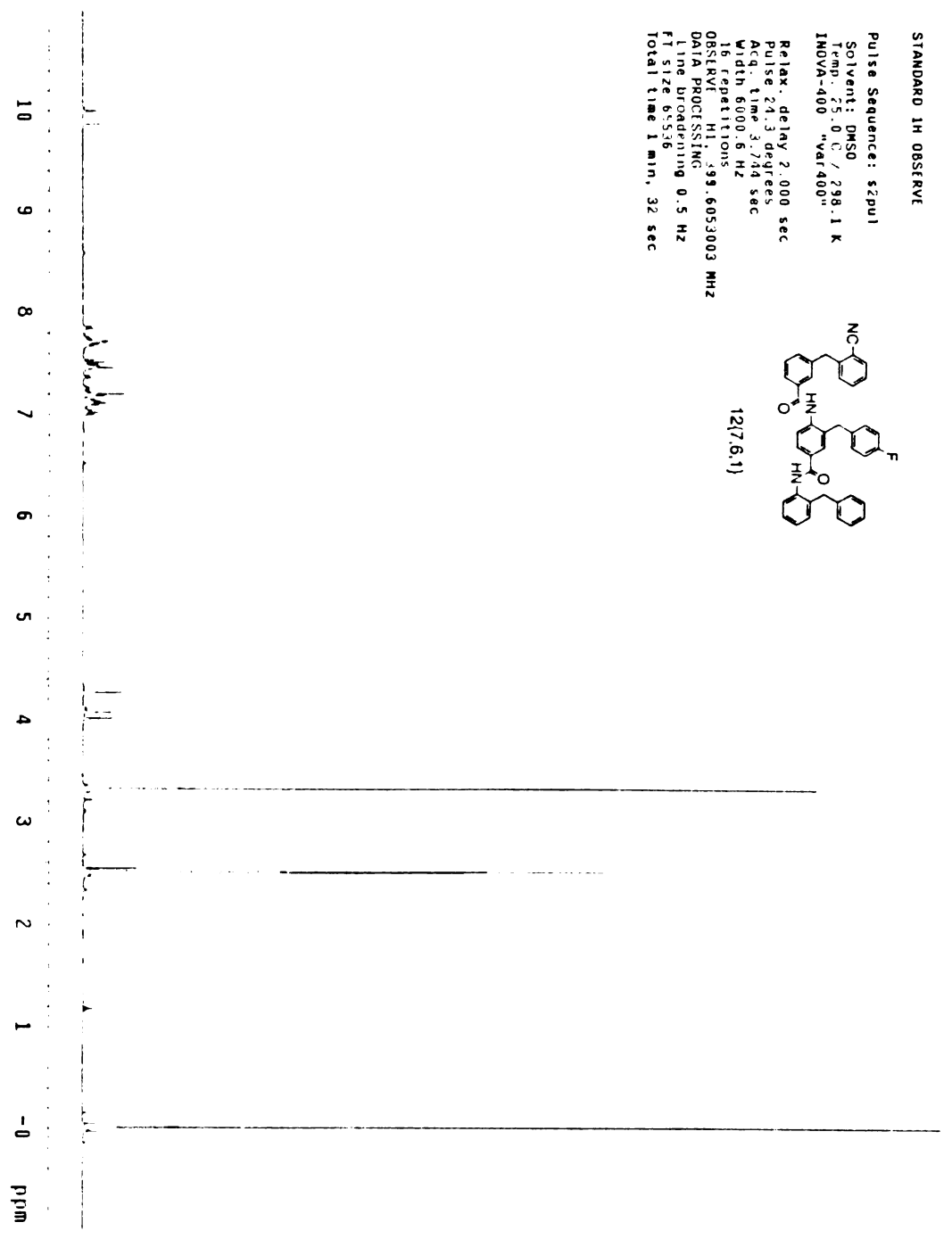


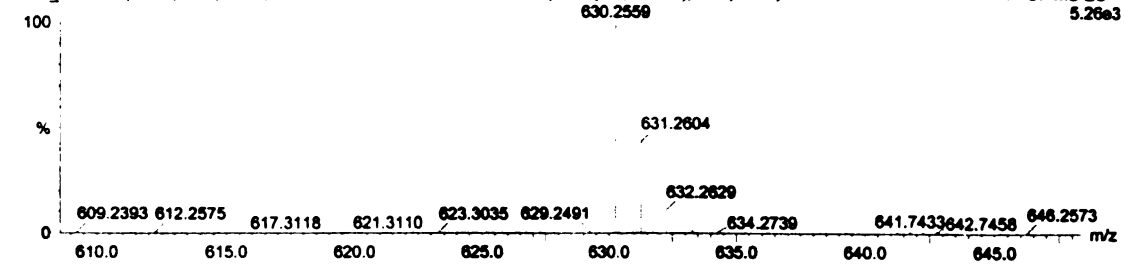
Figure A.12 HPLC and HRMS spectra for 12{7,6,1}

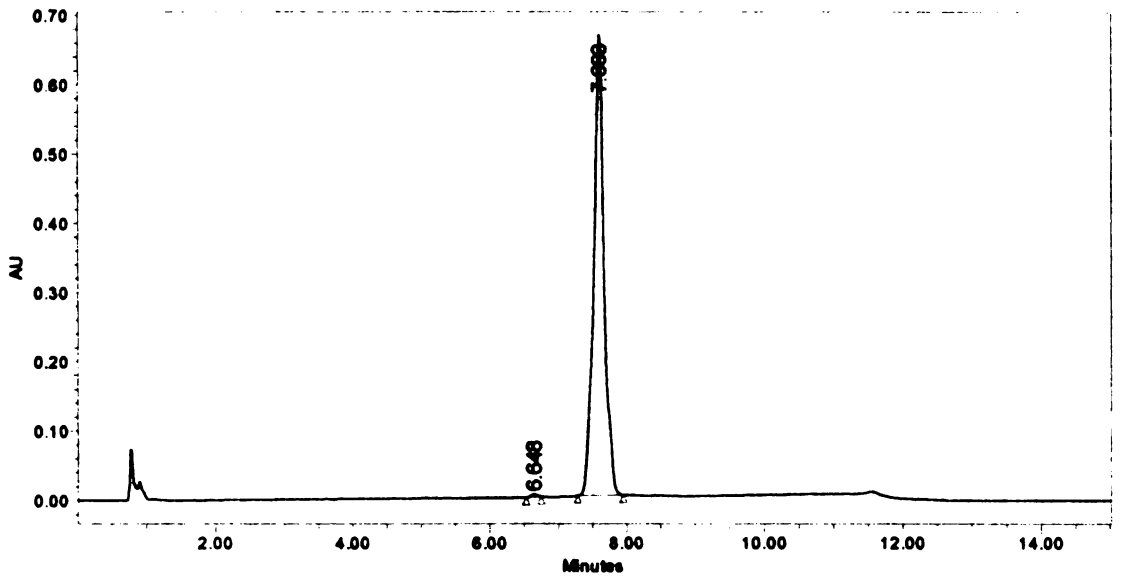
Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
630 formula(e) evaluated with 355 results within limits (up to 50 closest results for each mass)

Felice Lu, FLA7
Qtof_3437 35 (2.507) AM (Cen,3, 80.00, Ar,17500.0,716.46,0.70,LS 10); Sm (SG, 2x3.00); Cm (32:39) 1: TOF MS ES+
630.2559 5.26e3





	RT	Area	% Area	Height
1	6.648	23439	0.33	3689
2	7.608	7034555	99.67	664660

Figure A.13 ^{1}HNMR spectra for **12**{6,11,1}

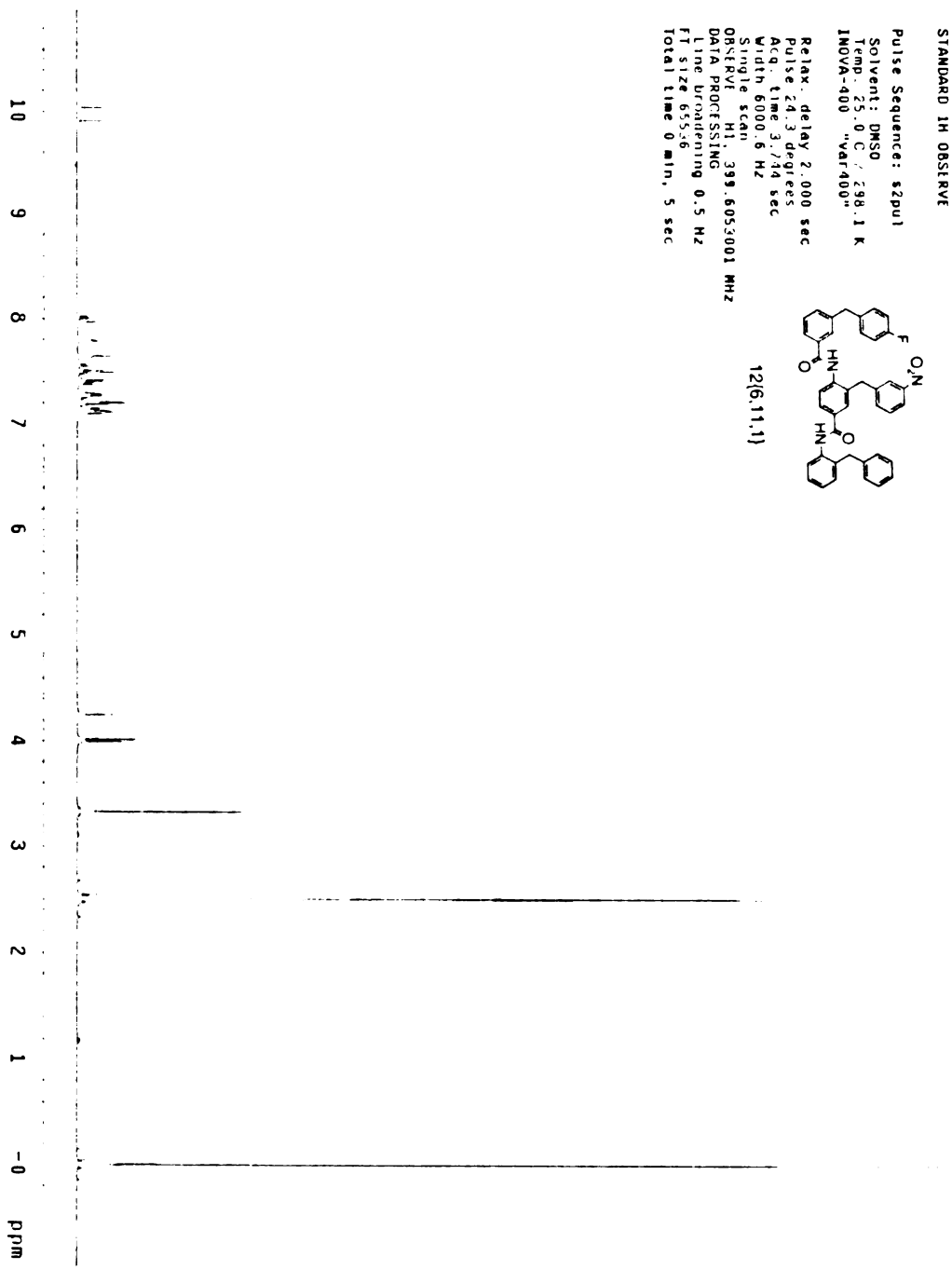
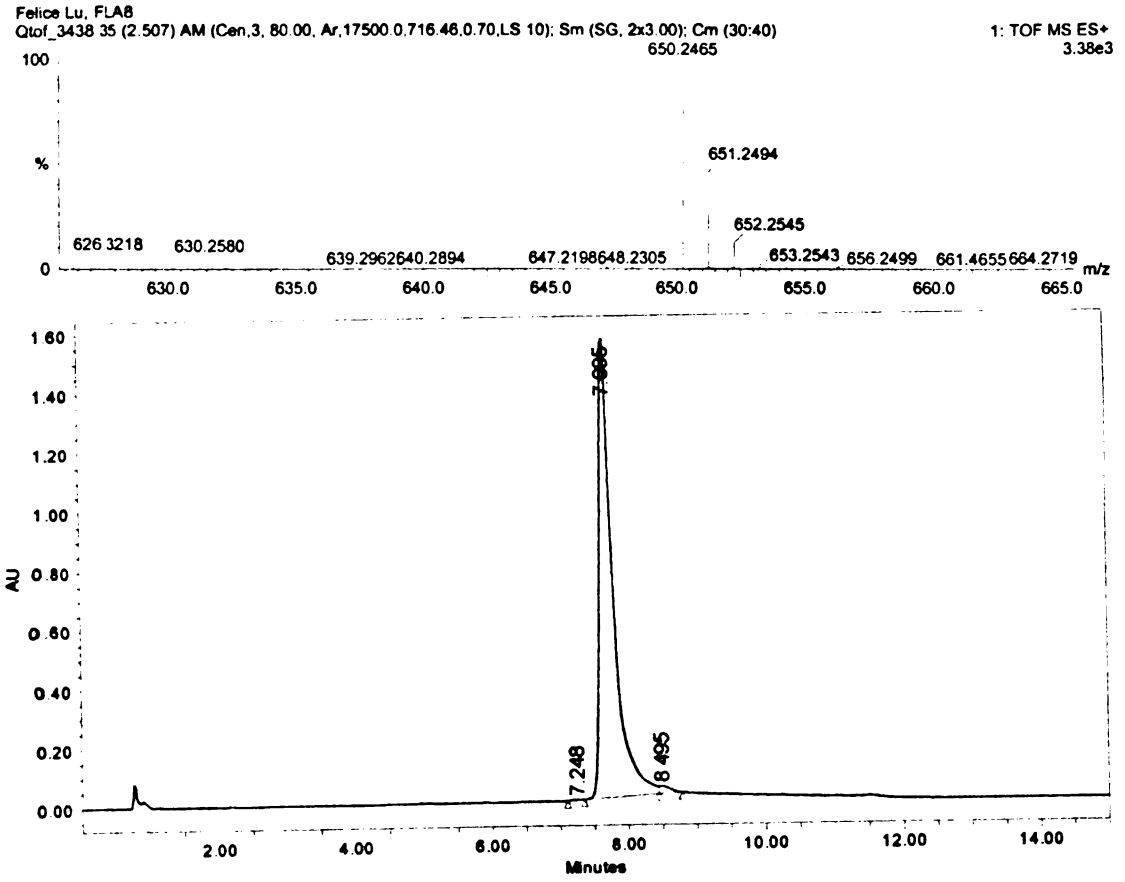


Figure A.14 HPLC and HRMS spectra for **12**{6,11,1}

Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
639 formula(e) evaluated with 348 results within limits (up to 50 closest results for each mass)



	RT	Area	% Area	Height
1	7.248	21323	0.10	2549
2	7.695	21608486	98.67	1551964
3	8.495	269986	1.23	25344

Figure A.15 ^{1}HNMR spectra for **12**{5,9,1}

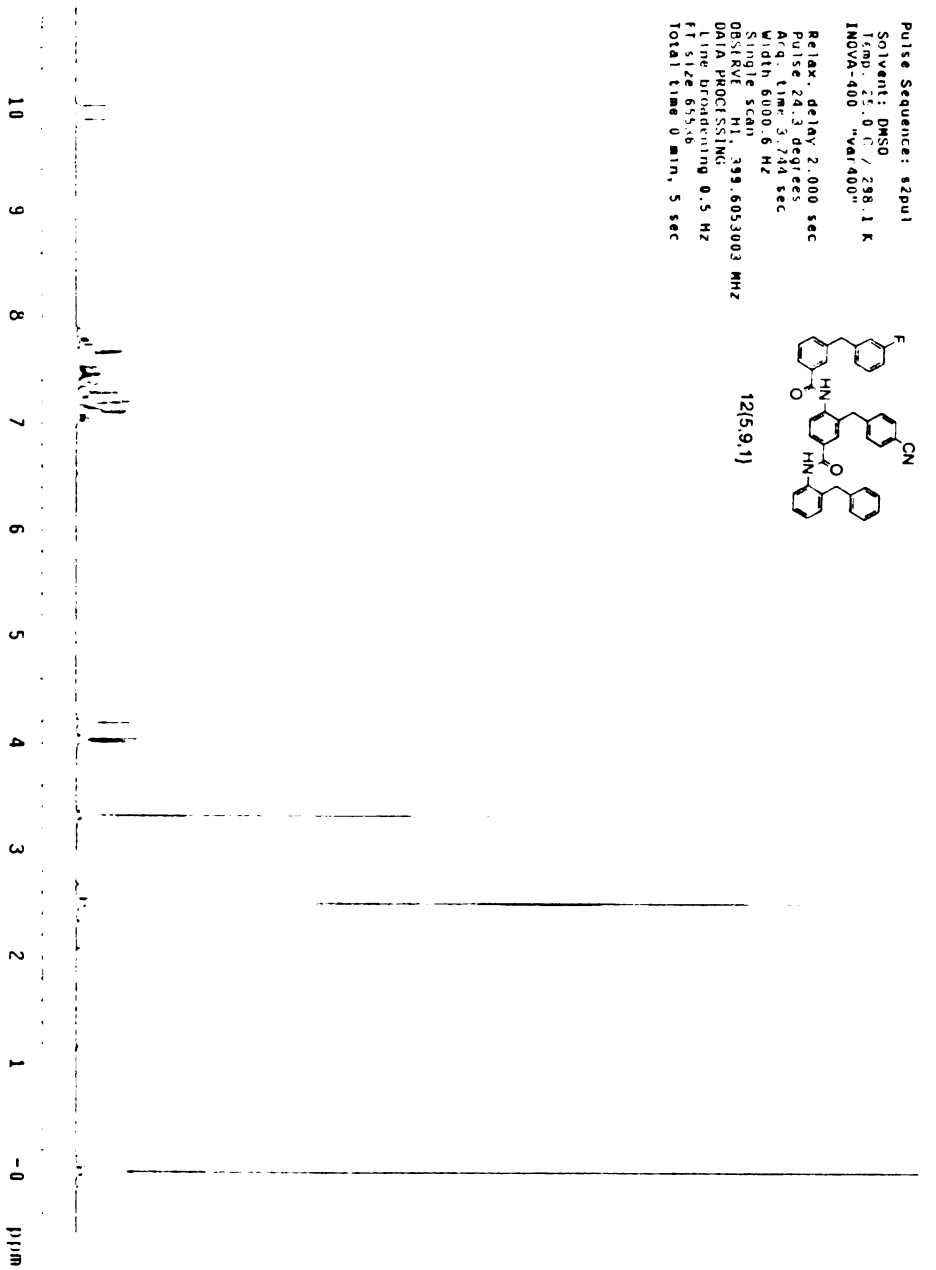
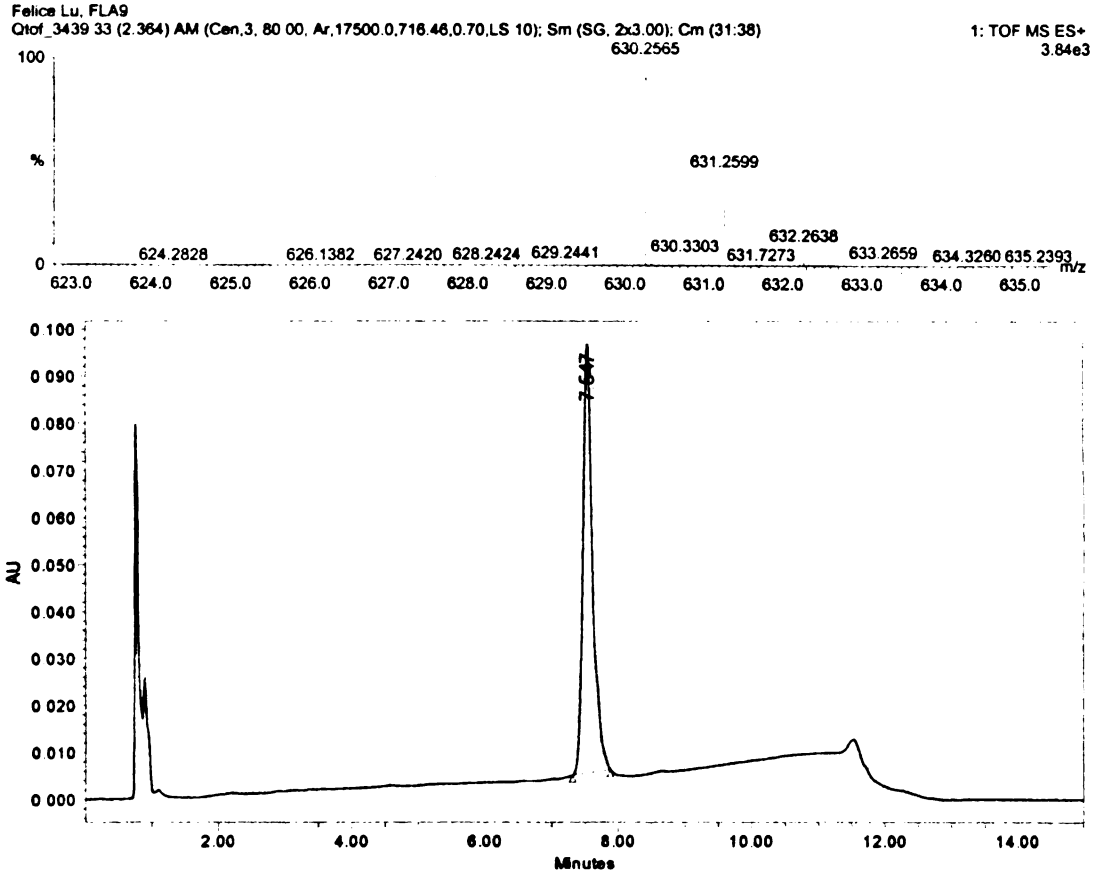


