Title
Molecular design using fullerenes

Author
Friedman, Simon Hilary

Publication Date
1996

Peer reviewed|Thesis/dissertation

Molecular Design Using Fullerenes:
Design, Synthesis and Testing of Fullerene Based Inhibitors of the HIV Protease and A Computational Strategy For Designing Regiospecific Syntheses

by

by Simon Hilary Friedman

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

For my mother, father and sister, who mean the world to me.

Acknowledgments

Firstly I would like to thank my advisor Professor George Kenyon, who has provided advice and guidance and always supported me in the tough places. I very much appreciate his faith in me and the freedom given to me to explore, while always backing me up. Thanks George.

I'd also like to thank the members of my committee Professor Norman Oppenheimer and Professor Tack Kuntz for their helpful comments and suggestions on this thesis.

I'd like to thank our collaborators for their fine synthetic chemistry: Dr. Rint Sijbesma, Dr. Gordana Srdanov and Professor Fred Wudl at UCSB, and Dr. Padma Ganapathi and Professor Yves Rubin at UCLA.

Making friends in graduate school is an unusual process, because most of your friends and people you socialize with are also those people from whom you learn (and hopefully teach). So I have much gratitude to the following people, for being my friends and teachers. When I arrived at UCSF, I was fortunate to have as classmates Marcello and Bill whose friendship and advice over the years has been great. I would like to thank Eric Petersen, aka MC Id, for being a mensch and confidant and general really smart guy. There have been 101 things you've helped me with over the years (general counsel, midas, perl, motorcycles, etc. etc.) and I really appreciate it. Elaine Meng, thanks for generally being sweet and for taking the time over the years to look at manuscripts and point out scientific flaws, as well as being encouraging regarding my science. What can I say about Chris Schafmeister? That it was pleasant being a part of the background of his life? Chris is the evil twin I never had. He is also a mensch at heart and I value his friendship. I'd like to thank Melanie Schafmeister for being tolerant of our groups' general scientific weirdness, as well as a sweet and generous person. I'd like to thank the two Sher(r)is, Sheri Cohen and Sherri Newmyer who have been good and supportive friends. I'd also like to thank Rob Cerpa, who is a rare thing: a super bright but humble guy. Thanks for all your good counsel. Finally I'd like to thank Diana Roe. Diana was my first friend at UCSF and was a real

teacher for me as well. I really admire and respect her intelligence and am grateful for interacting with her over the years.

In addition to making friends in the department, I've been fortunate to be in a group that is chock-a-block with kind-hearted and talented people: Carolyn Koo is the glue that holds the group together. I very much appreciate her friendship and all that she's done for the group and me over the years. When I started in the lab I benefited by interacting with QiFeng Ma, my rotation project advisor and Sean Kerwin who guided me at the start of my first project. Sean helped to start me off in organic chemistry and molecular modeling. Doug Beuchter and Pete Guidinger were the wise elder statesmen of the group, and provided much good counsel as well as being good humored guys. Vibhakar Shah: My tutor in synthetic techniques, co-founder of the Witzelsücht Support Group and Coffee Club, chief guru of the cult of DNA, and all around master of joie de vivre. It has been great knowing him and learning from his example. The Chen family, aka "The Hardest Working Family in Science", Susan, Lorenzo, Elise, and Christopher: They've been so generous and kind to me, as well as being warm, amusing and an inspiration in balancing science with life. I thank Patsy Babbitt for much counsel over the years and for her generosity with time. Dolan Eargle, another founder of the W.S.G.C.C. has also been an example to me of how to keep learning and experiencing life and how to balance it with science. I will never forget the Bear Dance on Mt. Madonna. Thanks for including me, Dolan. Its also been great knowing Meg and Camille. (killer cookies and bread are the least of it) Angelika Muscate, one of the Angel Island Four, has been a good friend and teacher. I thank her for sharing her enzymology and organic chemistry experience, as well as her exhortations to stay an extra day on Angel Island. She's another scholar of balance in life.

I'd like to thank Prof. Charly Craik, for being an example of someone who really enthusiastically digs science. Thanks also to members of his group, Fiona McFee, Diane DeCamp, Nancy and Jason Rose, for Q7K mutant protease, as well as advice on setting up the protease assay.

The folks in the Dept. office and Deans office, Judi Mozesson, Liza Crosse, Gloria Dela Cruz and Peggy Kaplan have always been helpful and

friendly way beyond the call of duty. I'd like to especially thank Peggy for providing approximately my weight in gummi worms over the past few years.

I've met a lot of good folks outside of my group as well. Thomas Fox and Wendy Cornell were also members of the Angel Island Four. I'd like to thank Thomas for helping me with forays into semi-empirical calculations as well as impressing upon me the Bavarian vision of the importance of living life. Wendy is thanked for being generally wonderful and a sweet cynic. Next door to the lab resides Steve Miller, who has been a great person to talk science with, as well as ponder the universe. I am hoping that one day he will explain to me what pompetoos means. I'm not sure why I am grouping Larry Wainchel, Neil Buckley and Gary Henehan together, but it seems like a reasonable grouping given their world views. Gary has been a great help in talking about enzyme kinetics. All three I think qualify as codgers. Finally, Lisa Kim-Shapiro has been a good friend and person to talk science and life with.

Over the years I have had the benefit of being instructed by some excellent teachers. Mrs. Harrison taught me how to organize and express my thoughts. Mr. Lewis, Mrs. Biddulph and Mr. Goldman taught me tools of mathematics and chemistry that I still use today. I very much appreciate the dedication their effort required, especially in the public school system. At M.I.T. I was fortunate to be instructed in organic chemistry by Professor Peter Lansbury who gave me a new perspective on chemistry and my abilities. Having the respect of someone who I admire greatly has been very rewarding. My research mentor at M.I.T. for two years was Dr. John Thomson, who had the rare combination of great skill and no arrogance. It was a really productive time for me, and I thank him for that opportunity and his friendship. Finally, I'd like to thank Professor Leonard Peller who has been a mentor and friend these past couple of years. Thank you for your guidance, encouragement and wisdom.

I'd also like to thank my extended family for all the concern they have shown for me while working on my degree. Aunt Roz, Uncle Alby, Aunt Sylvia, Aunt Maria, Uncle Harold, Thomas and Steven: thanks for pulling for me.

Rose Rodriguez believed in me and my ideas a long time ago, and also introduced me to the mystery of biochemistry, for which I thank her.

I can really say nothing about Professor Jens O.M. Karlsson here, as I would not want to offend his sensibilities as to the appropriate people to mention in an acknowledgments section. I will say this: That his friendship is treasured, even if he is a misanthrope.

* * * *

I have been truly blessed with a family who loves and believes in me. I can't fully express what a difference this has made, in grad school and in life. My parents did not have easy lives, and in many ways their dreams were deferred: by history, by family, by illness. I need to say that whatever I have achieved, for better or worse, it is due to parents who gave me time, love and the strongest belief in what I might do. They also taught me how to question the world around me and how much fun it was to create, and even to risk being silly. They also gave me an older sister who knows me better than anyone and is a true friend. Thanks for believing in me and making me believe in myself.

Molecular Design Using Fullerenes: Design, Synthesis, and Testing of Fullerene Based Inhibitors of the HIV Protease and A Computational Strategy For Designing Regiospecific Syntheses

by Simon Hilary Friedman

Abstract

The protease specific to the Human Immunodeficiency Virus (HIVP) has been shown to be a target for the treatment of Acquired Immune Deficiency Syndrome (AIDS). The first aim of the work in this thesis was to design and test a C60 fullerene based inhibitor of the HIVP. (Chapter 2) The design was based on the recognition that the non-polar and spherical C60 fullerene was highly complementary in shape and chemical nature to the non-polar, cylindrical active site of the HIVP. This observation was based on modeling of the C60/HIVP interaction, using interactive molecular graphics and the program DOCK. Analysis of the type of surface desolvated upon formation of the modeled complex was performed using molecular surfaces generated with the program MS and showed the desolvated surface to be >90% non-polar. Further modeling showed that a water soluble fullerene, di(phenethylaminosuccinate)-C60, would position the C60 sphere within the active site in a position nearly identical to the underivatized C60. Kinetic analysis of this compound with the HIVP showed that it bound competitively with a Ki of 5μM.

This initial design was refined, (Chapter 3) by analyzing C60 derivatives that could interact with non-polar regions of the HIVP active site that the initial compound could not. Hypothetical compounds were designed and docked into the active site, then analyzed for increased non-polar surface desolvation relative to the initially tested compound. Two compounds were selected for synthesis based on their ability to significantly increase the non-polar surface desolvation, relative to the originally tested compounds. These compounds, a trans di-isopropyl cyclohexanol C60 and a cyclopropyl C60 alcohol, showed increased affinity for the HIVP with Ki values of 150 and 103nM, respectively.

Chapter 4 details a method for designing regiospecific syntheses of fullerene derivatives. It utilizes a statistical mechanical analysis of conformations generated from systematic conformational searching to determine the site of reaction of two equivalents that have been linked with a flexible molecule. The method was validated through application to a known tethered directed regiospecific fullerene derivative synthesis originally performed by Francois Diederich and coworkers.

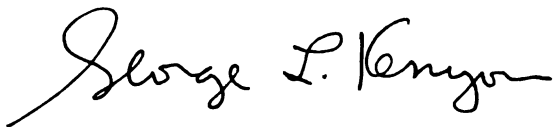


Table of Contents

	Page
Acknowledgments	iv
Abstract	viii
Table of Contents	x
List of Tables	xiii
List of Figures	xiv

Chapter 1: Introduction

Introduction	1
Fullerenes in the Biochemical Realm	1
The HIV Protease as Therapeutic Target	2
Ligands and Receptors	4
Computational Aspects of Work	5
Roots	7

Chapter 2: Design of Fullerene Based Protease Inhibitors

Introduction	8
Journal of the American Chemical Society Paper	9
Abstract	10
Introduction	11
Results and Discussion	11
Conclusions	22
Experimental Section	23
Modeling:	23
Enzyme Assays	23
Background to Journal of the American Chemical Society Paper	25
Hydrophobic Effect Discussion	27
Computational Tools	35

Chapter 3: Refinement of Design

Introduction 37
Journal of Medicinal Chemistry Paper 38
Abstract 38
Introduction 39
Modeling Strategy 40
Results and Discussion 44
Experimental 45
Compound 5 Synthesis 45
Enzyme Assays 46
Background to Journal of Medicinal Chemistry Paper 52
MINDOCK 53
Refinement Philosophy 53
Analysis of Hypothetical Derivatives 55
Bisadducts 55
Monoadducts 72
First Target Compound 76
Analysis Of Solubilizing Chain 79
Second Target Compound 90
Enzymology 93
Conclusions 100

Chapter 4: Computational Strategy for Regiospecific Syntheses

Introduction 101
Description of Approach 101
Computational Strategy 106
Modeling of Diamidines 109
Synthetic Attempts 111
Journal of the American Chemical Society Note 114

Chapter 5: Conclusions

Summary and Conclusions 120

References 122

List of Tables

		Page
Table 2.1	Surface Desolvation Breakdown; C60/HIVP Complex	14
Table 2.2	MS vs ΔG Replot Data	34
Table 3.1	Summary of Compounds/Desolvation Data	43
Table 3.2	Surface Desolvation Breakdown of Compound A	56
Table 3.3	Surface Desolvation Breakdown of Compound B	58
Table 3.4	Surface Desolvation Breakdown of Compound C	61
Table 3.5	Surface Desolvation Breakdown of Compound G	66
Table 3.6	Surface Desolvation Breakdown of Compound H	69
Table 3.7	Surface Desolvation Breakdown of Compound I	75
Table 3.8	Surface Desolvation Breakdown of Compound M	77
Table 3.9	Surface Desolvation Breakdown of Compound Pa	87
Table 3.10	Surface Desolvation Breakdown of Compound Pe	89
Table 3.11	Surface Desolvation Breakdown of Compound Ra	91
Table 4.1	Summation of Boltzmann Factors for Reactive Bonds	118

List of Figures

		Page
Figure 1.1	Life Cycle of HIV	2
Figure 2.1a	"Front" View of the HIVP	12
Figure 2.1b,c	Top Scoring Complex of HIVP and C60	13
Figure 2.2	Structure of Tested Compound	15
Figure 2.3	Top DOCK3 Complex of Compound 1	17
Figure 2.4	Double Reciprocal Plot of Compound 1	18
Figure 2.5	Compound 2	20
Figure 2.6	DOCK3 Complex of Compound 2	21
Figure 2.7	Figure Taken From Sharp, K. et al.	29
Figure 2.8	Result of Tuñon et al.	30
Figure 2.9	Comparison of SAS and MS	31
Figure 2.10	Replot of MS Versus ΔG (non-polar)	33
Figure 3.1a,b	Front and Top Views of Compound 1 Complex	48
Figure 3.1c	Cross Section of Compound 1 Complex	49
Figure 3.2a,b	Front and Top Views of Compound 5 Complex	50
Figure 3.2c	Cross Section of Compound 5 Complex	51
Figure 3.3	Cross Section of HIVP/C60 Complex	54
Figure 3.4	Sites of Attachment of Bis-Modified C60	55
Figure 3.5	Compound A	56
Figure 3.6	Cross Section of Compound A/HIVP Complex	57
Figure 3.7	Compound B	57
Figure 3.8	Model Complex of Compound B/HIVP	59
Figure 3.9	C60 Diels Alder Reaction	60
Figure 3.10	Compound C	60
Figure 3.11	Cross Section of Compound C/HIVP Complex	62
Figure 3.12	Compound D	63
Figure 3.13	Compound E	63
Figure 3.14	Compound F	63
Figure 3.15	C70 Reaction Pattern	64
Figure 3.16	Compound G	65
Figure 3.17	Model Complex of Compound G/HIVP	67

Figure 3.18	Cross Section of Compound G/HIVP Complex	68
Figure 3.19	Compound H	69
Figure 3.20	Cross Section of Compound H/HIVP Complex	70
Figure 3.21	Oxygen/Nitrogen Interaction in Cmpd. H Complex	71
Figure 3.22	C60 Monoadduct Motifs	73
Figure 3.23	Compound I	73
Figure 3.24	Model Complex of Compound I/HIVP	74
Figure 3.25	Compound J	75
Figure 3.26	Compound K	76
Figure 3.27	Compound L	76
Figure 3.28	Compound M	77
Figure 3.29	Model Complex of Compound M/HIVP	78
Figure 3.30	Cross Section of Compound M/HIVP Complex	79
Figure 3.31	Compound N	79
Figure 3.32	Model Complex of Compound N/HIVP	81
Figure 3.33	Compound O	82
Figure 3.34	Diels-Alder Reaction of C60 with TMS Derivative	83
Figure 3.35	Double Michael Addition to C60	84
Figure 3.36	Compound P	84
Figure 3.37	Enantiomer/Conformer Pairs in Compound P	85
Figure 3.38	Model Complex of Compound Pa/HIVP	86
Figure 3.39	NOE Effect in Compound P	88
Figure 3.40	Cross Section of Compound Pe/HIVP Complex	89
Figure 3.41	Compound Q	90
Figure 3.42	Compounds Ra and Rb	91
Figure 3.43	Model Complex of Compound Ra/HIVP	92
Figure 3.44	Cross Section of Compound Ra/HIVP Complex	93
Figure 3.45	Time Course Compound R (No Preincubation)	95
Figure 3.46	IC50 Plot of Compound R (High Substrate)	96
Figure 3.47	Time Course Compound R (With Preincubation)	97
Figure 3.48	Compound P Kinetic Replot	98
Figure 3.49	Compound R Kinetic Replot	99
Figure 4.1	Schematic of Regiospecific Synthetic Strategy	102
Figure 4.2	1,4 Diamino C60	103
Figure 4.3	Amine Reactions With C60	104
Figure 4.4	Alkyl Amine Synthesis Via Amidines	105

Figure 4.5	Linked Amidine Strategy	106
Figure 4.6	Singly Reacted C60	107
Figure 4.7	1,4 and 1,3 Diamidino Benzene	109
Figure 4.8	1,4 Diamidino Benzene Distance Distribution	110
Figure 4.9	1,3 Diamidino Benzene Distance Distribution	110
Figure 4.10	Tether Linked Diene and Bond Targets	116

Chapter 1

Introduction

The aim of the work described in this thesis is to develop and apply molecular modeling approaches for the design of small molecules with defined properties. This general description can be broken down into two specific areas: A) the identification and refinement of fullerene-based inhibitors of the HIV protease, using a combination of design, docking, and surface transfer analysis (Chapters 2 and 3), and B) the design of a strategy for regiospecific syntheses through application of statistical mechanical analysis of molecular mechanics energies. This is the most succinct way of describing the second project which is the focus of Chapter 4. These two projects are different in that in the former, predictions of binding between two molecules are made, a thermodynamic prediction, whereas in the latter, predictions of reactivity are made, i.e., a kinetic prediction. They are, however, alike in that for both, models of the molecules and their inter/intra molecular energies are used to predict macroscopically measurable quantities. This is what has motivated me most in this work: to build models, use them to make predictions and finally do the experiments that will test the predictions.

Fullerenes in the Biochemical Realm

The first part of this thesis describes the *denovo* design of a fullerene-based HIV protease inhibitor, followed by its refinement into higher affinity inhibitors, through the application of molecular modeling. At first glance, fullerenes appear unlikely candidates for biochemical investigation. There has been no precedent for their use in this realm. There are, however, several aspects of the fullerenes that make them both an ideal starting point for the design of HIV protease inhibitors, as well as general conformationally restricted scaffolds upon which pharmacophores may be positioned.

The C60 fullerene structure is highly complementary both in shape and in chemical nature to the active site of HIV protease, a target for anti HIV therapy. This complementarity translates into binding, enzyme inhibition and therefore potential for therapeutic application. Also, because of the

greater conformational rigidity of fullerenes relative to typical enzyme inhibitors, the modeling of their complexes with the protease (or any target protein) is simplified.

Perhaps the most generally interesting aspect of fullerenes in the biochemical realm is that they provide a conformationally restricted surface upon which groups required for target enzyme binding may be positioned. Typically, conformationally restricted scaffolds such as small cyclic peptides present only a few locations at which elements required for binding to a macromolecular target can be introduced. In this regard, a fullerene is more general, in that there are a large variety of group to group distances and angles that their surfaces provide. This can be further expanded by introducing spacers between the fullerene surface and the binding elements, or by utilizing higher fullerenes such as C70. The key requirement of applying fullerenes in this way is the development and improvement of regiospecific fullerene synthetic chemistry. Work to this end is the focus of Chapter 4.

The HIV Protease as Therapeutic Target

Infection by the HIV-1 has been shown to be the cause of the Acquired Immunodeficiency Syndrome (AIDS)[1]. For this reason, the life cycle of the HIV-1 has been studied in detail in an effort to locate weaknesses that may be exploited by drug or other therapeutic intervention.

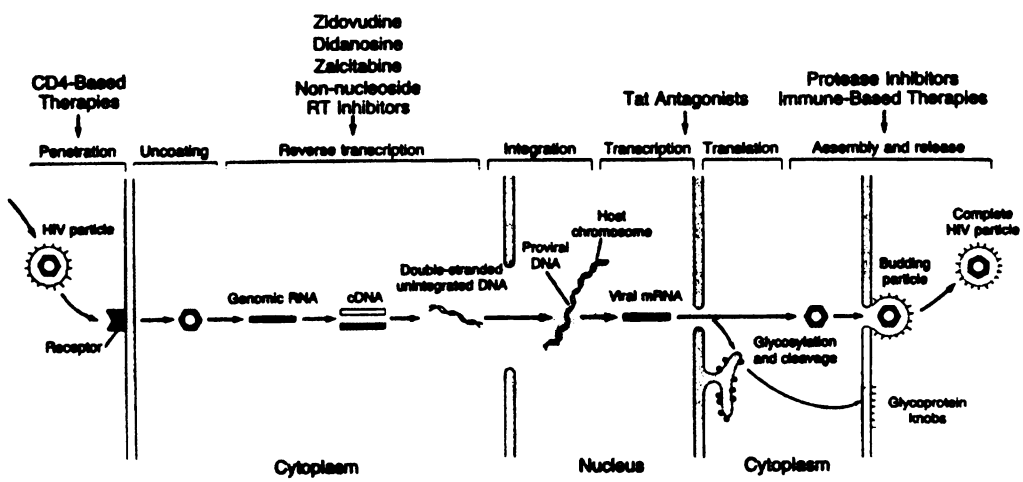


Figure 1.1 Life cycle of the HIV, Showing Site of Potential Therapeutic Action (from Hirsch, M.S. et al. [2])

These weaknesses, in part, take the form of the many unique proteins that are involved in the virus's life cycle, proteins whose action can, in theory, be interfered with without harm to the infected host. (Figure 1.1)

One of the unique proteins of the HIV-1 that has been found to be crucial for its life cycle is the HIV-1 protease (HIVP)[3]. The HIVP is responsible for cleaving large proteins that the virus produces into smaller proteins that are then used to make new infectious viruses. These smaller pieces of protein are used both as structural elements of the new viruses as well as enzymatic elements. It has been established that compounds which block the HIVP (i.e., HIVP inhibitors) also block proliferation of HIV infection[3]. Within the last year, three HIVP inhibitors have been approved by the FDA for clinical use. These compounds are proving to be extremely efficacious, increasing T-cell counts and decreasing viral loads in infected patients, sometimes to undetectable levels. A major problem with these inhibitors, and indeed with all anti-HIV chemotherapy, is the development of resistance to these compounds by the highly mutable HIV. Combination therapy appears to be the best way around this problem, by forcing the virus to adopt multiple mutations faster than may be possible.

It was the desire for non-peptide based HIVP inhibitors that led us to consider the modeling of fullerenes as possible inhibitors. The promise of structure-based drug design is that hypothetical inhibitors can be "tested" using computer generated models before synthesis and biochemical evaluation, thus leading to a more efficient identification and optimization of lead compounds. Because of the potential therapeutic application of HIVP inhibitors, a number of groups have determined atomic resolution structures of the HIVP, both with and without inhibitors present[4]. These allow the computer-aided, structure-based design of compounds that will bind to a portion of the protease that performs its function (i.e., the active site) and thereby block that function.

Ligands and Receptors

There are two elements that are key to predict the binding of a small molecule with a target macromolecule. These are the shape and the chemical nature of the ligand and target. When one examines crystallographically determined complexes of tightly bound ligand/macromolecule pairs, the surfaces of the respective molecules approach closely, usually within < 1Å. (A notable exception to this observation is for solvent molecules and/or ions which may be present in the ligand/macromolecule complex. They of course also contribute to the binding energetics and make their own surface contacts). If one is interested in mimicking nature with the intent of binding to a site on a macromolecule, it would seem prudent to take this lesson to heart, and to design molecules which are able to match closely the shape of the site on the macromolecule.

In addition to the need to match the shape of the target site, it is also important that the surfaces that come in contact with each other during complexation are appropriate. By appropriate I mean that they will contribute in a beneficial way to the free energy of binding. The term "chemically appropriate" can be summarized with the following three rules: a) non-polar inhibitor surfaces are in contact with non-polar protein surfaces, leading to association driven by the hydrophobic effect , b) hydrogen bond donating species are in contact with hydrogen bond accepting species and c) positively charged species are in contact with negatively charged species. Real world complexes contain these features, as well as a mixture of "non-ideal" interactions, i.e., non-polar surfaces being buried in polar environments, charged groups being partially desolvated in non-polar environments, etc.

A guide to the experimentally observed magnitudes of these interactions are the studies of Fersht and coworkers[5]. Fersht utilized enzymology to examine the binding of amino acids with their appropriate t-RNA synthetases, and, in so doing, determined ranges of energies for different types of interactions. Specifically, he examined the binding affinity of amino acids with their natural t-RNA synthetase target, as well as the binding of small molecules which varied from the amino acid in a specific feature (e.g., a methyl group). By utilizing this small molecule equivalent to site directed mutagenesis, Fersht was able to determine the binding free energy contribution of various types of interactions. These experiments

provide data which act as a guide when considering the introduction of specific interactions, and which were invaluable in estimating the potential affinities of ligand and target protein. This will be discussed in more detail in Chapter 2.

Computational Aspects of Work

The work described here has relied on a variety of computational resources. These include DOCK3[6], MINDOCK[7] Midas-Plus[8], MS, Amsol, Flymol and SYBYL[9]. DOCK3, MINDOCK and MS were central to the design and refinement work. The DOCK program is established as a tool for rapidly identifying lead compounds as ligands to receptor sites[10]. In addition, however, it has great utility as an interactive design tool, an application for which I have not seen it used. DOCK's great strength is its ability to rapidly identify possible binding modes of small molecules with a target receptor site. It has been used to identify lead compounds which bind to a range of targets. In this work, I extensively use DOCK's 'single mode', i.e., the mode which allows an individual molecule to be oriented in the target site, and scored. The utility of this for design is to determine how a small change to a lead compound will be tolerated by the active site, and, more importantly, how this change may introduce interactions that were not in place prior to the addition. Because it is quick, the process can be iterated over a reasonable period of time. Starting from a real or modeled complex of a lead compound, a molecule is 'drawn up' with the purpose of introducing a new interaction in the complex. It is then docked and top scoring complexes are then analyzed. From these complexes it is determined if the criteria for improved interaction through the introduction of specific interaction has been achieved. The details of the approach and the criteria actually used are given in Chapter 3.

An outgrowth of the work on designing fullerene based inhibitors of the HIV protease was the development of a computational strategy for designing syntheses to control the regiospecificity of organic reactions. Controlling regiospecificity, i.e., controlling the reaction of a specific site on a molecule when two or more reactive sites are present, is a major concern of synthetic organic chemistry. This concern is particularly great in the realm of fullerene chemistry, where nucleophiles can react with a large and varying

number of the 60 carbons on the surface. If one wishes to utilize fullerenes for applications where precise structure is required (i.e., as therapeutics or in applications of nano-technology), one needs precise control of the sites of reactant addition. This is key when the addition of two or more substituents to the surface is required.

During the initial phases of this project, I conceived of a possible improved fullerene based inhibitor: a C60 with two amino groups attached to the surface with a defined distance separating them. The solution I devised to positioning two or more groups on the surface of the fullerene was to tie the reactive equivalents with a conformationally restricted, cleavable linker that could then be removed after the two equivalents had reacted. The initial reaction is an *inter*molecular reaction with the C60 and one of the tied equivalents on the reagent molecules reacting. This then sets up an intramolecular reaction with the second equivalent on the reagent with the C60 surface. The conformational nature of the cleavable linker defines the isomers that will be produced. The computational aspect of this approach enters in the analysis of what isomers a specific linker will end up producing. What is modeled is the species that exists after a single equivalent on the reagent has reacted. The entire conformational ensemble of the singly reacted fullerene/reagent adduct is generated using systematic conformational searching. With this, a molecular mechanics based energy is generated for each conformation. A 'reactive subset' of the conformations is then examined, i.e., the subset of conformations in which the second reactive group is within some arbitrary distance from the surface. The energies for this subset are converted to Boltzmann factors, essentially probabilities. These probabilities can then be analyzed, in conjunction with the geometric properties of the conformations that they represent, to make predictions of the expected separation of the two reactive groups on the surface after the second has reacted with the C60 surface.

The purpose of developing this methodology was in order to design the synthesis of a specific inhibitor. However, the computational analysis was also applied to a fullerene derivative that had already been synthesized in the laboratory of Francois Diederich[11] and shown to predict the observed reaction product. The details of this work will be given in Chapter 4.

Roots

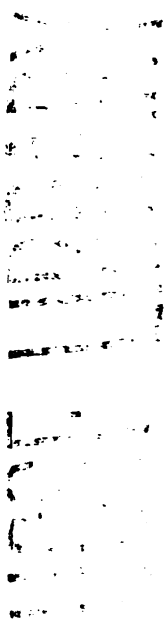
Although structure based ligand design is considered a modern endeavor of the 1980s and 90s, it is interesting to see how the roots of this work lie five or more decades in the past. Within a few years of the first electronic computers being constructed (ENIAC, MANIAC), their creators were already using them to model physical chemical systems. In the classic paper by Metropolis et al., [12] a methodology for application of "high speed computing devices" was used to model and predict the behavior of a simple two dimensional non-ideal gas. This paper demonstrated the possibility of using a non-exact solution to an equation of state using an appropriate approximation. What is particularly interesting about this work are the number of approaches that it utilized that are still used today: Monte-Carlo searching methods, lattice models and periodic boundary conditions, to name a few.

Paul Ehrlich initiated the idea of drug design in recognizing the possibility of targeting compounds to specific pathogens. A more specific insight came with the advent of the concept of anti-metabolites in the late thirties and early forties. This concept was largely explored in work done by D.W. Woolley and others. [13]The idea behind anti-metabolites was that a small molecule inhibitor could be made for a specific target enzyme by using features of substrates of the enzyme which were known to be responsible for binding, while altering the molecule in such a way as to prevent its use as a substrate. This work preceded knowledge of any crystal structures of proteins, and yet identified the *key* elements that would be required of a potential enzyme inhibitor: to have the appropriate shape and chemical nature. With the advent of protein crystallography, it was no longer necessary to be confined to substrate analogs. The guide could now be the structure of the binding site alone. Despite this, knowledge of substrate structure remains a key component of drug design and development, both when the structure of the target enzyme is known and when it is not known.

Chapter 2

Introduction

The following chapter consists initially of the Journal of the American Chemical Society paper in which the original concepts of the fullerene based inhibitors of the HIV protease were delineated. (Figure numbers and reference numbers have been changed to be consistent with the rest of the thesis.) Because of the constraints of the original format, additional information has been added to the end of the chapter to expand on issues mentioned briefly in the text. These include an elaboration of the ideas that preceded the testing of the compounds discussed in the Journal of the American Chemical Society manuscript text, as well as an discussion of the treatment of the hydrophobic effect and the computational tools used to quantitate it.



Inhibition of the HIV-1 Protease by Fullerene Derivatives: Model Building Studies and Experimental Verification

Simon H. Friedman[†], Dianne L. DeCamp[†], Rint P. Sijbesma[§], Gordana Srdanov[§], Fred Wudl[§], George L. Kenyon[†,*]

[†]Department of Pharmaceutical Chemistry, University of California, San Francisco, 94143
[§]Department of Chemistry, University of California, Santa Barbara, 93106

Abstract: The ability of C_{60} fullerene ("Bucky-Ball") derivatives to interact with the active site of HIV1-protease (HIVP) has been examined through model-building and simple physical chemical analysis. The model complexes generated via the program DOCK3 suggest that C60 derivatives will fit snugly in the active site, thereby removing 298$Å^2$ of primarily non-polar surface from solvent exposure and driving ligand/protein association. The prediction that these compounds should bind to the active site and thereby act as inhibitors has been borne out by the experimental evidence. Kinetic analysis of HIVP in the presence of a water soluble C60 derivative, di(phenethylaminosuccinate)C60, suggests a competitive mode of inhibition. This is consistent with and supports the predicted binding mode. Di-amino C60 has been proposed as a "second-generation" C60 derivative that will be able to form salt bridges with the catalytic aspartic acids in addition to Van der Waals contacts with the non-polar HIVP surface, thereby improving the binding relative to the tested compound.