Figure A.16 HPLC and HRMS spectra for **12**{5,9,1}

Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
630 formula(e) evaluated with 356 results within limits (up to 50 closest results for each mass)



	RT	Area	% Area	Height
1	7.547	913991	100.00	91274

Figure A.17 ^{1}HNMR spectra for **12**{2,4,1}

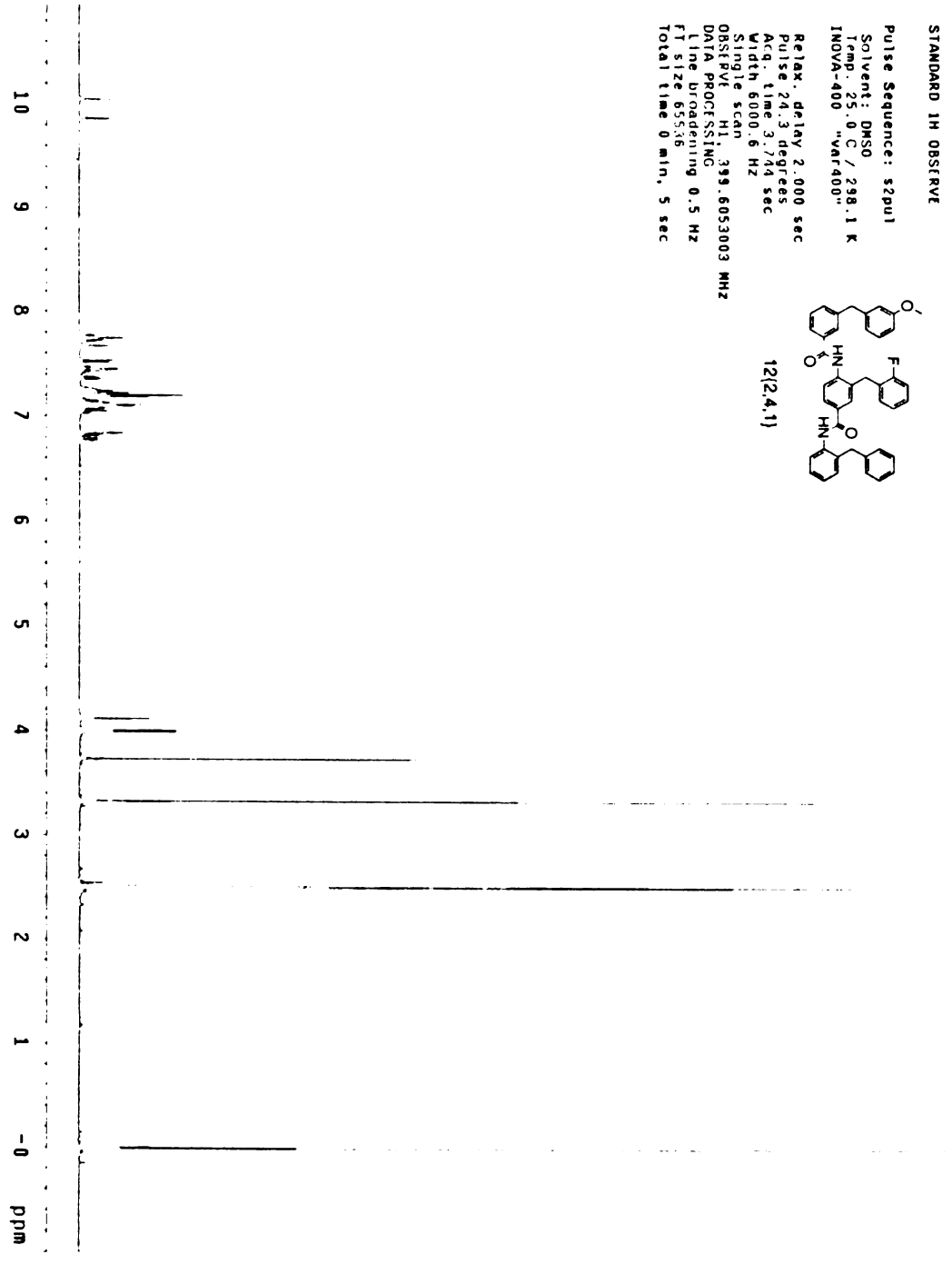
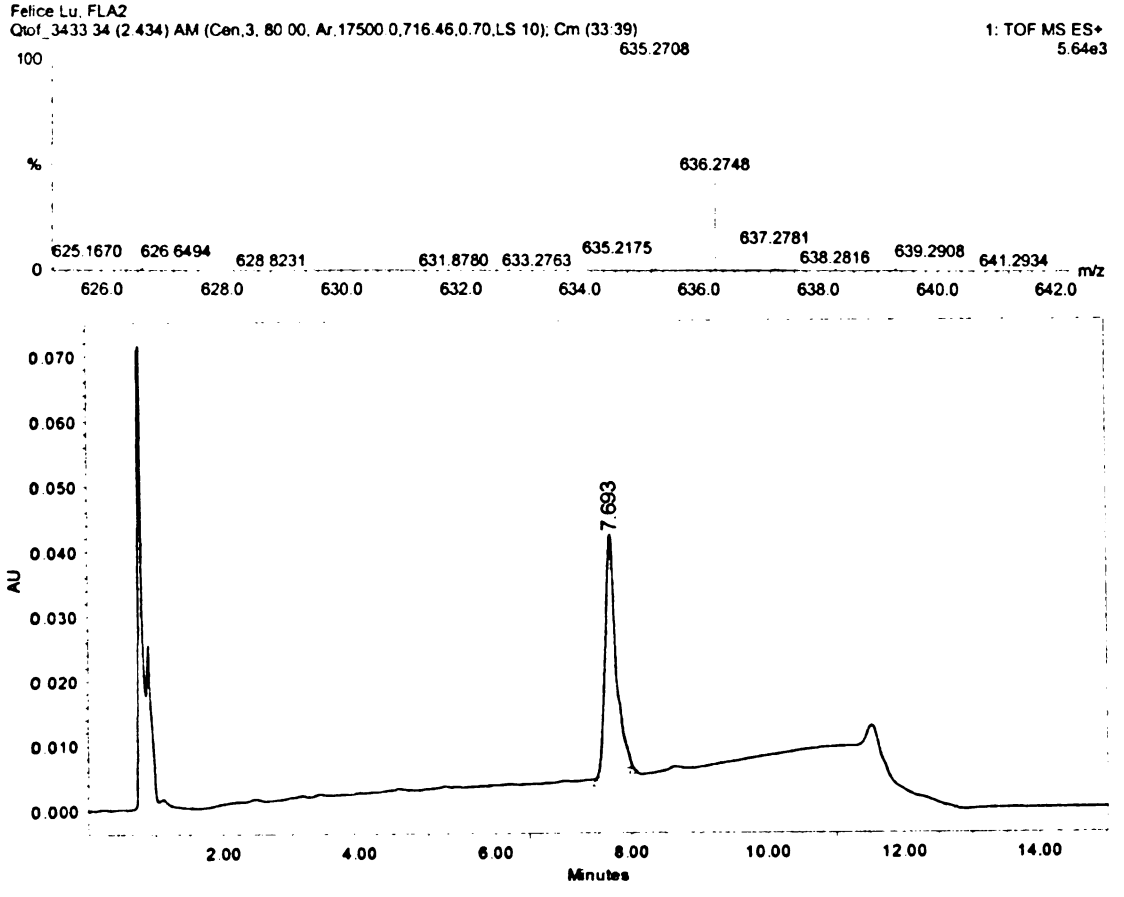


Figure A.18 HPLC and HRMS spectra for **12**{2,4,1}

Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
250 formula(e) evaluated with 134 results within limits (up to 50 closest results for each mass)



	RT	Area	% Area	Height
1	7.693	407029	100.00	37080

Figure A.19 ^{1}HNMR spectra for **12**{1,1,1}

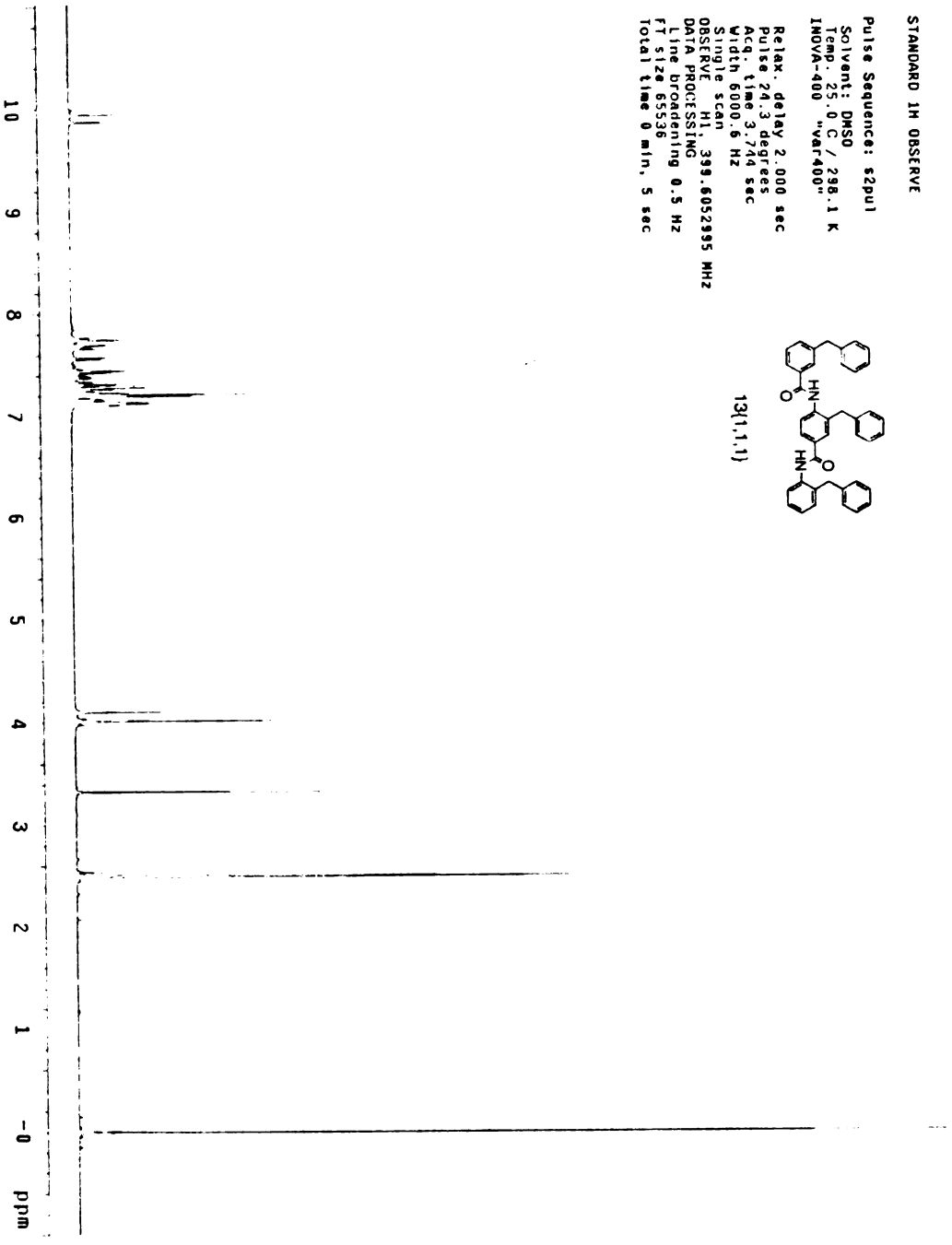


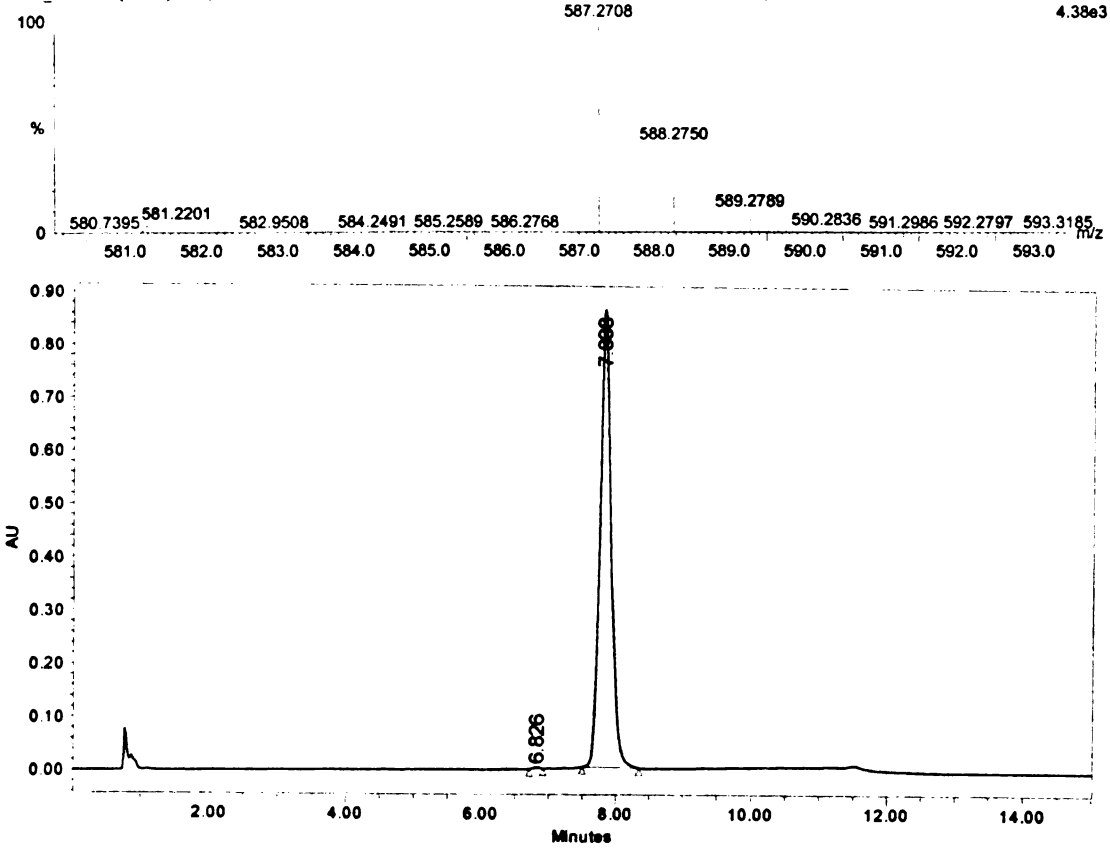
Figure A.20 HPLC and HRMS spectra for **12**{1,1,1}

Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 200.0 mDa / DBE: min = -1.5, max = 50.0
Isotope cluster parameters: Separation = 1.0 Abundance = 1.0%

Monoisotopic Mass, Odd and Even Electron Ions
592 formula(e) evaluated with 360 results within limits (up to 50 closest results for each mass)

Felice Lu, FLA11
Qtof_3441 35 (2.510) AM (Cen,3, 80.00, Ar,17500.0,716.46,0.70,LS 10); Sm (SG, 2x3.00); Cm (32:38) 1: TOF MS ES+
4.38e3