Introduction

The protease specific to the human immunodeficiency virus 1 (HIVP) has been shown to be a viable target for anti-viral therapy[3]. The active site of this enzyme can be roughly described as an open ended cylinder which is lined almost exclusively by hydrophobic amino acids (Figure 2.1a). Notable exceptions to this hydrophobic trend are the two catalytic aspartic acids (Asp 25,Asp 125) which catalyze the attack of water on the scissile peptide bond of the substrate. We hypothesized that since a C_{60} molecule (i.e., fullerene) has approximately the same radius as the cylinder that describes the active site of the HIVP and since C_{60} (and its derivatives) are primarily hydrophobic, an opportunity therefore exists for a strong hydrophobic interaction between the C_{60} derivative and active site surfaces. This interaction should make C60 derivatives inhibitors of the HIVP. In this work we describe model complexes of C60 and HIVP generated via the program DOCK3. The surface desolvated due to complex formation is shown to be almost exclusively hydrophobic. In addition, kinetic analysis is presented that supports a competitive mode of inhibition of a tested C60 derivative, consistent with our intuition and the complexes generated. Finally, we propose and validate the design of an amino derivatized C60 as a reasonable next step in improving the binding energy of C60 derivatives to the HIVP.

Results and Discussion

To test the hypothesis regarding the complementarity of the C_{60} with the HIVP active site, a model of C_{60} was created and minimized using the SYBYL package (Version 5.4, Tripos Associates, Inc.). The model produced had a diameter within 0.2Å of a spectroscopically determined C_{60} structure[14]. This model was fitted into the active site of the so-called "open" (i.e., uncomplexed) form of the HIVP [4] using the program DOCK3[6]. DOCK3 finds optimal orientations of a ligand with its receptor, scoring on the basis of van der Waals contacts and complementary electrostatics. This procedure produced complexes with the C_{60} squarely in the center of the active site

forming good van der Waals contacts with the active site surface, thereby reinforcing our model. Figures 2.1b and 2.1c show the highest-scoring complex of C_{60} with HIVP in "front" and "side" views, showing the van der Waals surface contacts.

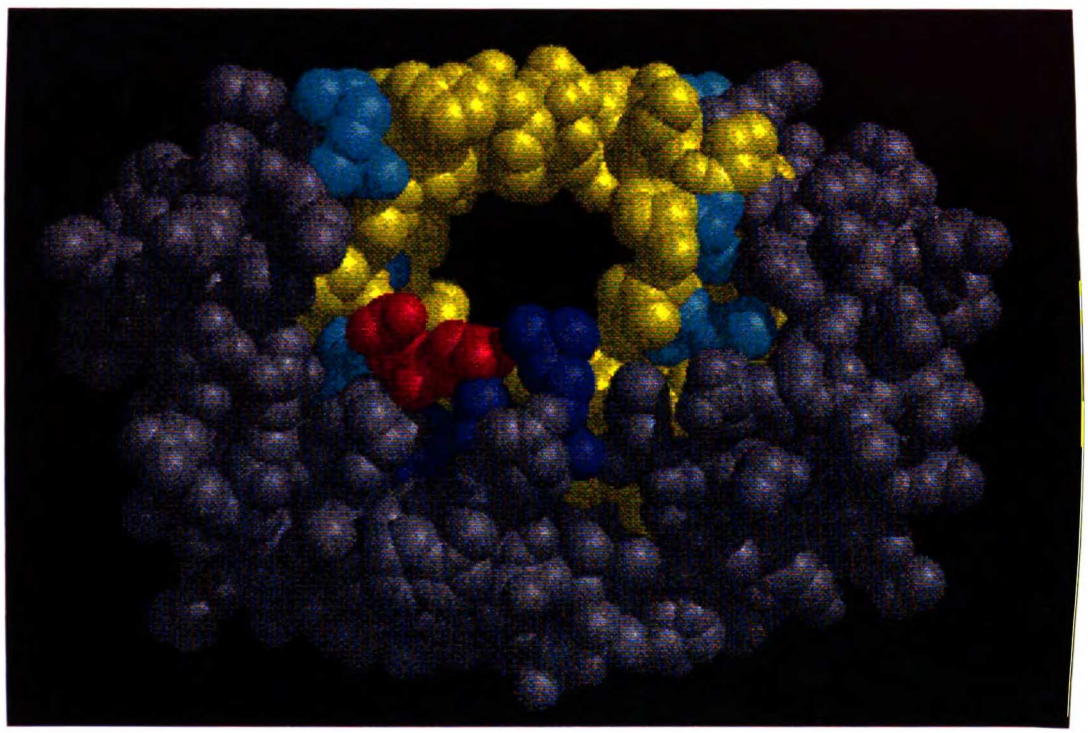


Figure 2.1. . (a) "Front" view of the HIV-1 protease. Color coding is as follows:
Yellow: LEU,ILE,PHE,TYR,VAL,TRP,PRO,GLY,ALA
Blue: LYS,ARG
Red: ASP,GLU
Cyan: THR,SER,GLN,ASN,CYS,MET,HIS
Gray: regions greater than 10 Å from the center of the active site.

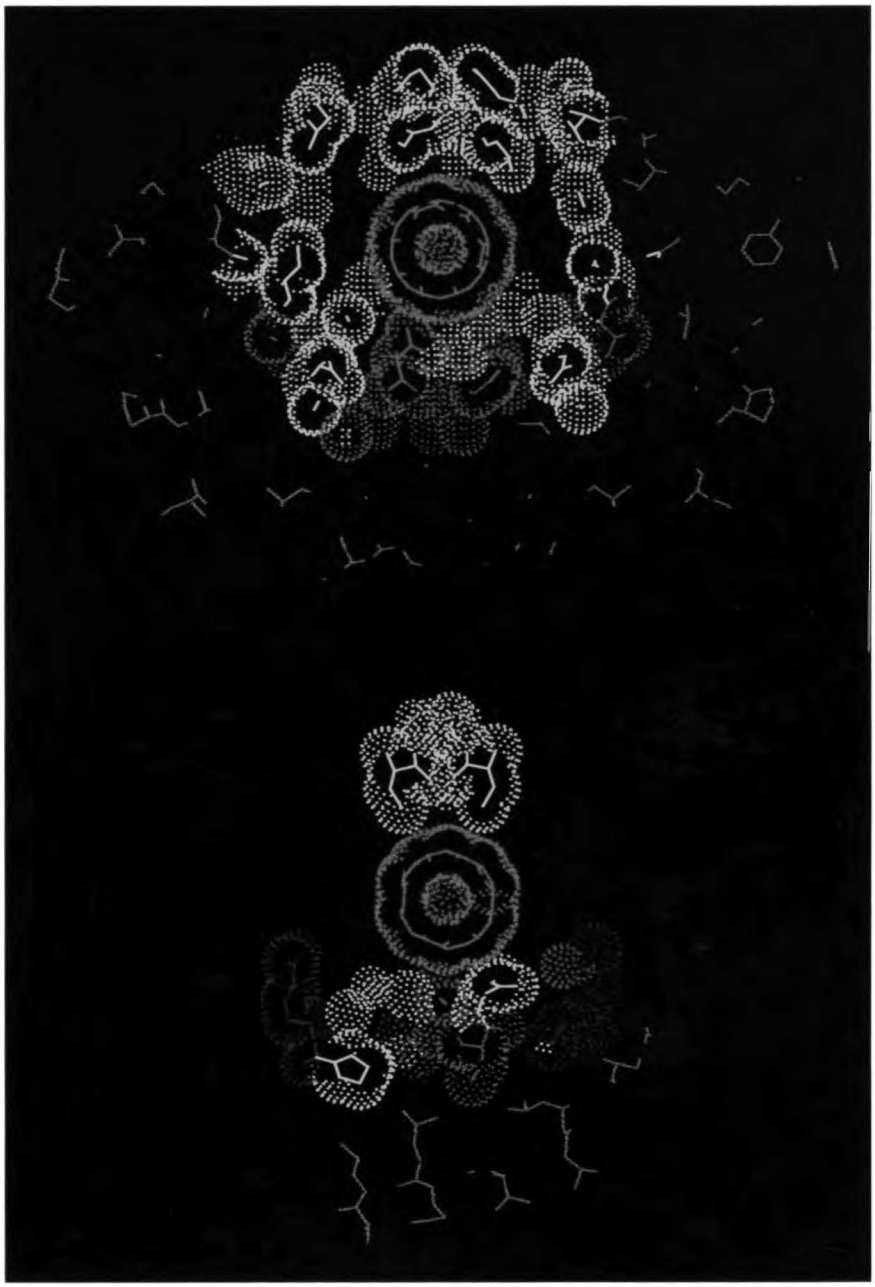


Figure 2.1b,c. (b) Same view as 2.1(a) with the top scoring C60 orientation shown. The C60 is colored magenta and the van der Waals surface of the active site and ligand are shown. (c) Same complex as (b) seen at a 90° cross section.

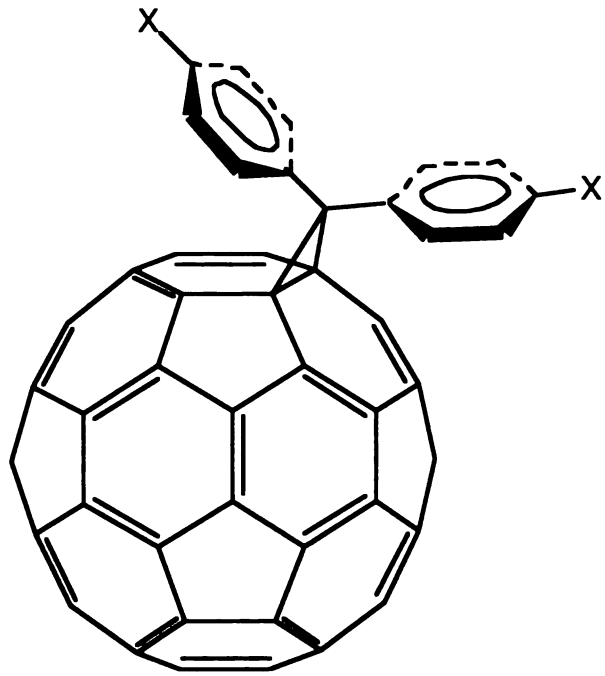
The change in solvent-exposed surface upon binding was determined in order to approximate the maximum magnitude of hydrophobic interactions. This was accomplished by first determining the total surface area of the active site and C_{60} molecules separately and then subtracting the total surface area of the highest scoring DOCK3 C_{60}/HIVP complex. All surface areas were determined from molecular surfaces generated by the program MS (Michael Connolly, University of California, San Francisco). The calculation indicates that 298 $Å^2$ of primarily hydrophobic surface is removed from solvent exposure. This total desolvated surface was further characterized by summing the individual surface elements according to atom type. The result of this summation (table 2.1) is that the large majority (273$Å^2$ or 92%) of the desolvated surface is due to C60-carbon/HIVP-carbon atom contact. The small amount of oxygen desolvation (7%) is due primarily to the partial blockage of the catalytic aspartates.

	CARBON	NITROGEN	OXYGEN
COMPLEX (HIVP + C60)	1537.64	109.272	266.456
HIVP	1402.55	112.504	287.898
C60	408.95	0.	0.
TOTAL CHANGE	-273.31	-3.232	-21.442

Table 2.1 Breakdown Of Molecular Surface Changes Upon C60/HIVP Complexation According To Atom Type Note: The surface areas of the complex and of HIVP were determined for an identical subset of the total protein structure which contained and flanked the active site.

Using the figure of 69.2 cal/mol/$Å^2$ recently shown to accurately describe the free energy released upon desolvation of hydrophobic molecular surface [15], the resultant free energy gain upon binding due to the carbon

surface that is desolvated would be 19 kcal/mol. In order to estimate an approximate binding constant of a C_{60} derivative, this value has to be corrected for the free energy cost due to loss of translational/rotational entropy that accompanies binding. This value has been estimated to be on the order of 7-11 kcal/mol [16]. Taking this energetic cost into account, the result is a total ΔG_{bind} of 8-12 kcal/mol. Converting this to Kd values using the expression $\Delta G°=-RTlnK_d$ results in dissociation constants on the order of 10^{-6} - 10^{-9} M.



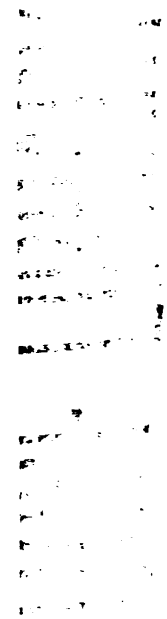
1, X = $HOC(O)(CH_2)_2C(O)NH(CH_2)_2$-

Figure 2.2. Structure of Tested Compound; di-phenethylaminosuccinate C60

Several factors have been left out of this analysis, for example, rotational entropy persistence of the C60 in the active site, conformational energy of the HIVP and interaction of the catalytic aspartates with the C60 surface. The purpose of this analysis is to account for the factors influencing

binding that are reasonably estimated from our understanding of protein-ligand interactions.

A relatively synthetically accessible water-soluble C_{60} derivative, di(phenethylaminosuccinate)C_{60} (Compound **1**, Figure 2.2), was the first test of our hypothesis. The synthesis and characterization of this compound are described in the accompanying manuscript[17]. The highest scoring DOCK3 complex of this compound with the HIVP again positions the core C_{60} in the center of the active site, with the charged side chains extending through the mouth of the active site into solution (Figure 3).



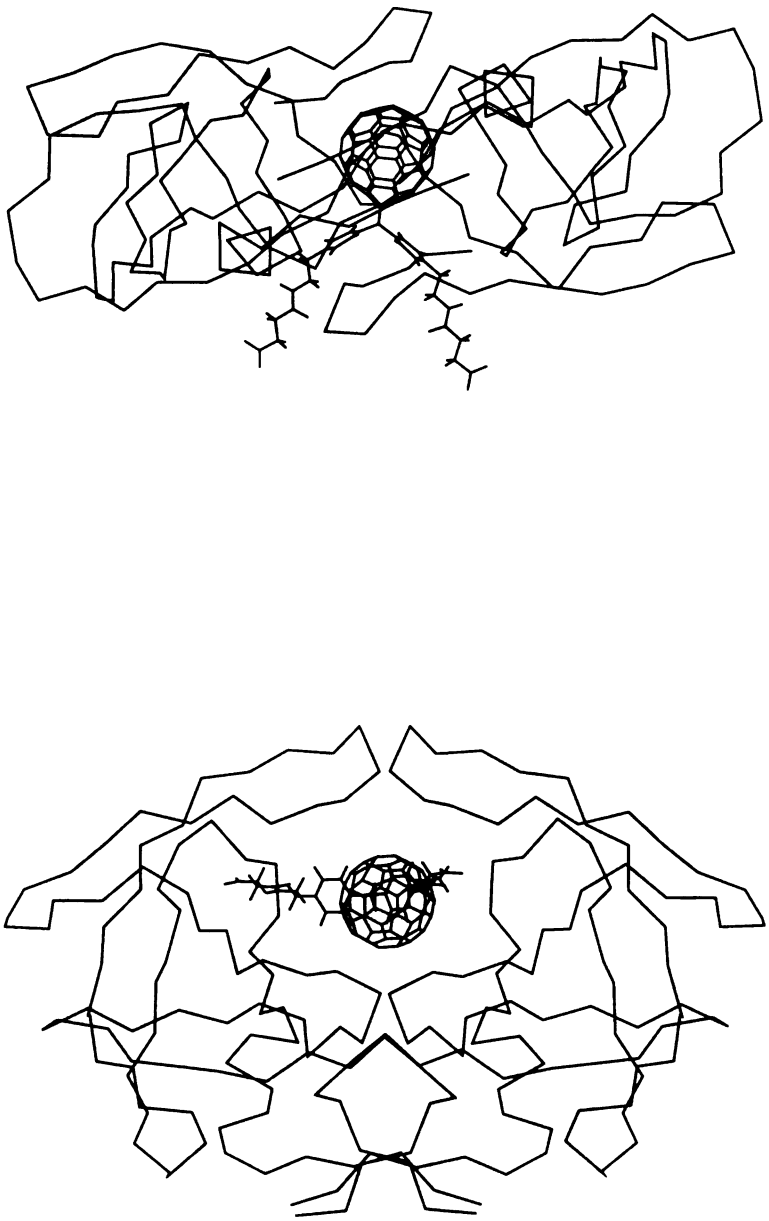


Figure 2.3 Top DOCK3 complex of Compound **1** in (a) "front" and (b) "top" views. For clarity, only the alpha carbon chain trace of HIVP is shown.

The ability of this compound to inhibit the HIVP was assayed with an HPLC method [18], and its K_i value was found to be 5.3μM (s.e. 0.98). The kinetic data fits the pattern of competitive inhibition well (Figure 2.4). This supports the proposed model complex, as the predicted binding mode of the C60 core should preclude any inhibitor binding while substrate is bound. It is of interest to note that compound **1** has been found to inhibit acutely and chronically HIV-1 infected human peripheral blood mononuclear (PBMC) cells with an EC50 of 7μM while showing no cytotoxicity in uninfected PBMC cells.(Raymond F. Shinazi, et. al., manuscript in preparation)

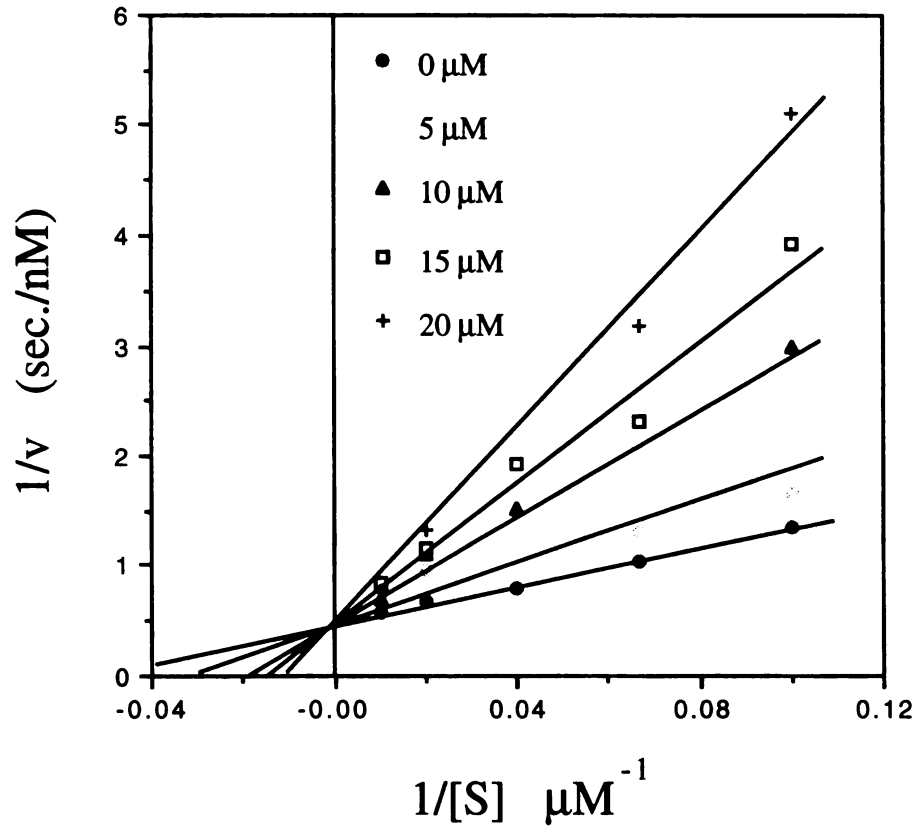


Figure 2.4. Double reciprocal plot of Compound **1** inhibition of HIVP. Assays were performed in buffer containing 50mM Na Acetate pH 5.5, 1.0 M NaCl, 5% Glycerol, 1% DMSO, 2mM EDTA. Kinetic constants were determined by fitting of the data to the equation $v=V_m \bullet S/[K_m \bullet (1+I/K_i)+S]$ which describes competitive inhibition. Ki:5.3μM [0.98], Km 15.9μM [2.9], Vm: 1.9 nM/sec. [0.1] Standard errors are indicated in brackets.

This introductory example demonstrates the potential for C_{60} based inhibitors of the HIVP. As a point of comparison, the best peptide-based inhibitors are effective in the sub-nanomolar range, and the best non-peptide inhibitors are effective in the nanomolar range [19-21]. The main driving force behind the association of the HIVP and the fullerene derivative examined is presumably hydrophobic interaction between the non-polar active site surface and C_{60} surface. In addition, however, there is an opportunity for increasing binding energy by the introduction of specific electrostatic interactions. An obvious possible electrostatic interaction is a salt bridge between the catalytic aspartates on the floor of the active site and a cationic site on the C_{60} surface. It has been found that several dicationic metals are effective inhibitors of the HIVP[22,23]. The authors of this work hypothesized that the principal mode of inhibition may be through a tight electrostatic interaction of the dication with the catalytic aspartates. K_i values for these dications are in the μM range, corresponding to ~8kcal/mol binding energy, over and above the Gibbs energy loss due to freezing out translational entropy. If even a fraction of the binding energy due to this type of interaction can be added to the existing C60 core binding energy, improvements of several orders of magnitude should be realized. It has been shown that the introduction of a single amine/carboxylate salt bridge can increase the binding energy of a ligand to its receptor by ~4 kcal/mol[24], leading to a thousand fold improvement in binding. With these ideas in mind we have investigated whether the active site of the HIVP can accommodate a C60 that has been derivatized with two amino groups and, in addition, if these groups can be positioned so that they are close enough to the catalytic aspartates to form effective salt bridges. Direct amino adducts can be prepared by reacting a neat amine with the C60, although controlling the stoichiometry of this reaction poses practical problems[25]. Synthetic issues aside, we have modeled compound 2, "1,4 diamino C60" (Figure 2.5) and examined its possible interactions with the HIVP active site.

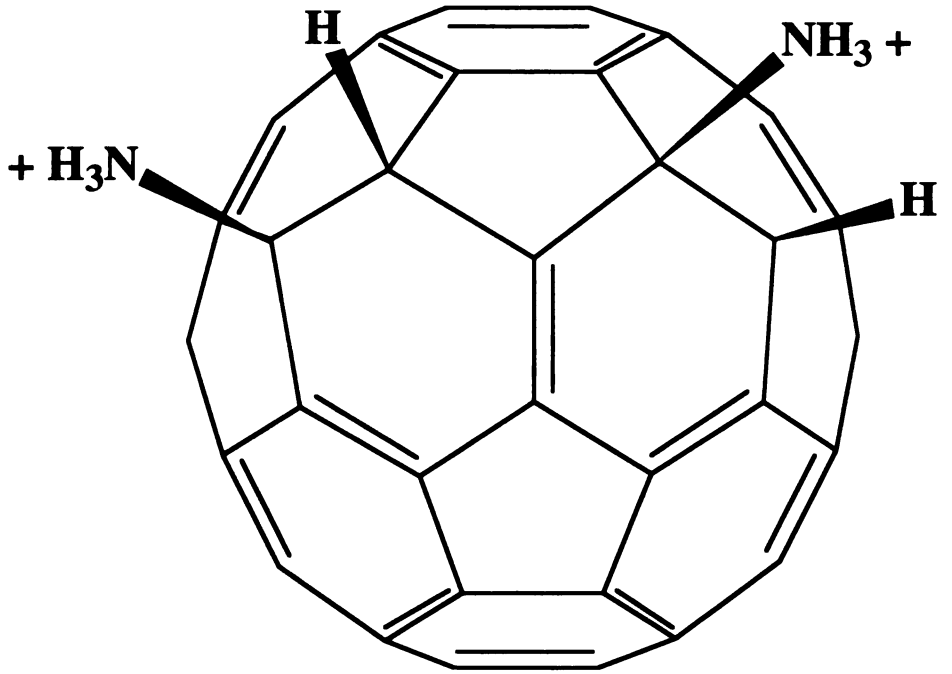


Figure 2.5. Compound **2**

DOCK3 is able to orient compound **2** within the active site placing the core C60 in a similar position to that of compound **1**, again allowing extensive non-polar van der Waals surface interaction. In addition, the two amino groups can effectively bridge the oxygens of the catalytic aspartates, approaching within 2.7 Å and 3.4 Å, respectively, (N-O distance), thus making these amino/carboxyl interactions good candidates for improving overall binding. (Figure 2.6)

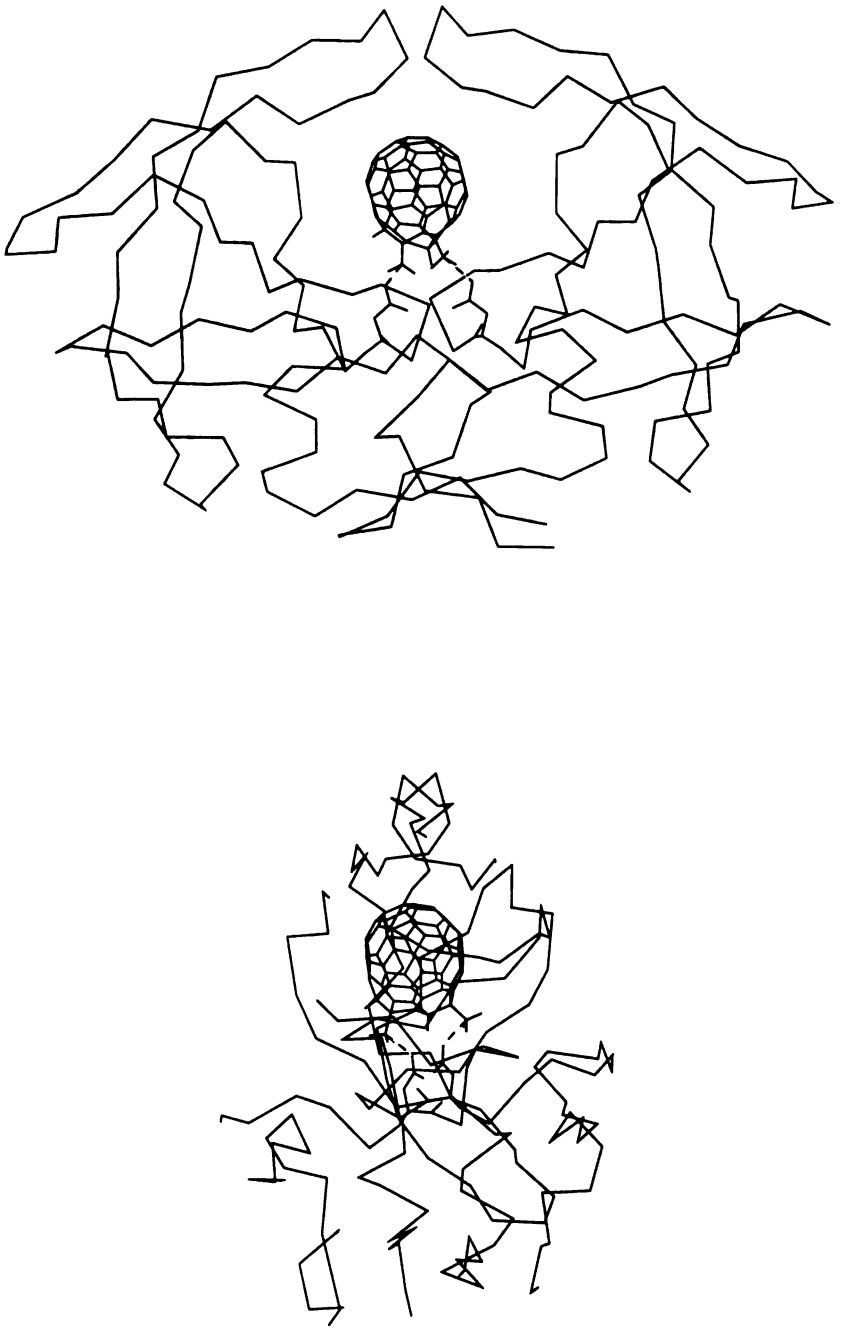


Figure 2.6. DOCK3 complex of Compound **2** with the HIVP in (a) "front" and (b) "side" views. Close approach of Compound **2** amino groups and HIVP catalytic aspartate oxygens is highlighted with dashed lines. For clarity, only the catalytic aspartates and the alpha carbon chain trace of HIVP are shown.

Conclusions

We have suggested through modeling that C60 derivatives should be inhibitors of the HIVP, due to their steric and chemical complementarity with the active site. We have demonstrated that this is the case and that this behavior is consistent with a qualitative analysis of hydrophobic surface transfer in model complexes.

Our interest in C60 derivatives is two-fold: First of all, they represent non-peptide based lead compounds that through careful modeling may result in effective tightly binding HIVP inhibitors. Secondly, they represent a rigid, conformationally restricted scaffold with which we can examine the nature of protein-ligand binding. Because of the steric bulk of C60 and its complementarity to the active site surface, there are severe limitations to the orientations it can adopt within the active site. Essentially, the principal degree of freedom of a C60 derivative within the active site is rotation around its center. This simplifies the problem of predicting the binding modes of various derivatives. The key to exploiting this system will be the development of the synthetic methodology to faciley and specifically modify the C60 surface.

Experimental Section

Modeling:

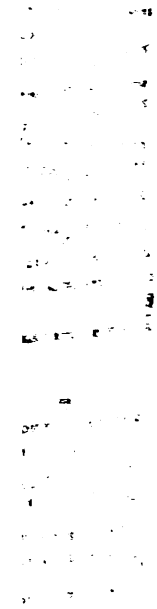
All modeled compounds were generated using the SYBYL 5.4 package. Atomic point charges were calculated using the Gasteiger-Huckel method. For conformationally flexible ligands, torsions were initially set to anticipated low energy conformers. Minimization to the used model structure was done using the Maximin2 minimizer and Tripos force field and parameters. Docking to the active site of the studied protein was done using the program DOCK3[6]. Grids required by DOCK3 were generated against the dimer formed from the Protein Data Bank file 3hvp, using the standard AMBER united atom charges and van der Waals parameters. Single mode runs of modeled compounds against the active site were performed using the following parameters: dislim 1.500, nodlim 5, ratiom 0.0000, lownod 4, lbinsz= 0.4000, lovlap= 0.1000, sbinsz= 0.8000, sovlap= 0.2000 . All molecular graphics were produced using the MIDAS Plus system. Molecular surfaces were generated using the program MS, written by Michael Connolly. A probe sphere diameter of 1.4Å and default values for vdw radii were used.

Enzyme Assays

Compound **1** was assayed against recombinant, affinity-purified HIV-1 Protease produced by Bachem Biosciences (.16mg protein/ml) at 25°C. Assays were performed in 100µl volume under final conditions of 50mM Na Acetate pH 5.5, 1.0 M NaCl, 5% Glycerol, 1% DMSO, 2mM EDTA. Inhibitor was pre-incubated with ~.05 µg of enzyme for 5 minutes at which time the reaction was initiated by addition of substrate. The reaction was quenched at <15% product formation by the addition of 15µl of 10% TFA. The cleavage products of the substrate peptide H-Lys-Ala-Arg-Val-Tyr-p-nitro-Phe-Glu-Ala-Nle-NH_2 (made by Bachem) were assayed by HPLC using a 10-40% (acetonitrile,.1%TFA):(water,.1%TFA) gradient over 30 minutes at 1ml/min.. Product was quantitated by integration of peak areas followed by comparison to product standard curves. Determination of kinetic constants was done with the program KinetAsyst (IntelliKinetics).

Acknowledgments: This work was supported by NIGMS grant GM39552 and NSF grant DMR9111097. The support of an American Chemical Society. Medicinal Chemistry Fellowship (S.H.F.) is acknowledged. We thank Diana C. Roe and Elaine C. Meng and Rafael Salto for valuable discussions and Charles S. Craik for providing recombinant protease.

Note Added in Proof: The parent compound to compound **1**, where X=(CH2)2NH2, was tested and found to have a Ki of ~2μM. This insensitivity of binding to the nature of the C60 side chain supports the predicted binding mode, which positions the side chains away from the active site into full solvent contact.



Background to the Journal of the American Chemical Society Paper

The original inspiration for thinking about fullerenes in the realm of the HIV protease lay in a conversation I had with Diane Roe in which we were discussing the state of the art in HIV protease inhibitor design. We were discussing the variety and strangeness of some of the molecules that had been used in an attempt to inhibit the HIV protease. Diane then said "What are they going to try next, Buckyballs?". My reaction was that, given the non-polar, cylindrical nature of the HIVP, that the spherical non-polar C60 fullerene might indeed be a reasonable candidate. I went to the white board and drew up a schematic protease and fullerene, with amino groups on the fullerene interacting with the catalytic aspartates. The key to this possibility lay in the relative size of the C60 and HIV protease active sites.

Several weeks passed. One evening while taking a break from another project, I drew up the C60 fullerene using the SYBYL molecular modeling package. Following minimization, I saved it in PDB format and opened it in the molecular graphics package MIDAS-plus. I also opened up the crystal structure of the "open" (uncomplexed) HIV protease (Brookhaven code 3hvp) and immediately saw that the cross section of the fullerene was in the right ball park to interact with the active site. By utilizing "hand docking" (simply moving one structure relative to the other, while observing surface contacts) and the 'watch' option in MIDAS (which indicates when the atoms of two molecules are less than their combined Van der Waal radii from each other), I was able to see that indeed the C60 fullerene was very complementary in shape to the active site of the HIVP.

For binding to take place, two criteria must be met, as described in the introduction: Shape complementarity and chemical complementarity. In this case, the C60 satisfied both of those criteria. In addition to the shape complementarity described above, there also exists the required chemical complementarity. The active site of the HIVP is lined predominantly by non-polar or hydrophobic amino acids. Because the C60 surface is entirely carbon, the surface is non-polar and prefers to be in relatively non-polar environments. Thus, the two criteria for potential binding were apparently satisfied for a C60.

Following this initial analysis of the feasibility of a C60 fullerene derivative, I did a "back of the envelope" calculation, estimating the potential binding affinity that could be expected from a C60 derivative. I used the radius of an underivatized C60 as determined from MIDAS to estimate the surface area of the C60. Taking two thirds of this value as an estimate of the actual amount of the surface of the C60 that can be desolvated upon complex formation, I then doubled this value to account for the desolvation of the protein as well. Taking this total desolvated surface area, I converted it into a free energy of transfer, using a value of 24cal/mol/Å^2, which I thought was a standard value.

There were several errors in this initial analysis. First of all, the value of 24cal/mol/Å^2 for the free energy contribution of hydrophobic desolvation is based on determinations of surface area where the *solvent accessible surface* is used. The solvent accessible surface is the surface of a molecule which is traced out by the center of a solvent molecule which is rolled along the surface of the molecule. Because the back of the envelope calculation I performed used the approximate surface of the C60 sphere, it was significantly underestimating the solvent accessible surface. This error would decrease the apparent potential binding affinity of a C60 molecule. However, there is another significant error in the estimation, that of not including the loss of rotational/translational entropy of the C60 derivative upon complexation. This is a significant quantity and not considering it will increase the apparent potential binding affinity. The combination of these two errors cancel somewhat, leading to a result that suggested a reasonable potential binding affinity. Reasonable enough to acquire a molecule to test. These errors show naiveté on my part; however, they do form the conceptual basis for a more accurate and sensible analysis of interactions between hypothetical C60 derivatives and the HIV protease that I later applied. Many of the methods are introduced at the end of this chapter and in the following chapter.

The original 'concept' molecule was an amino modified fullerene: one that would be able to form salt bridges with the catalytic aspartates while simultaneously desolvating a large amount of non-polar surface.

We contacted Professor Fred Wudl at UCSB because of his group's experience at adding amino groups to the surface of C60. However, while they had been able to add amino groups to the surface of C60, they had not been able to control the regiochemistry or stoichiometry to the point where an

appropriate molecule could be tested. Indeed, the typical result of the Wudl groups amine/C60 reactions were >= 6:1 amino:C60 adducts, which would not have allowed the hydrophobic desolvation that was expected to drive the association. Professor Wudl was, however, able to provide the water soluble fullerene derivative described in the Journal of the American Chemical Society manuscript, the so called di-phenethylaminosuccinate C60. This compound did not contain positively charged groups that could access the catalytic aspartates as originally envisioned. The result is that the partial negative charge of aspartate oxygens would not have an optimal interaction, i.e., full electrostatic complementarity. I felt that despite this loss of a potential stabilizing interaction, and the introduction of a potentially destabilizing interaction (the quenching of a partial negative charge with a non-polar C60 surface), the hydrophobic desolvation that would still accompany the complex formation could well be enough to drive the association. As shown in the Journal of the American Chemical Society manuscript, this was indeed the case.

The synthetic issues that prevented the regiospecific and stoichiometric synthesis of the envisioned C60 diamine were addressed and will be discussed in Chapter 4.

Hydrophobic Effect Discussion

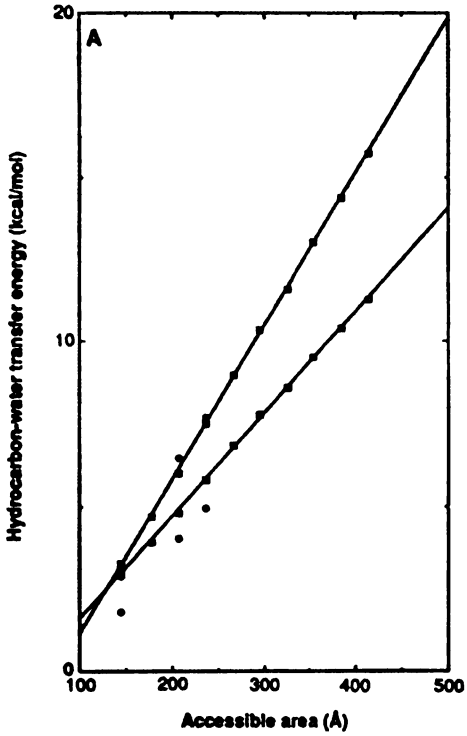
The purpose of this section is to elaborate on some of the concepts that were introduced in the preceding manuscript. Specifically, it will deal with the methods used for assessing hydrophobic surface transfer using the program MS and an awk based script, as well as the theory behind the magnitude of binding free energy that is contributed to by these surface transfers.

A key element in the initial analysis of fullerene inhibitors and, to a greater degree, the subsequent refinement of the design, was a reasonably accurate assessment of the free energy of transfer of non-polar surfaces from aqueous to non-aqueous environments, i.e., the hydrophobic effect. The literature in this field is vast and of varying complexity. My interest is in both the theoretical underpinnings of the phenomenon and in the practical consequences. I have therefore attempted to analyze the theory to the extent

necessary to make reasonable real world predictions. This has led me utilize the *molecular surface* (MS) to sum hydrophobic surface transfers, as opposed to the more prevalently used *solvent accessible surface* (SAS). It has also lead me to use a value of 70cal/mol/Å^2 as a value for the hydrophobic effect as determined from these molecular surfaces.

Over the past several decades there has been a large amount of work aimed at understanding the magnitude of the hydrophobic effect, i.e., the tendency for non-polar species to associate with each other in aqueous environments. Early workers showed a proportionality between the size of the molecule's surface area being transferred from an aqueous to non-polar solvent and the proportion of the molecules found in the non-polar solvent. Analysis of thermodynamic data determined for these transfer experiments indicated that at room temperature there was little enthalpic change in the system upon transfer of non-polar molecules from aqueous to non-polar environments. Instead, the transfer was driven by entropy. Commonly, the cause of this entropic change is envisioned as being the 'freezing out' of rotational degrees of freedom of water molecules which are in contact with non-polar surfaces. Because they cannot form effective hydrogen bonds or some other electrostatically based interaction with the non-polar surface, the orientations available to them with adjacent water molecules are limited. This restriction leads to a decrease in entropy of the solvent when non-polar species are present, and an equivalent free energy increase of the solvent when the non-polar species leaves the aqueous environment and enters a non-polar one.

It is reasonable to expect that the magnitude of hydrophobic effect would be some function of the amount of surface area of the transferred non-polar molecule. If the effect is indeed driven by entropic factors whose source is water interacting with non-polar surface, it would seem reasonable that as the transferred surface was increased, there would be a commensurate increase in solvent entropy and therefore free energy of transfer. In addition, one would expect that as the non-polar surface were decreased to zero, the transfer free energy would also tend towards zero. Graphs of transfer free energy succeed in this regard to varying degrees (Figure 2.7)



(**A**) Free energy of transfer of *n*-alkanes from hydrocarbon solvent to water. Solubility data from McAuliffe (*12*). (□) Analyzed with mole fractions (Eq. 1), (■) analyzed with volume fractions with a molar volume correction (Eq. 2), (○) uncorrected cyclohexane-to-water transfer energies (*20*), and (●) corrected cyclohexane-to-water transfer energies. Lines indicate best fits to the solubility data.

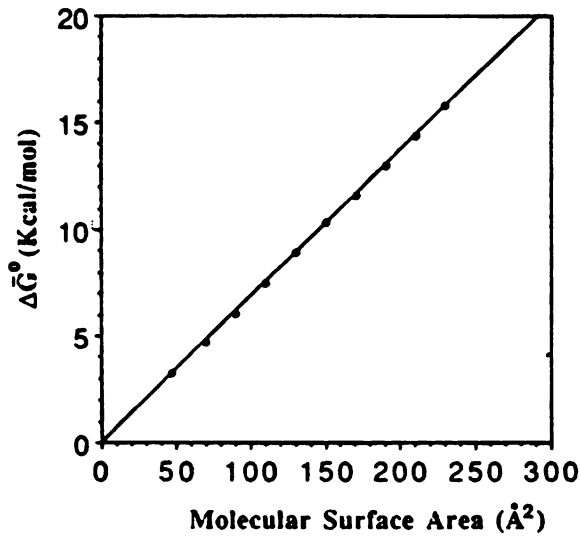
Figure 2.7. Figure taken from Sharp, K. et al. Science, April 5, 1991, pp106-109

In addition, the value of the hydrophobic effect derived from these transfer experiments is quite different (lower) from that determined from interfacial curvature experiments, which look at the curvature of the interface between a non-polar solvent and water. This gives an indication of the contact free energy, and another indication of the hydrophobic effect. Sharp and Honig have examined these systems in more detail and have suggested several factors have been ignored in the past which give better agreement between the transfer experiments (microscopic) and interfacial experiments (macroscopic).

Sharp and Honig have brought several points to the discussion. First of all they have introduced corrections to the free energies of transfer for non-polar solute molecules between aqueous and non-polar solvents, based on the relative sizes of the solvents and solute molecules. A more profound effect that they have discussed is the importance of the *curvature* of the solute surfaces. If the source of the hydrophobic effect is indeed the pinning down of

orientations of water which is in contact with a non-polar surface, the curvature of the surface should have an effect on the orientations available. For example, if a water is in contact with a highly concave surface, a smaller portion of its "orientational sphere" is limited than if it was in contact with a highly convex surface. Sharp and Honig attempted to work these types of corrections into the original transfer data, and succeeded in bringing the 'microscopic' value for the hydrophobic effect more in line with the 'macroscopic' value[26].

Tuñon and co-workers took the results of Sharp and Honig and found an interesting result: using Sharp and Honigs transfer data, which had been corrected for solvent and solute sizes, they plotted transfer free energy versus surface area. However, instead of using the area of the *solvent accessible surface* they used the area of the *molecular surface.* It is not clear from the publication why they did this. What is clear is that the resultant plot is linear and has an intercept at the origin. Furthermore, the slope of the line is within 3% of the value of the hydrophobic effect determined from solvent contact curvature experiments (i.e., the macroscopic value of 72-75 kcal/mol/Å^2). (Figure 2.8)



Plot of the transfer free energies of hydrocarbons from pure liquid to solution versus molecular surface areas.

Figure 2.8. Result of Tuñon et al. [15] showing correlation of hydrophobic surface transfer with the molecular surface.

What is the difference between the *solvent accessible surface* (SAS) and the *molecular surface* (MS)? The SAS is formed from the set of points traced out by the center of a probe solvent sphere as it is traced out upon the van der Waals surface of the molecule. The MS is formed from the points of contact between the probe solvent sphere and the van der Waals surface of the molecule. (Figure 2.9) It is clear that for a convex surface, SAS > MS, for a concave surface MS > SAS and for a flat surface MS = SAS.

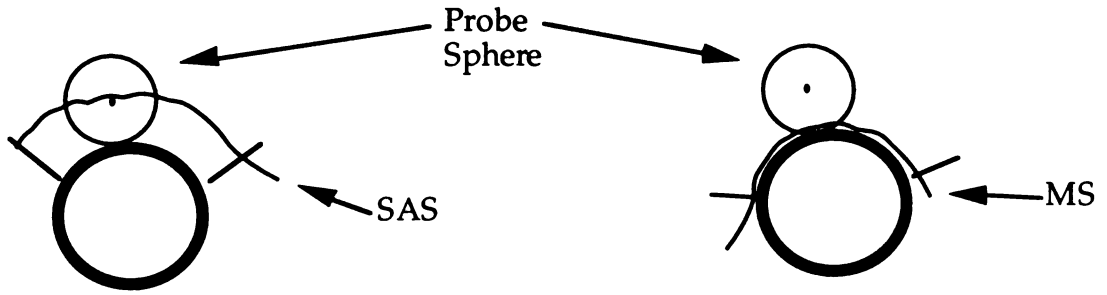


Figure 2.9. Comparison of the Solvent Accessible Surface (SAS) and Molecular Surface (MS)

Why does the MS work better than the SAS at giving reasonable assessment of the transfer free energy? Tuñon and coworkers suggested that the MS may be doing the correction that Sharp and Honig did 'by hand' to correct for differences in solute curvature. Why would this be so? If one accepts the idea that the source of the hydrophobic effect is the limitation of orientations of water molecules in contact with non-polar surfaces, then it is reasonable that the extent of this limitation is proportional to the *contact* surface between the water molecule and the non-polar molecule. It is helpful to think about this at the limits: As the curvature of a surface increases towards infinity, it eventually becomes a point. A non-polar point will have a finite SAS but a zero MS. Using the MS, one would predict no hydrophobic effect for transferring this point from an aqueous to non-polar solution. This is what is expected, as the individual water molecules surrounding the point have complete orientational freedom to contact adjacent water molecules. The SAS has failed in this limiting case to properly assess the hydrophobic effect by attributing a finite value to it. What the SAS is proportional to is the *number* of water molecules that can be in contact with a non-polar surface. What is important when assessing the hydrophobic effect, however, is not the number of water molecules that can pack against a non-polar surface, but

rather how that surface impairs the water molecule's interactions with each other. The MS will more closely assess that feature because it is proportional to the amount of area that the water molecule itself will have blocked from other water molecules.

In addition to questioning the theoretical basis for computational tools, it is necessary to test them against experimental data. This is as or more important. As well as the Tuñon data behave, they reduce a complex physical phenomenon to analysis by a single variable. Before relying heavily on the MS as a tool for assessing the free energy for transfer of non-polar surface from aqueous to non-polar environments, I checked its ability to reproduce some experimentally determined data for specific ligand/protein binding interactions.

I used a summary of hydrophobic binding free energy contributions found in Thomas Creighton's "Proteins" (1st edition, p365). This summary was from work by Fersht and Jencks. Typically in their studies, the affinities of small molecule enzyme substrates were measured, with and without a hydrophobic portion of the molecule present, in an attempt to estimate the binding contribution due to that hydrophobic group. I drew up the hydrophobic portions that had been analyzed, using SYBYL, generated low energy conformers and calculated the molecular surface using the program *ms*, which also provides the amount of the total surface area. The binding free energies as reported were corrected for conformational entropy using a simple statistical assessment of the 'active conformation'. In other words, if a given hydrophobic portion had several conformers, the free energy of binding that had been experimentally determined was corrected by $k\ln\Omega$, where Ω represents the number of conformers. The purpose of the correction is to make the free energy represent as closely as possible *only* that part of the binding energy that is a result of the hydrophobic effect between the non-polar portions of the amino acid and the enzyme, not the confounding effect of freezing out of a single conformation by the enzyme.

The resultant corrected free energies of binding were plotted against the determined molecular surface area. (Figure 2.10)

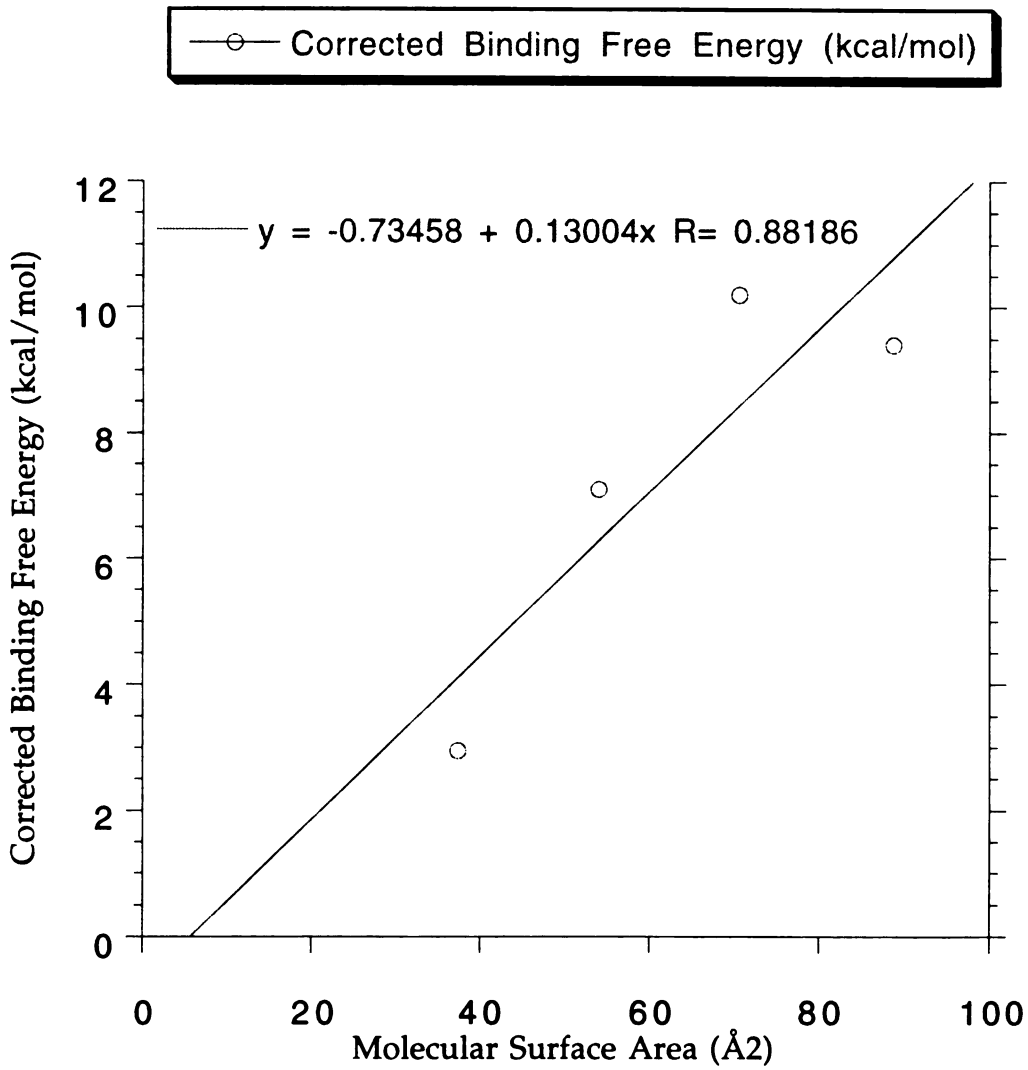


Figure 2.10. Replot of molecular surface area versus non-polar contribution to binding energy.

Non-polar substituent	Molecular Surface Area (Å2)	Conformers	Literature Binding Free Energy	Conformational Entropy Correction	Corrected Binding Free Energy (kcal/mol)
methyl	37.300	1.0000	-2 to -3.9	0.0000	2.9500
ethyl	54.000	3.0000	-6.5	0.60000	7.1000
i-propyl	70.600	3.0000	-9.6	0.60000	10.200
n-butyl	88.700	27.000	-7 to -8	1.9000	9.4000

Table 2.2. Data used in figure 2.10.

Although the data are noisy, the plot gives a value for the hydrophobic effect of 130 cal/mol/Å^2 with the intercept close to the origin. The value of 130 needs to be divided by 2 to give a proper assessment per total surface area contact, in order to account for the surfaces of the protein which are desolvated during binding, and which should be approximately equal to the desolvated surfaces of the hydrophobic amino acids. This results in a value of 65cal/mol/Å^2, a value which closely matches the value determined by Tuñon and coworkers from solvent transfer experiments.

The purpose of examining so closely the free energy gain expected from a given hydrophobic desolvation is so that during the design process increases in hydrophobic desolvation between known and unknown derivatives can then be assessed for their ability to improve binding. Being able to quantitate the amounts of surface desolvated during complexation and assess the proportion of this that was due to non-polar, carbon surface was not critical in the initial lead compound identification (this chapter). However, it was useful in the subsequent refinement of the designs (Chapter 3).

Computational Tools

The purpose of the following section is to briefly elaborate on how the molecular surfaces were generated and analyzed in the Journal of the American Chemical Society paper. The original purpose of the "surface transfer analysis" was to demonstrate that, as hypothesized, the majority of the surface of the protein and of the C60 derivative that was desolvated upon complexation was non-polar carbon surface, which would then result in the hydrophobically driven association. In the subsequent refinement work, discussed in the next chapter, the quantification of desolvated non-polar surface was the key guide in designing improved inhibitors.

The program MS was used to generate molecular surfaces of the protein, the ligand and of the complex. For surfaces involving the protein, a subset of the amino acids that line and flank the active site were used to speed the computation while simultaneously covering all appropriate areas. MS creates ascii surface files with the following format:

```
ARG    8   CD  36.311   53.196   -3.938 SS0  0.130 -0.436  0.849 -0.298
ARG    8   CD  36.432   53.889   -2.989 SS0  0.134 -0.205  0.694 -0.690
ARG    8   CD  36.797   53.449   -3.615 SR4  0.164 -0.281  0.865 -0.415
ARG    8   CD  36.809   53.683   -3.279 SR0  0.164 -0.289  0.698 -0.655
ARG    8   CD  36.910   53.370   -3.997 SR0  0.164 -0.362  0.921 -0.142
ARG    8   NE  36.127   51.290   -0.874 A
ARG    8   NE  35.057   52.312   -0.375 SR4  0.164 -0.534  0.534  0.655
ARG    8   NE  35.399   52.432   -0.154 SS0  0.140 -0.517  0.545  0.660
ARG    8   NE  35.589   51.903    0.392 SC4  0.188 -0.354  0.409  0.841
ARG    8   NE  35.318   52.092    0.112 SC9  0.188 -0.534  0.534  0.655
ARG    8   NE  35.759   52.494    0.001 SS0  0.140 -0.284  0.585  0.759
ARG    8   NE  36.080   52.434    0.170 SS0  0.130 -0.247  0.509  0.825
ARG    8   CZ  35.240   50.293   -0.997 A
ARG    8   CZ  33.358   50.165   -1.260 SS0  0.150 -0.998 -0.053 -0.033
ARG    8   CZ  33.384   50.066   -0.622 SS0  0.150 -0.993 -0.061  0.100
ARG    8   CZ  33.389   50.582   -1.330 SS4  0.150 -0.981  0.179 -0.072
```

Columns 4 5 and 6 indicate the cartesian coordinates of a surface point, and the 8th column indicates the amount of surface area that is associated with it. The point referred to can also be one of the original input atoms, as opposed to a surface point. This is signified by the presence of an 'A' or 'S'

flag in column 7. Because each surface point is identified as being associated with a specific atom type , which is listed in column 3, the total surface can be summed according to atom type.

Summation of atom types was performed using a simple awk script. (awk is a UNIX based pattern scanning language that can quickly identify elements within a file and perform simple mathematical manipulations on them.) The awk script used is as follows:

```
{if (substr($3,1,1)=="C" && substr($7,1,1)=="S")  t+=$8 }
{if (substr($3,1,1)=="N" && substr($7,1,1)=="S")  n+=$8 }
{if (substr($3,1,1)=="O" && substr($7,1,1)=="S")  o+=$8 }
END{print "carbon: ",t," nitrogen: ",n," oxygen: ",o}
```

Essentially, the awk script looks at each line, asks the question "Is this a surface point?". If it is, it identifies the element that is associated with it, carbon, nitrogen or oxygen, and sums the area in the appropriate category, giving a single output line:

carbon: 1494.41 nitrogen: 109.519 oxygen: 244.041

This approach only examines heavy atoms, i.e., no hydrogen. The reason for this is to avoid the ambiguous nature of hydrogen surface in a molecule. If the desolvated hydrogen is attached to a nitrogen, it is polar. If it is attached to a carbon, it is non-polar. The protein crystal structure starts off without hydrogens, so the small molecule ligand is examined without hydrogens, to allow for the assessment of the proportions of non-polar (carbon) and polar (nitrogen and oxygen) surface transfer. By using the result of the awk script scan of the surface files for the ligand, protein and ligand/protein complex, the total change in surface solvation upon complexation by atom type was analyzed.

Chapter 3

Introduction

The focus of this chapter is the refinement of the original fullerene based inhibitors of the HIV protease, by introduction of interactions between ligand and protein previously not available to the tested compounds.

The initial work on the fullerene derivatives was encouraging. A hand built model of interaction with the HIV protease was supported by a more objective analysis by DOCK. This model was then supported by the experimental results, which showed a competitive kinetic pattern with a Ki value in a range expected from an estimation of hydrophobic desolvation. My next interest was in refining these designs and introducing elements that could improve binding. The originally tested compounds described in Chapter 2 were not the original compounds I had in mind as the basis for inhibitors of the HIVP. My original interest was in amino derivatized C60, which could introduce interactions of the amines on the C60 surface with both the catalytic aspartates and anionic sites on the mouth of the active site. When it became clear that controlling the stoichiometry of the amine addition reaction had not been achieved by Wudl and coworkers (the original contact for fullerene derivatives), the alternative compounds described in the last chapter were selected for initial testing. Although they could not introduce the salt bridge to the catalytic aspartates originally envisioned, I felt that there were enough features (i.e., large amounts of hydrophobic desolvation) to potentially compensate, a belief borne out by experiment.

An alternative to introducing the salt bridge interactions with the catalytic aspartates was the identification of other sites in the active site which the previously tested fullerene derivatives had not been able to interact with. This is the focus of a manuscript submitted to the Journal of Medicinal Chemistry, which is reprinted verbatim in the following pages. This manuscript is a brief summary of the modeling of a large number of molecules, and touches on some of the pitfalls encountered. Additional details of the modeling and experimental analysis follow the Journal of Medicinal Chemistry manuscript.

Optimizing the Binding of Fullerene Based Inhibitors of the HIV-1 Protease Using a Novel Modeling Strategy

Simon H. Friedman, Padma Ganapathi, Yves Rubin, George L. Kenyon
Dept. of Pharmaceutical Chemistry, University of California, San Francisco 94143-0446
Dept. of Chemistry, University of California, Los Angeles, 90095-1569

Abstract

We have developed and applied a computational strategy to increase the affinity of fullerene based inhibitors of the HIV protease. The result is a ~50-fold increase in affinity from previously tested fullerene compounds. The strategy is based on the design of derivatives which may potentially increase hydrophobic desolvation upon complex formation, followed by the docking of the hypothetical derivatives into the HIV protease active site and assessment of the model complexes so formed. The model complexes are generated by the program DOCK and then analyzed for desolvated hydrophobic surface. The amount of hydrophobic surface desolvated was compared with a previously tested compound, and if the amount was significantly greater than the previously modeled, synthesized, and assayed compound, it was selected as a target. Using this approach, two targets were identified and synthesized, using two different synthetic approaches: a diphenyl C60 alcohol based on a cyclopropyl derivative of Bingle et al., and a di-isopropyl cyclohexyl C60 alcohol as synthesized by Rubin et al. Both showed tighter binding than the originally tested compound, (di-phenethylaminosuccinate C60, Ki=5μM) with 103 and 150 nM as Ki values respectively. In addition to demonstrating the utility of this approach, it shows that simple modification of fullerenes can result in high affinity ligands of the HIV protease, for which they are highly complementary in structure and chemical nature.

Introduction

For many years the protease specific for the Human Immunodeficiency Virus (HIVP) has been a target for anti-HIV drug development. We were drawn to the fullerene core as the basis for inhibitors of this enzyme because the C60 fullerene's structure is highly complementary, both sterically and chemically, to the HIVP, thus potentially leading to its binding to the HIVP active site[27]. The active site of the HIVP is approximately cylindrical and lined predominantly with hydrophobic amino acids. Our original modeling suggested that an appropriately derivatized C60 would fit snugly in the active site. Furthermore, the modeled complex indicated that the majority of the surface that becomes desolvated upon complexation would be hydrophobic. An estimate of the binding affinity of a fullerene based on the calculated hydrophobic desolvation indicated a potential affinity constant approximately in the low micromolar to nanomolar range. This was confirmed with experiment, and compound **1** (Table 3.1) was shown to bind with a Ki value of 5μM[27]. The modeled complex of this tested compound, generated using the program DOCK, showed the core C60 fitting snugly in the active site, and the derivatizing "arms" extending out of the active site into solvent (Figures 3.1a, 3.1b). The modeled complex was supported in several ways: 1) The binding kinetics were competitive. 2) The binding affinity of a closely related derivative, the diamine parent compound of compound **1**, had a nearly identical Ki value. This insensitivity of the binding affinity to the nature of the solubilizing arms supported minimal interaction of them with the active site. 3) The approximate estimated binding affinity was consistent with the amount of desolvated hydrophobic surface.