	RT	Area	% Area	Height
1	6.826	19495	0.20	3021
2	7.838	9931954	99.80	863211

Figure A.21 ^{1}HNMR spectra for **11**{11,6}

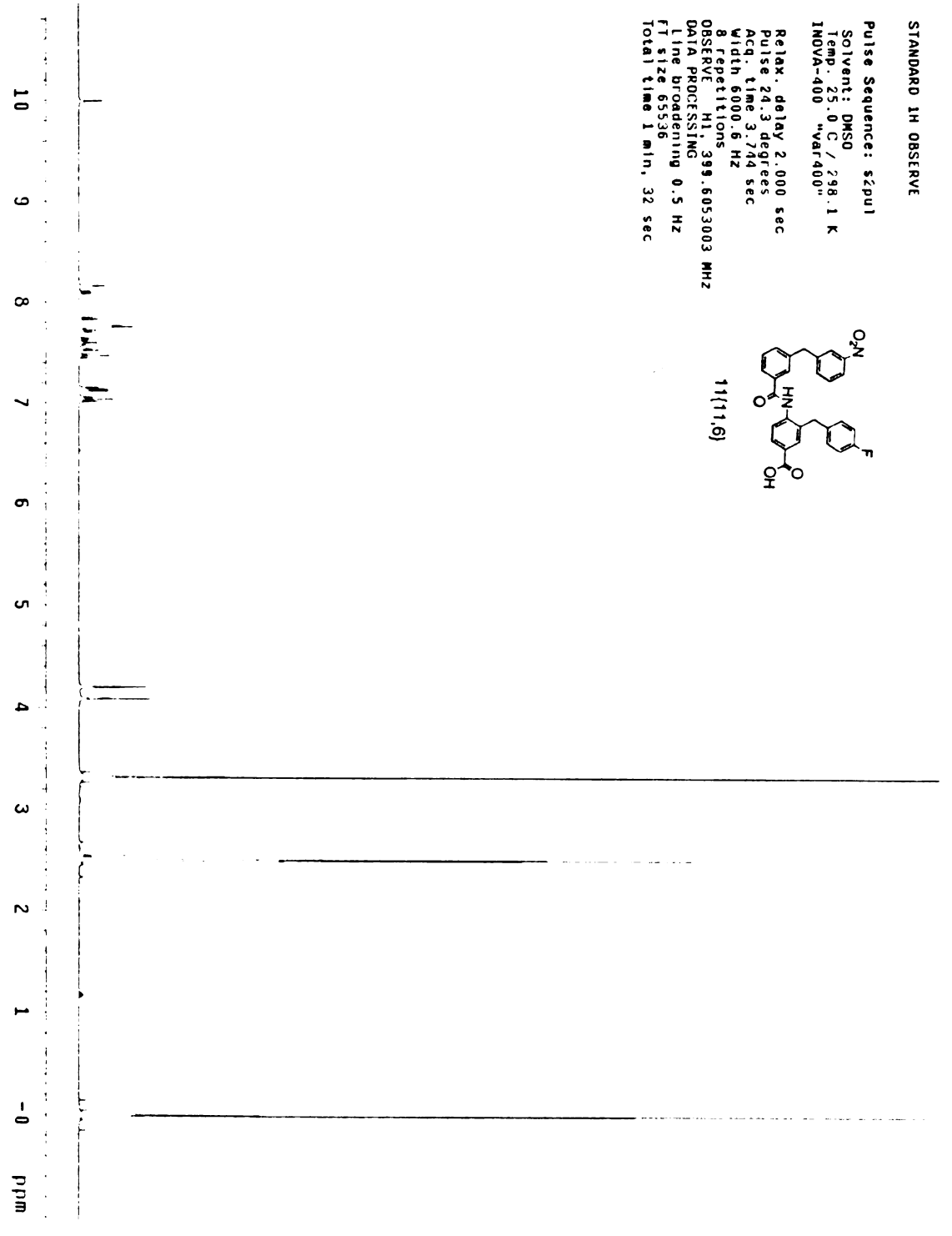
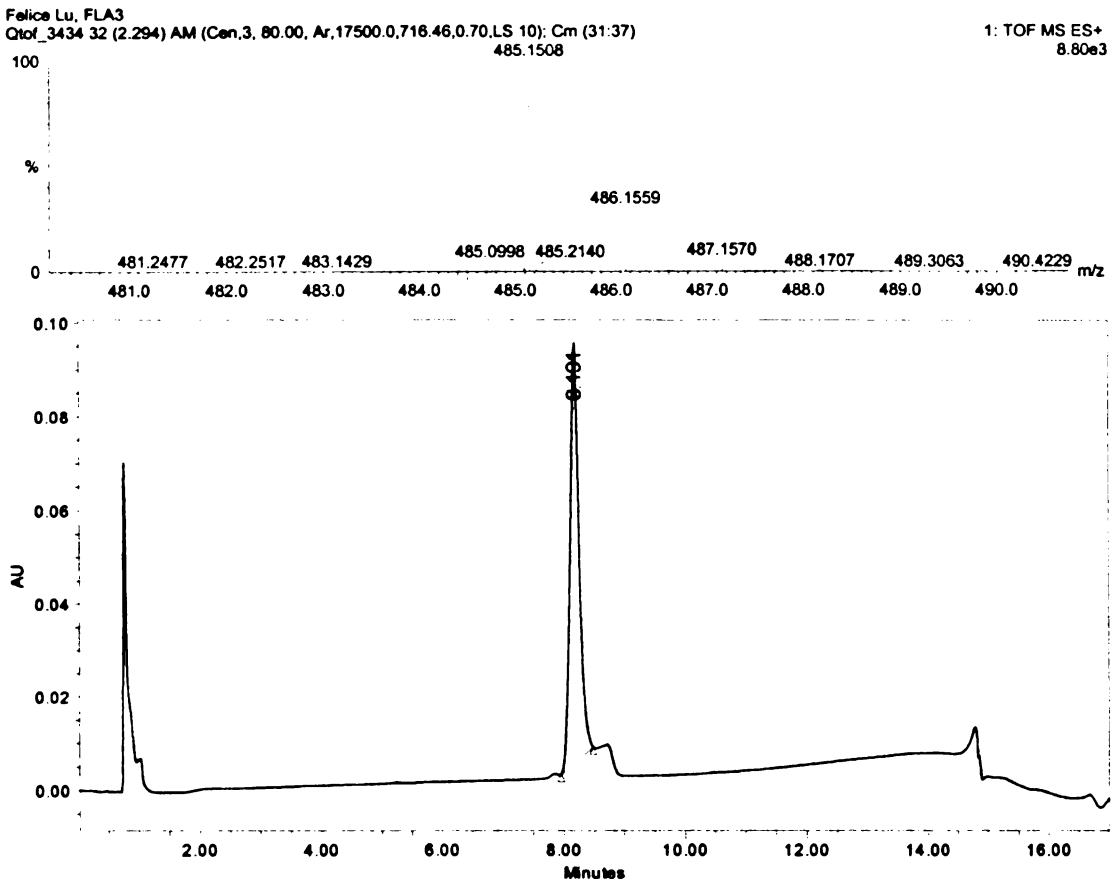


Figure A.22 HPLC and HRMS spectra for **11**{11,6}



	RT	Area	% Area	Height
1	8.191	921234	100.00	89495

Figure A.23 ^{1}HNMR spectra for **11**{2,4}

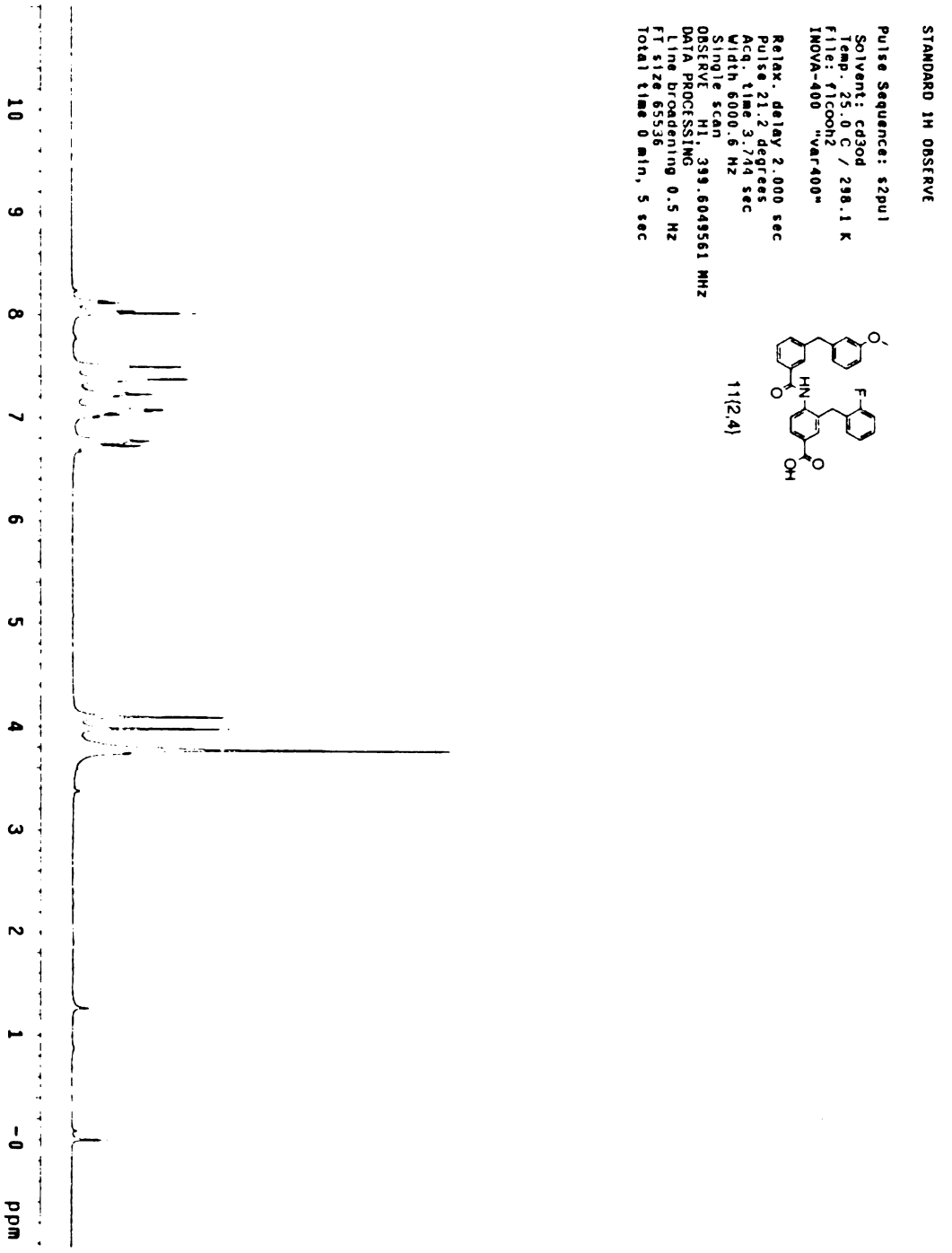
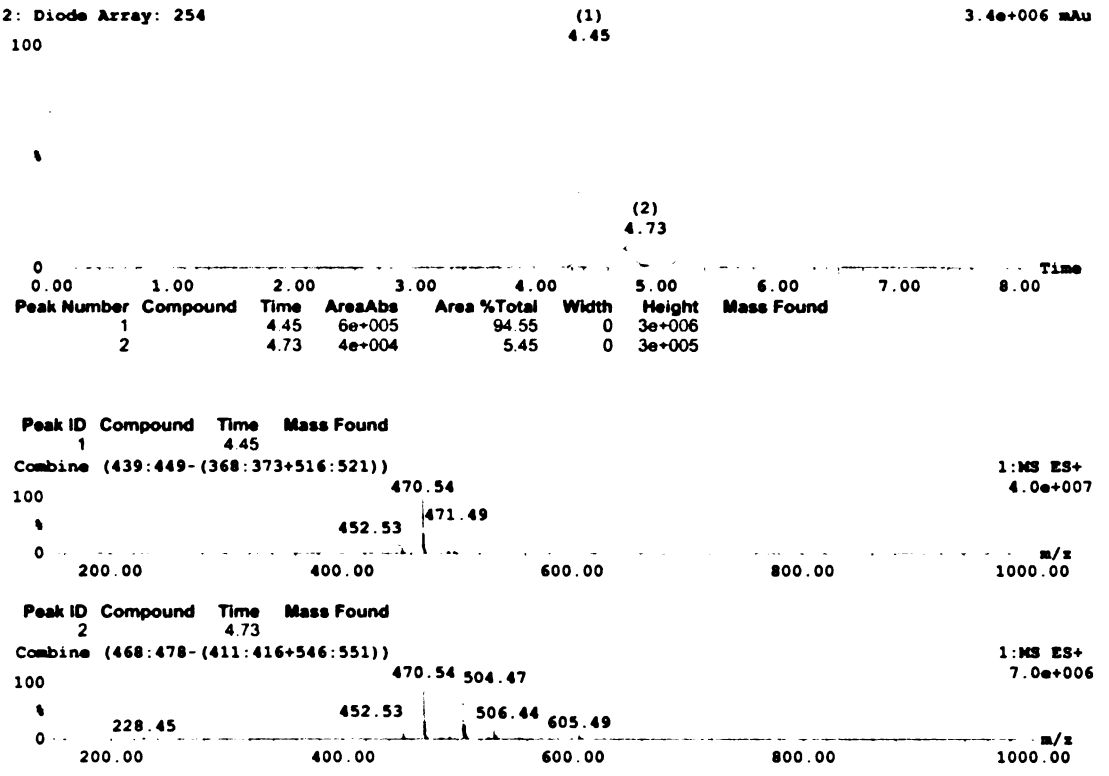


Figure A.24 LCMS spectra for **11**{2,4}

Sample Report:

Sample 7 Vial 1:2,B ID File fl-b2 Date 06-Jun-2005 Time 15:44:20 Description



Peak Number	Compound	Time	AreaAbs	Area %Total	Width	Height	Mass Found
1		4.45	6e+005	94.55	0	3e+006	
2		4.73	4e+004	5.45	0	3e+005	

Peak ID	Compound	Time	Mass Found
1		4.45	

Peak ID	Compound	Time	Mass Found
2		4.73	

Figure A.25 ^{1}HNMR spectra for **11**{13,1}

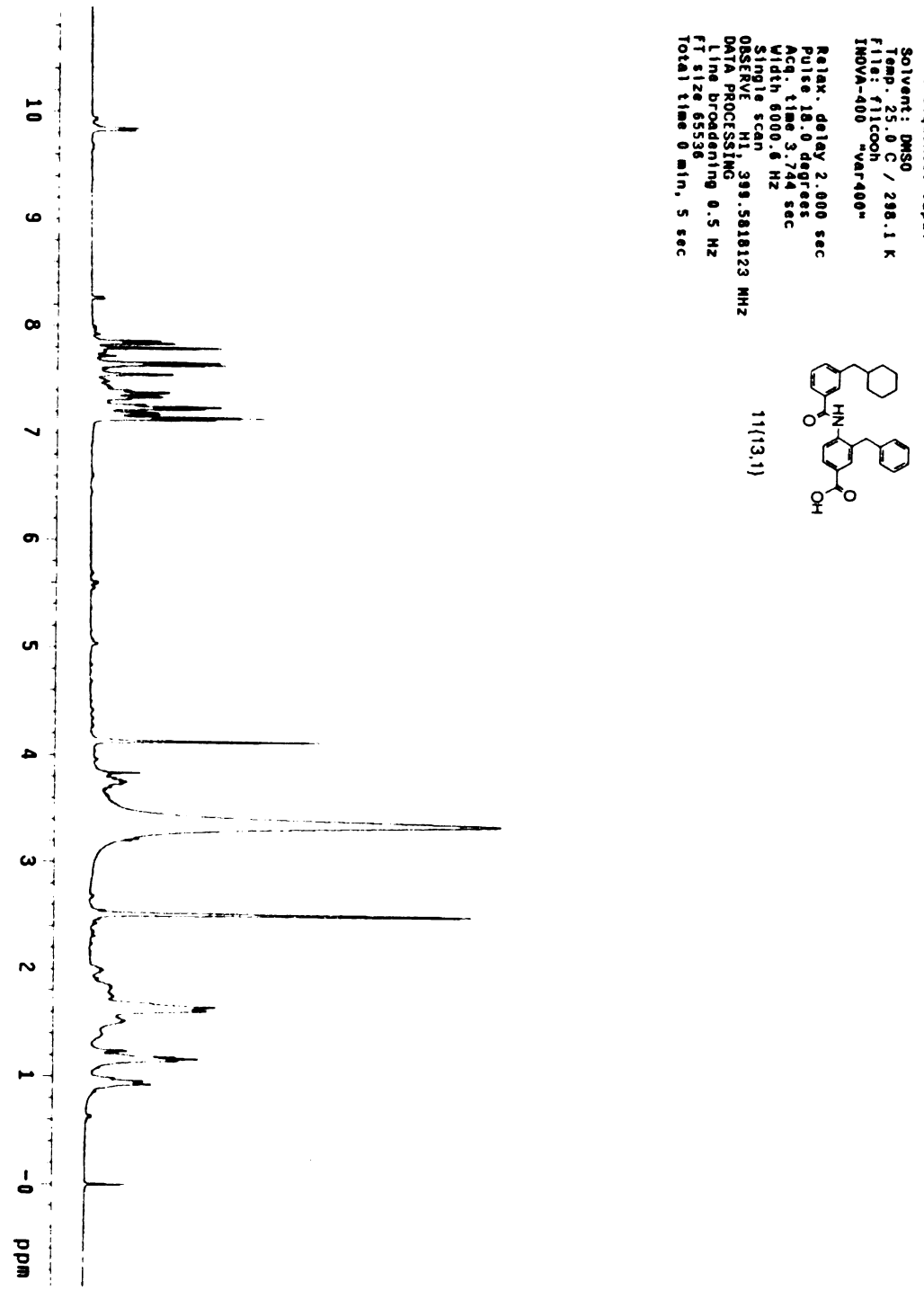


Figure A.26 LCMS spectra for **11**{13,1}

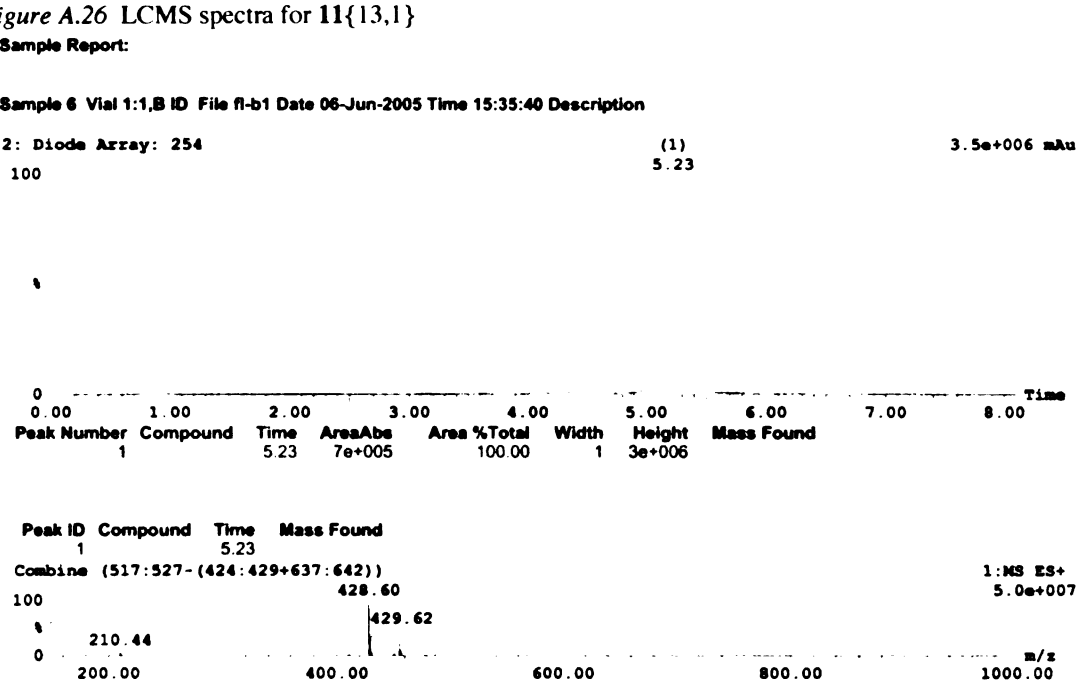


Figure A.27 ^{1}HNMR spectra for **11**{1,7}

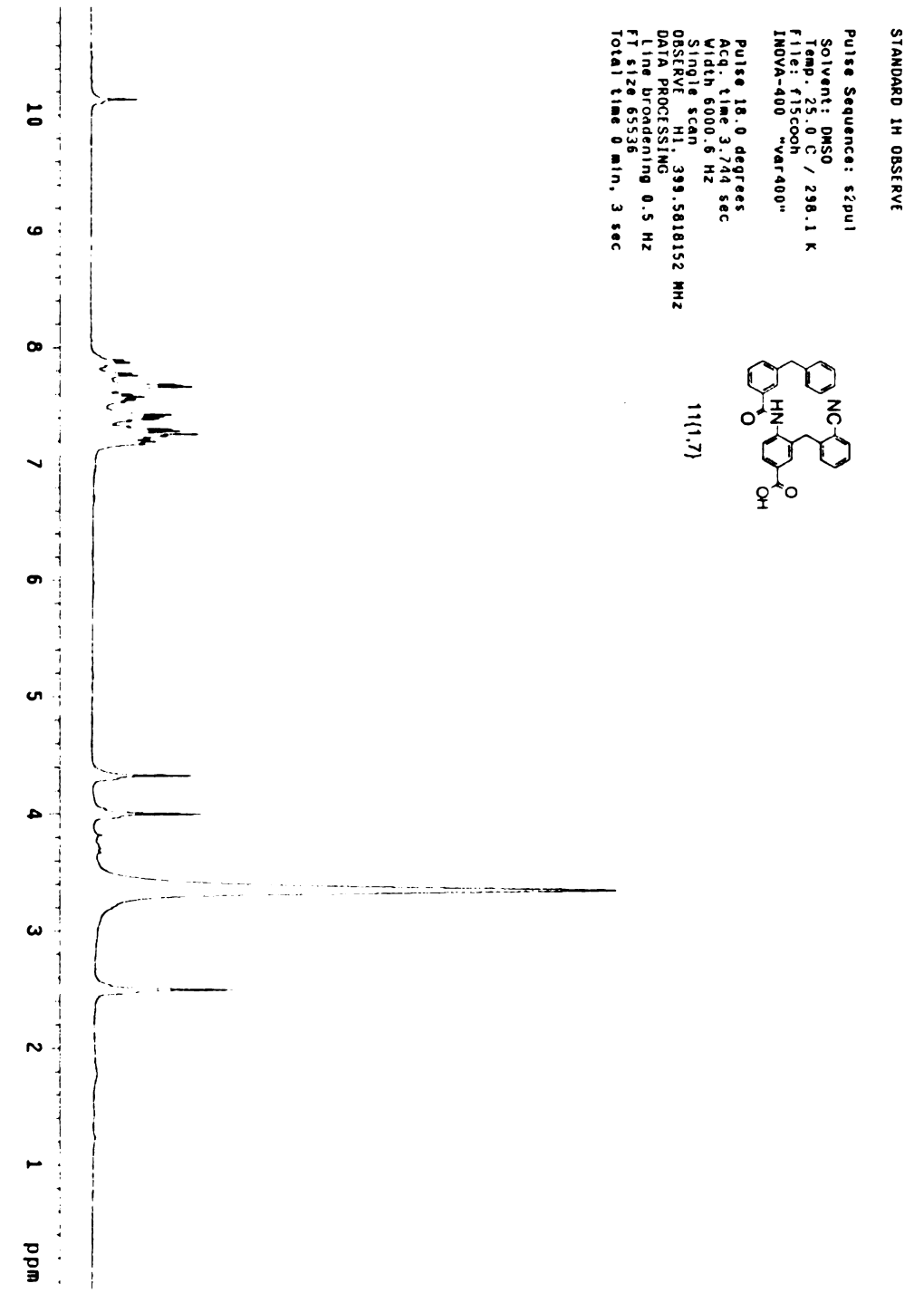


Figure A.28 LCMS spectra for **11**{1,7}

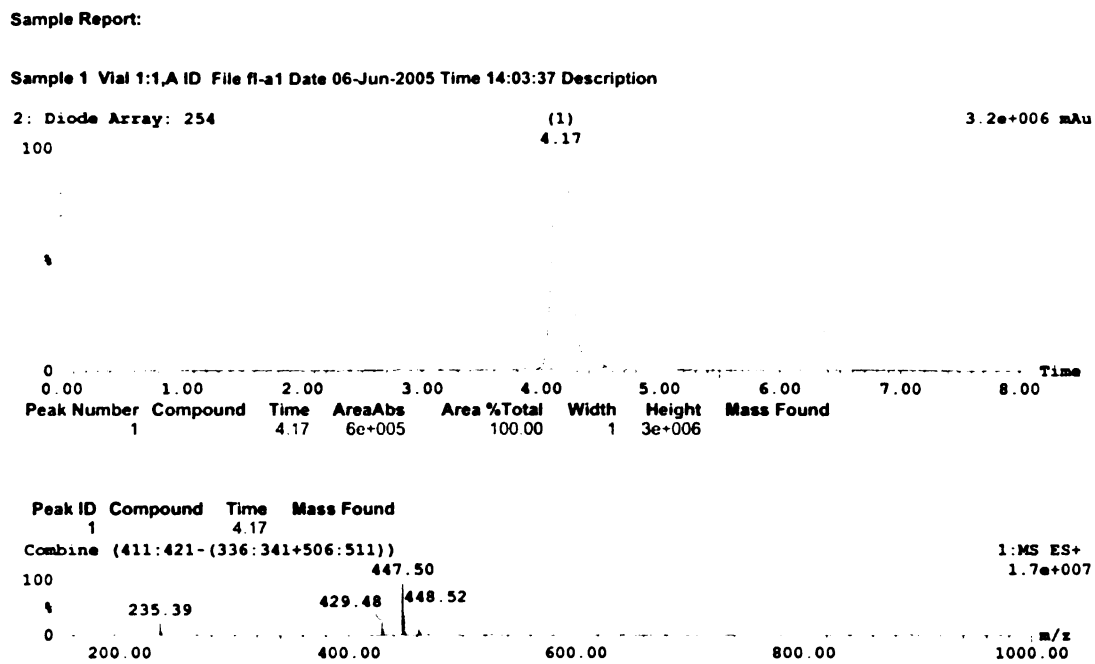


Figure A.29 ^{1}HNMR spectra for **11**{9,8}

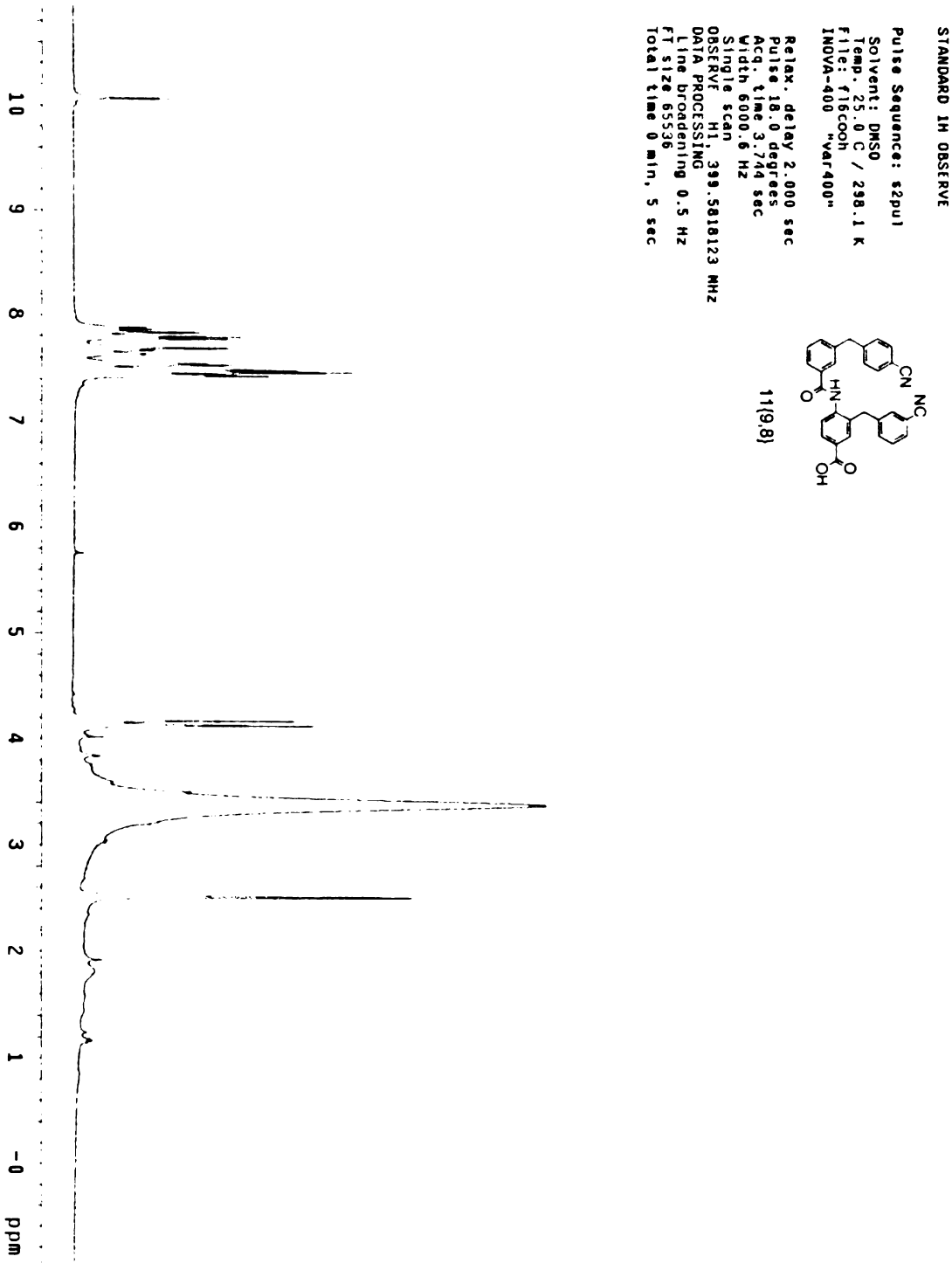


Figure A.30 LCMS spectra for **11**{9,8}

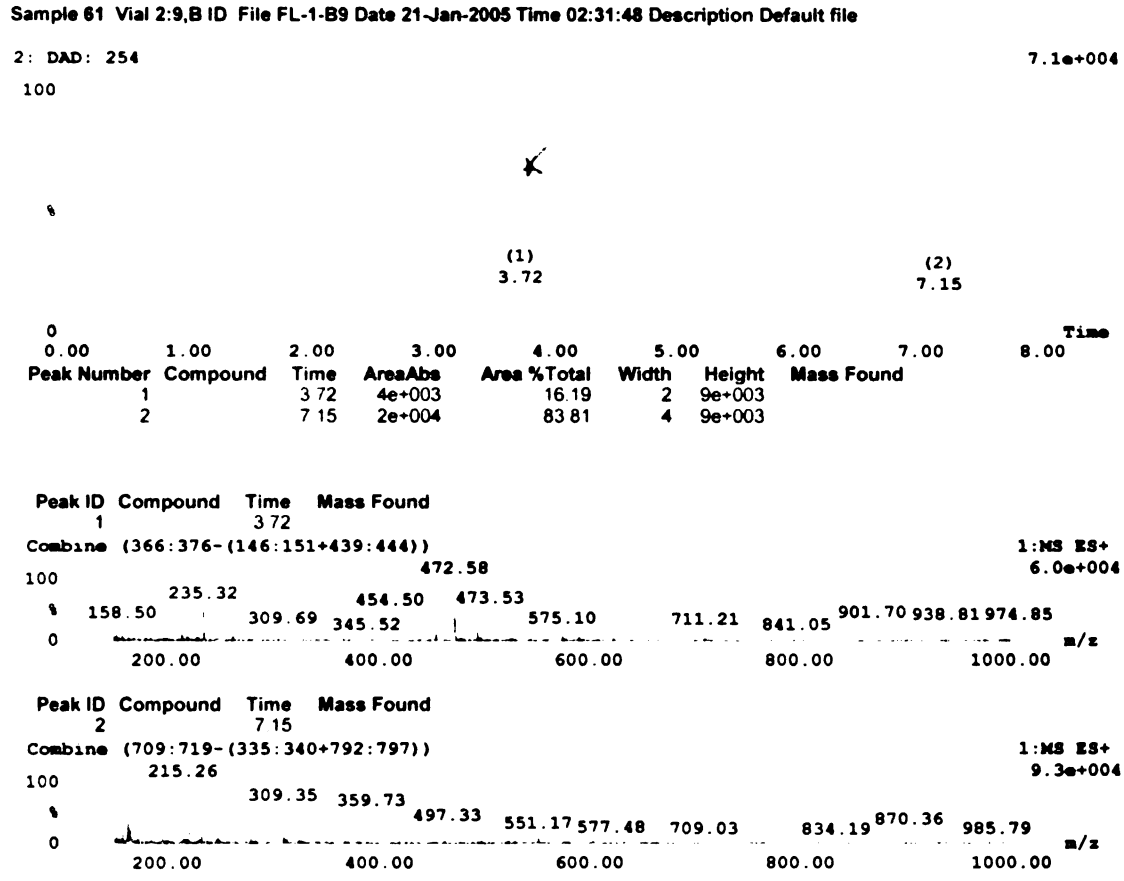


Figure A.31 ^{1}HNMR spectra for **11**{8,9}

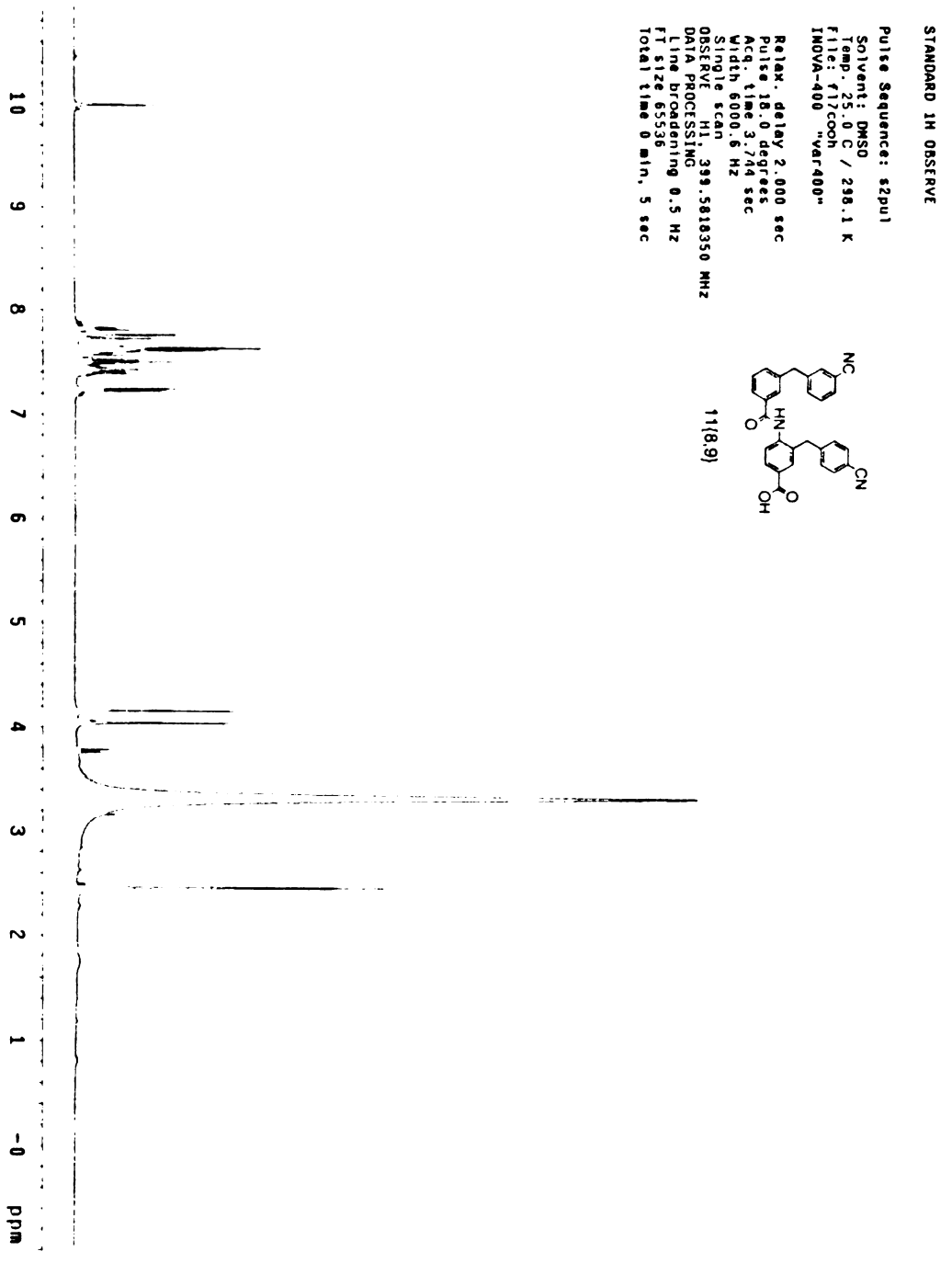
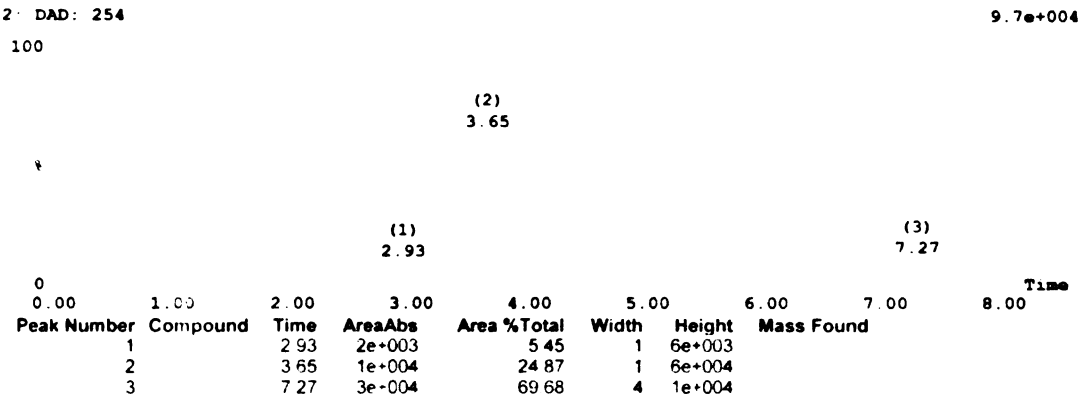


Figure A.32 LCMS spectra for **11**{8,9}

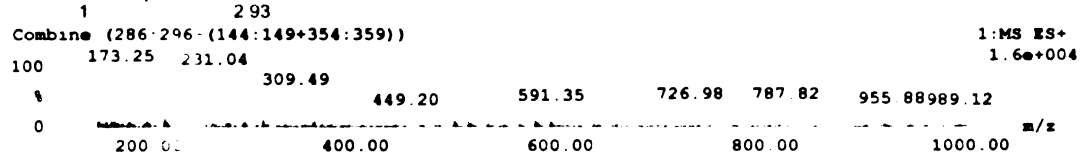
Sample Report:

Sample 130 Vial 2:10.C ID File FL-3-C10 Date 24-Jan-2005 Time 14:36:19 Description



Peak Number	Compound	Time	AreaAbs	Area %Total	Width	Height	Mass Found
1		2 93	2e+003	5 45	1	6e+003	
2		3 65	1e+004	24 87	1	6e+004	
3		7 27	3e+004	69 68	4	1e+004	