Because of the success of the initial model building and consistency of the model with experimental data, we have sought to improve on the binding affinity of these compounds in a rational manner and use modeling techniques to introduce substituents to the C60 surface that are able to exploit regions of the HIVP which did not interact with the initial lead compounds. The method employed involves identification of regions unexploited in model complexes of previously tested compounds. This is followed by design of hypothetical derivatives which are potentially able to interact with these regions. The hypothetical derivatives are then fitted into the active site,

using the program DOCK[6]. These modeled complexes so generated are then analyzed for the amount of hydrophobic surface desolvation brought about through complex formation. If this amount is a significant increase over the hydrophobic surface desolvation achieved by the lead compound **1**, it is selected as a synthetic target. Because the C60 is highly complementary to the HIVP, and is a completely conformationally restricted scaffold, the modeling of hypothetical compounds and their interactions with the HIVP active site are simplified.

Modeling Strategy

The first step in the design strategy was the identification of regions of the HIVP that were unable to interact with the previously tested compound **1**. Examination of modeled complexes of compound **1** with the HIVP and their associated molecular surfaces indicated large non-polar regions in the middle of the active site with which compound **1** was unable to interact. Figure 3.1c shows a thin cross-section through the middle of the DOCK generated model complex (i.e., containing the C2 axis and catalytic aspartates) of compound **1** and the protease. It shows two symmetrical water-exposed channels lined predominantly with non-polar surfaces. These channels form because the active site is not a perfect cylinder but instead is elliptical in cross-section, and the bulkiness of the side chains of **1** prevents them from accessing these areas. The strategy then was to design C60 derivatives that could fill one or both of these regions with a non-polar group, and, in so doing, increase the amount of hydrophobic desolvation relative to compound **1** and therefore increase the binding affinity.

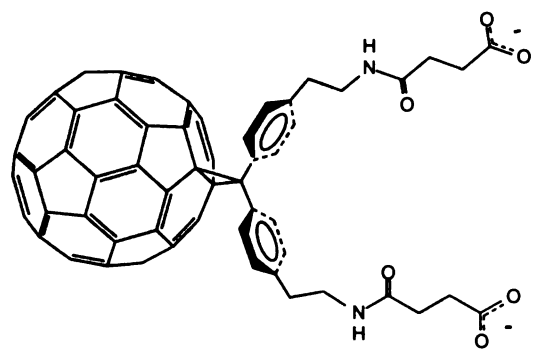
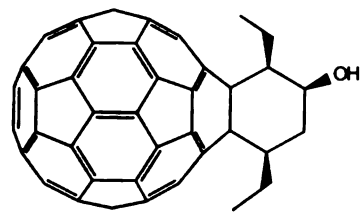
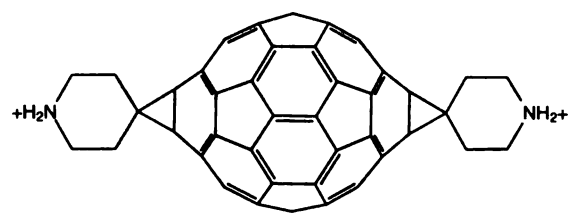
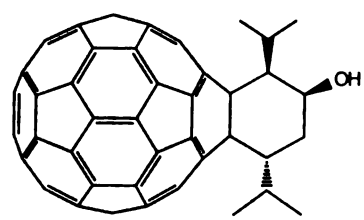
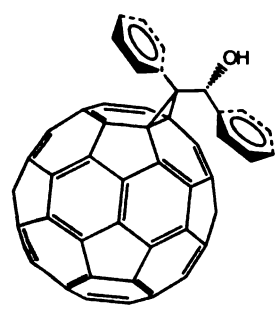
The structure of hypothetical derivatives were modeled using the SYBYL package[9]. The program MINDOCK was then used to generate hundreds of possible orientations of the hypothetical derivative with the active site. MINDOCK is a version of DOCK which uses rigid body minimization to relieve unfavorable contacts within the complexes[7]. The complexes were then scored within MINDOCK using a previously described approach which sums "good" and "bad" van der Waals interactions and electrostatics[6].

Following the initial screen of the compounds using DOCK's scoring, top scoring complexes from each run were further analyzed. Specifically, the

amount of non-polar surface that was desolvated during complex formation was analyzed. Molecular surfaces for the inhibitor alone, the uncomplexed HIVP and the inhibitor/protease complex were generated using the program MS[28]. MS generates surfaces which form at the contact point between a probe water sphere and the vdW surface of the molecule being analyzed. MS generates files which contain atom types for each surface element. Using these files and an 'awk' script, the total surface for each atom type N O and C, were separately summed. The total of C atom areas was used as an estimate of non-polar surface. Using the difference between the final complex surface and the sum of the uncomplexed surfaces, it was possible to determine the amount of surface desolvated and the proportion of it due to non-polar desolvation. Typically, the desolvation due to non-polar surface was in the 90% range.

When we analyzed the originally tested compound **1**, we found that it formed a complex that could desolvate 333-352 $Å^2$ in non-polar surface depending on slight variations in the complex. The diamine parent of compound **1** created similar complexes (i.e., with the solubilizing arms extending into solvent), and desolvated 381 $Å^2$ of non-polar surface. The value of 381$Å^2$ became a floor upon which hypothetical compounds would have to improve substantially in order to be selected as synthetic candidates. Table 3.1 shows some of the 40 compounds analyzed using this method, together with their desolvated surface area breakdowns. The amount of increase in hydrophobic desolvation relative to compound **1** was used to estimate the potential improvement in binding relative to this compound. Many of the analyzed compounds are chiral and are most readily synthesized as racemates. The modeling required analysis of each enantiomer separately, because the different enantiomers should interact differently with the chiral environment of the protease active site. In Table 3.1, therefore, the surface desolvation listed is for the enantiomer that gave the largest amount of desolvation, which is the actual enantiomer pictured. Figure 3.2c shows the cross section of the HIVP, now with the presence of compound **5** in a DOCK generated configuration. The previously water exposed channel (from Figure 3.1c) is now partially occluded by the substituent on the C60, producing a decrease in non-polar surface exposure. This is supported by the total amount of non-polar surface desolvation for this compound indicated in Table 3.1, which is 472 $Å^2$ or ~90$Å^2$ improvement over compound **1**.

Table 3.1. (Following Page) Fullerene derivatives which were analyzed for their ability to desolvate non-polar surfaces in model complexes formed with the HIV-1 protease. Molecules 2 and 3 were not synthesized, but are shown to represent some of the range of derivatives that were analyzed and rejected as possible synthetic targets. (s.e. : standard error)

Compound		Hydrophobic Desolvation ($Å^2$)	Ki (s.e.)
	**1**	-332.8, -352.1	5μM (0.98)
	**2**	-367.6	
	**3**	-457.8	
	**4**	-444.7[29]	150nM (48)
	**5**	-472.4	103nM (37)

Compounds **4** and **5** were selected as targets for synthesis and testing because of the relatively large increases in hydrophobic solvation they produced, as well as the relative ease of synthesis. Our initial interest was in bifunctional compounds such as compound **3** which potentially could interact with both solvent exposed channels simultaneously. For synthetic ease, however, mono-functional compounds were examined closely too, and were found to satisfy the design criteria (i.e., a large increase in hydrophobic desolvation relative to compound **1**.)

Compound **4** was prepared as previously described[30]. Compound **5** was prepared by DIBAL-H reduction of the ketone precursor originally synthesized by Bingel and coworkers[31]. The original intent behind the OH groups on these molecules was to use them as a branch point to introduce additional functionalities, primarily to aid in solubilization and possibly aid in binding. Top scoring complexes of these two molecules show the OH being partially buried in the active site, meaning that additional groups could not fit onto the OH, within the context of the modeled complex. However, further derivatization from the OH could produce a molecule which binds in a different orientation, but with equal hydrophobic desolvation and affinity. We are currently exploring this possibility.

Fortunately, both of the compounds **4** and **5** were soluble enough to assay as is, in the presence of 1% organic solvent. (N-methyl pyrrolidone in the case of **4** and dimethyl sulfoxide in the case of **5**). The results are shown in Table 3.1. Compound **4** has a Ki of 150 nM, and compound **5** has a Ki of 103nM, representing an increase in affinity of about 50 fold relative to the lead compound **1**. The assay conditions used were identical to those originally used for compound **1**[27], (50mM sodium acetate pH 5.5, 1.0 M NaCl, 5% glycerol, 1% DMSO, 2mM EDTA) except that the relatively stable mutant enzyme Q7K was used[32]. A control showed that the Ki of compound **1** with the mutant enzyme was identical with its originally determined Ki. The use of standard high salt conditions for the assay no doubt contributes to the affinity of the tested compounds, which is primarily hydrophobically driven. However, of primary interest to this work is the *relative* change in affinity, in comparison to the compounds that were tested in an identical manner, i.e., compound **1**. As before, the inhibitors were preincubated with enzyme for 5

minutes. There was evidence for a slow-on kinetics without preincubation; however, with 5-minute preincubation, time courses for the evolution of product were linear with respect to time, with 0 or near 0 intercepts.

The data were best fit to a competitive model of inhibition, although the standard error as shown is high. Typical standard deviation in repeated determinations of the inhibited velocities was 30% of the mean with these compounds, whereas the identical assay with a control inhibitor (acetyl pepstatin) showed standard deviations of 8% from the mean. This may be due to solubility problems and possibly surface absorption, leading to scatter in assay results.

Because large regions of the fullerene derivative surface are both unmodified and in solvent contact in modeled complexes, there is an opportunity to add new elements to the unmodified regions that can potentially add binding energy as well as modulate solubility, toxicity and bioavailability. In general the analysis of the interaction of fullerene derivatives with the HIVP active site is simplified, in part due to the nature of the fullerene sphere and its complementarity to the HIVP active site. The starting point for design is an active site in large contact with a completely conformationally restricted C60 sphere. Additional elements that are introduced to the C60 surface are positioned by the sphere, which can orient in a limited number of ways, thereby simplifying the prediction of binding mode. The ability of the fullerene surface to position multiple elements required for binding to a target, and to do so in a conformationally restricted manner, may make them of wider utility in pharmaceutical applications as generic, conformationally-restricted pharmacophore scaffolds.

Experimental

Synthesis of Compound 5:

Compound 5 was produce via DIBAL-H reduction of the ketone precursor. The ketone was prepared by the method of Bingel[31]. Ketone (52mg, 5.7×10^{-5} mole) was dissolved in 30ml of toluene which was sparged with argon and wrapped with aluminum foil. DIBAL-H (87 µl of 1M solution) in toluene (Aldrich) was added to the stirring mixture. After 3.5 hours an additional 40ml of toluene was added, and 239 µl of 1M DIBAL-H in

toluene was added. After 2.5 hours and additional 582 µl of 1M DIBAL-H was added. After ~24 hours reaction, TLC showed quantitative conversion to a lower Rf product (Rf 0.65 in toluene.) The toluene was extracted with water, and dried with MgSO4, then filtered and taken to dryness on a rotoevaporator, yielding 45mg of crude product (87%). This material was purified by silica column. The 45mg product was dissolved in 10ml of 3:1 toluene:hexanes and loaded onto a 28x2.5cm silica column. This was eluted with 3:1 toluene: hexanes and yielded 6.5mg. ^{1}H NMR (300MHz, CDCl3) 2.67 (d, 1 H, J=5Hz), 6.44 (d, 1 H J=5Hz), 7.5 (m, 10 H) ^{13}C NMR (10sec. relaxation) 4 resonances <100ppm at 57.49, 72.35, 78.42, 79.10 corresponding to the 4 non aromatic carbons in the structure. A cluster of 56 resolvable peaks was observed ~122-151ppm corresponding to most of the peaks expected from the 70 aromatic carbons in the molecule. MS (LSIMS) 915.9. (Exact mass expected 916.1)

Enzyme Assays

Compounds were assayed at 25°C using recombinant HIV protease (mutant Q7K)[32]. Assays were performed in 50µl volume under final conditions of 50mM Na acetate, pH 5.5, 1.0 M NaCl, 5% Glycerol, 1% DMSO, 2mM EDTA. Inhibitor was pre-incubated with enzyme for 5 minutes at which time the reaction was initiated by addition of substrate. The reaction was quenched at <15% product formation by the addition of 10µl of 10% TFA. The cleavage products of the substrate peptide H-Lys-Ala-Arg-Val-Tyr-p-nitro-Phe-Glu-Ala-Nle-NH_2 (made by Bachem) were assayed by HPLC using a 10-40% (acetonitrile, 0.1%TFA):(water, 0.1%TFA) gradient over 30 minutes at 1ml/min.. Product was quantitated by integration of peak areas followed by comparison to product standard curves. Determination of kinetic constants was done with the program KinetAsyst (IntelliKinetics)

Acknowledgments

We wish to thank Dr. Dan Gschwend and Professor Irwin Kuntz for access to MINDOCK in early stages of development. This research was supported by the NIH (Grant GM 39552)

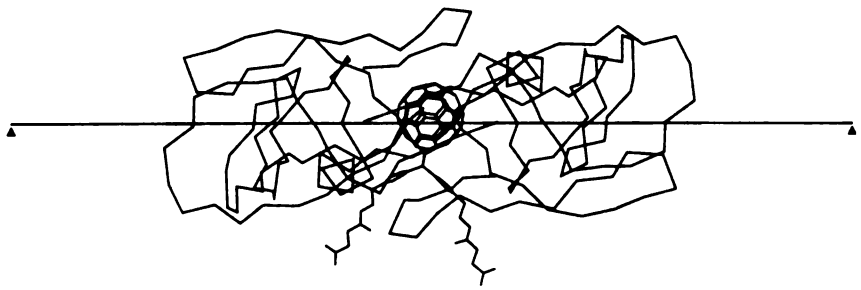


Figure 3.1. a) DOCK generated complex of compound **1** with the HIVP, in "top" view. The thin line in Figure 3.1a indicates the plane of the cross-section depicted in 3.1c.

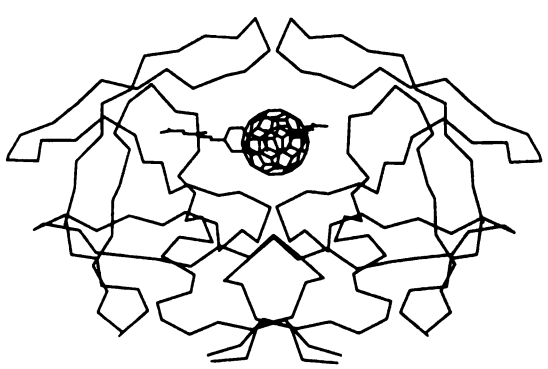


Figure 3.1. b) This same complex viewed from the "front".

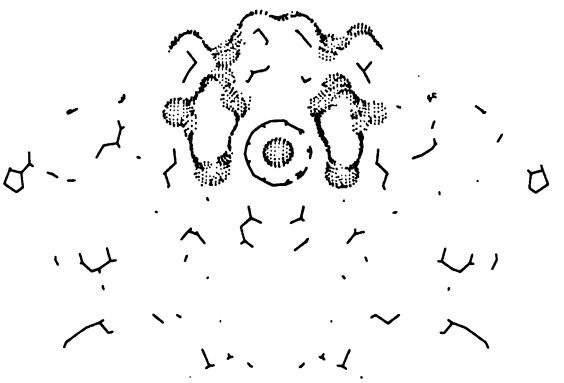


Figure 3.1. c) The identical front view of the compound **1** complex, but with only a thin cross section of the complex showing. The cross-section contains the C2 axis and both catalytic aspartates. The dots are the points of a solvent accessible molecular surface generated from the complex.

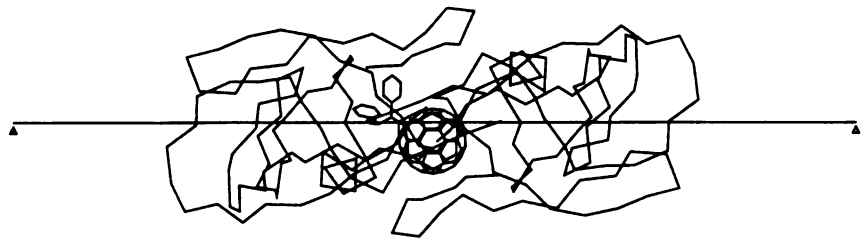


Figure 3.2. a) DOCK generated complex of compound **5** with the HIVP, in "top" view. The thin line in Figure 3.1a indicates the plane of the cross-section depicted in 3.1c.

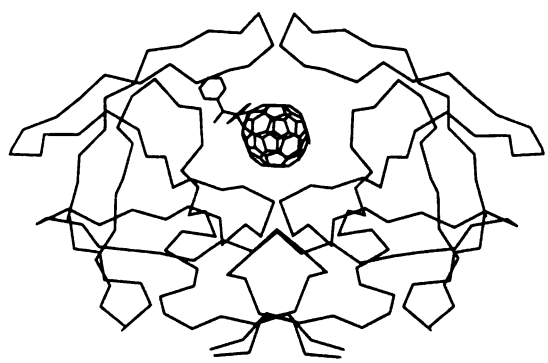


Figure 3.2. b) This same complex viewed from the "front".

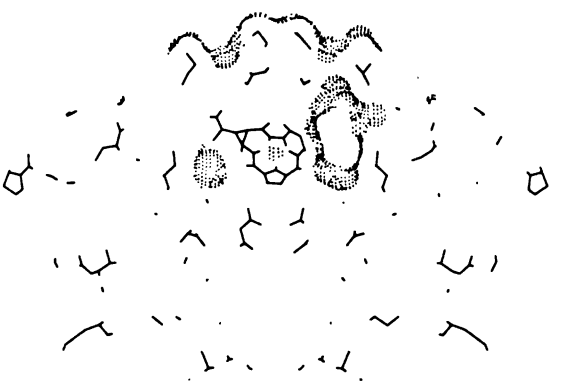


Figure 3.2. c) The identical front view of the compound **5** complex, but with only a thin cross section of the complex showing. The cross-section contains the C2 axis and both catalytic aspartates. The dots are the points of a solvent accessible molecular surface generated from the complex.

Background to Journal of Medicinal Chemistry Paper

The previous manuscript briefly summarized the work in refining the original fullerene designs. The following pages elaborate on that description and describe in detail the somewhat tortuous route to the final two target molecules.

The analysis of these hypothetical fullerene derivatives was done via single mode docking, in which a single molecule is oriented by DOCK in the active site. More typically, DOCK is used in database mode where a large database of molecules is screened for good fit. In the case of the fullerene refinement, I used the single mode docking as a design tool, by iteratively examining how small changes in a given molecule would improve or abrogate binding. This was an approach I had also started using in 1990 in a project to design molecules that would specifically bind to the minor groove of RNA/DNA duplexes over the DNA/DNA duplexes, based on the difference in width in their minor grooves. I designed molecules that potentially could fit into the wider RNA/DNA minor groove but not the DNA/DNA minor groove. I then single mode docked the hypothetical molecule to each of these targets, and examined, both visually and by scoring, how well the molecule could differentiate between the two structures.

In the case of the fullerene derivative refinement, I pursued the same approach, designing hypothetical molecules that could potentially access the non-polar, water accessible channels described in the last chapter, then docking these in the single mode. These were assessed by a single criteria: did these complexes increase the hydrophobic desolvation relative to the known fullerene inhibitors? This is a drastic reduction of a multidimensional problem into a single dimensional one. Obviously, this is an approximation, and eliminates the subtleties of the ligand-protein interaction. I felt this was a reasonable approximation because the vast majority (~90%+) of the surfaces typically desolvated during complexation of these molecules were non-polar. Also, some of the energy costs of binding had already been paid by the original fullerene inhibitors. These include the rotational translational entropy loss as well as the desolvation of the catalytic aspartates in order to come in contact with the fullerene surface. By reducing the problem to a *relative change* in hydrophobic desolvation compared to the original inhibitors (Chapter 2), which had the central C60 in the same location, complicating

binding energies were reduced. So while DOCK provided a rapid generation of reasonable complexes, with good vdW interactions indicated in the force field score, the final arbiter of the inhibitor quality was the increase in hydrophobic desolvation as measured by surfaces generated by the program MS (described in Chapter 2). The top scoring DOCK generated orientation was not necessarily used always. The top 2kcal of complexes were examined visually, and a variety of closely scoring complexes were analyzed for hydrophobic desolvation. In general, it was visually readily apparent when one or both of the non-polar lined channels had been filled by the molecule, which would then lead to a large amount of hydrophobic desolvation.

MINDOCK

Near the beginning of the refinement design effort, I switched from using DOCK3 to MINDOCK. MINDOCK is similar to DOCK3 in the method of scoring the complexes generated, but differs somewhat in the method of generation. Like DOCK3, MINDOCK starts by matching atom centers of the ligand to sphere centers, but follows this with rigid-body minimization of the ligand with respect to the docking site. What this does is allow forces on the ligand, caused, for example, by bad contacts, to be relieved. This can potentially save a complex that might otherwise be productive except for a few bad contacts. Dan Gschwend made the code available early on in his development process which was helpful. My qualitative observation was that MINDOCK could 'find' complexes that intuition told me were there, but which DOCK3 could not find. Again, qualitatively, it seemed to me that MINDOCK more thoroughly searched the possible ways of binding than did DOCK3. MINDOCK no longer exists as such, but the features of rigid body minimization have been introduced into the DOCK package.

Refinement Philosophy

The intention of this refinement effort was to introduce elements to the fullerene that could interact with regions of the active site "adjacent" to the central C60 portion of the original inhibitor modeled complexes, which could not interact with the inhibitor due to the bulkiness of the solubilizing side chains. (Figure 3.3)

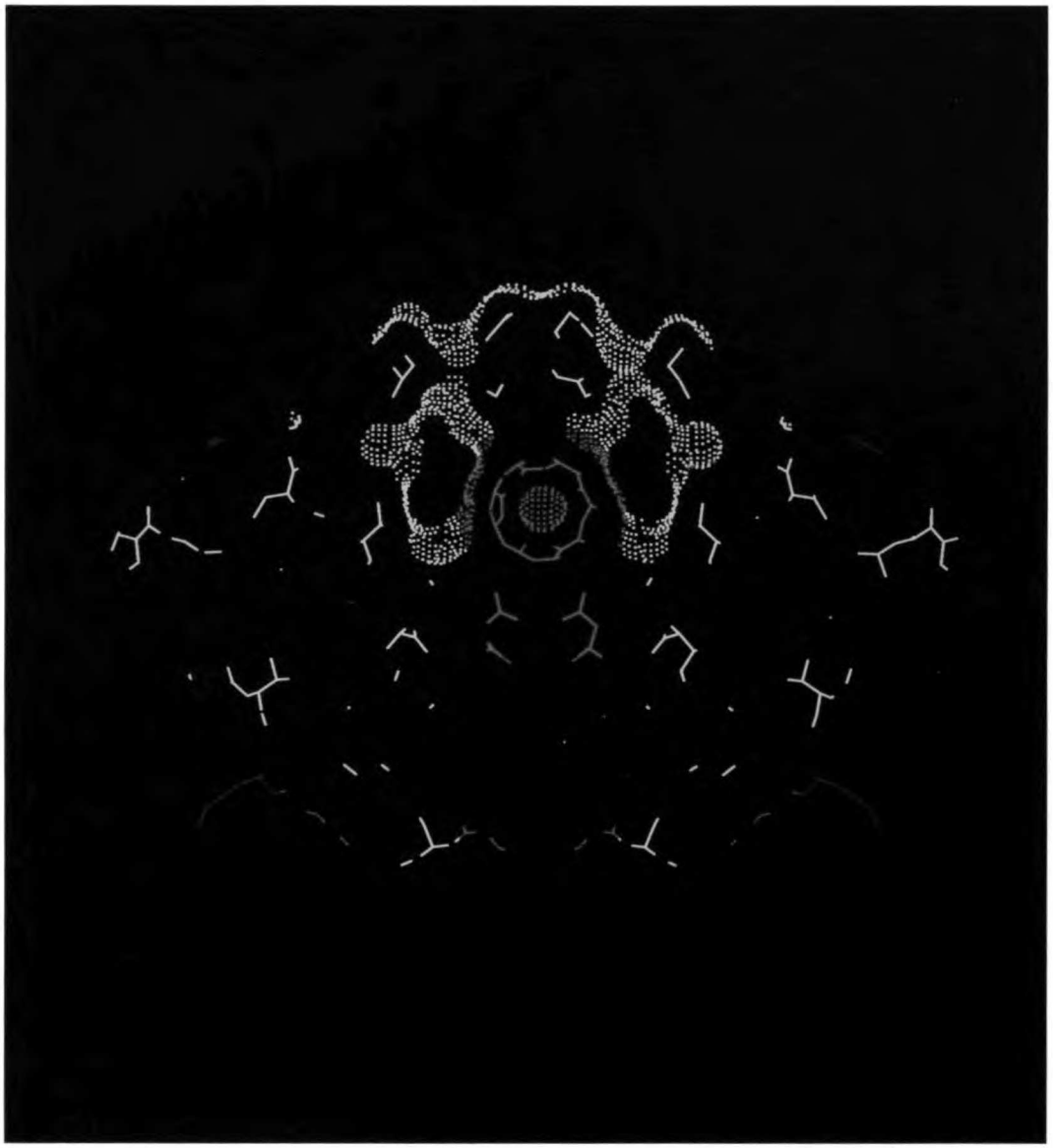


Figure 3.3. Cross Section of the HIV Protease/C60 Complex With Molecular Surface. Yellow Surface Points Represent Hydrophobic Residues.

The initial modeling focus was on bis-functionalized C60, i.e., fullerenes with two attachment points, the two 6-6 bonds (meaning between two six membered rings) related through a reflection through the center of the C60 sphere. (Figure 3.4)

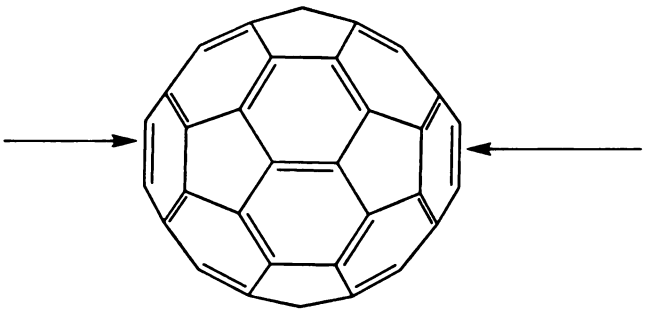


Figure 3.4. Site of Attachment of Analyzed, Bis-Functionalized C60 Derivatives

The promise of such molecules is the ability to potentially gain twice the free energy of binding increase as when a single site is interacted with. There is a potential problem with this idea relating to symmetry, and realized in retrospect, which I will discuss later in the description of the monoadducts. A possible advantage of bisadducts is that by introducing charged or polar elements at either end of the fullerene, the fullerene is prevented from entering membranes and staying there as a pseudo-fatty acid. The charges, full or partial, separated by 15-20 Å, will not be able to span the width of a bilayer, which is ~40Å, and will have a preference for solvation over presence in the hydrophobic parts of the bilayer. This could be quite an important issue in practically using fullerenes in the realm of pharmaceuticals.

Analysis of Hypothetical Derivatives

Bisadducts

As described in the manuscript at the beginning of this chapter, the aim was to find compounds that could increase the hydrophobic desolvation level significantly above the 353-381Å^2 achieved in model complexes by the originally tested compounds from Chapter 2. One of the first bis-modified molecules analyzed was compound A. (Figure 3.5) This is truly a "hypothetical" derivative because the lack of polar groups would make it insoluble, although further derivatization could improve solubility. This molecule is based again on the cyclo propyl ring, created upon reaction of a 6-6 bond with a di-azo compound (or other approaches).

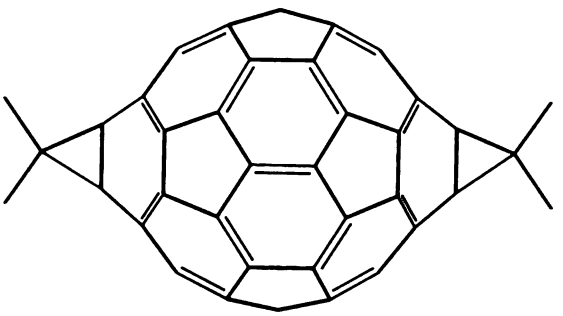


Figure 3.5. Compound A

The top scoring MINDOCK orientation of compound A with the HIVP achieved a very large increase in hydrophobic desolvation as determined by analysis of the MS surfaces:

	Carbon	Nitrogen	Oxygen
Complex ($Å^2$)	1390	110	248
Protease ($Å^2$)	1403	113	288
Ligand ($Å^2$)	452		
Desolvation Due to Complexation ($Å^2$)	**-465**	-3	-40

Table 3.2. Surface Desolvation Breakdown of Compound A

This represents the possibility of a several order magnitude increase in binding energy, based on the previously described 70 cal/mol/$Å^2$ value for desolvation of non-polar surfaces, and was encouraging in that simple modifications to the surface could introduce large increases in interaction. The top scoring complex of this compound which produces this numerical result visually demonstrates the "filling in" of the previously water filled areas adjacent to the central C60 sphere. Both of these regions are now almost completely occluded by the new substituents. (Figure 3.6)

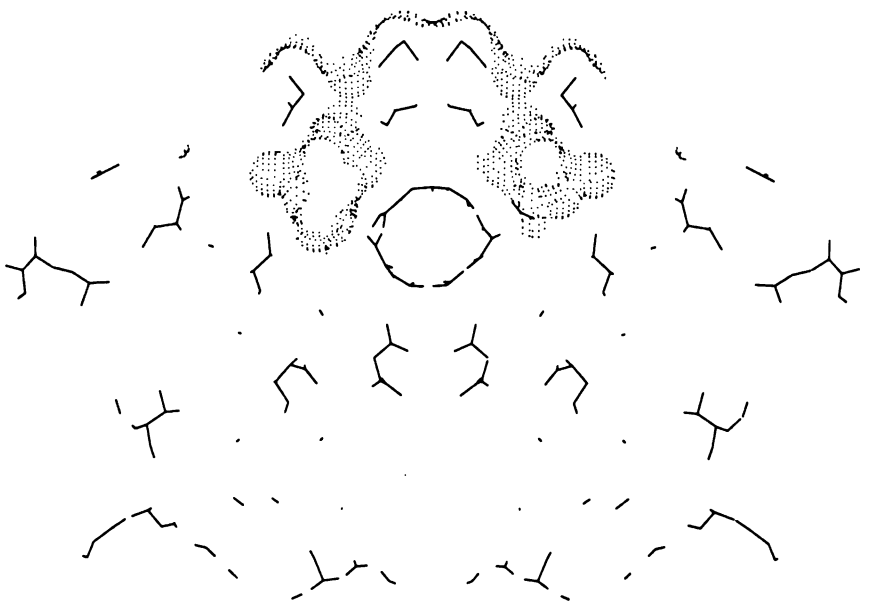


Figure 3.6. Top Scoring Complex of Compound A With the HIVP, Showing the Occlusion of Previously Water Exposed Non-Polar Regions

Following the analysis of compound A, I examined a related command, compound B, which repeats the bis-cyclopropyl arrangement and introduces two charged moieties. (Figure 3.7)

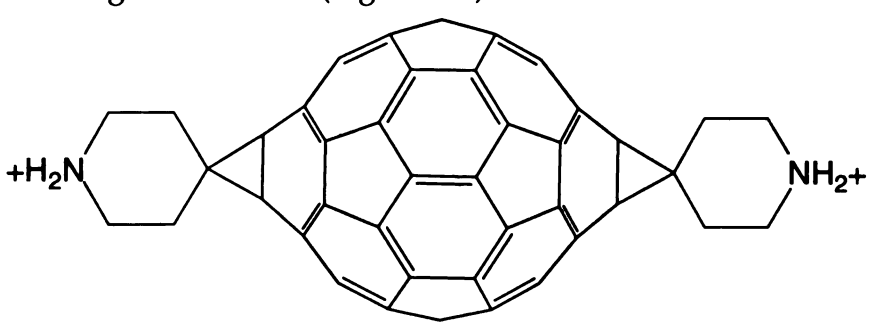


Figure 3.7. Compound B

The purpose of these was primarily to introduce water solubility to the hypothetical compounds, as well as introducing potential salt bridges between the amines and anionic groups at the mouth of the active site. The top

scoring complex of this compound with the HIVP effectively increases the level of hydrophobic desolvation as summarized in the surface transfer, although not as much as compound A.