Peak ID 1 **Compound** **Time** 2 93 **Mass Found**



Peak ID 2 **Compound** **Time** 3 65 **Mass Found**

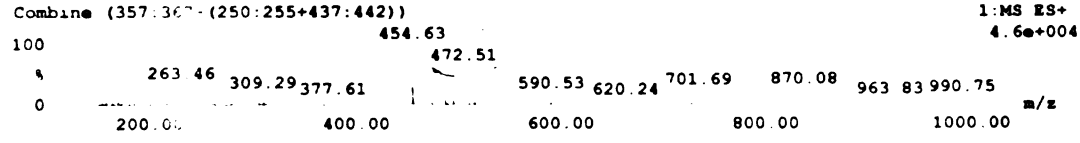


Figure A.33 ^{1}HNMR spectra for **11**{5,9}

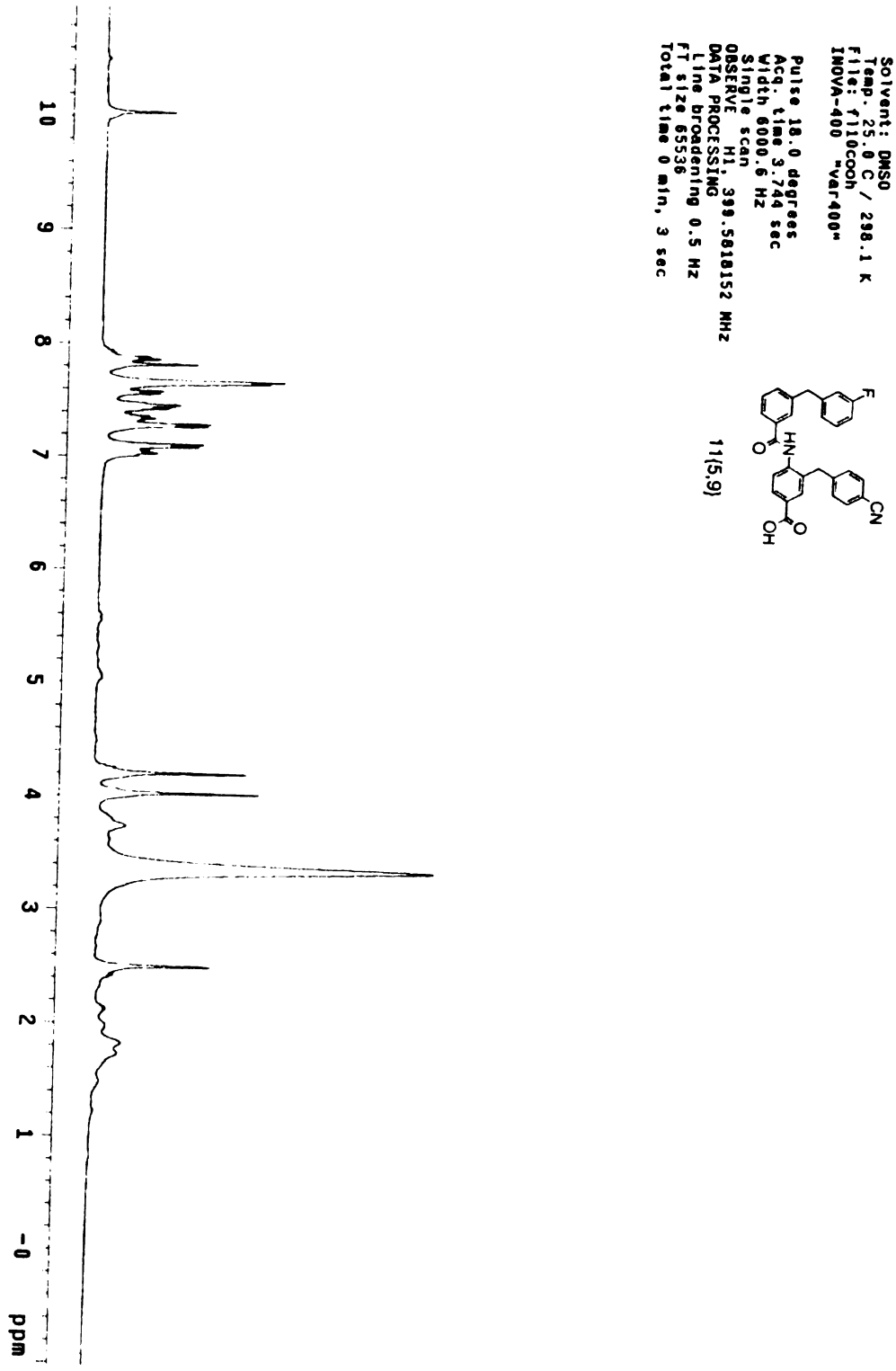


Figure A.34 LCMS spectra for **11**{5,9}

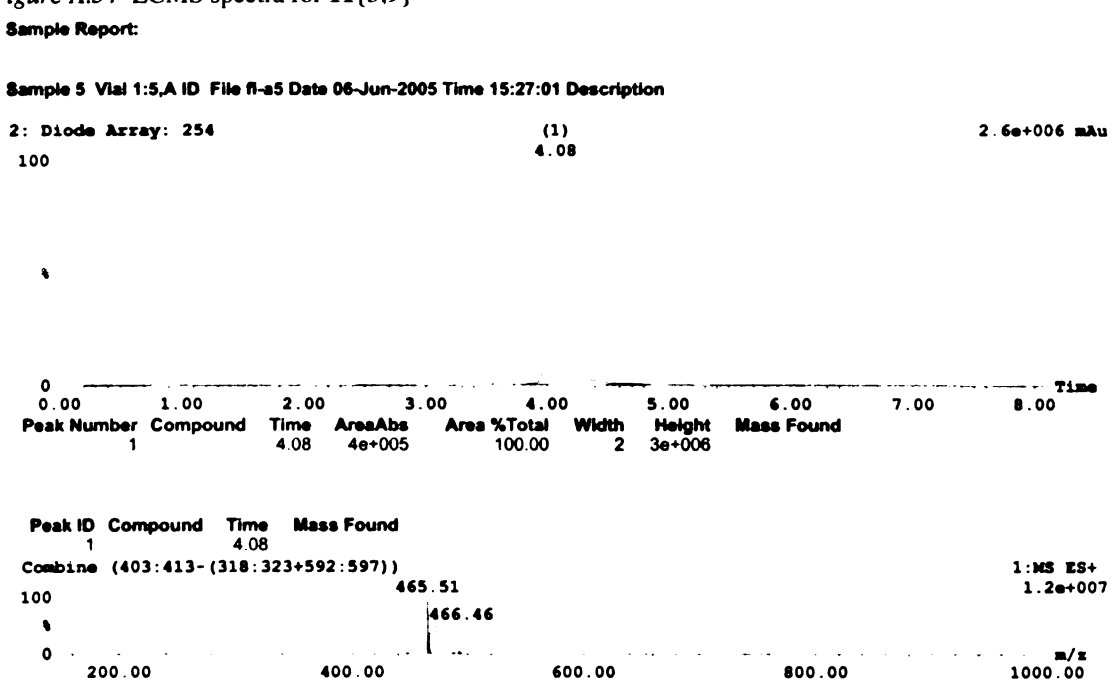


Figure A.35 ^{1}HNMR spectra for **11**{4,2}

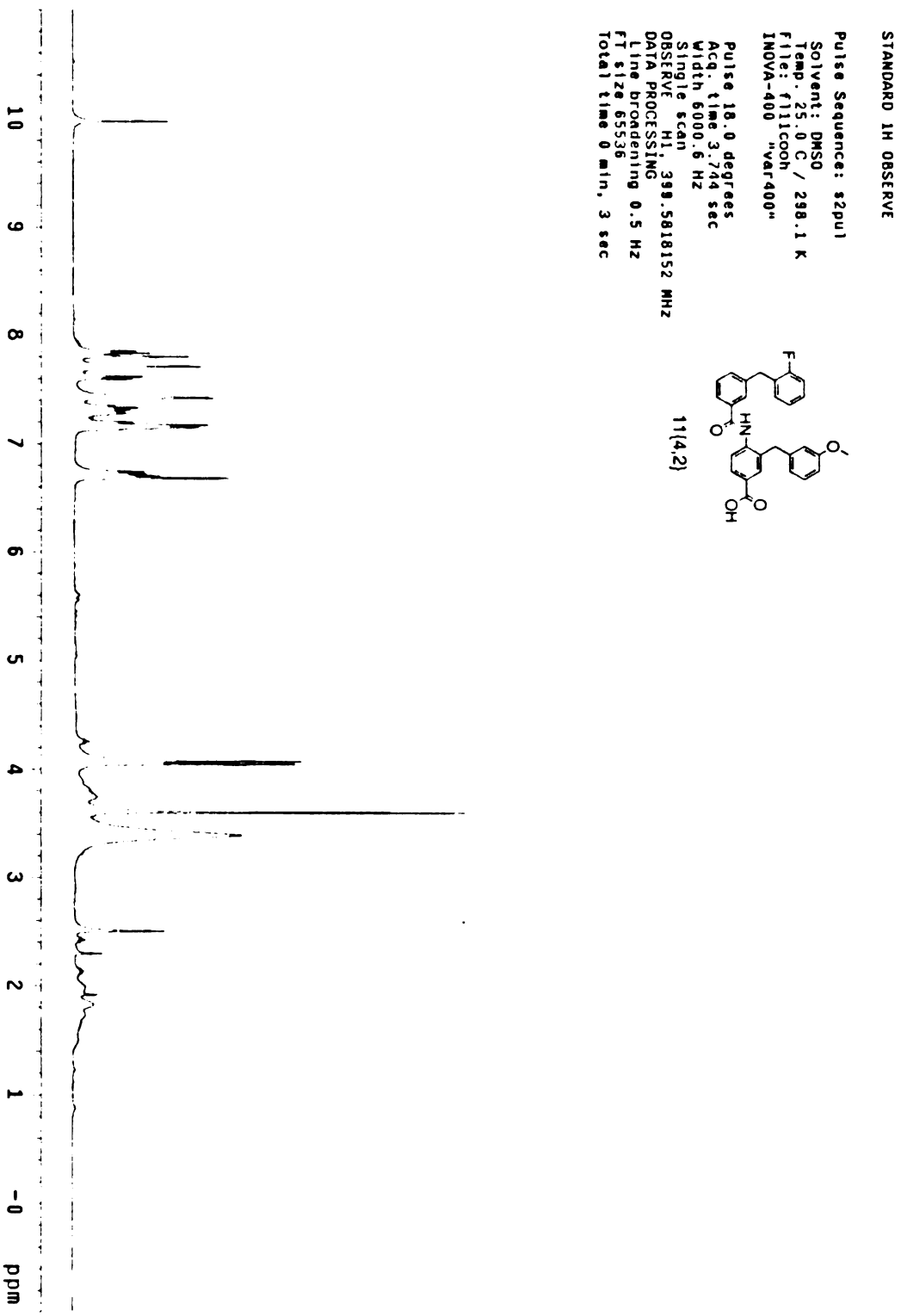


Figure A.36 LCMS spectra for **11**{4,2}

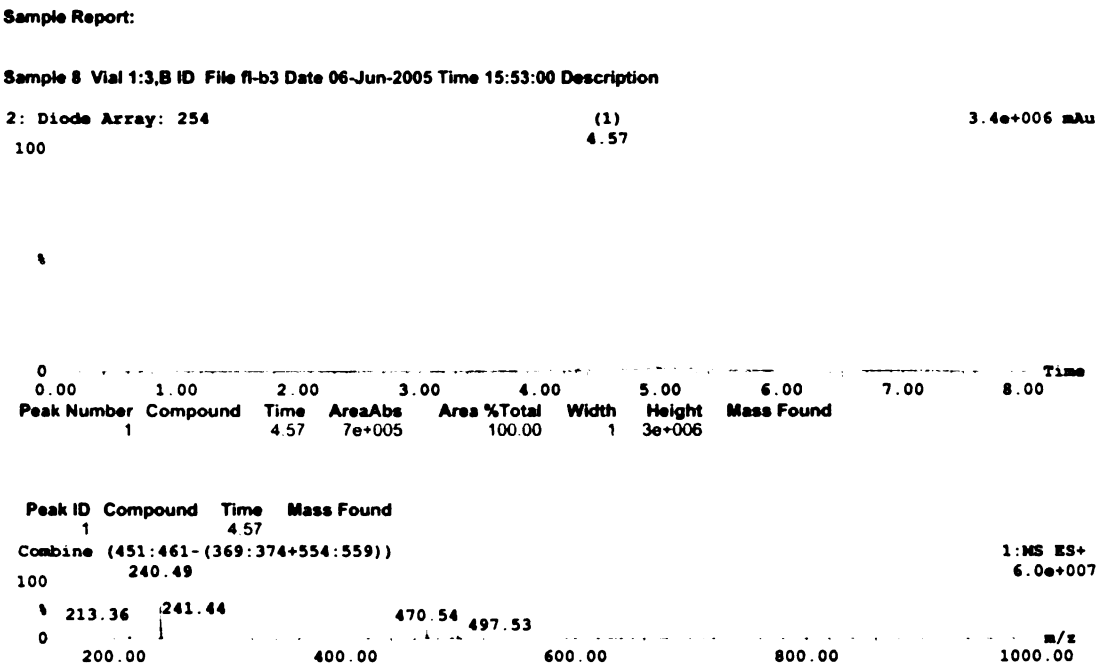


Figure A.37 ^{1}HNMR spectra for **11**{7,6}

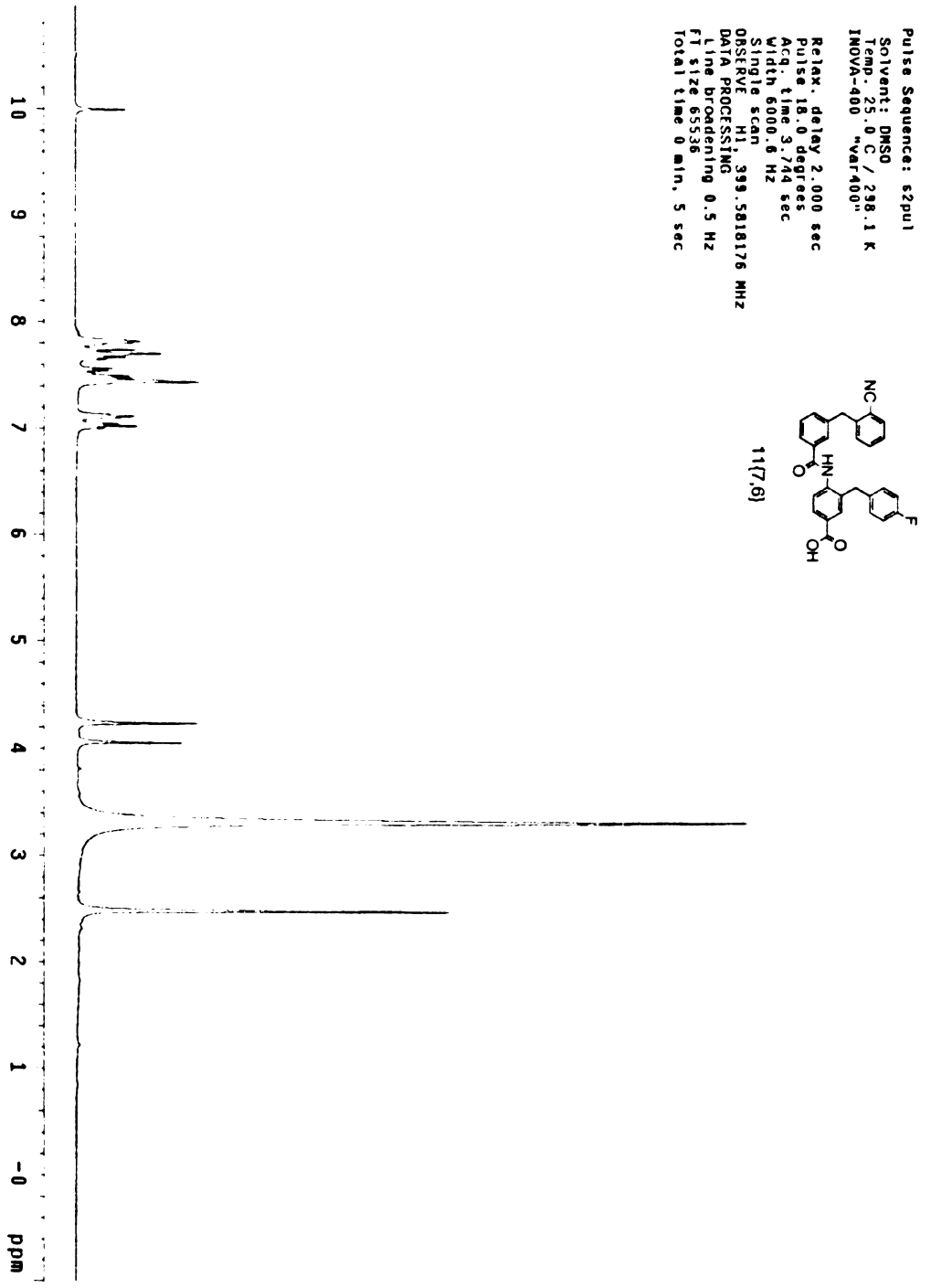


Figure A.38 LCMS spectra for **11**{7,6}

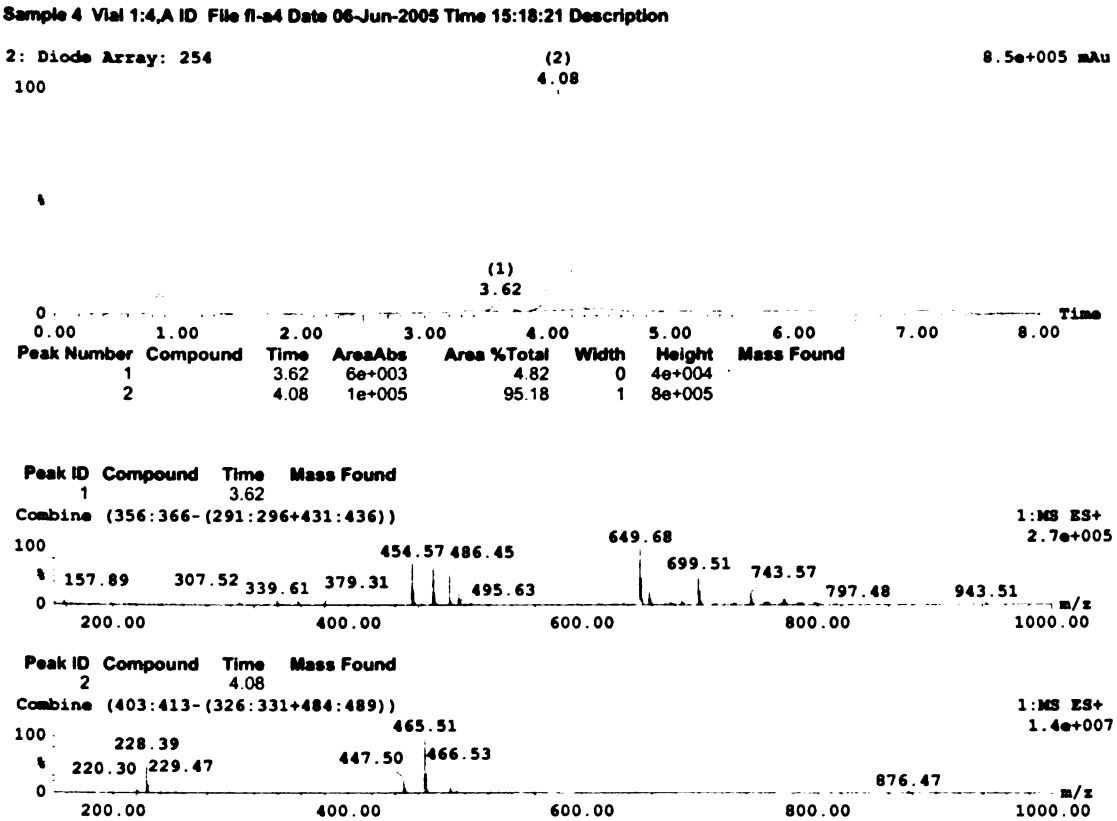


Figure A.39 ^{1}HNMR spectra for **11**{1,1}

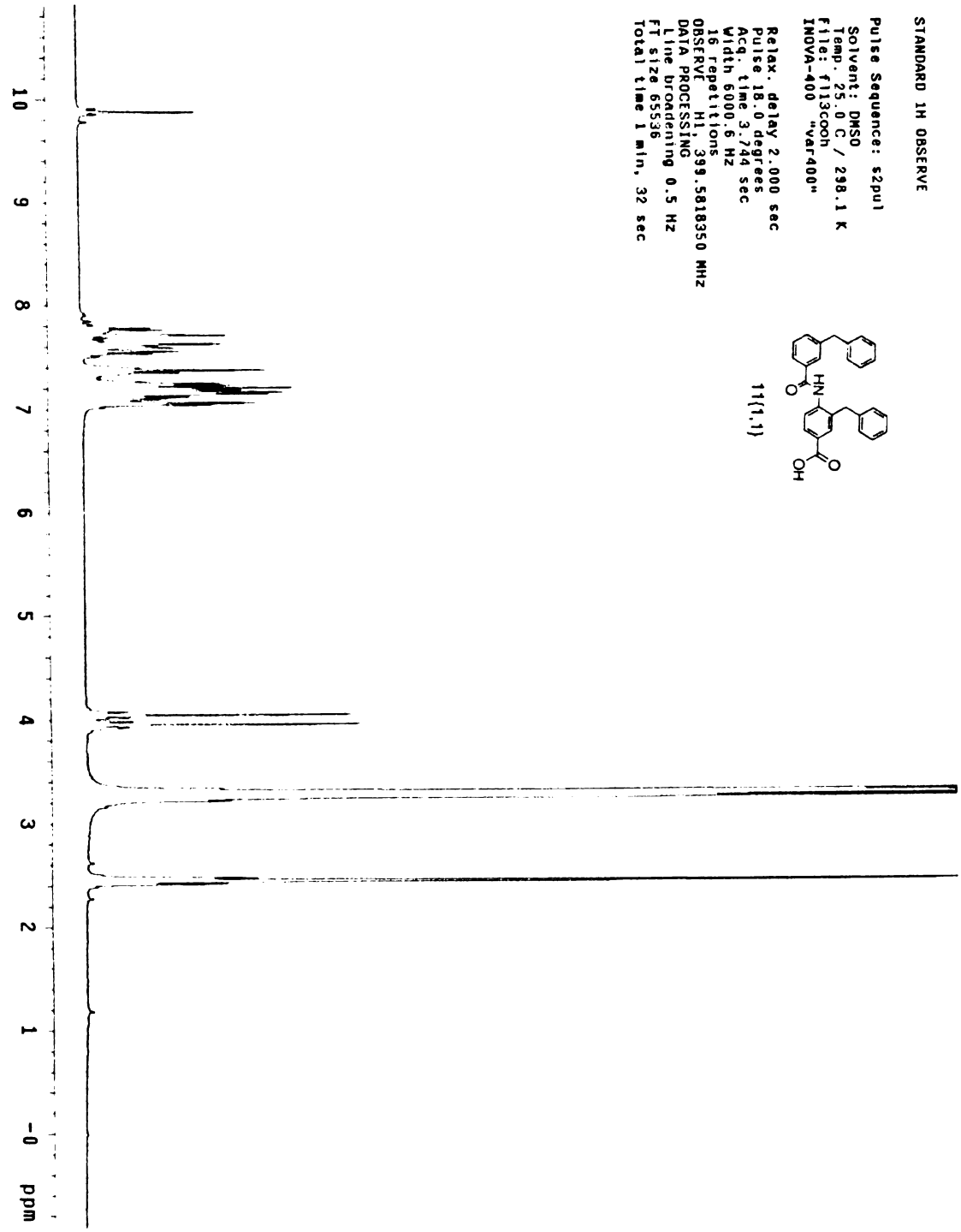