	Carbon	Nitrogen	Oxygen
Complex ($Å^2$)	1430	111	261
Protease ($Å^2$)	1403	113	288
Ligand ($Å^2$)	485	16	
Desolvation Due to Complexation ($Å^2$)	**-458**	-18	-27

Table 3.3. Surface Desolvation Breakdown of Compound B.

In addition to this hydrophobic desolvation, this complex positioned the amine groups close to the acidic side chains at the mouth of the active site as shown in Figure 3.8 .

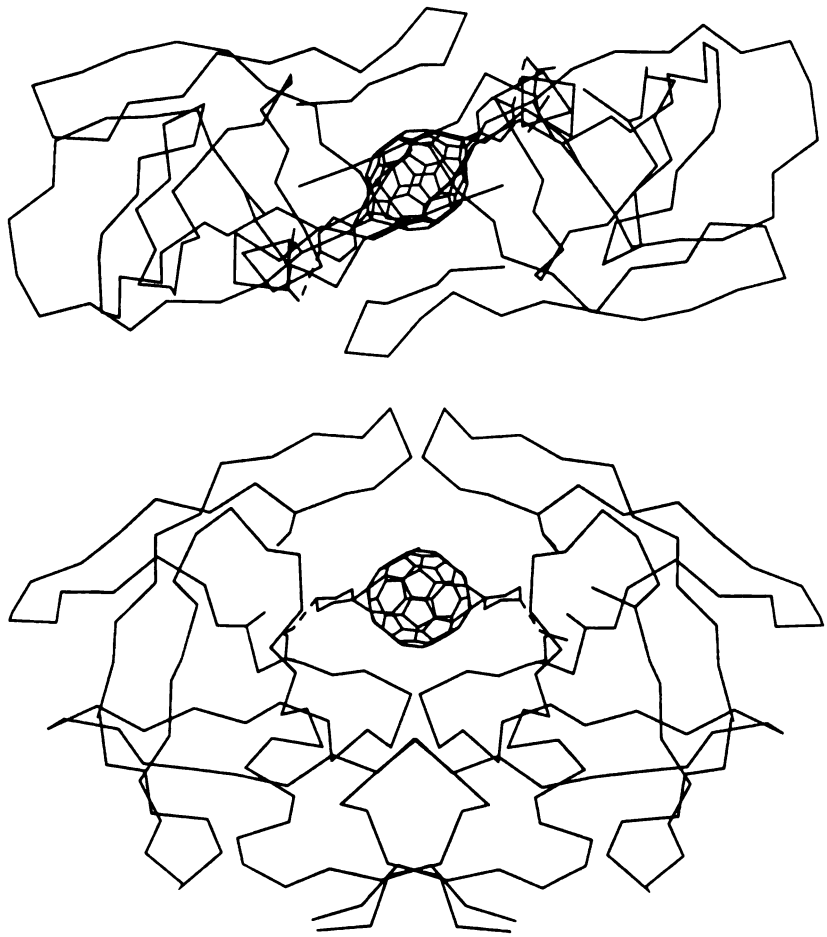


Figure 3.8. Model Complex of Compound B with the HIVP, Showing Potential Anionic Interaction Sites.

Compound C is a similar motif to compound B, with charged solubilizing groups diametrically opposed on the surface of the C60. This compound is based on a different type of chemistry, that of Diels-Alder addition of dienes to a 6-6 double bond on the C60 surface. (Figure 3.9) This fullerene synthetic chemistry had been fairly well developed and offered alternative routes to target molecules.

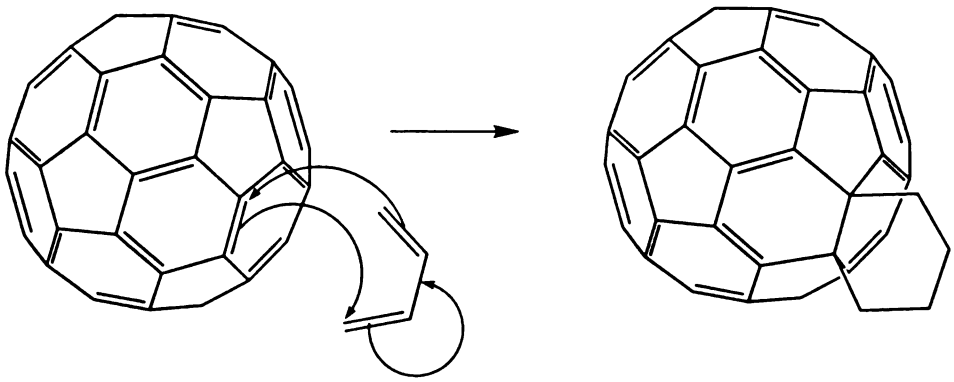


Figure 3.9. Depiction of the Diels-Alder Reaction of Butadiene with a 6-6 Bond of the C60.

The majority of this work had been pursued in the lab of Prof. Yves Rubin at UCLA. The Rubin group had produced cyclohexyl alcohol adducts to the C60, using a silyl protected enol diene, and the products were well characterized[33]. The conversion of the alcohol group to an amine should not pose a great problem. Compound C (Figure 3.10) does illustrate an important problem which is introduced with this compound, and that is the problem of stereochemistry.

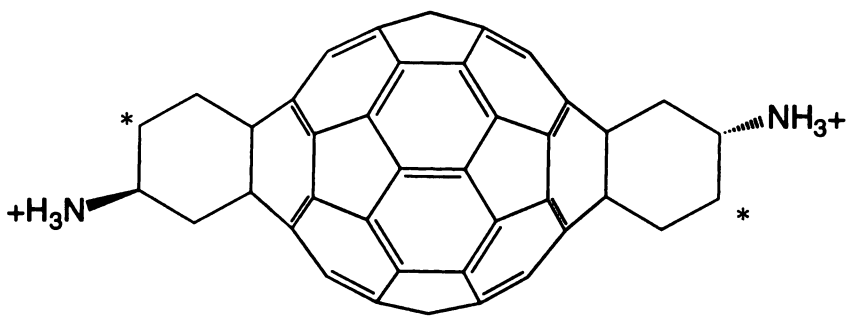


Figure 3.10. Compound C

In "drawing up" the molecule using SYBYL, I had to decide on the stereochemistry at the two chiral centers (the point of attachment of the amino groups). In addition, I had to decide which orientation the second reacted group would have relative to the first. In other words, the amino group could have also been on the carbons indicated with the asterisk. In general, my aim has been only to analyze molecules that can be synthesized, and synthesized relatively simply. These early designs do not really qualify as easily synthesized, especially compound C. The purpose of analyzing this

type of molecule was to gain an understanding of how different motifs could potentially interact with the protease.

The top scoring MINDOCK generated orientation of this compound with the HIVP resulted in a modest improvement in hydrophobic desolvation of 440 Å^2 (Table 3.4) and this lower level of improvement relative to compound A can be seen in the way that the non-polar regions adjacent to the central C60 are partially, but not fully, occluded. (Figure 3.11) As with the other compounds, the large majority of the desolvated surfaces are non-polar carbon surfaces, consistent with the nature of the active site and ligand.

	Carbon	Nitrogen	Oxygen
Complex (Å^2)	1425	128	266
Protease (Å^2)	1403	113	288
Ligand (Å^2)	462	30	
Desolvation Due to Complexation (Å^2)	**-440**	-15	-22

Table 3.4 Surface Desolvation Breakdown of Compound C

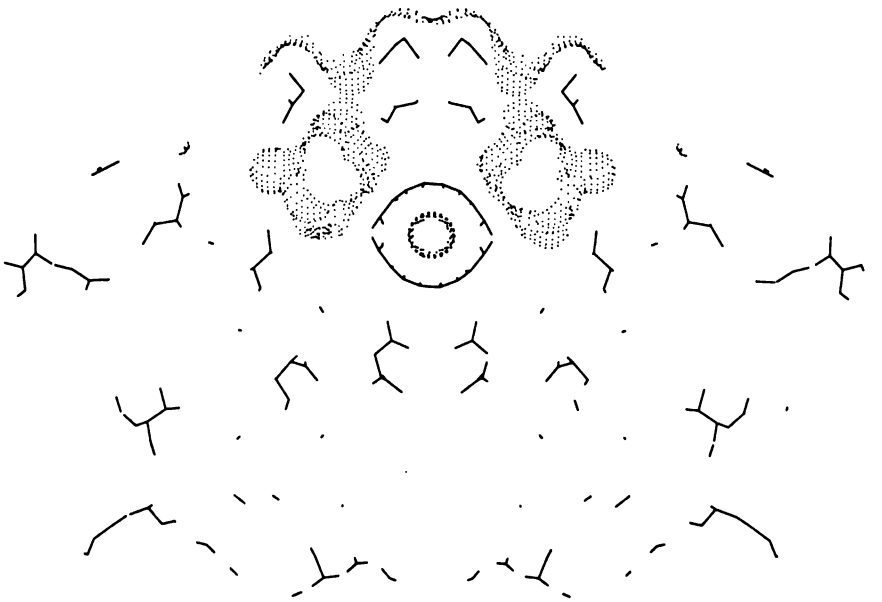


Figure 3.11 Cross Section of MINDOCK Generated Complex of Compound C, Showing Molecular Surfaces.

Following these and other initial analyses, we were contacted by Prof. Yves Rubin who expressed an interest in collaborating, which was fortuitous considering my prior analysis of compounds based on the chemistry he had developed.

I met with Prof. Rubin in order to find out the variations on the cyclohexyl motif which were reasonably synthetically accessible. These variations would then form the basis of the designs and the analysis of them to find a good target. The "good" target would satisfy the design criteria (increased hydrophobic-desolvation) and be synthetically accessible.

This resulted in the design of compounds D and E. (Figures 3.12,3.13) One advantage of these compounds is that they have eliminated all chiral centers, thereby making characterization simpler. The results with these two compounds were not encouraging, however, both with hydrophobic desolvations less than 400Å^2.

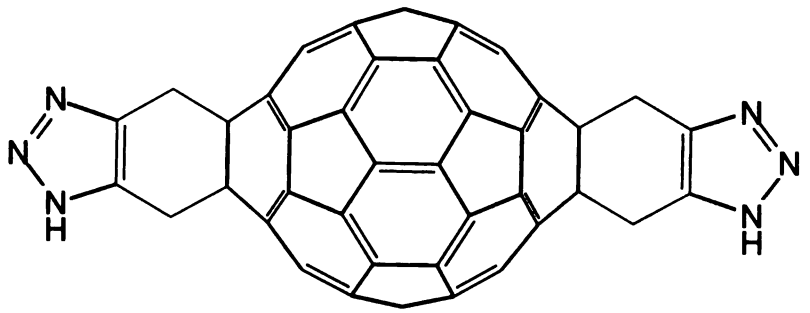


Figure 3.12. Compound D

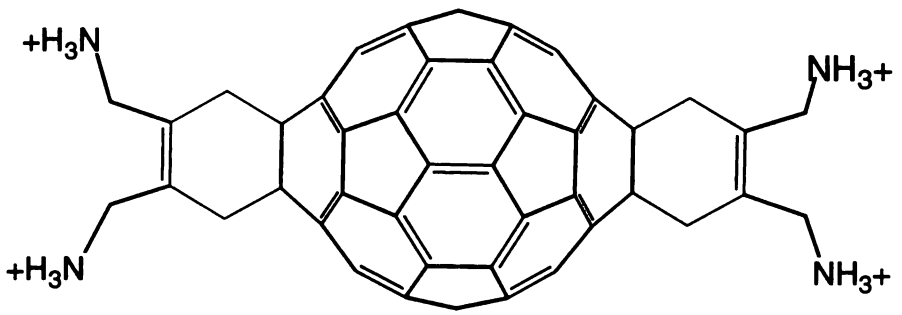


Figure 3.13. Compound E

Compound D could be "salvaged" by the addition of methyl groups to the carbons which attach to the fullerene surface, making compound F. (Figure 3.14) This molecule, due to the extra methyl group surface, produced a large amount of hydrophobic desolvation, 483 $Å^2$. A major synthetic problem with molecule is that the steric crowding at the site of reaction introduced by the extra methyls might make the reaction sluggish or nonexistent.

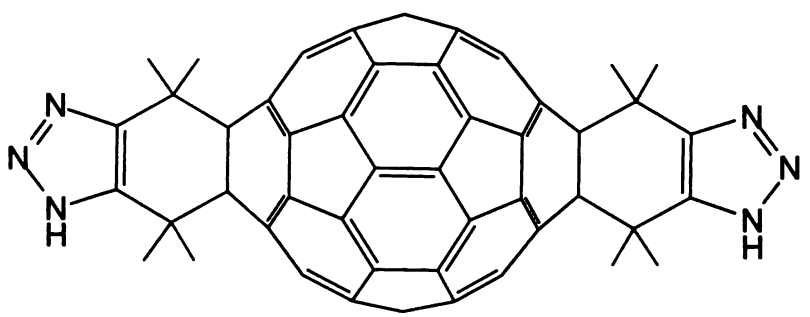


Figure 3.14. Compound F

Another approach Prof. Rubin suggested was the use of the C70 molecule as a base. I had already examined C70 to see if it could, underivatized, fill the non-polar regions adjacent to the C60 sphere and had found that it could not very effectively. The reason Prof. Rubin suggested the C70, however, was for a different purpose: to control regiospecificity. With all of these bis-modified C60 derivatives outlined above, there is the problem of regiospecificity. Once the first equivalent has reacted, the second may add in any number of places, leading to a mixture of products. The C70 on the other hand, had a preference for reaction at specific sites, those that were under the greatest strain in the molecule. Balch and co-workers had shown that they could generate a single bisadduct of C70, in which two specific bonds had reacted with a organo-metallic complex[34]. These two reactive bonds are indicated in Figure 3.15 . Balch and coworkers noted that after the first addition, there are actually three 6-6 bonds at the other end of the C70 that have equivalent strain. However, they observed only a single one of these actually reacting.

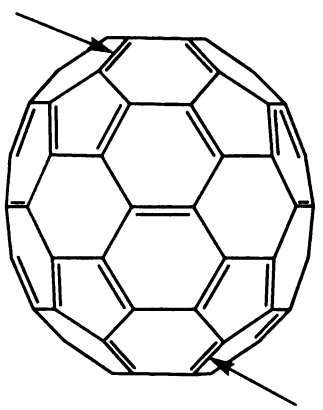


Figure 3.15 C70 Showing positions of reaction observed by Balch et al.

I began analyzing potential adducts of C70 that could be formed by analogy to the Diels-Alder additions observed for C60, assuming that the regio-specificity would be the same as observed by Balch and coworkers, because of a similar release in strain. The first adduct I analyzed was compound G, which is the analog to compound C. As with compound C, I made an arbitrary decision as to the stereochemistry of reaction, and this is depicted in Figure 3.16.

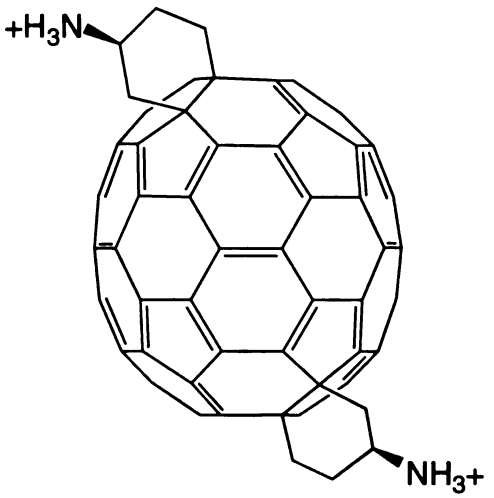


Figure 3.16. Compound G

This compound docked well, and the complex that was analyzed is shown in Figure 3.17 in "front" and "top" views, which show the C70 effectively orienting the two cyclohexyl groups into the target regions. It should be noted that this complex was not the highest scoring complex as identified by MINDOCK. Typically, I assessed the top 20 scoring complexes to see if the molecule could orient in a productive manner, i.e., with the substituents placed in the target regions. My final appraisal of a complex was based on its ability to improve on the hydrophobic desolvation of the originally tested fullerene inhibitors. Because the version of DOCK that I used did not correct for solvation effects, the scoring can be considered an estimate of potential affinity. I took the top twenty complexes to be of roughly similar probability and then queried whether or not they could satisfy the design constraint: increased hydrophobic desolvation. In the case of compound G, this hydrophobic desolvation was 478 $Å^2$, or 89% of the total surface desolvated. (Table 3.5)

	Carbon	Nitrogen	Oxygen
Complex (Å^2)	1441	136	233
Protease (Å^2)	1403	113	288
Ligand (Å^2)	514	30	
Desolvation Due to Complexation (Å^2)	**-476**	-7	-55

Table 3.5 Surface Desolvation Breakdown of Compound G.

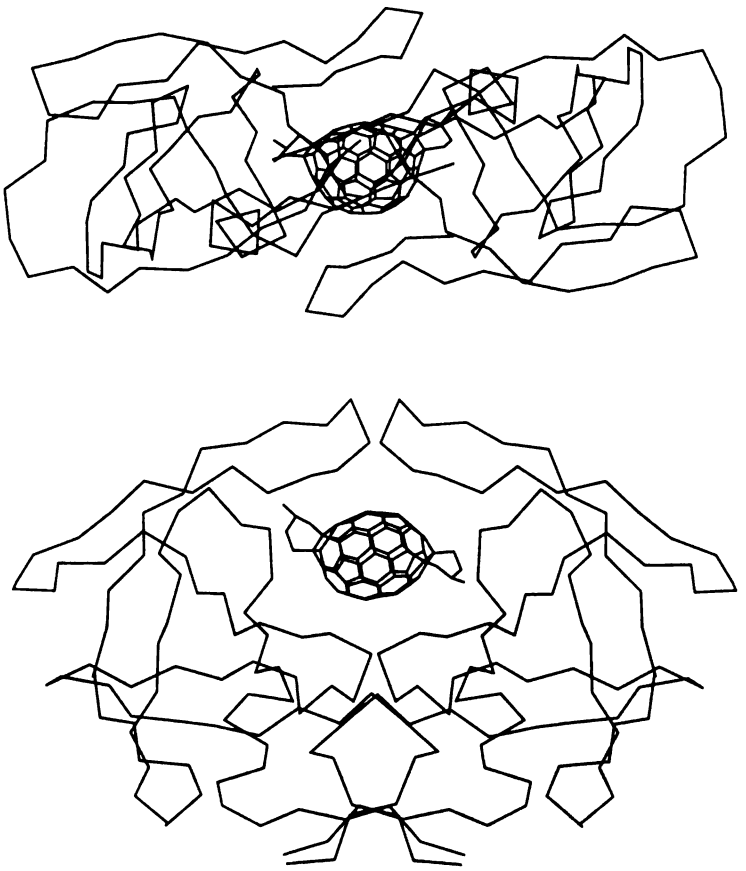


Figure 3.17 Model Complex of Compound G, Shown in "Top" and "Front" Views.

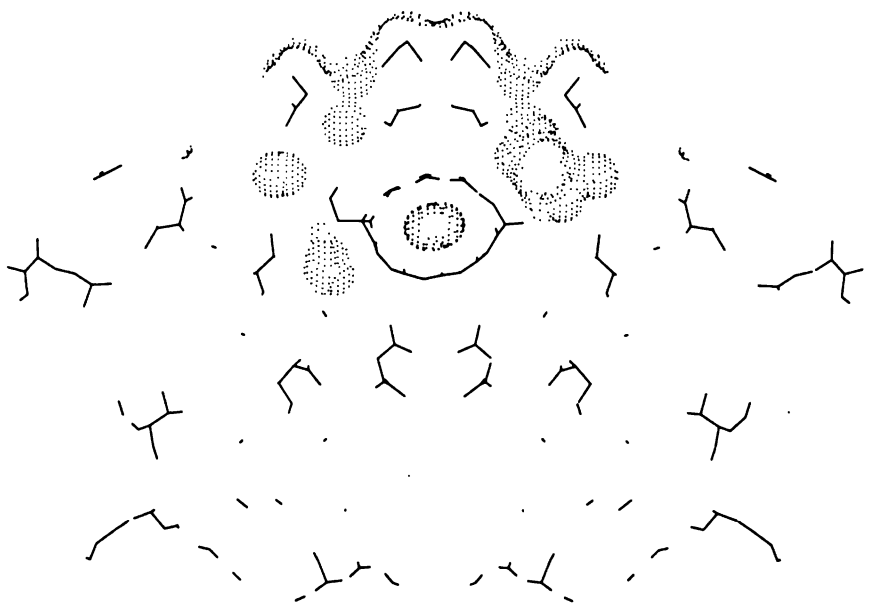


Figure 3.18 Cross Section of Model Complex of Compound G with the HIVP, Showing Molecular Surfaces.

This was very encouraging in that it showed that indeed the C70 could position the two substituents in the target regions. Compound G, however, suffered from the same problems that the C60 version did (compound C), in that it introduced synthetic complexity in the form of undetermined stereochemistry of the product. There would be a large number of potential products, varying in the placement of the amino groups relative to each other. In order to alleviate this problem, I examined several derivatives without the chiral centers present in compound G. One of these is compound H, which is the C70 analog of compound D. (Figure 3.19)

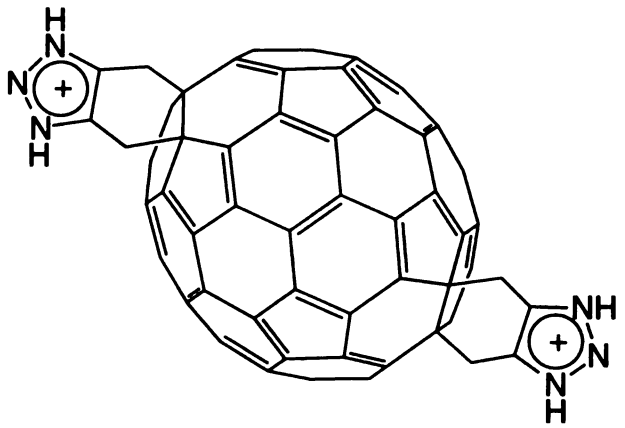


Figure 3.19. Compound H

	Carbon	Nitrogen	Oxygen
	Carbon	Nitrogen	Oxygen
Complex ($Å^2$)	1428	129	234
Protease ($Å^2$)	1403	113	288
Ligand ($Å^2$)	492	66	
Desolvation Due to Complexation ($Å^2$)	**-467**	-50	-54

Table 3.6 Surface Desolvation Breakdown of Compound H.

The complex of compound H with the HIVP effectively increases the amount of hydrophobic desolvation relative to the original derivatives, with about a 90$Å^2$ improvement. (Table 3.6) (Figure 3.20)The proportion of desolvation that is hydrophobic, however, dropped from the 90% range to 82%. Bringing non-polar surfaces in contact with polar surfaces is generally not energetically favorable, so this decrease in the proportion may have represented some of the non-polar carbons being desolvated through contact with polar atoms. An alternative explanation is apparent when examining the complex that was analyzed.

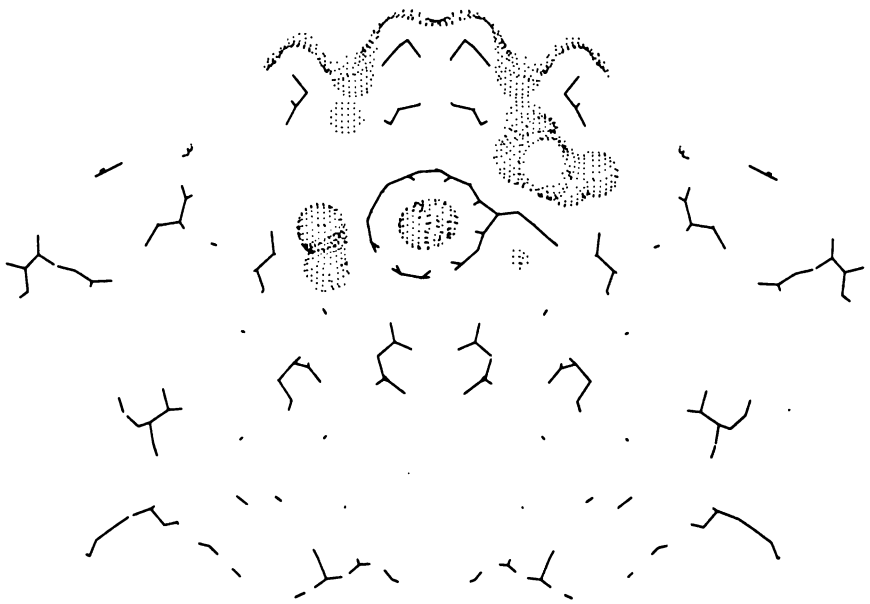


Figure 3.20 Cross Section of Model Complex of Compound H with the HIVP, Showing Molecular Surfaces.

Both positively charged triazole systems are in close contact with a large number of partially negatively charged oxygens. In fact, the *only* heteroatoms <5Å from the triazole nitrogens are oxygens. (Figure 3.21) It is not surprising that MINDOCK placed the molecule in such a way as to have an effective charge-charge interaction with these oxygens, as favorable electrostatic interactions are accounted for. What is exciting about this molecule is that it can simultaneously desolvate large portions of non-polar surface, in both target regions of the protease, and simultaneously form two sets of charge-charge interactions which can in principle add to the overall binding energy.

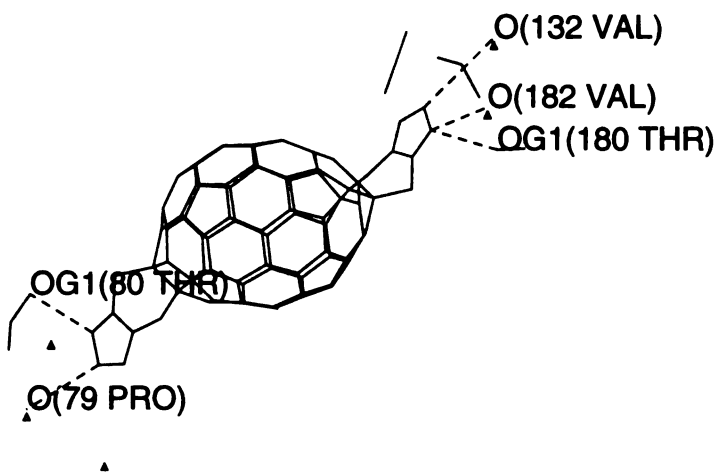


Figure 3.21. Triazole Nitrogen/HIVP Oxygen Distances in Compound H Model Complex.

By adding a single methyl group to each of the substituent attachment carbons of compound H, hydrophobic desolvation could be increased to 533Å^2. This, however, again introduced synthetic complexity in the introduction of chiral centers. I decided that the benefit of 66Å^2 improvement over compound H was probably not worth the synthetic problems associated.

At the end of the analysis of the bis-substituted fullerene compounds described to this point, compound H appeared to be the target of choice. 1) It satisfied the basic design criteria: significant increase in hydrophobic

desolvation relative to the originally tested derivatives. 2) It achieved this through interaction with the two target regions of the active site. 3) It had two solubilizing groups that would probably confer water solubility, and it was in a good position to add salt bridges with anionic sites at the mouth of the active site. 4) It had no additional chiral centers. 5) The regio-specificity required would be directed by the known reactivity of specific sites.

Prof. Rubin, however, was concerned that the regiospecificity of the C70 reaction would not go as cleanly at it had for Balch (despite the fact that he had recommended the C70 approach to begin with), and preferred to work with the C60, making monoadducts before attempting any bis-chemistry. This was reasonable, but required re-initiating the analysis with monoadducts. My belief was that designing a single substituent to optimally bind at a single site would be different from finding a substituent that would make a bisadduct that could bind effectively simultaneously at two sites. In other words, chopping a "good" bisadduct in half would not turn it into a "fair" monoadduct.

Monoadducts

I wanted now to limit my designs to compounds I could be sure were easily synthetically feasible and asked Prof. Rubin to rank in terms of synthetic ease the various motifs that had been introduced in our meeting. These motifs are shown in Figure 3.22.

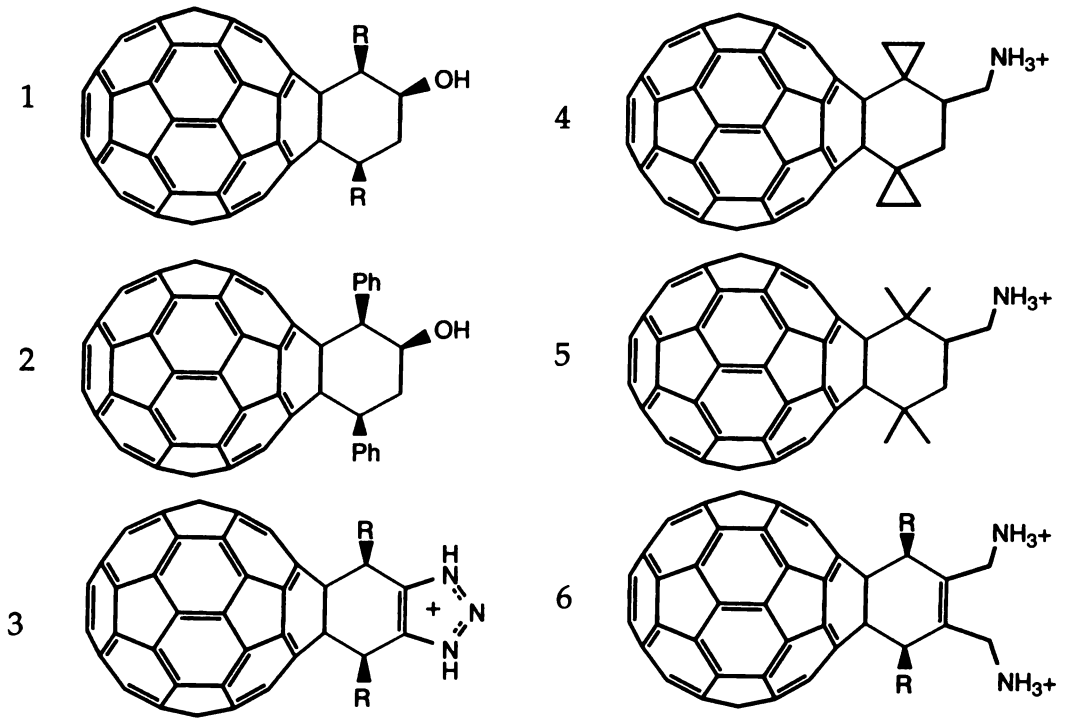


Figure 3.22. Possible Motifs Available for Cyclohexyl C60 Adducts

The ordered preference for synthesis was I/II/VI. Also the R groups could be any alkyl group so long as they were not too bulky; i.e., tertiary butyl groups were ruled out. This then became the pallet from which I designed an appropriate monoadduct.