Figure A.40 LCMS spectra for **11**{1,1}

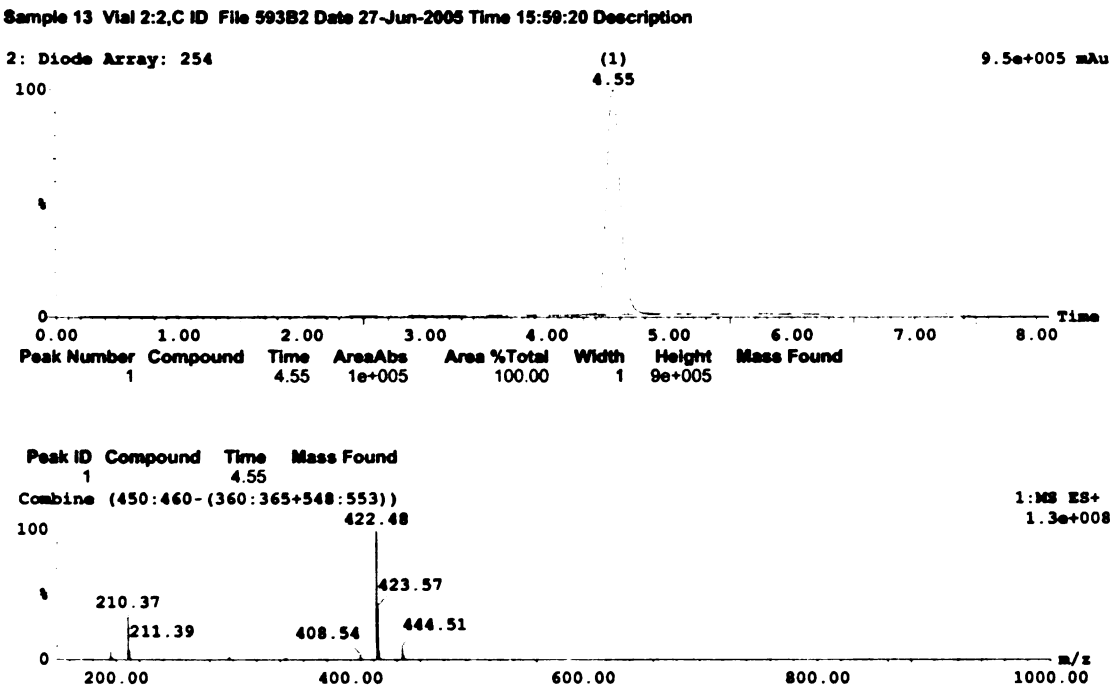


Figure A.41 ^{1}H NMR spectra for **2**{1}

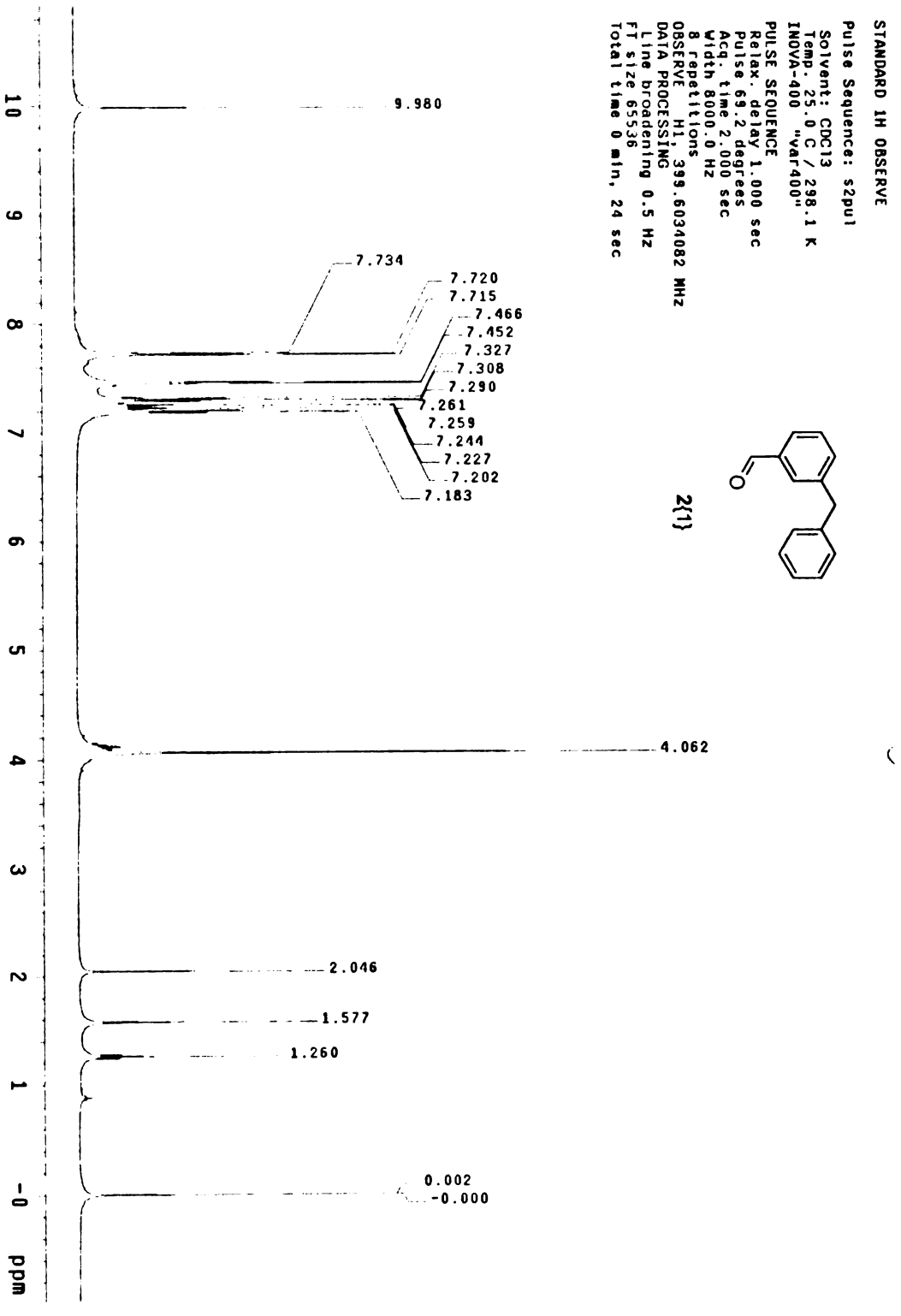


Figure A.42 ^{1}H NMR spectra for 3{1}

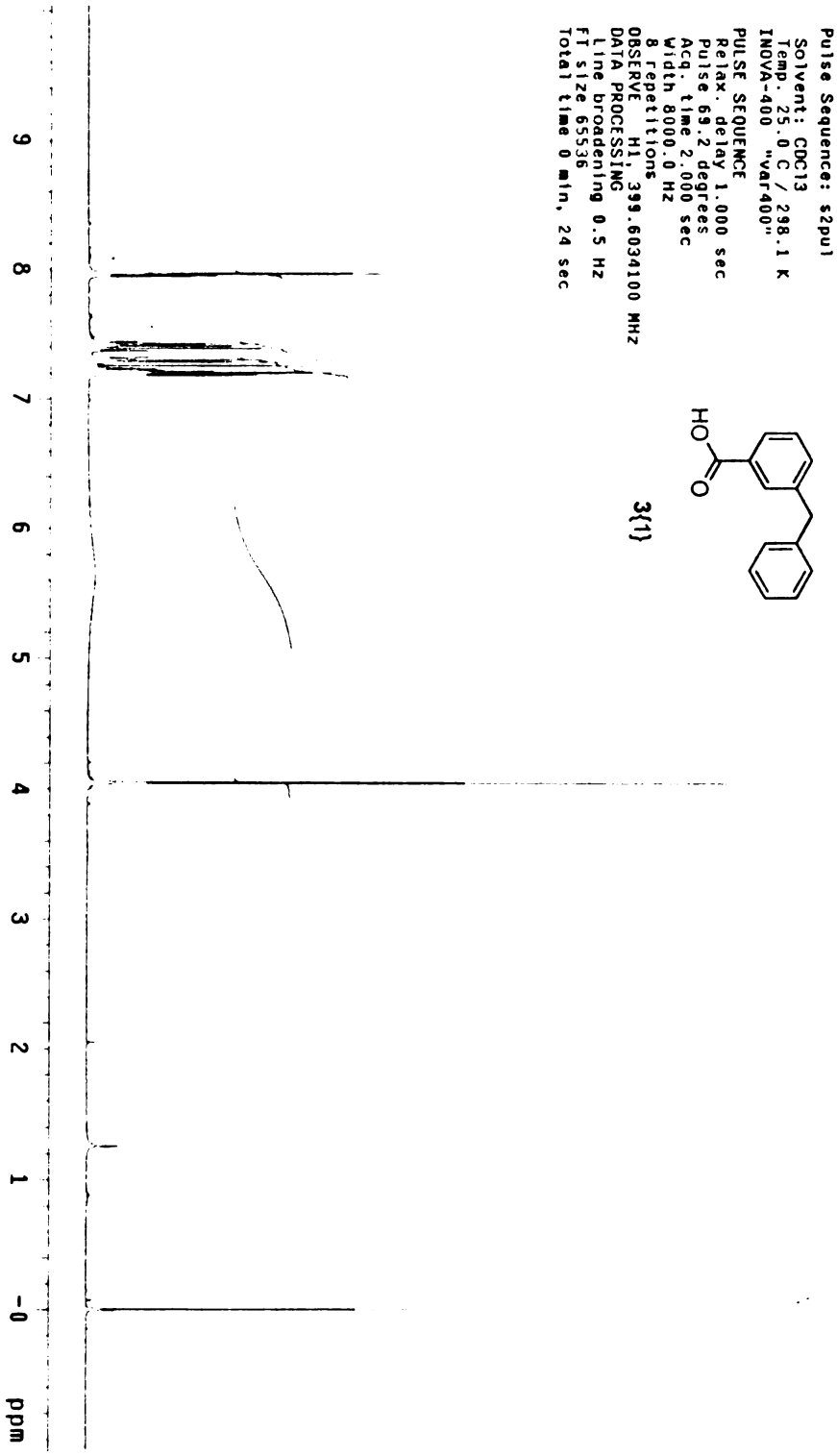
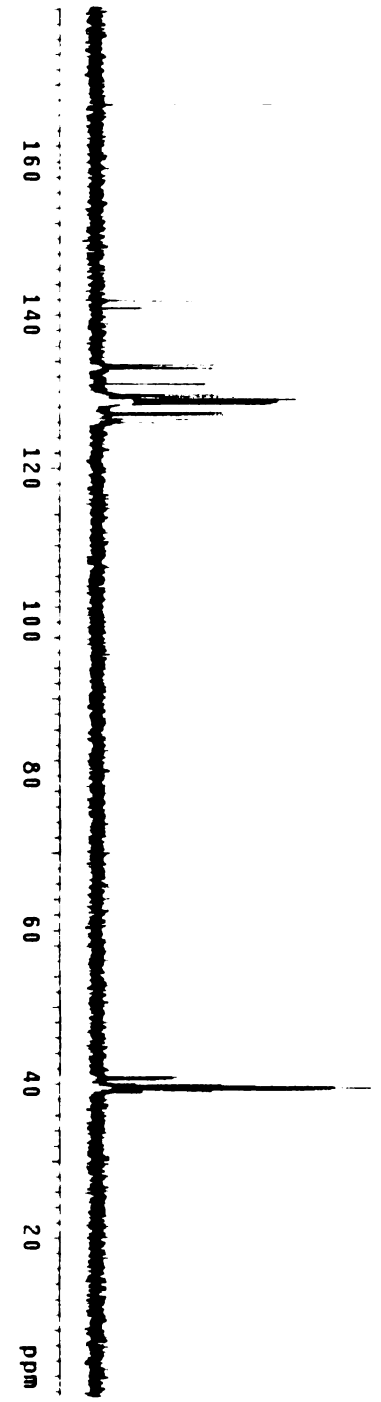


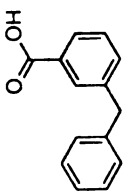
Figure A.43 ^{13}C NMR spectra for **3**{1}



13C OBSERVE

Pulse Sequence: s2pul
Solvent: DMSO
Temp. 25.0 C / 298.1 K
INOVA-400 "var400"

Pulse 60.0 degrees
Acq. time 1.199 sec
Width 25000.0 Hz
448 repetitions
OBSERVE C13, 100.4749710 MHz
DECOUPLE H1, 399.5838332 MHz
Power 30 dB
continuously on
WALTZ-16 modulated
DATA PROCESSING
Line broadening 1.0 Hz
FT size 65536
Total time 20 min, 37 sec



3{1}

Figure A.44 ^{1}H NMR spectra for **5**{1}

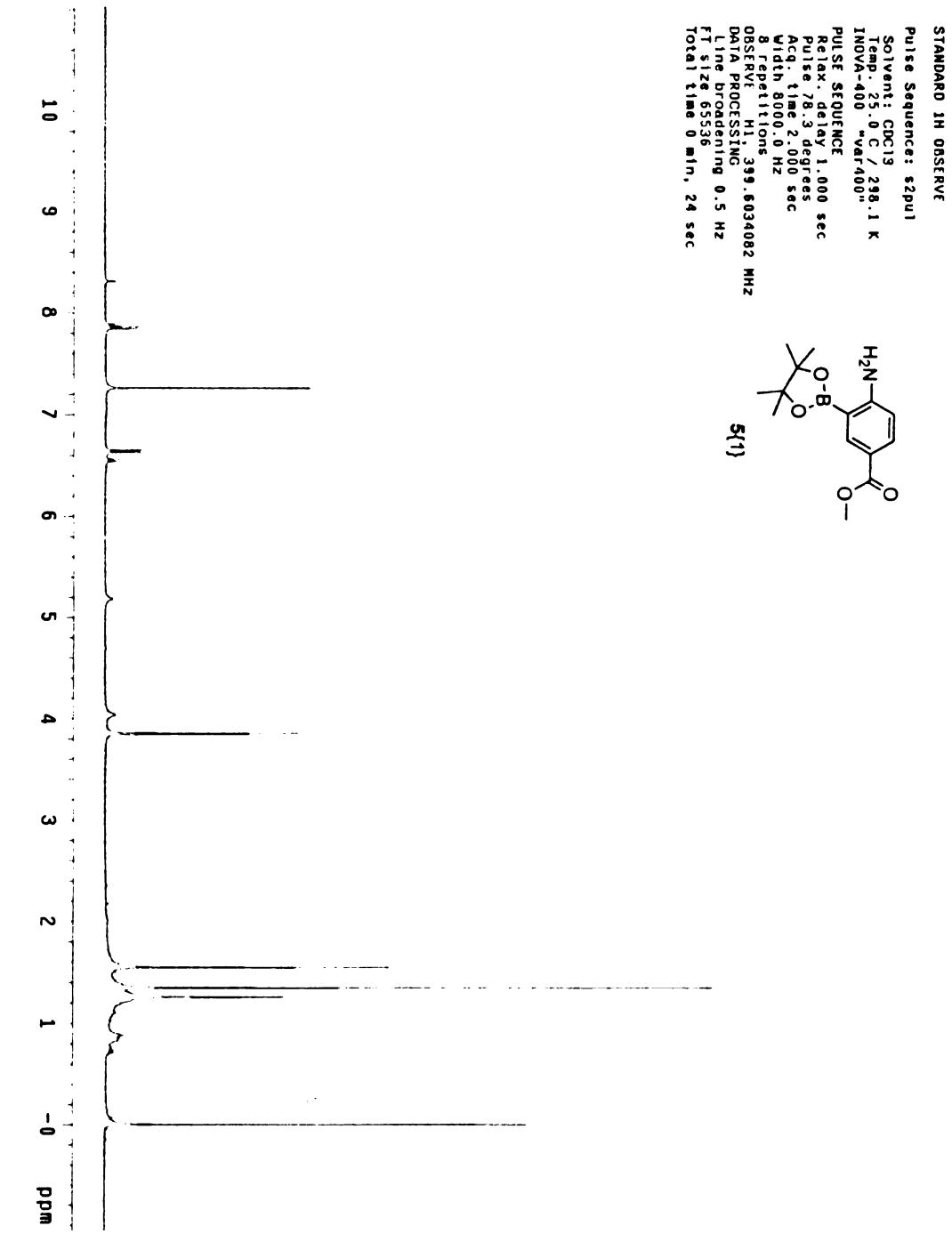


Figure A.45 ^{13}C NMR spectra for **5**{1}

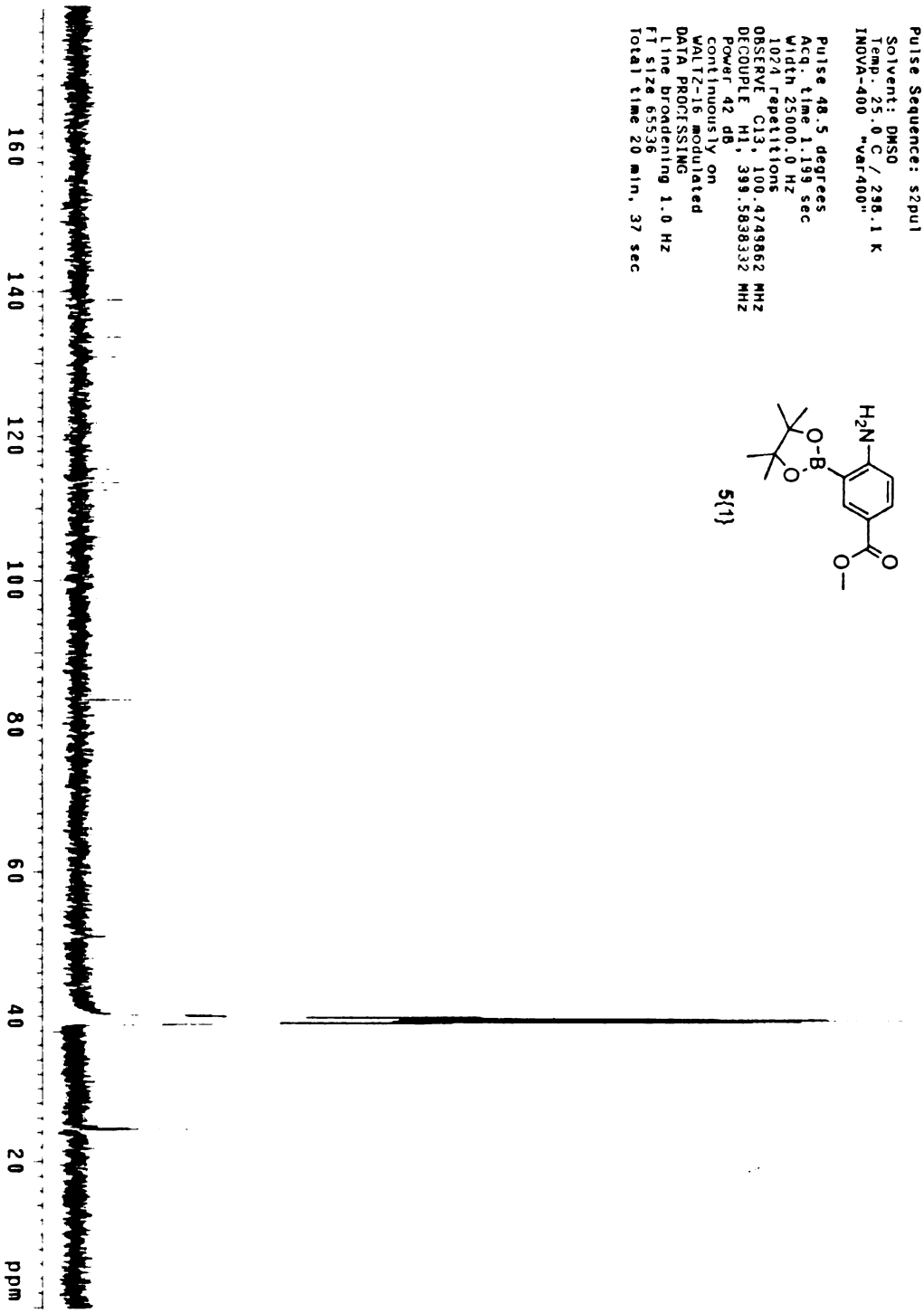


Figure A.46 ^{1}H NMR spectra for **6{1}**

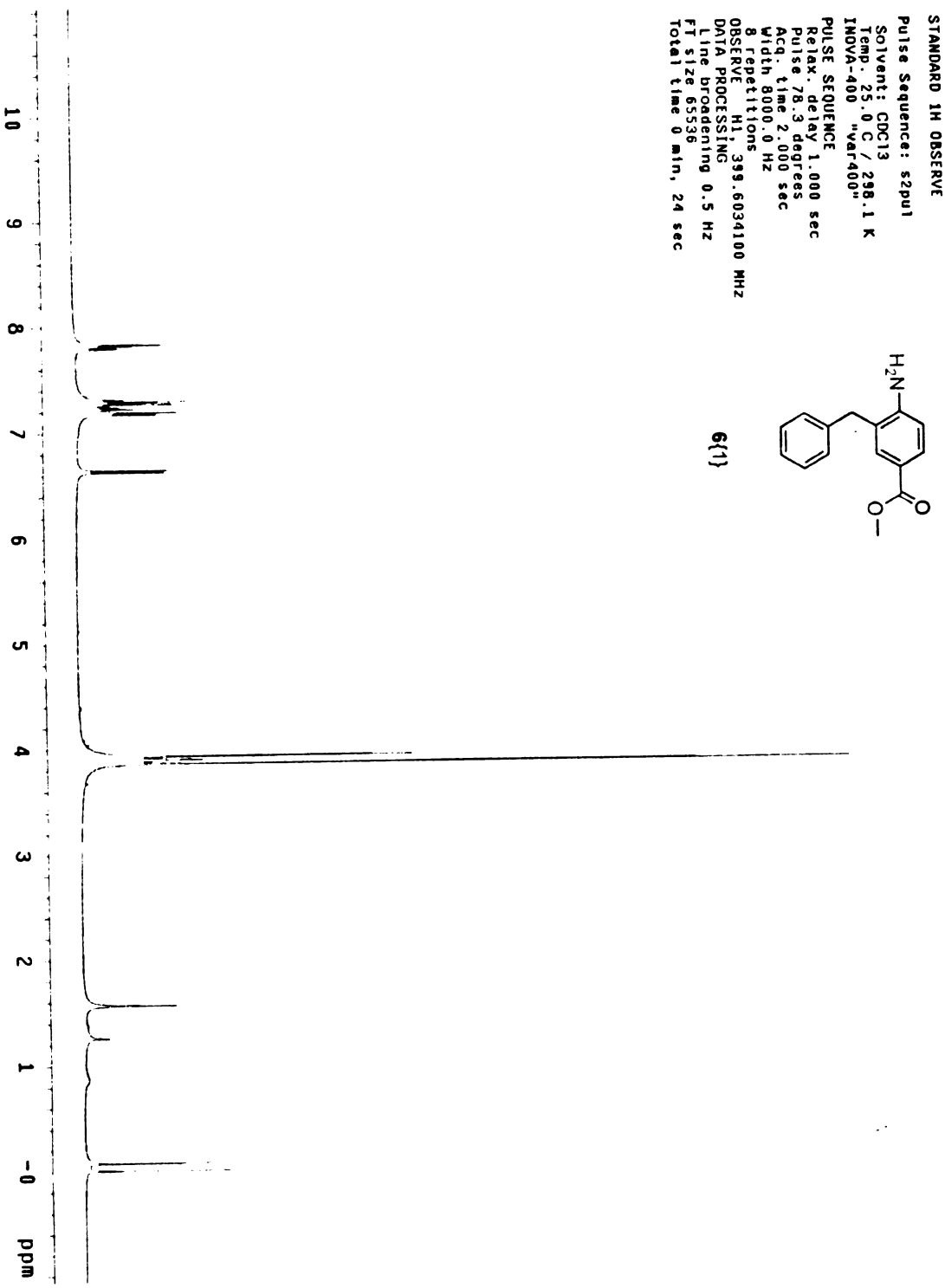


Figure A.47 ^{13}C NMR spectra for **6**{1}

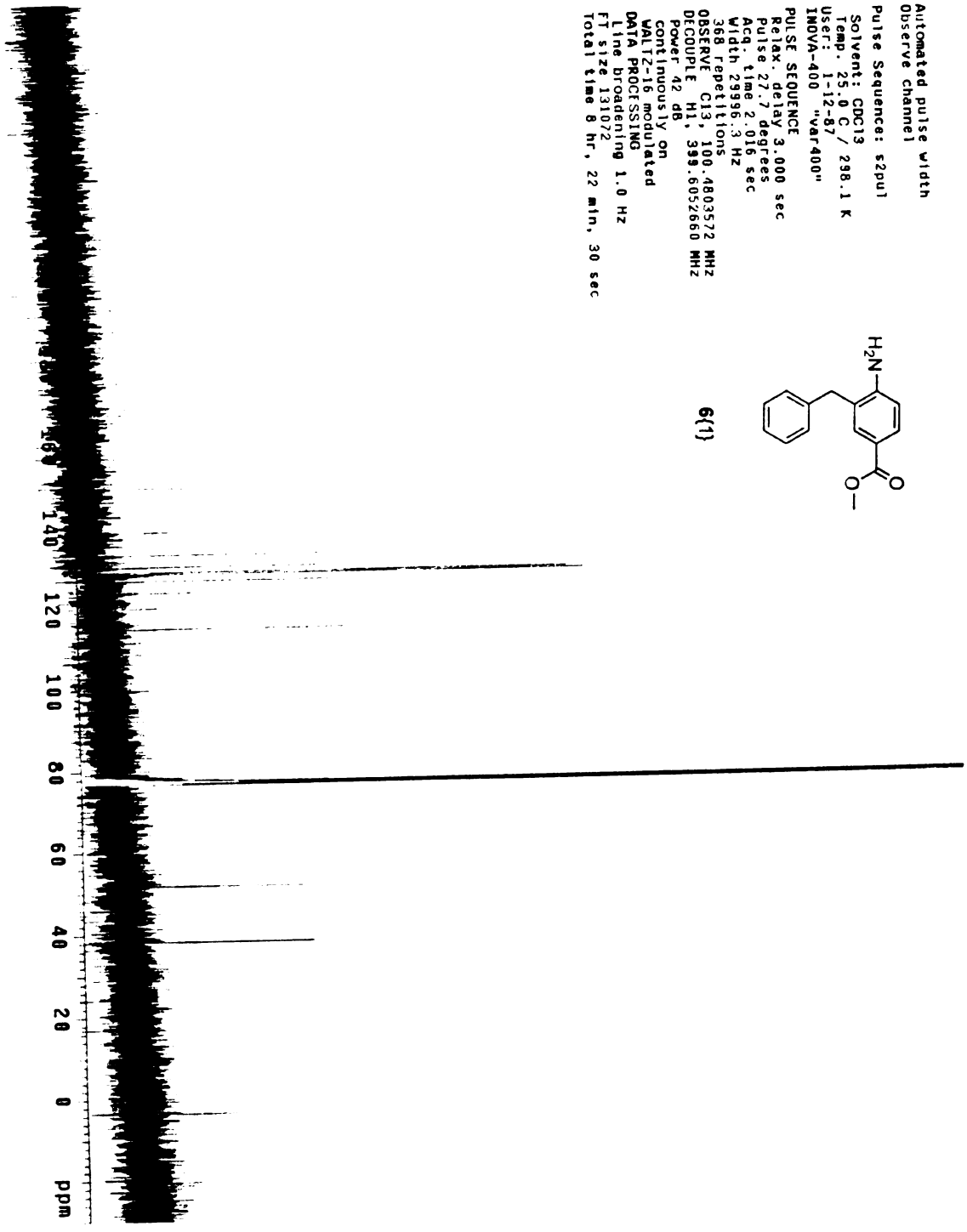


Figure A.48 ^{1}H NMR spectra for **8**{1}

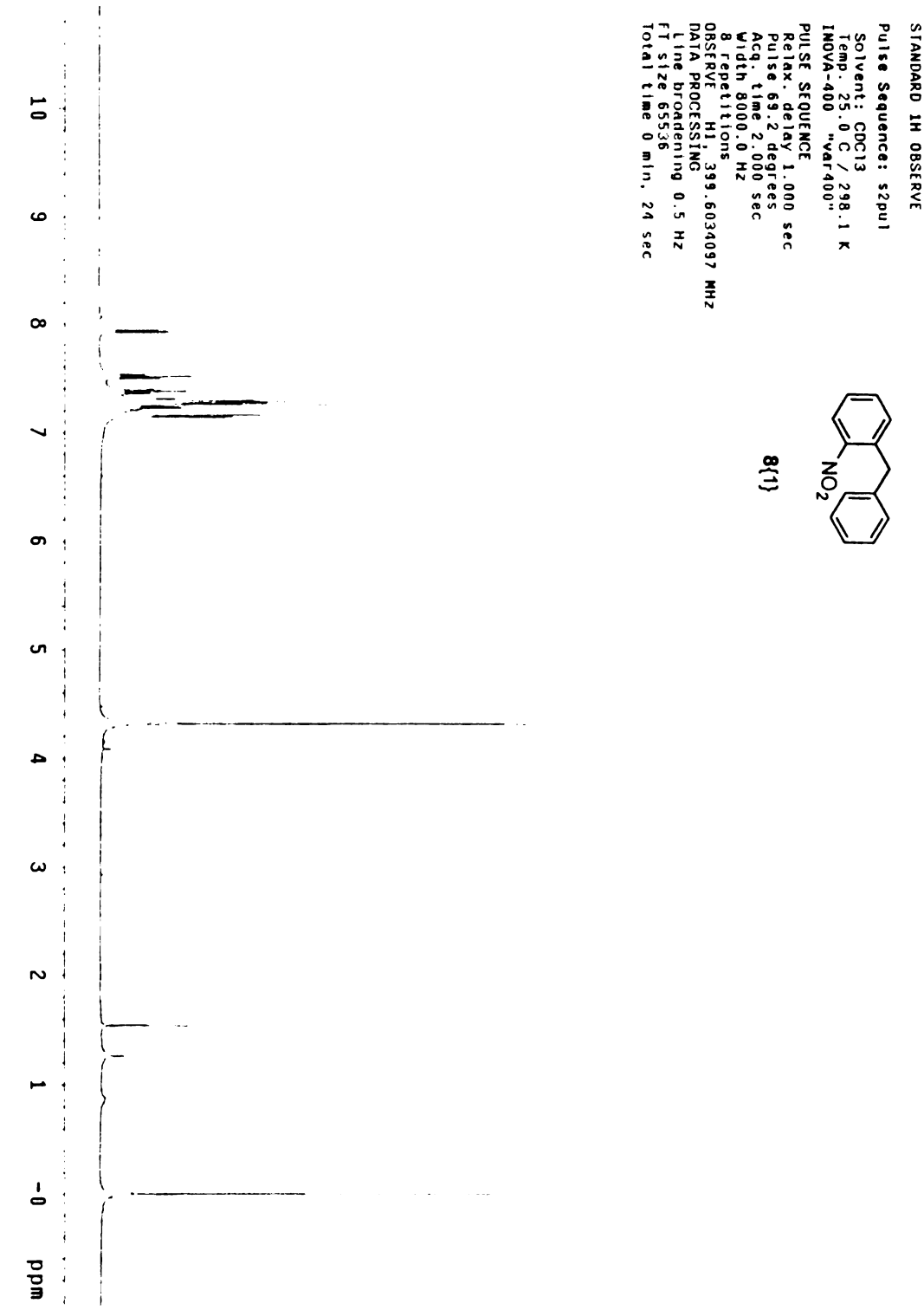


Figure A.49 ^{1}H NMR spectra for **9**{1}

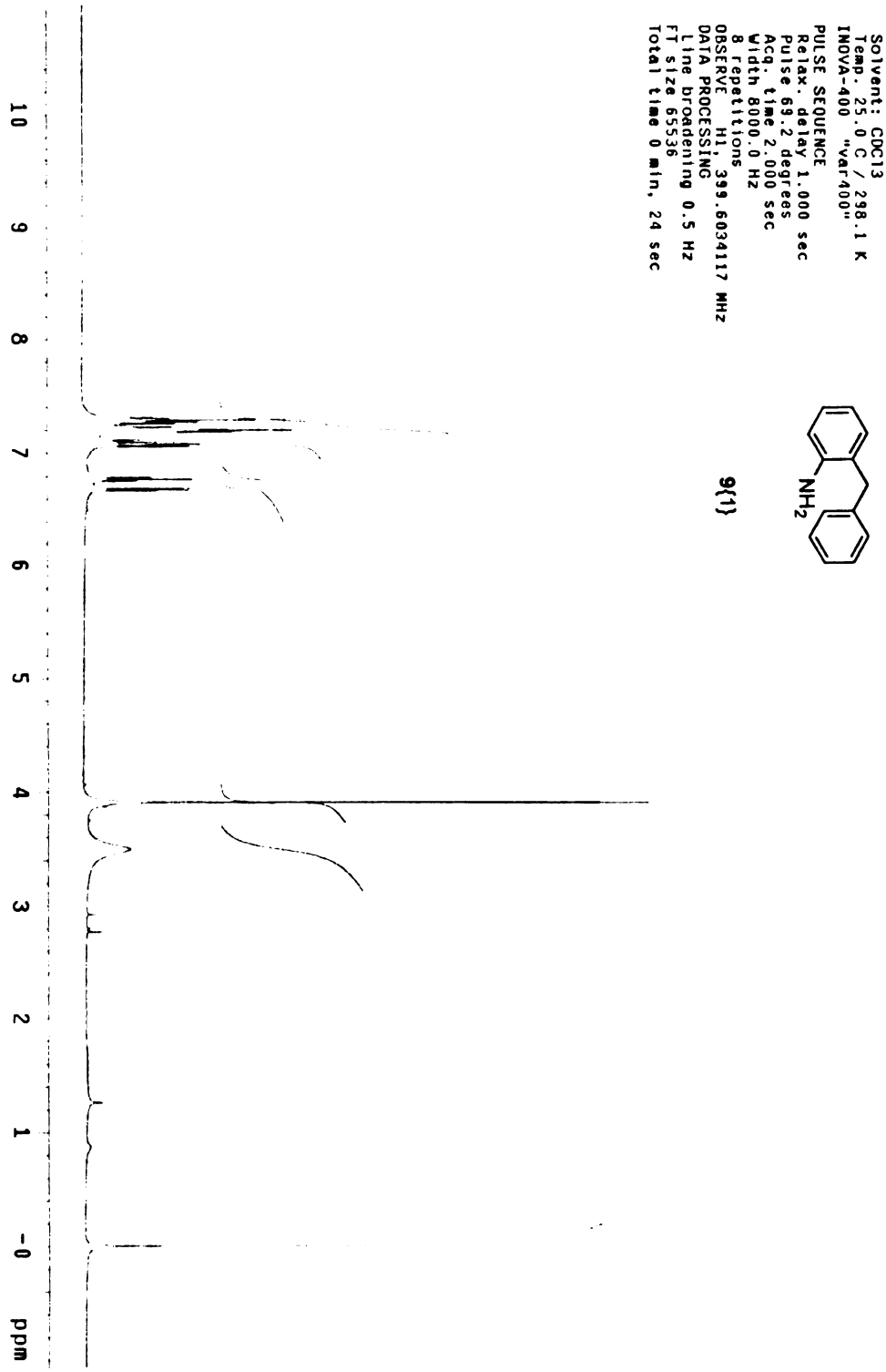


Figure A.50 ^{13}C NMR spectra for **9**{1}

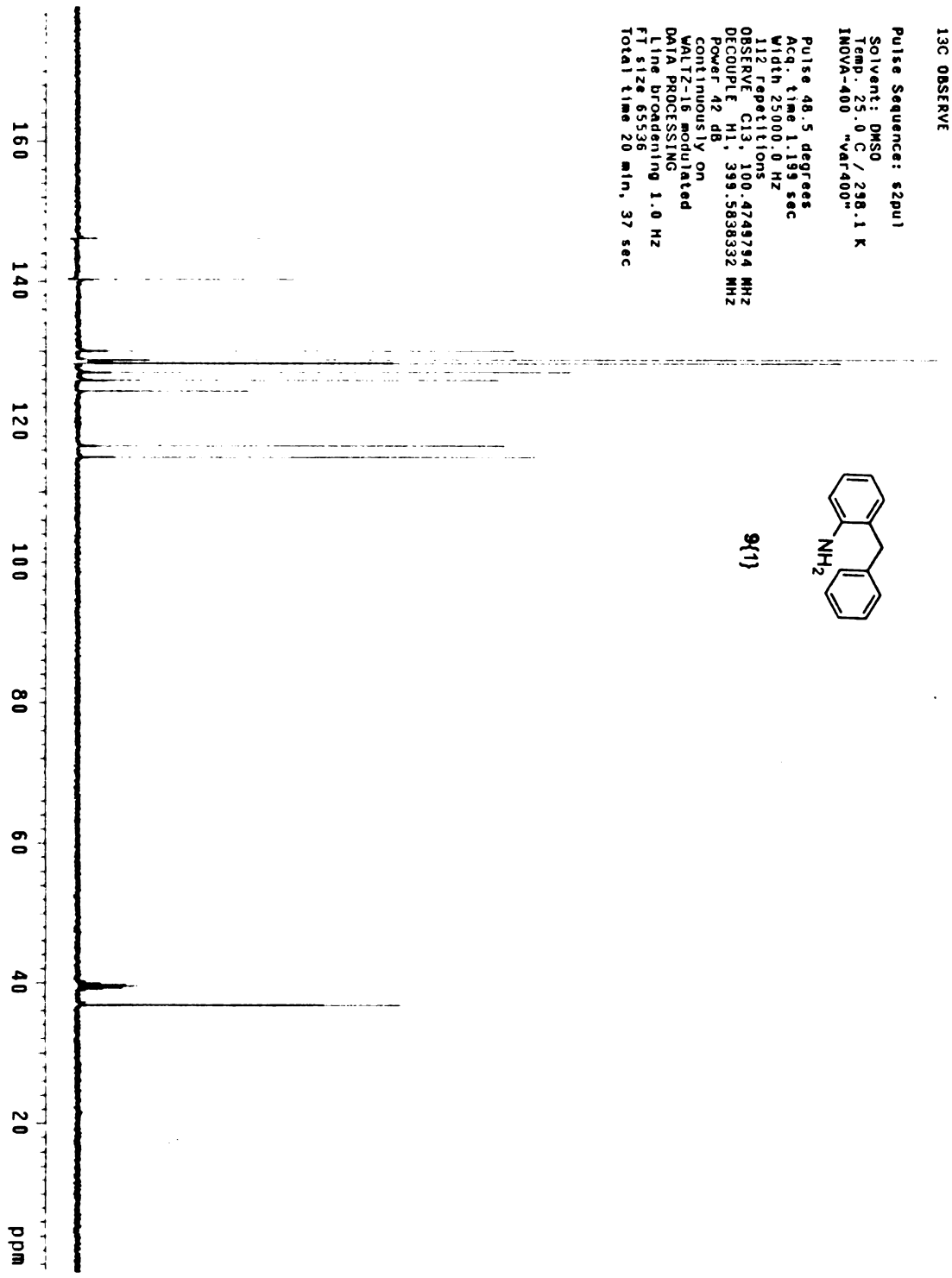
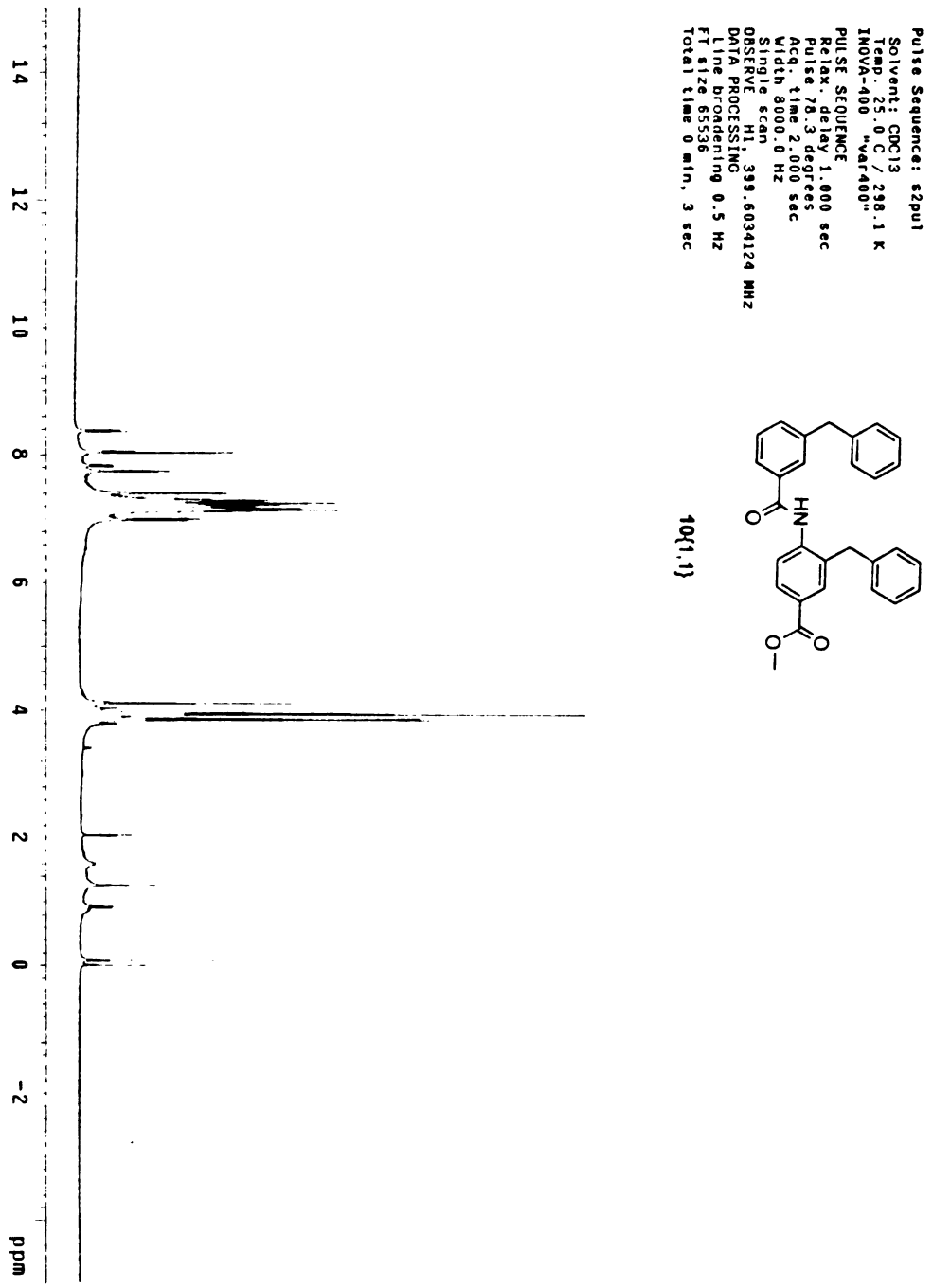


Figure A.51 ^{1}H NMR spectra for **10**{1,1}



10{1,1}

Figure A.52 ^{13}C NMR spectra for **10**{1,1}

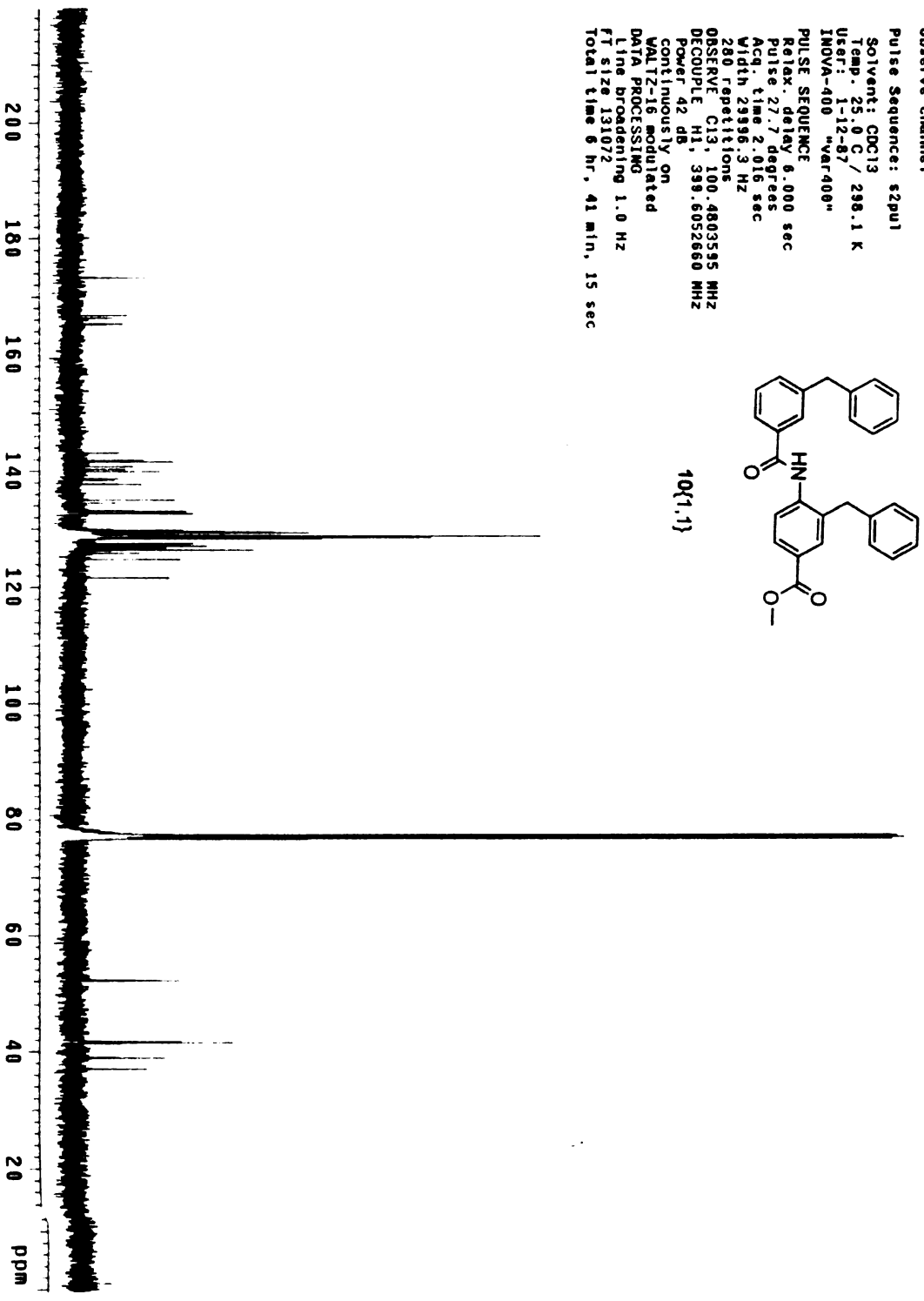
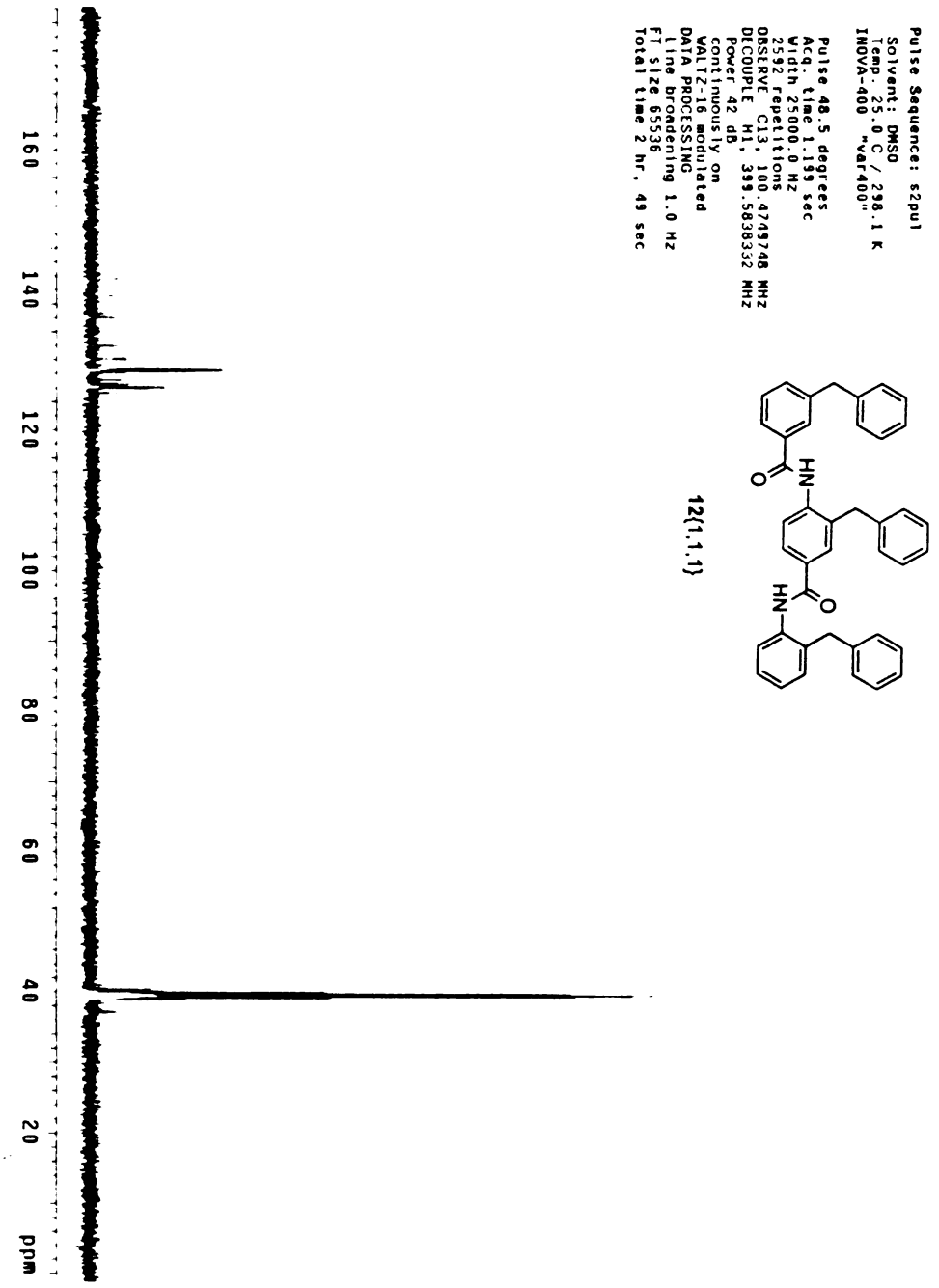


Figure A.53 ^{13}C NMR spectra for **12**{1,1,1}



Appendix 2

Computational Parameters

OMEGA Parameters

GP_RMS_CUTOFF	0.8
GP_ENERGY_WINDOW	3.0
GP_NUM_OUTPUT_CONFS	50
GP_MAX_ROTORS	17
GP_MAX_TRIES	500000
MAX_POOL_SIZE	20000
GP_FIX_FROM_FILE	false
GP_OUTPUT_RIGID	false
GP_SELECT_RANDOM	false
GP_INCLUDE_INPUT	true
OUTPUT_TO_SINGLE_FILE	true
NUKE_HYDROGENS	true
FIX_LARGEST_CYCLIC	true
MIN_RIGID_FRAGMENT	6
VERBOSE	true
UPDATE_INTERVAL	100
COPY_COM_FILE	false

CAVEAT Input File

Vector_: cvectors: 3 nvectors: 3
Vector: id: "vec1" base: "18" tip: "21" angle: 0.20 distance: 0.24
Vector: id: "vec2" base: "51" tip: "54" angle: 0.20 distance: 0.24
Vector: id: "vec3" base: "82" tip: "85" angle: 0.20 distance: 0.24