One of the first hypothetical monoadducts I analyzed was compound I. (Figure 3.23)

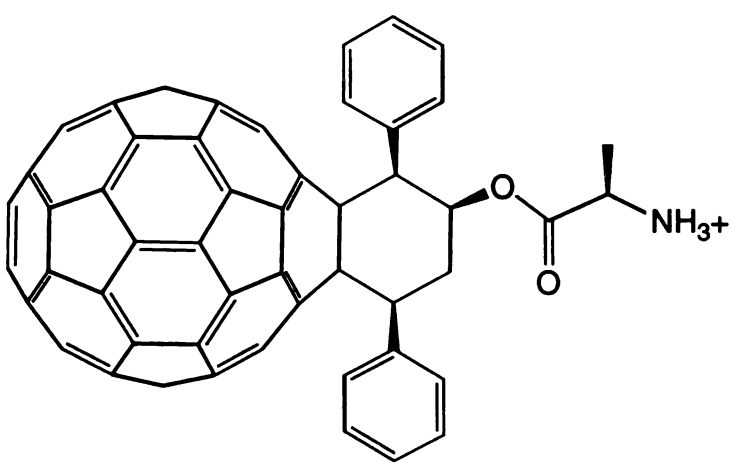


Figure 3.23. Compound I

The idea behind the hydroxyl group in the motifs 1 and 2 is to act as a attachment point for a variety of solubilizing groups. In the case of compound I, the solubilizing group was alanine, whose charged amino group was anticipated to improve water solubility.

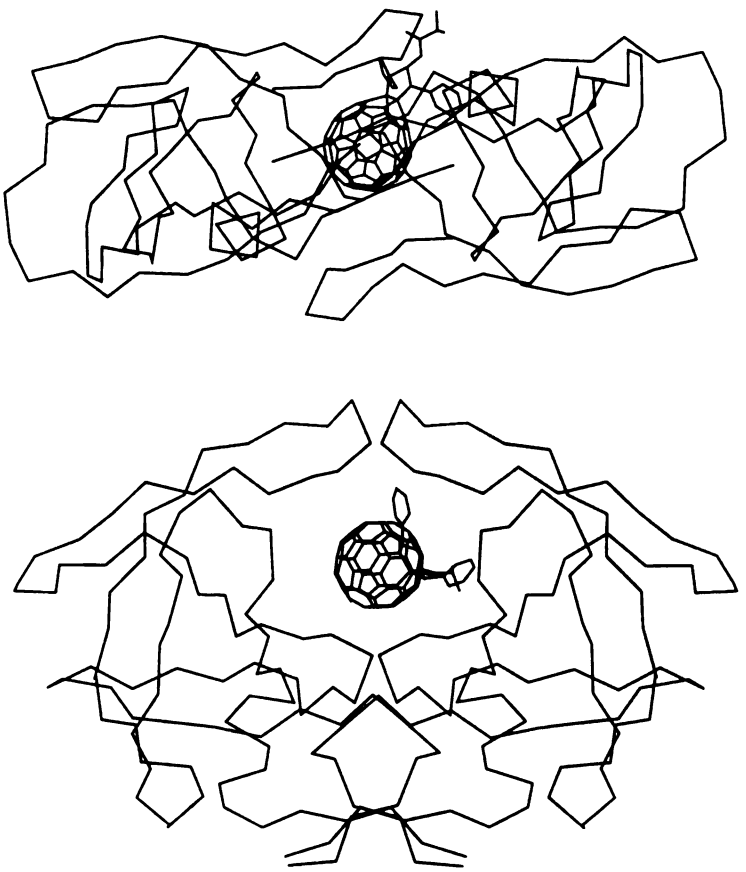


Figure 3.24 Model Complex of Compound I Shown in "Top" and "Front" Views.

The top ranking complex of this compound with the HIVP (Figure 3.24) (in "top" and "front" views) shows the inability of the substituents to interact with the target region, in part due to their bulkiness. This is supported by the desolvation breakdown, which shows 412 $Å^2$ desolvation of non-polar carbon

surface; only a 30 Å^2 improvement over the originally tested fullerene derivatives. (Table 3.6)

	Carbon	Nitrogen	Oxygen
Complex (Å^2)	1577	114	270
Protease (Å^2)	1403	113	288
Ligand (Å^2)	588	17	13
Desolvation Due to Complexation (Å^2)	**-414**	-16	-31

Table 3.7 Surface Desolvation Breakdown of Compound I.

Compound J, K and L were all analyzed and showed similar complexes and desolvations to compound I. (-396 Å^2 for compound K, -412 Å^2 for compound L. Compound J complexes assessed visually only. Figures 3.25,26,27)

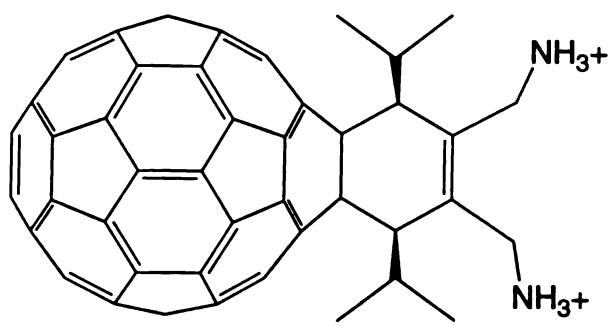


Figure 3.25 Compound J

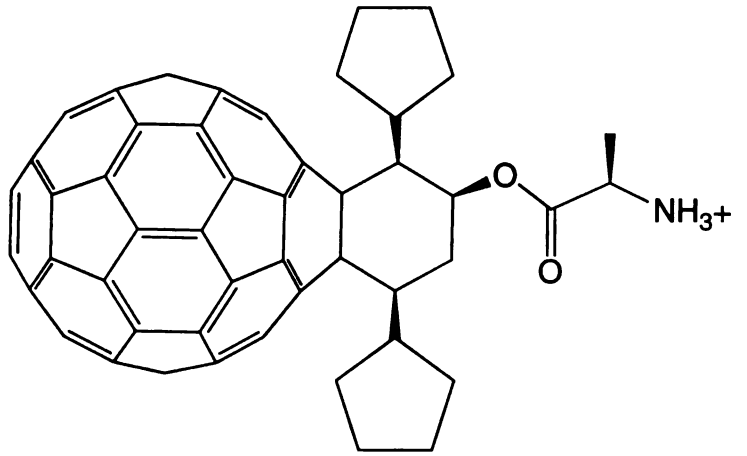


Figure 3.26 Compound K

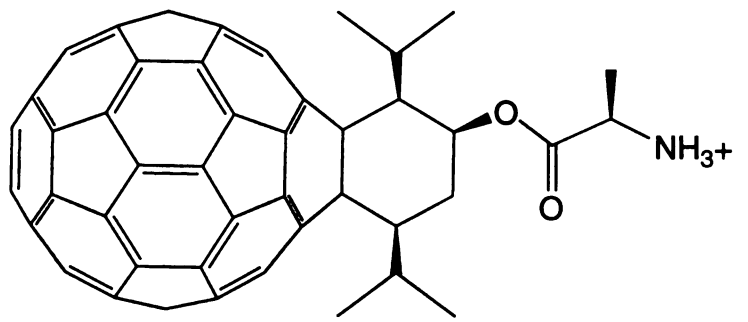


Figure3.27 Compound L

First Target Compound

All of these monoadducts were having difficulty orienting in such a way as to access the target non-polar regions of the active site. I believed that this could be caused by the bulk of the solubilizing groups: the alanine substituent in the cases of compounds K, L and I, and the two methyl-amine substituents in compound J. There should be many available (i.e., of similar energy) conformations of these solubilizing groups possible; however, only a small percentage could possible dock effectively and access the target regions. There were two possible ways to go around this problem. Either generate many conformers of the solubilizing chains on the derivative, or dock the *underivatized* compound, and then build off of the (hopefully exposed) hydroxyl group. In this way, the arbitrarily set conformation of the solubilizing side chain would not impede the formation of the ideal complex of the core molecule. This approach necessitates further analysis after docking to assess the feasibility of solubilizing chain attachment. For

simplicity's sake, I chose the latter approach, of docking the un-modified derivative first, and then analyzing the possibility of chain attachment.

The first molecule so analyzed, compound M, (Figure 3.28) was the parent compound of compound L, i.e., with just a free hydroxyl group and no esterified alanine attached.

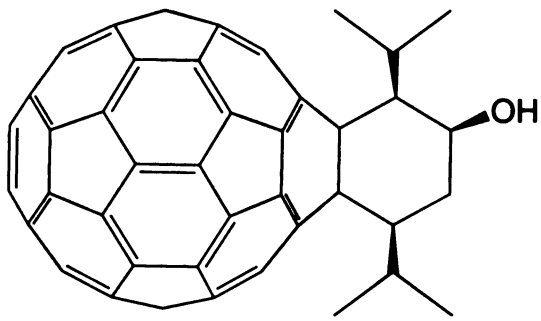


Figure 3.28. Compound M

The results were very good. The top scoring complex (shown in Figure 3.29) showed a non-polar surface desolvation, -486 Å^2, (Table 3.7) that was almost as good as the best bisadducts. One of the target non-polar regions in the HIVP active site was completely occluded (Figure 3.30) as indicated by the thin cross section. Furthermore, the hydroxyl group appeared to be significantly solvent exposed, which meant that adding a solubilizing chain should be possible.

	Carbon	Nitrogen	Oxygen
Complex (Å^2)	1417	106	246
Protease (Å^2)	1403	113	288
Ligand (Å^2)	500		11
Desolvation Due to Complexation (Å^2)	**-486**	-7	-53

Table 3.8 Surface Desolvation Breakdown of Compound M.

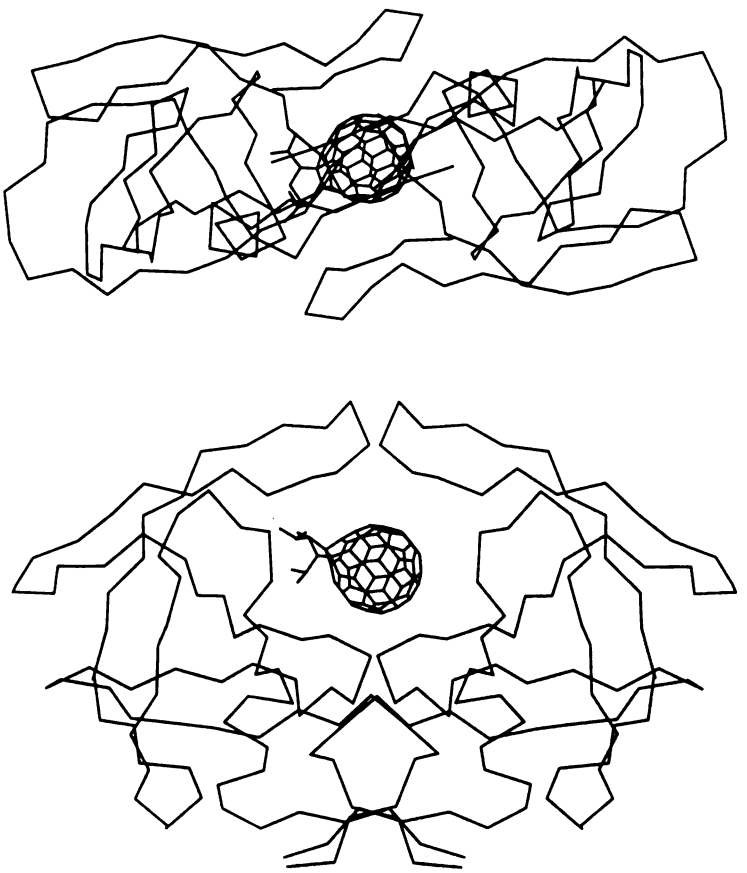


Figure 3.29 Model Complex of Compound M, Shown in "Top" and "Front" Views.

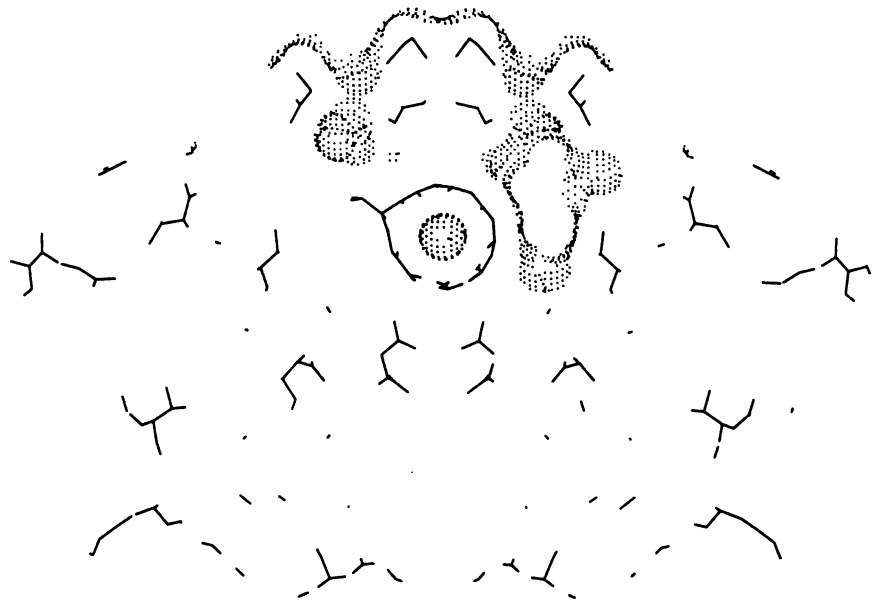


Figure 3.30 Cross Section of Model Complex of Compound M with the HIVP, Showing Molecular Surfaces.

Analysis of Solubilizing Chain

Now that I had demonstrated that compound M was very effective at increasing the amount of hydrophobic desolvation, it was necessary to determine if the attachment of a solubilizing chain would be accommodated by the complex. This was a necessary result of separating the docking of the central portion of the molecule from the attachment. The first chain examined is shown in Figure 3.31.

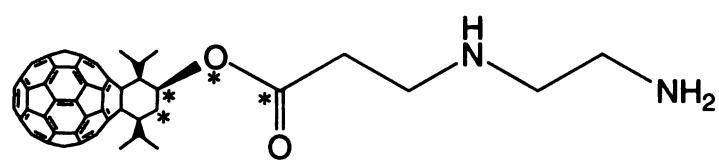


Figure3.31 Compound N

This chain was attached through ester linkage with the carboxyl group. The intention, as with the alanine adduct, is to introduce polar and charged groups to improve water solubility. This chain was built off of the hydroxyl of the top scoring complex using SYBYL. The dihedral formed by the asterisked atoms was driven, while monitoring the steric energy of the ligand protein complex using SYBYL's interactive docking feature. All other dihedrals in the chain were set at 180°. There were large steric clashes between the chain and the protein; however, in the dihedral range of 220° to 120°, these were relieved and the steric energy (between ligand and protein) dropped to -60 kcal/mol. The low steric energy conformer observed which had a dihedral of 177° was written out.

I was then interested to see if MINDOCK would be able to dock this identified adduct conformer in a similar position to the starting point, i.e., the top complex of compound M. The top scoring MINDOCK complex of compound N (Figure 3.32) places the core di-isopropyl C60 in virtually the same position as it did with compound M, which lacked the extra chain. This supported the SYBYL modeling and the idea that there was at least one conformer of this solubilizing chain that would not prevent the core from binding in the manner shown for compound M.

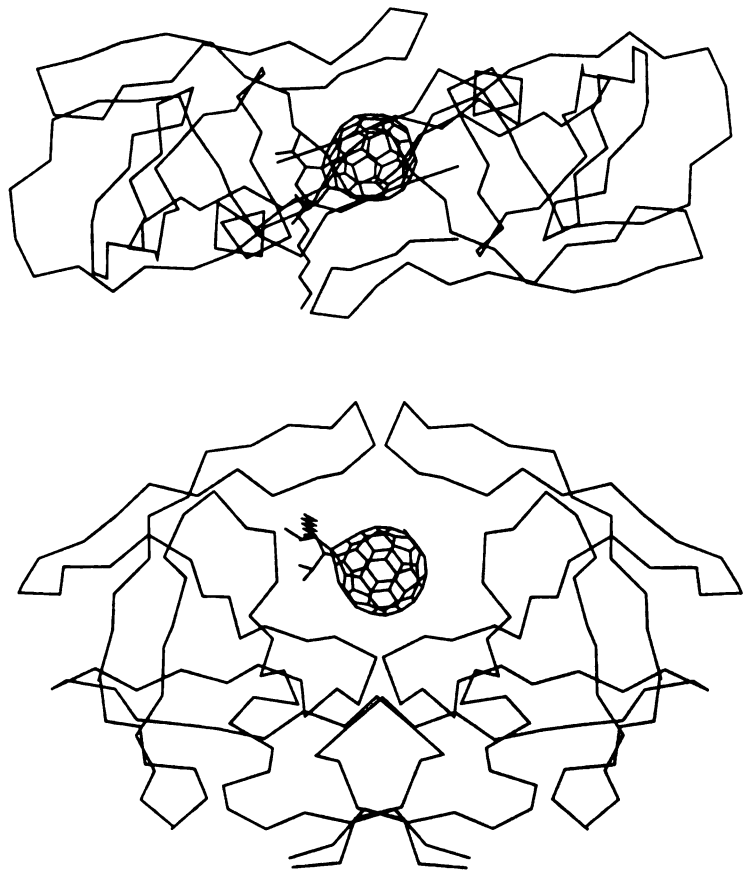


Figure 3.32. Model Complex of Compound N, Shown in "Top" and "Front" Views, and Depicting the Solubilizing Chain Extending into Solution

In addition to showing that a single conformer of the solubilizing chain existed that would not abrogate binding, I was interested in determining the number of other conformations of the chain that could also be formed without major steric clashes with the protein, and how this number compared to the number of conformations available to the uncomplexed ligand. This comparison would give an estimate of the conformational entropy loss of the solubilizing chain due to complex formation. The method used to approach this question was systematic conformational searching. Starting with the top scoring MINDOCK complex

of compound N, the 7 dihedrals of the solubilizing chain were driven, the asterisked dihedral with a 30° resolution, the ester bond with a 90° resolution and the remaining bonds with a 60° resolution. The theoretical limit of conformations is 373,248. When this simulation was performed on the uncomplexed compound N, 1694 conformations were within 10kcal/mol of the lowest conformer identified, which had an molecular mechanics energy of 1626.0 kcal/mol. The same simulation performed on the complex of compound N with the HIVP, produced 148 "available" conformations, i.e., within 10 kcal/mol of the lowest conformation found. This low energy conformation was 1626.3 kcal/mol (energy of the ligand conformation alone), virtually identical with the low energy conformation of the unliganded compound. It should be noted that the bottom 10 kcal/mol filtering was of the total energy, including the interaction energy of the protein and the link, which is probably not accurately assessed with this approach. This is why a 10kcal/mol window was used for analysis. The main purpose of this simulation was to determine roughly how much of the conformational space of the link was off limits due to the presence of the protein. This is estimated above at about 90%, meaning from conformational entropy alone, the loss in binding would be about a factor of 10. I thought this was a reasonable trade-off, considering the large amount of non-polar desolvation available to the core of the molecule, as well as possible charge-charge interactions available to the link amines with anionic sites at the mouth of the active site.

Another potential solubilizing chain that were analyzed was ornithine, which like the original alanine, had alpha branching. (Figure 3.33) This solubilizing group, when analyzed, could not produce any conformations that would not clash with the HIVP surface, due most likely to the presence of the alpha branching close to the area of most congestion in the active site.

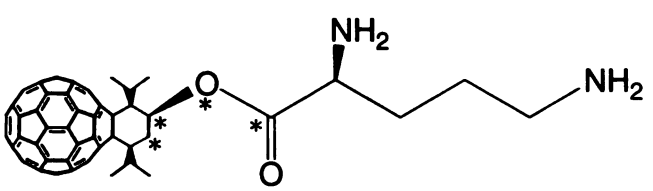


Figure 3.33 Compound O

For completeness, other monoadducts from the motifs listed in Figure 3.22 including the di-phenyl, di-ethyl, di-n-propyl were analyzed. None of these were able to desolvate as much non-polar surface as compound M, or present the hydroxyl group to allow effective attachment of a solubilizing chain as in compound N.

At this point in the analysis, I had a core compound (compound M) that achieved the design requirements of accessing the target non-polar regions of the active site. In addition, I had identified an adduct of this which would be accommodated by the protein, not prevent binding, and add water solubility to the molecule (compound N).

Unfortunately, not soon after the identification of this target, it became apparent that the chemistry was not proceeding as anticipated. The original synthetic methodology developed by Prof. Rubin was a Diels Alder addition across a double bond by the various substituted dienes (Figure 3.34). With the iPr substituent, however, this reaction was not proceeding effectively, probably due to steric crowding.

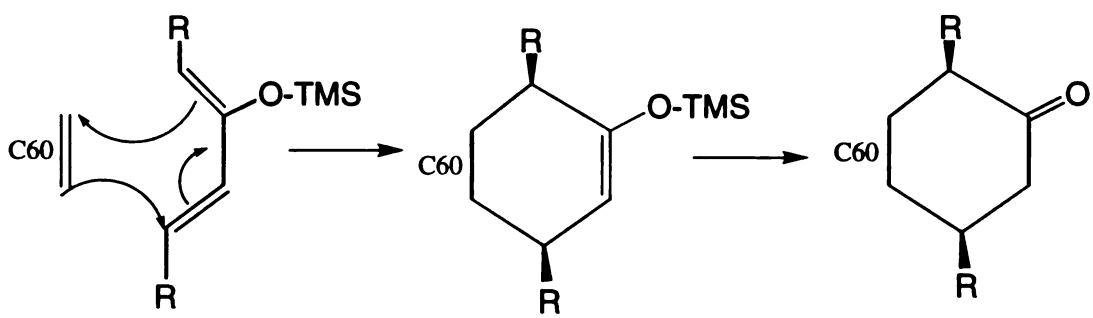


Figure 3.34 Diels-Alder Reaction of TMS Protected Butadiene Derivatives With C60

To get around this problem, he attempted the reaction using a LDA generated enolate, which would attack the double bond of the C60 in a Michael fashion. The electron pair generated would then attack the double bond on the reactant, leading to cyclization (Figure 3.35). This is illustrated as in this figure as a concerted process, but is most likely a stepwise process. The result of these two sequential Michael additions is that the two R groups are now trans to each other, as opposed to the original placement cis, which would have resulted from the Diels Alder addition. The alcohol formed

upon reduction of the ketone was expected to be cis to the adjacent R group (due to nucleophilic attack by hydride on the face of the carbonyl carbon opposite to the R group), and this was confirmed experimentally[30].

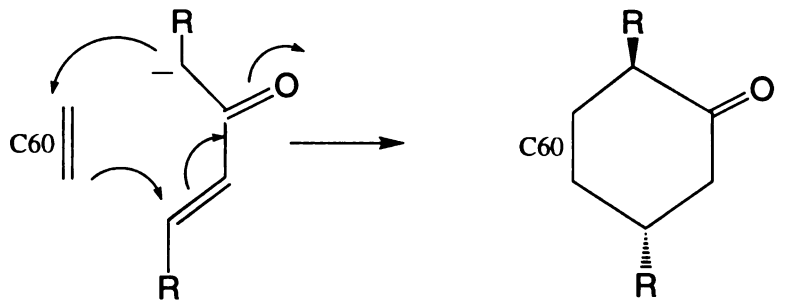


Figure 3.35. Double Michael Addition to C60

This necessitated re-examining the ability of the iPr derivative to achieve the design criteria, now that it appeared that the iso-propyl groups would be trans and not cis relative to each other. It was in the process of drawing up the trans version of compound M (i.e., compound P) (Figure 3.36) that I realized that I had been analyzing only a single enantiomer of each hypothetical derivative up to that point. This was not a concern for compound such as D, E and J, which had symmetry elements that made them optically inactive. For other derivatives, their chirality meant that the different enantiomers should interact differently with the chiral environment of the active site. This did not negate the positive results for compound M, but it did mean that it was possible that other compound's enantiomers may have given good results.

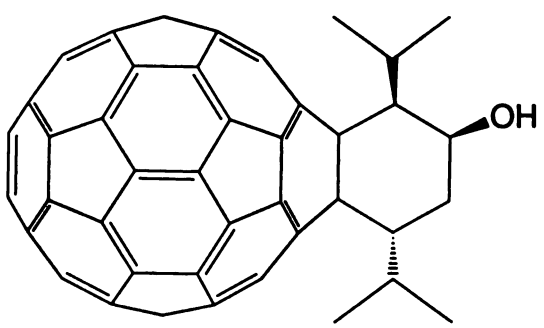


Figure 3.36. Compound P

An additional variable that I became aware of was varying ring conformations of the cyclohexyl group attached to the C60 in compound P. In the process of drawing up the two enantiomers, it was apparent that there were at least two slightly different chair forms, one which placed the OH axial and one which placed the OH equatorial. These two forms differed only slightly in their molecular mechanics energies, and so I used them both as the basis for docking. The result of these two variables, which enantiomer and which conformer, was that there were four species to dock. For each of these, I did a systematic conformational search on the iPr torsion to identify the lowest energy conformer with respect to that torsion. These four species are depicted in Figure 3.37 which shows the view looking "down" toward the external surface of the C60.

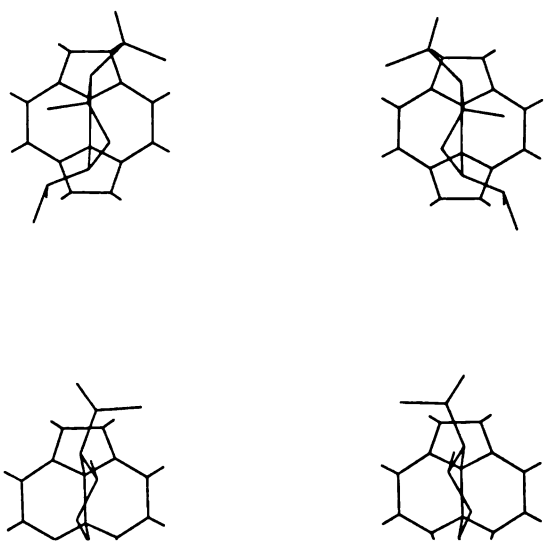


Figure 3.37 Comparison of four enantiomer/conformers of compound P. The species are designated Pa, Pb, Pc, Pd, starting from the upper left and going clockwise.

These four species were analyzed as all previous hypothetical derivatives were, i.e., generation of possible complexes with MINDOCK, followed by visual inspection and finally analysis of surface transfer, Of these

four isomers, compound Pa was significantly superior to the other 3 species in achieving large increases in non-polar surface desolvation. Figure 3.38 shows the 5th highest scoring complex of compound Pa with the HIVP, followed by the surface desolvation table (Table 3.8). This table indicates this species to be virtually as good as compound M in increasing non-polar surface desolvation.

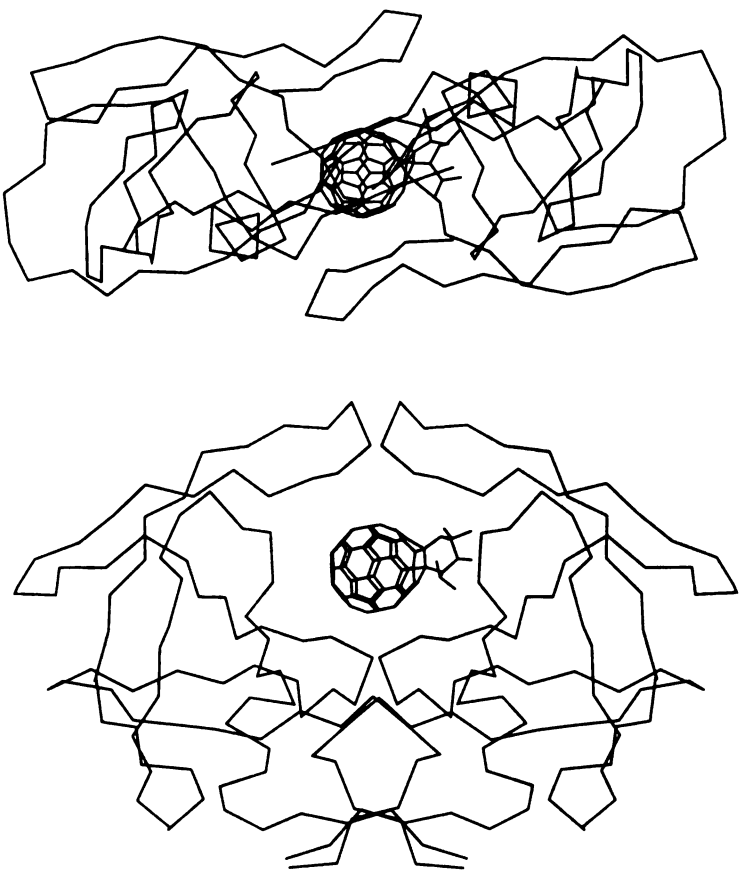


Figure 3.38 Model Complex of Compound Pa, Shown in "Top" and "Front" Views.

	Carbon	Nitrogen	Oxygen
Complex ($Å^2$)	1433	107	242
Protease ($Å^2$)	1403	113	288
Ligand ($Å^2$)	505		11
Desolvation Due to Complexation ($Å^2$)	**-475**	-6	-57

Table 3.9 Surface Desolvation Breakdown of Compound Pa.

A major problem, however, with this complex and the others with the trans compound is that they tended to bury the hydroxyl group. This makes the hydroxyl group inaccessible as an attachment point for a solubilizing chain. This was confirmed by repeating the conformational analysis previously done for compound N. This confirmed that for the above complex (Figure 3.38) there were no conformations for the tag that did not produce major steric clashes with the active site. It was possible that the hydroxyl group on the compound could be converted to an amino group directly, and thereby confer greater water solubility. Another possibility was that the hydroxyl alone may have provided sparing solubility enough to at least assay the compound.

Soon after this analysis a final problem cropped up. In the course of analyzing the NMR of the target compound P, Prof. Rubin observed an NOE effect that was consistent with the ring conformation being a boat, not a chair as I had modeled. This interpretation was based on the presence of an NOE between two protons as shown in Figure 3.39 which would be difficult to achieve without the boat conformation forming. Furthermore, using MM2 and a conformational search method, Prof. Rubin showed this conformation to be the lowest in energy, as ascribed by this force field.