Pair_: cpairs: 1 npairs: 1
Pair: id: "pair1"
vector_: nvectors: 3
vector: "vec1" vector: "vec2" vector: "vec3"
angle: 0.00
vector_: nvectors: 3
vector: "vec1" vector: "vec2" vector: "vec3"
angle: 0.00
distance: 0.24
mode: CV_MANUALPAIR basedist: -1.00 disttol: -1.00 angletol: -1.00 tolfactor: 1.00
Class: id: ""
Core: id: "" mode: CV_NTIPC
Score: id: "" mode: CV_ERRS cmode: CV_NOCSCORE
Filter_: cfilters: 0 nfilters: 0
Comp: id: "" mode: CV_SCOMP
Clus: id: "" mode: CV_SKIPCL crit: CV_GRAPHCL
Uniq: id: "" mode: CV_SCOMP

DOCK 4.0 Parameters, Scaffold Docking

flexible_ligand	no
orient_ligand	no
score_ligand	yes
minimize_ligand	no
multiple_ligands	yes
parallel_jobs	no
intermolecular_score	yes
gridded_score	yes
grid_version	4
contact_score	no
chemical_score	no
energy_score	yes
atom_model	a
vdw_scale	1
electrostatic_scale	0
energy_maximum	1000
ligands_maximum	50000
initial_skip	0
interval_skip	0
heavy_atoms_minimum	3
heavy_atoms_maximum	<infinity>
rank_ligands	no

DOCK 4.0 Parameters, Full Molecule Scoring

flexible_ligand	no
orient_ligand	no
score_ligand	yes
minimize_ligand	yes
multiple_ligands	yes
parallel_jobs	no
random_seed	4803
intermolecular_score	yes
gridded_score	yes
grid_version	4
contact_score	no
chemical_score	no
energy_score	yes
atom_model	u
vdw_scale	1
electrostatic_scale	0
energy_maximum	0
energy_minimize	yes
initial_translation	1
initial_rotation	0.1
maximum_iterations	100
energy_convergence	0.1
maximum_cycles	1
ligands_maximum	1000
initial_skip	0
interval_skip	0
heavy_atoms_minimum	3
heavy_atoms_maximum	100
rank_ligands	yes