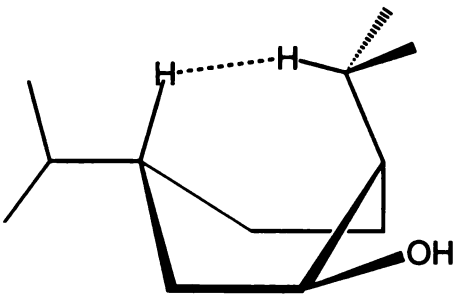


Figure 3.39. Protons Exhibiting NOE in NMR of compound P.

I minimized the structure found by Prof. Rubin which was consistent with the NOE data, and it minimized to a similar boat conformation. However, the molecular mechanics energy of this conformation was several kcal/mol higher than the chair conformations that I had previously used. In other words, the TRIPOS force field had ascribed a higher energy to the conformation that was experimentally observed.

This demonstrates a real problem with modeling of small molecule/macromolecule interactions. That is, correctly identifying the probable conformers. It is probably true that the conformers that I originally analyzed (Compounds Pa,Pb,Pc,Pd) are accessible in solution, but not in as high a population as the experimentally observed conformer. Another problem is that the experimentally observed NOEs were done in organic solvent, the molecular mechanics energies were determined in gas phase, but the final molecule is tested in aqueous solution. Aqueous conformational preferences may be different than those identified via NMR and molecular mechanics.

To determine if this more probable conformer of compound P was also effective at increasing non-polar surface desolvation, both enantiomers of the boat version of were docked into the HIVP using MINDOCK. Compound Pe and Pf have the same chirality as Pa and Pb, respectively, but with the boat conformation shown in Figure 3.39. Of these two, compound Pe was able to effectively increase non-polar desolvation, to a level of 445 $Å^2$ as shown in Table 3.9 . Figure 3.40 shows the cross section of the top scoring MINDOCK complex which generated this level of surface desolvation, and illustrates the desolvating of one of the two non-polar target areas in the HIVP active site. Although 445$Å^2$ is not as large a desolvation as the chair conformation, it still represents an approximate 65$Å^2$ increase in non-polar surface desolvation

relative to the originally tested inhibitors. This converted into an additional free energy of binding using the figure of 70cal/mol/Å^2 for molecular surface described earlier, suggesting a possible improvement in binding of 3 orders of magnitude.

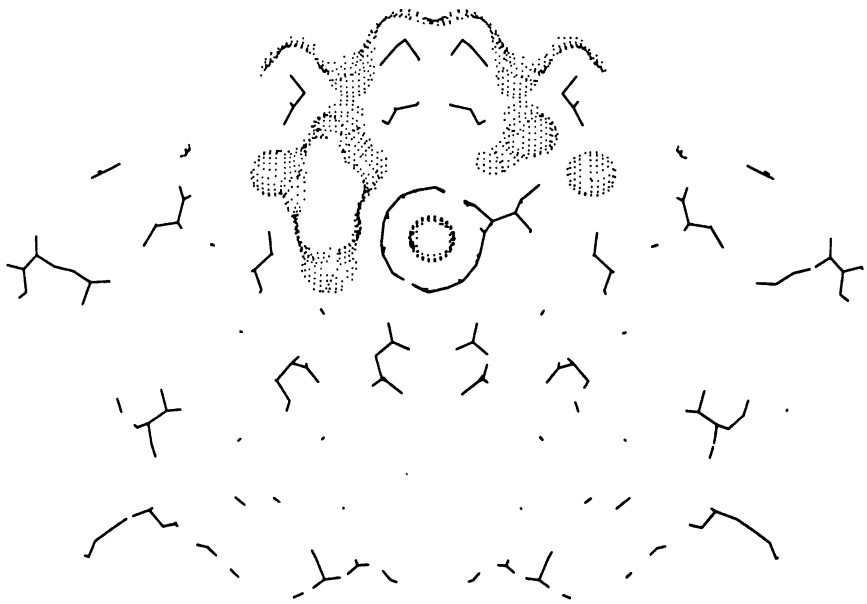


Figure 3.40. Cross Section of Model Complex of Compound Pe with the HIVP, Showing Molecular Surfaces.

	Carbon	Nitrogen	Oxygen
Complex (Å^2)	1464	108	243
Protease (Å^2)	1403	113	288
Ligand (Å^2)	506		12
Desolvation Due to Complexation (Å^2)	**-445**	-5	-57

Table 3.10 Surface Desolvation Breakdown of Compound Pe.

Unfortunately, as with the other trans iPr derivatives, the hydroxyl group was not well exposed to solvent to allow the attachment of a solubilizing group. However, I felt it was worth testing because if the affinity for HIVP was very high, it would not be necessary to get a lot of the compound into solution to determine its affinity. The subsequent test of this compound is described in the enzymology section that follows shortly.

Second Target Compound

During the analysis of the cyclohexyl compounds, I kept an eye on the fullerene synthetic literature, looking for compounds that could be synthesized and/or derivatized easily, and which had roughly the right morphology to interact with the target regions of the HIVP active site. In particular I found compound Q to be interesting (Figure 3.41). It had two phenyl groups fairly close to the C60 surface, and a ketone which potentially could be reduced to form a hydroxyl group which could in turn act as an attachment point for solubilizing groups. This compound was synthesized by Bingle in good yield[31].

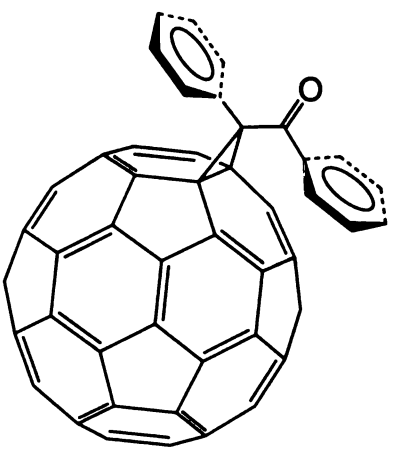


Figure 3.41 Compound Q

I analyzed compounds Ra and Rb, the two enantiomeric reduction products expected upon reaction with a non-chiral hydride source (Figure 3.42). As before, both of these were drawn up and docked to the HIVP using MINDOCK. Compound Rb did not produce complexes that looked

interesting; i.e., the substituents were not placed in or near the target regions of the HIVP. Compound Ra, however, produced a high scoring complex that was able to access the target regions. The eighth highest scoring complex resulted in a non-polar surface desolvation of 473 Å^2, as high as some of the better bisadducts (Table 3.10). Again, as with compound P, the hydroxyl group of compound Ra in this complex is sufficiently buried such that attachment directly is not possible. The hope was that the solubility would be great enough and that the affinity for the HIVP could be determined without the addition of solubilizing groups to the hydroxyl group. Thus, compound Ra also became a target molecule. The model complex and its cross section are shown in Figures 3.43 and 3.44.

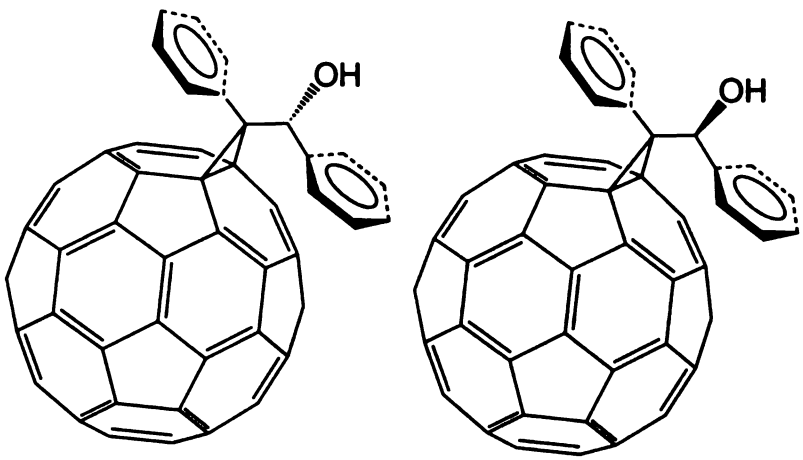


Figure 3.42 Compounds Ra and Rb

	Carbon	Nitrogen	Oxygen
Complex (Å^2)	1469	109	250
Protease (Å^2)	1403	113	288
Ligand (Å^2)	539		11
Desolvation Due to Complexation (Å^2)	**-473**	-4	-49

Table 3.11 Surface Desolvation Breakdown of Compound Ra.

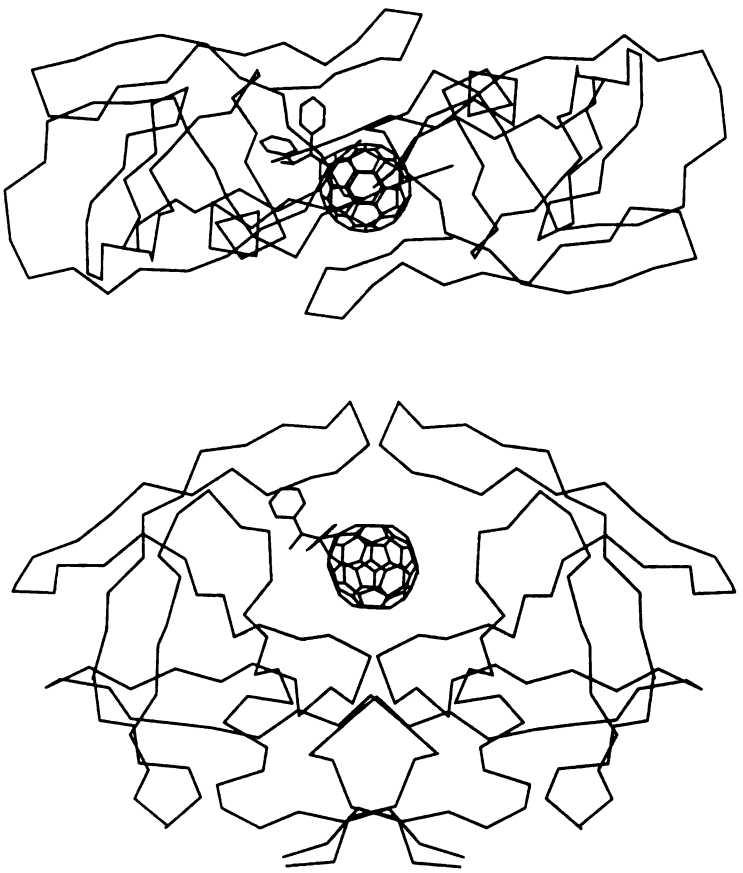


Figure 3.43 Model Complex of Compound Ra, Shown in "Top" and "Front" Views.

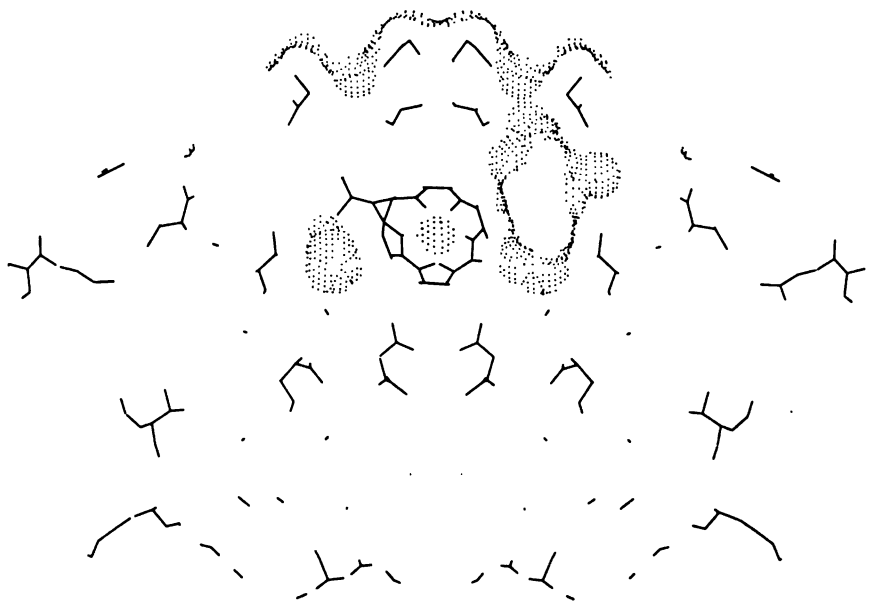


Figure 3.44 Cross Section of Model Complex of Compound Ra with the HIVP, Showing Molecular Surfaces.

Enzymology

The HIV protease is a remarkable enzyme. Consisting of two chains of only 99 amino acids each, it forms a C2 symmetric dimer in which an aspartic acid from each subunit combine to form the catalytic core which activates water for attack on the scissile peptide bond. The kinetics of reaction of the protease can be surprisingly complex. Several dynamic processes can take place with the protease in addition to the normal catalytic processing of substrate peptides and proteins. These are as follows: 1) monomer-dimer equilibrium, whereby the dimeric form of the protease is active and the monomer is not. The Kd, as well as k1 and k2, for this complex is affected by temperature, pH, ionic strength, presence of substrate and protein concentration[35]; 2) Autoproteolysis, whereby the enzyme will catalyze its own cleavage resulting in time dependent inactivation of protein; 3) A time-dependent loss of activity that is not related to the dimer-monomer equilibrium or autoproteolysis[36]; 4) A recently described time-dependent loss of activity of the protease dependent on *stirring* of the enzyme assay

mixture[37]. Early on in the enzymology I switched to the mutant protease Q7K designed and expressed by Professor Charles Craik and coworkers[32]. This mutant has similar kinetic properties to wild type but has been engineered to be stable to autoproteolysis. This meant that the enzyme's activity would be stable over the course of an evenings experiments, and, in fact, upon repeated thawings of a given aliquot over several days, which was not necessarily the case with the wild type protein.

In the previous testing of fullerene inhibitors, they were dissolved in DMSO which was then used in 1% in the assay mixture. Unfortunately, but not surprisingly, compound P was virtually insoluble in DMSO. Compound R however was soluble to low mM levels. An alternative solvent, N methyl pyrollidone (NMP) was found to solvate compound P well, to similar stock concentrations as compound R. Both of these stock solutions when added in 1% to aqueous solutions appeared soluble, i.e., no detectable precipitate, although this is a qualitative assessment.

Initial testing of the compounds used the approach described in Chapter 2, in which inhibitor and enzyme were preincubated for 5 minutes to allow a possibly slow binding inhibitor time to bind. The enzyme reaction was initiated with substrate addition and quenched. HPLC was used to quantitate the peptide reaction product and compare to a standard curve as described in Chapter 2. The initial inhibition data indicated mid to low nM binding affinity, although the data were quite noisy. In order to better determine causes of unclean data, the assay was set up to be a multi-point assay. Whereas previously, a single point was used to determine the enzyme reaction rate, multiple aliquots were now used and quenched at time-intervals to construct a full time-course for the enzyme reaction. This was important to detect possible slow-on kinetics of the inhibitor as well as possible monomer dimer equilibration kinetics.

In an attempt to eliminate a possible monomer-dimer equilibrium problem, a series of assays were run on compound R at high substrate concentration (60μM) and with no pre-incubation. The idea here was to saturate the enzyme with substrate (and inhibitor) from t=0, thus preventing a possible time-dependent loss of activity due to dimer converting to monomer upon dilution of stock enzyme to the assay mix. The time courses for inhibitor concentrations of 700 and 1300nM for the reaction are shown in Figure 3.45 and show a definite time dependence.

Compound R; No preincubation, 60μM substrate (~3x Km)

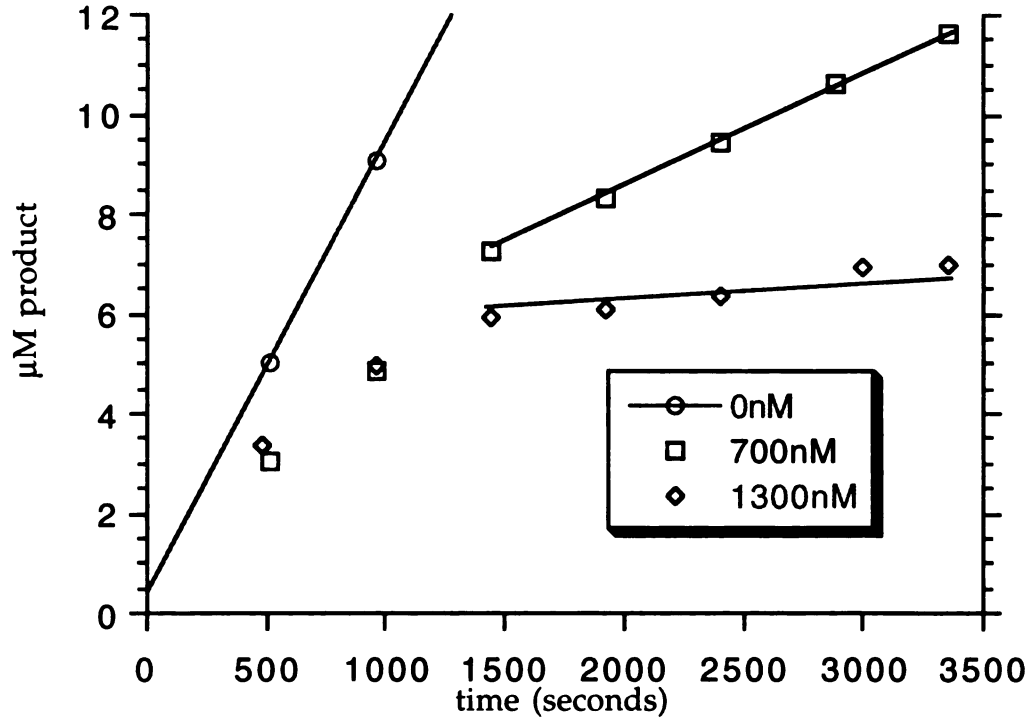


Figure 3.45 Compound R; No preincubation reaction course

Note this is not due to slowing of the reaction due to substrate consumption as the substrate concentration is several times above Km (15-20μM) and less than 20% of the substrate has been consumed. Figure 3.46 shows the IC50 plot of the range of these time courses, showing a mid nM IC50. These vi values were obtained by a straight linear fit to the time courses, a procedure which overestimates the reaction velocity (due to slowing of reaction over time). The relatively high velocity at the lowest inhibitor concentrations may be due either to a slower on rate or to a possible inhibitor-substrate interaction. The IC50 in a rough fit to these data is again in the mid nM range, which corresponds to a mid to low nM Ki, assuming competitive kinetics.

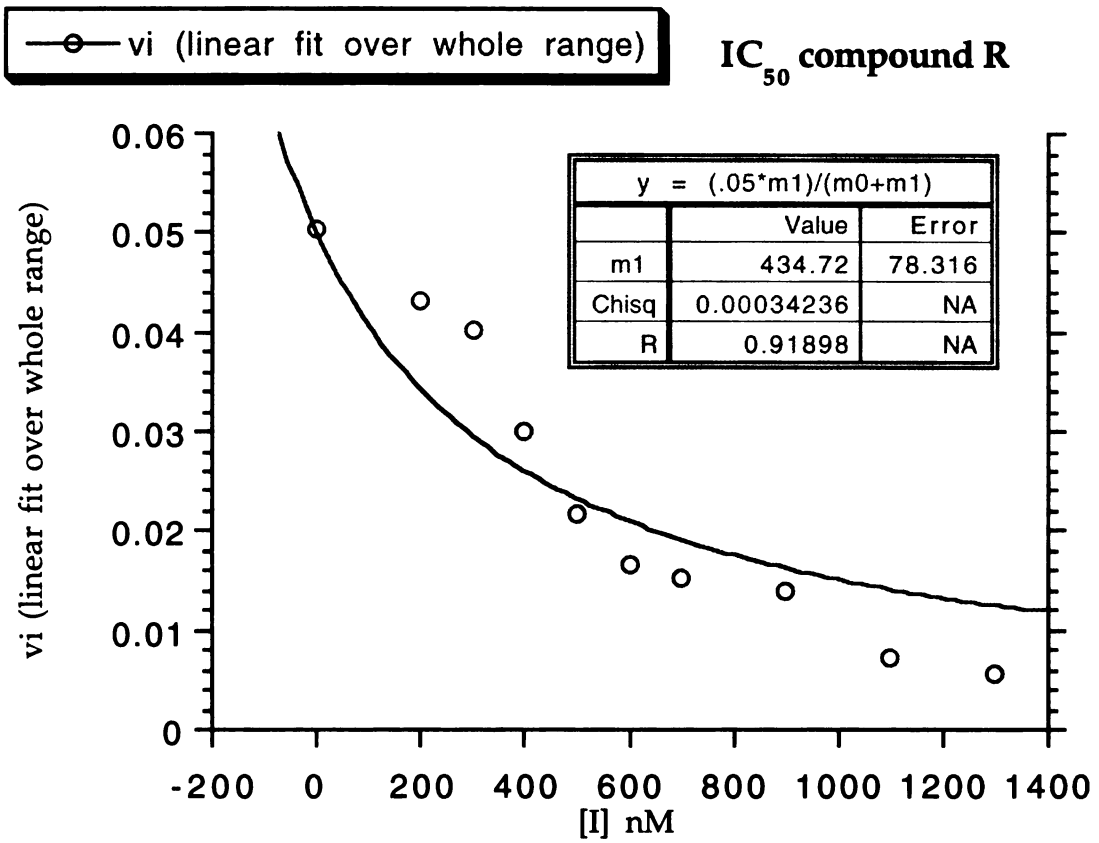


Figure 3.46 IC_{50} Plot of Compound R at High Substrate (~3x Km)

By eliminating the pre-incubation, in an attempt to prevent a possible monomer dimer equilibrium problem, I had introduced another problem, namely that of varying inhibitor enzyme association times. To assess whether or not the monomer-dimer equilibrium would be an actual problem, a control was done. The concern behind the monomer-dimer equilibrium problem is that during the 5 minute preincubation, the inhibitor will bind and stabilize the enzyme to loss of activity due to the dimer falling apart. The result of this would be an additional complicating factor added to the kinetics. A simple control was used to examine the change in activity of an assay mix when the enzyme was added five minutes before initiation of reaction with substrate vs initiating the assay with the enzyme addition to a mix in which substrate was already present. In this control, the activity of the two experiments varied by 10%, and the time course was linear with a close to 0 intercept, suggesting that the monomer-dimer equilibrium was not a

problem during a 5 minute preincubation. Henceforth all of the assays were performed with a 5 minute preincubation.

The result of this is that the time course of the enzyme reactions were linear, not curved as in the case of no preincubation, and had close to 0 intercepts. This indicated a normal enzymatic process. A sample progress curve for compound R is shown in Figure 3.47.

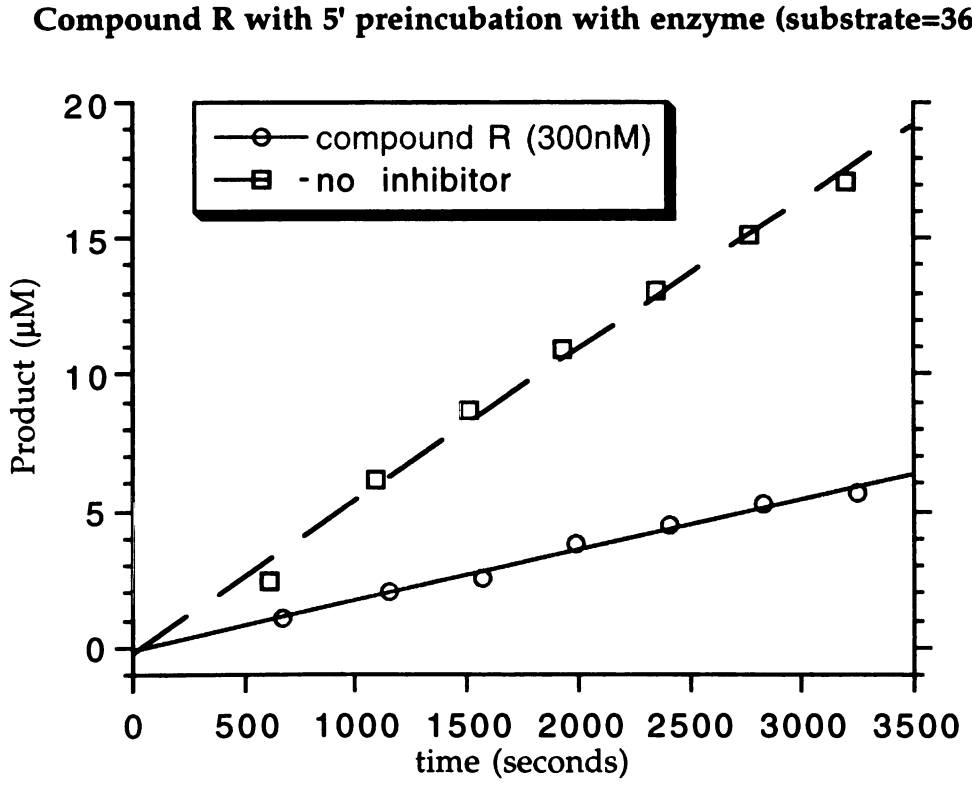


Figure 3.47 Sample Progress Curve for Compound R, With Enzyme/Inhibitor Preincubation

There was a problem in plotting of the data to extract Ki values, however. Although the data fit the general shape of competitive inhibition, they were noisy. Several factors were analyzed as potential sources of noise, including light and oxygen. There is some chance that fullerenes in the presence of light may promote the creation of singlet oxygen, which could be the inhibitory species. This was shown not to be the case, when inhibition persisted in assays performed in argon sparged solutions, while wrapped in

aluminum foil in subdued lighting. The assay itself was examined, and multiple controls were run with a known inhibitor of the protease, acetyl pepstatin. Acetyl pepstatin inhibited the enzyme, but more importantly, multiple determinations of vi with this inhibitor clustered closely, giving a standard deviation of 8% of the mean value of vi. In the case of compound P for example, scatter was much wider, approximately 35% of the mean. This scatter may be related to a low solubility of these compounds in water, even in the presence of organics. It is possible that when the organic solution of inhibitor is added to the buffered assay mixture and disperses, it is supersaturated, and will at a random interval come out of solution.

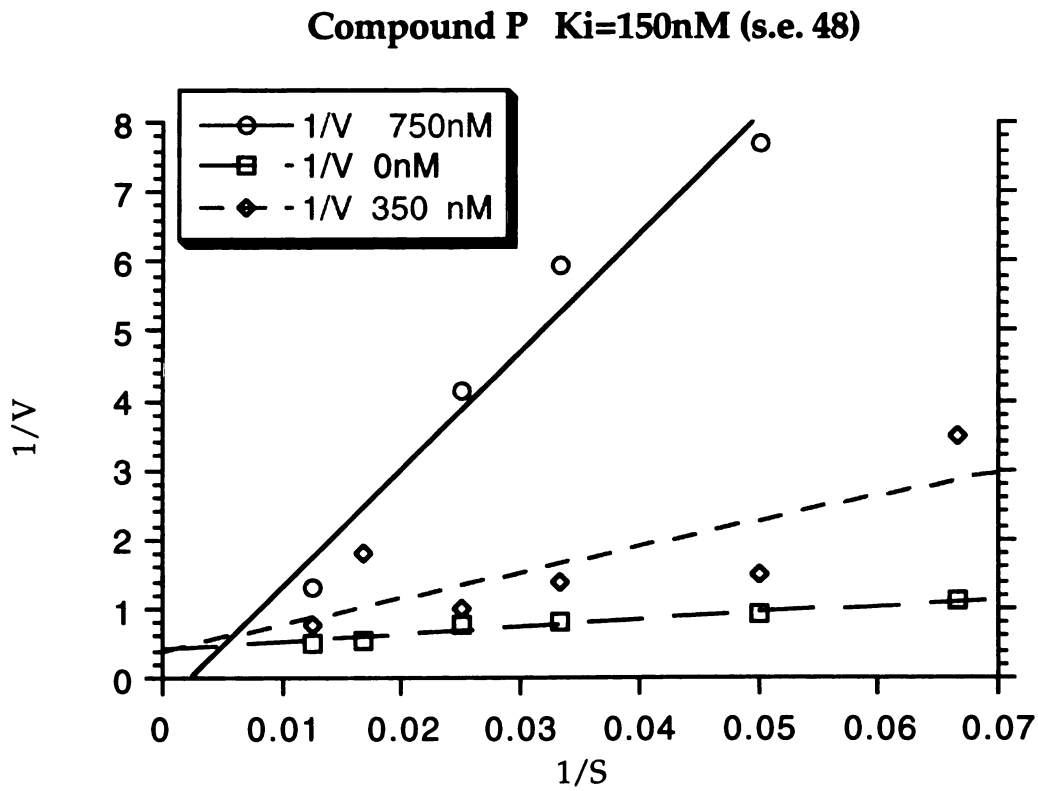


Figure 3.48 Compound P Kinetic Replot

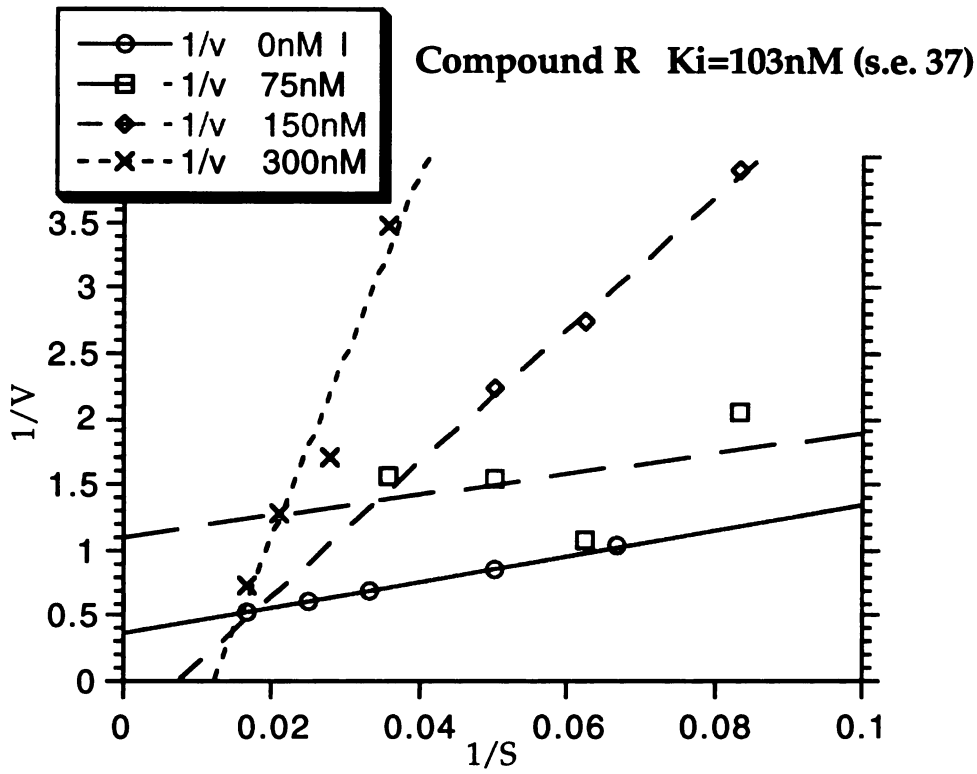


Figure 3.49 Compound R Kinetic Replot

Kinetic replots of vi values for compound P and R are shown in Figures 3.48 and 3.49 . The data are best fit by a competitive model, although even with this the standard error value is high, approximately 30%. Even with this relatively high error, it represents a large increase in affinity in comparison to the originally tested compounds.

Conclusions

These results are encouraging, in that the originally developed models continue to be predictive, and correctly identified two new compounds of higher affinity. The compounds solubility is a source of concern for further application. There are numerous ways in which these molecules can be further derivatized, however, and polar groups introduced which will increase water solubility. The phenyl rings of compound Ra are partially solvent exposed in its top scoring complex, which could offer a site of derivatization. Furthermore, although attachment to the hydroxyl of both tested compounds is not possible in the top scoring complexes (as previously described), it is possible that a hydroxyl adduct could dock in a different, but similarly successful, orientation.

Chapter 4

Introduction

Adding single functional groups to the surface of the fullerene is not difficult to do. However, after this initial addition, there are myriad theoretical locations for subsequent additions. The ability to place a second functional group on the surface in a defined orientation relative to the first is of great utility, especially in the field of molecular design, which relies heavily on defined spatial arrangements of functionalities.

In this chapter I describe the development of an approach to the design of regio-specific fullerene derivative syntheses, as well as a computational methodology developed for it. This is followed by a verbatim reprint of a manuscript submitted to Journal of the American Chemical Society which describes the application of this approach in the analysis of a specific system. This methodology was developed in order to solve a specific problem, the synthesis of a diamino fullerene described in Chapter 2 and in the 1993 Journal of the American Chemical Society article. Despite the specific objectives of this development, this approach may be generally applicable to designing regiospecific fullerene synthesis as well as to understanding reaction dynamics.

Description of Approach

The developed approach is summarized as follows: to bring about the addition of two groups to the surface of the fullerene with a specific final separation, the two reactive groups are joined by a flexible linker molecule. After the first reactive group adds to the surface the second is directed to specific locations by the conformational tendency of the linker. (Figure 4.1)

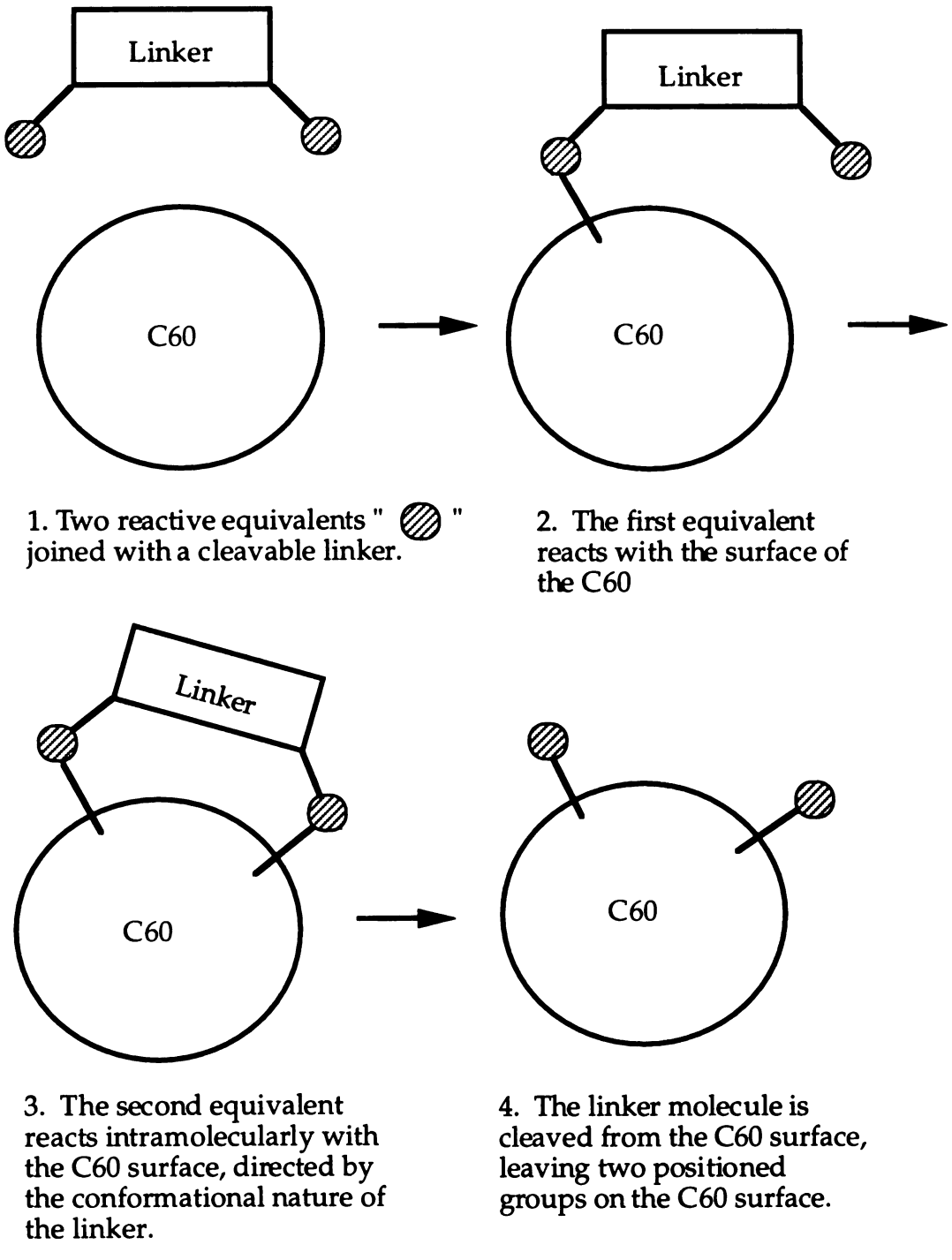


Figure 4.1 Schematic Depiction of Regiospecific Fullerene Synthetic Strategy.

To analyze which locations are expected to be 'hit' by the second reactive group, a simulation is performed, where the torsions of the linker are systematically driven, to create the entire ensemble of rotational conformers. The energies of each of these conformers is determined via molecular mechanics, from which a Boltzmann factor is determined. This factor is proportional to the probability of observing a specific conformation. These conformations are then analyzed for their ability to position the second reactive group in a productive way (i.e., with reasonable geometry) next to a reactive site on the fullerene surface. The conformations that satisfy the criteria for productive reaction have their Boltzmann factors summed, thus providing an estimate of the probability of observing reaction at specific sites.

As alluded to in the introduction to this chapter, the issue of regiospecific fullerene syntheses became important following my design of a 1,4 diamino C60 (Figure 4.2) as a possible tighter binding derivative. The intent was to provide two amino groups to form salt bridges with the active site aspartic acid carboxyl groups, thereby adding up to 6kcal/mol binding energy above the core interaction. The 1,4 diamino compound was selected after a series of different diamino C60 isomers were modeled, and single mode DOCKed into the protease. The 1,4 diamino isomer was the only one able to both fit into the active site and position the amino groups within ~3Å of the carboxylate groups.

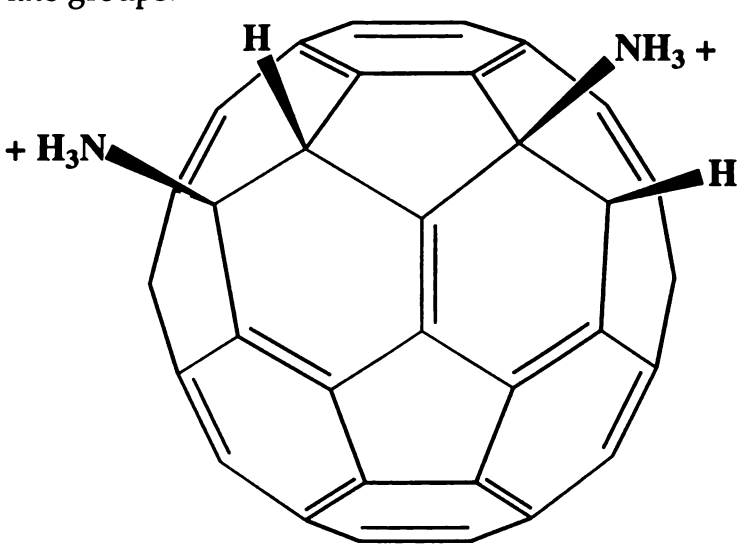
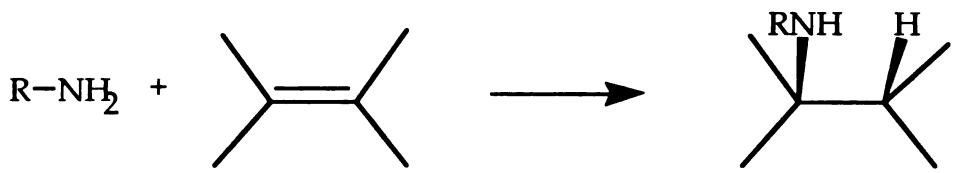
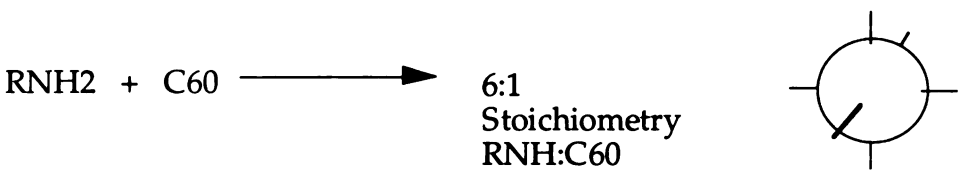


Figure 4.2 "1,4 Diamino C60"

Early in the development of fullerene chemistry it was discovered that amines add to the surface of C60. (Figure 4.3)



A) Bulky amine case (e.g., t-butyl amine)



B) Unhindered amine case (e.g., ethyl amine)

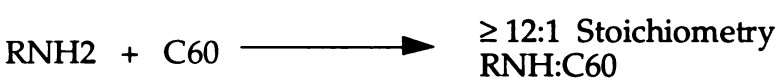


Figure 4.3 Amine reactions with C60

If these were bulky amines, a stoichiometry of 6:1 was observed. An obvious interpretation of this is a symmetrical octahedral addition pattern. If, however, the amine was not bulky, many more amines could be added to the surface (over 12). This suggested that the reaction was being sterically limited, not electronically; i.e., as long as additional amines could reach the surface of the C60 after the first 6 equivalents had reacted, they could then react. This supported the idea that two underivatized amines could react in the relatively close separation required of the 1,4 diamino C60 envisioned. The main problem was then, how to limit the reaction to two equivalents, and, furthermore, how to ensure that they would react to produce a product with a specific separation.

The devised solution to this problem was to link the two amine equivalents with a cleavable linker, that could deliver both equivalents to the surface and then be removed, leaving the two groups with a specific orientation. The specific tool for delivering the equivalents was the amidino group. The idea was an analogy to a technique used to generate alkyl amines. In this case, an amidino group acts as a nucleophile and attacks an alkyl halide, producing an alkyl amidine. The amidine can then be hydrolyzed, releasing the alkyl amine. (Figure 4.4)

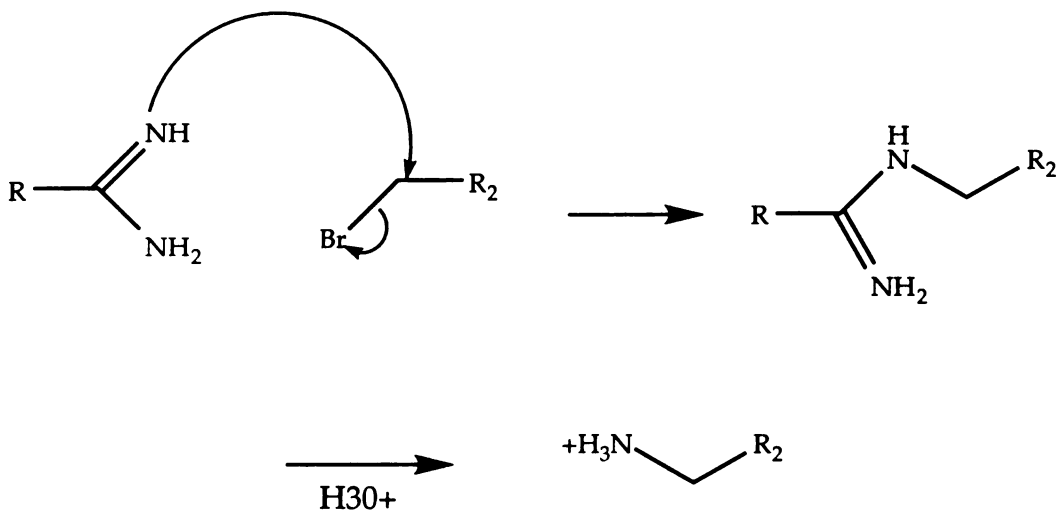


Figure 4.4 Method of using amidines to produce alkyl amines.

In the case of the C60 synthesis, the C60 itself would act as the electrophile (in place of the alkyl halide). It would react with the amidine, in a manner analogous to the previously described amine addition reactions. The amidine could then be hydrolyzed, leaving the amino group on the surface. More importantly, two amidino groups could be linked to a conformationally restricted linker. After an initial intermolecular reaction of the C60 with the first amidino group, the second amidino group would react in an intramolecular fashion. (Figure 4.5)

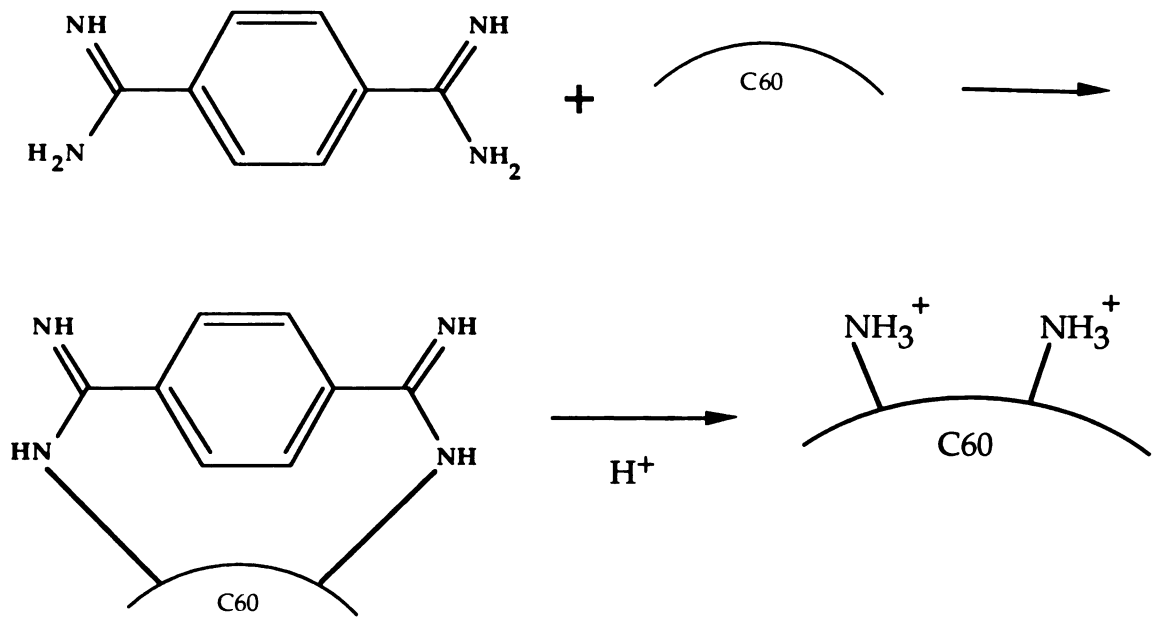


Figure 4.5 Strategy of Linked Amidines to Produce a Specific Diamino C60

Computational Strategy

A major question I had was, which linker should be used to produce a final molecule with the separation of amino groups originally envisioned? How could the final separation of amino groups in the final product be predicted from a given linked reagent molecule? The solution devised utilized a statistical mechanical analysis of conformational energies as derived from molecular mechanics. This allowed a determination of the probable distance observed between the two amino groups following reaction.

First, a model of the singly reacted species was modeled using the modeling package SYBYL. (Figure 4.6)

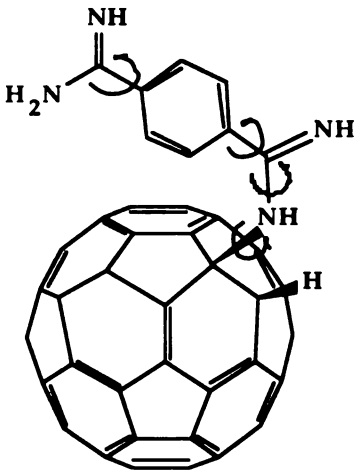


Figure 4.6 Singly reacted C60, showing bonds rotated during simulation.

Then the full ensemble of available conformations of this molecule were generated using the technique of systematic conformational searching which drives all rotatable bonds with a specified resolution thereby generating $(360/r)^n$ conformers. For each of these, a conformational energy was assigned using the Tripos force field. These conformers were then filtered for the ability to place the second amidino group within close vdW contact distance of the C60 surface. This filtering is facilitated by a feature of the systematic conformational search module of SYBYL which allows distance constraints to be put on conformers. In this case, the constraint was that the distance between the reacting amidino nitrogen and the centroid (center of mass) defined by the atoms of the C60 cage had to be within 7Å. The conformers that passed this filter were then further analyzed. During the conformational search, a so-called 'distance-map' was created. A distance map is simply a measure of distance between two identified centers for each conformation that is generated. In this case the distance map was for two nitrogens: one already reacted with the surface of the C60, and the other sweeping out over the surface. These distance were binned with a 0.1Å resolution.

At the end of the simulation, one is left with 1) a group of conformations that can position the second amidino group within a 'nominal reactive distance' from the C60 surface, (henceforth called 'productive conformers'), 2) the energy of each of these productive conformers, and 3) the N-N distance that they produce. It is the combination of this information that gives a probability distribution for the separation of the amino groups that the linker can deliver. How this is so is now described.

For each of these productive conformers, its conformational energy is converted into a Boltzmann factor using the expression $p=e^{-[(Ec-Eg)/RT]}$ where Ec is the energy of the conformation, Eg is the global energy minimum discovered during the complete conformation search, R is the universal gas constant and T is the temperature. The Boltzmann factor represents a quantity that is proportional to the probability of observing a state of a given energy. The trend is that as the energy of a state increases one will observe it less. Also, as the temperature is increased, states tend to equal representation. The lowest energy conformational state energy is subtracted from the observed conformational energy because of the way that such systems are defined: one examines how the energy partitions among states starting with ground state. At absolute zero, all particles in a system will be in this lowest state, the ground state, or global conformational minimum.

Within SYBYL the results of the initial conformational search, a table consisting of three columns: a conformation number, the N-N distance defined above and a molecular mechanics energy. This table was then dumped to a text file with these same columns. Using the following awk script

```
{print exp(-($1-1710.87)/.591),$2}
```

this file was converted into another file which contained the N-N distance and the Boltzmann factor in pairs for each conformer. This file was then processed using a PERL script, which summed all of the Boltzmann factors for each possible N-N distance bin. The PERL script follows, and was written by Eric Petersen.

```
#!/usr/local/bin/perl
while(<>) {
    chop;
    ($val, $cat) = split(' ');
    $total{$cat} += $val;
}

while(($cat, $tot) = each %total) {
    print "$cat: $tot\n";
}
```

The output of this script is a file containing a list of number pairs: a specific N-N separation, and the sum of the Boltzmann factors for this separation. These pairs were then exported and plotted. The plot result is a probability distribution showing the likely hood of observing a specific N-N separation when the second nitrogen is in contact with the surface of the fullerene.

Modeling of Diamidines

Two principal molecules were modeled: 1,3 diamidino benzene and 1,4 diamidino benzene (Figure 4.7).

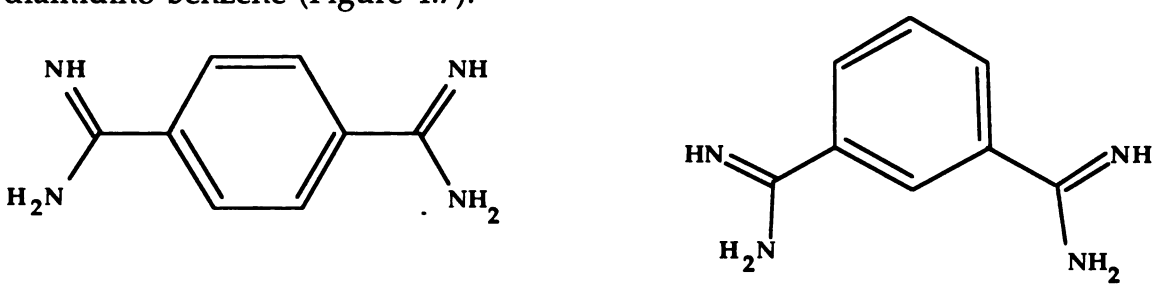


Figure 4.7 1,4 Diamidino Benzene and 1,3 Diamidino Benzene

Figure 4.6 shows the singly reacted species of 1,4 diamidino benzene as well as indicating the torsions that were driven in the systematic conformational search at a resolution of 10 degrees. The results of the analysis following the procedure described above are shown in figures 4.8 and 4.9

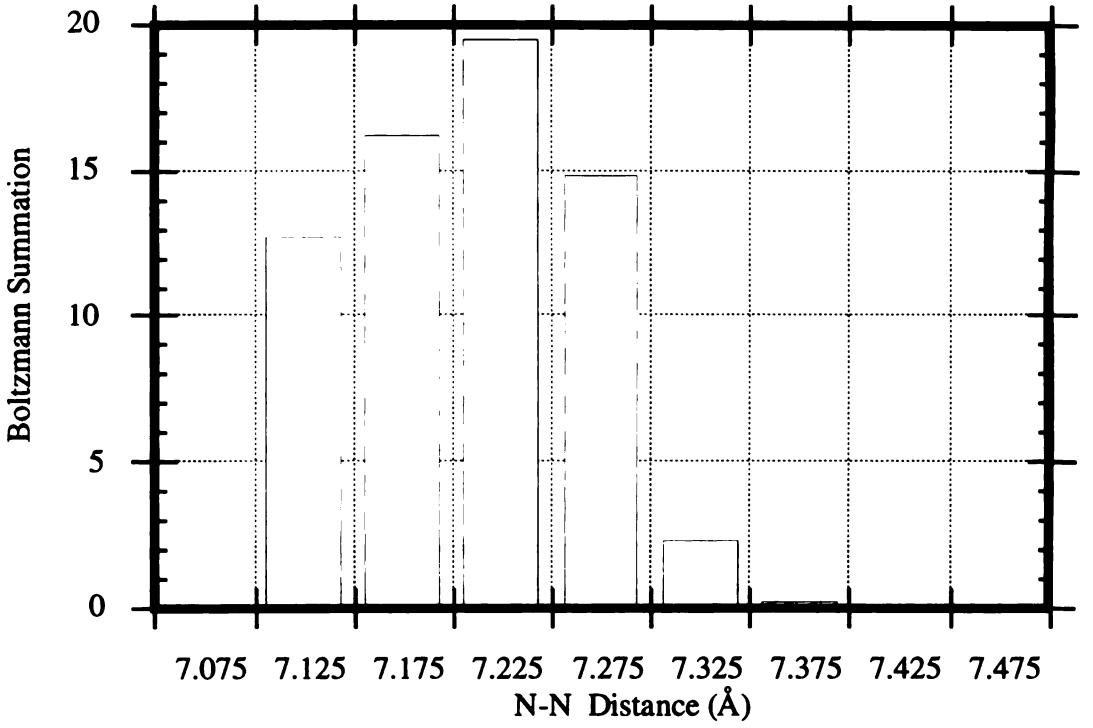


Figure 4.8 1,4 Di-Amidino Benzene [7Å,centroid-N subset] N-N Distance Distribution

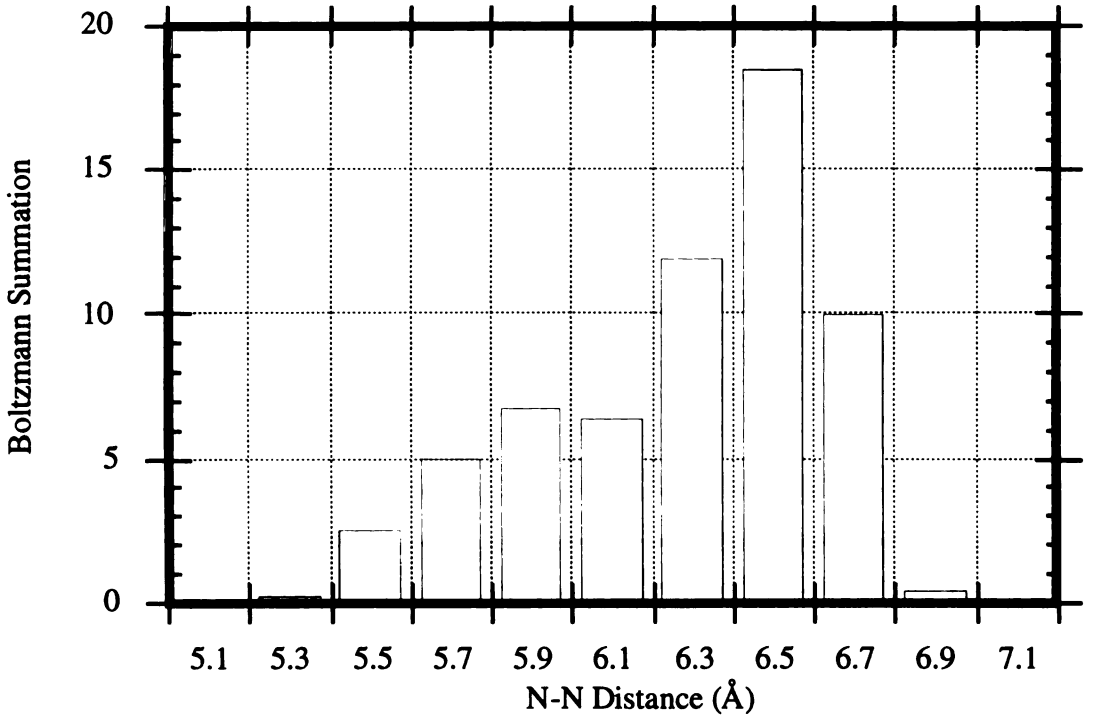


Figure 4.9 1,3 Di-Amidino Benzene [7Å,centroid-N subset] N-N Distance Distribution

These show that each of these molecules delivers the second equivalent of amine to the surface with a very narrow possible range of final nitrogen to nitrogen distances, 7.1-7.4 Å in the case of 1,4 diamidino benzene and 5.3-6.9Å in the case of 1,3 diamidino benzene. Also it shows that this distance can be finely modulated depending on the conformational nature of the linkage. It should be emphasized that in this particular application, the result does not indicate that there *will* be reaction of the second equivalent or which exact atom it is likely to add to. It indicates only that if the reaction *does* take place, it will be confined to a very narrow range of possible nitrogen to nitrogen distances. Although it was not done in this simulation, it is a relatively simple matter to query individual reactive sites on the fullerene surface to determine the probability of individual specific site reactivity, and this will be described in some detail in the manuscript that follows the first half of this chapter.

Because the simulation indicated that 1,4 diamidino benzene would produce an isomer, or series of isomers, with the correct range of amine to amine distances required of the C60 derivative, it was selected as the reagent to use to attempt this approach. Unfortunately, the synthesis was not able to be completed due to problems with the final cleavage of the linker portion. The following briefly summarizes the synthetic attempts.

Synthetic Attempts

Initially, to test whether amidines would react with C60 in an analogous manner to that of simple amines, a test reaction of benzamidine, i.e., 1-amidino benzene, with C60 was performed. First, benzamidine free base was liberated from the hydrochloride salt by extraction from aq. 5N NaOH with diethyl ether. The ether solution was dried with sodium sulfate and taken to dryness on a rotoevaporator, yielding a yellow oil. This oil was vacuum sublimed at 50-110° C, yielding white crystals. Sixty seven mg of this purified benzamidine was treated with 50mg of C60 in 35ml toluene, under nitrogen for 13 hours at room temperature. The normally violet C60/toluene solution turned reddish brown in this time, with a large amount of precipitate. The red-brown precipitate was washed with toluene and hexanes and dried to yield 60mg product (58mg expected from 1:1

stoichiometry reaction). NMR of this crude material in d6-DMSO showed an aromatic cluster at 7-8ppm and a broad shallow hump from 5-6ppm, due probably to the amidine proton(s). Forty mg of the crude material was loaded in CH_2Cl_2 to a silica column (1.5 x 25cm) and eluted first with CH_2Cl_2, during which unreacted C60 eluted in a violet band, and then with 1% methanol/CH_2Cl_2 during which two streaky red-brown band eluted. Both of these band showed at least two spots by TLC. They were combined and submitted for analysis using mass spectrometry. LSIMS MS showed the expected mass of 841 corresponding to the monoadduct. In addition it showed peaks at 961, 1081 and 1201, each corresponding to higher adducts of benzamidine. The 1081 peak had the highest signal of all the peaks in the mass spectrum, with about 5 times the signal of the other 3 observed adducts. Although the stoichiometry of the reaction was not controlled (no doubt due to the use of a 5 fold molar excess of benzamidine), this showed that the benzamidine would react with the C60 surface in the expected manner, and could potentially form the basis for the molecule that could deliver the two amine equivalents to the C60 surface.

The generation of the final di-amine target compound of course depends on the hydrolytic cleavage of the di-amidino reacted product to the free amine and 1,4 benzene carboxylate. This was the major stumbling block of this synthesis. A variety of possible hydrolytic conditions were attempted. The initial attempt used the material described above that was shown to be a mixture of benzamidine/C60 adducts in a stoichiometry of (1-4):1. Ten mg of this material was dissolved in 6ml of glacial acetic acid and 0.4ml of conc. HCl, in which partial solubility was attained. This was stirred for 15 hours, after which no product was detected by TLC, and the starting material spot remained. This solution was then heated to 100°C for 24 hours, after which, again, no reaction was detected by TLC. Another approach used HBr instead of HCl. HBr was suggested by Fred Wudl. Ten mg of the benzamidine:C60 adduct was stirred with 8ml acetic acid, 2ml of 30% HBr/acetic acid and 0.6ml of water. This was stirred for 4 days, at which time it showed no shift in the starting material spots via TLC. It was then heated to 100-110°C and fitted with a condenser. After 48 hours, 3ml of HBr solution was added. At the end of the reaction, the solution was clear and TLC showed a loss of the starting material. The solvent was removed, and a sample was run on NMR and sent to mass spectral analysis. NMR was ambiguous, showing two apparent

singlets at 2.1 and 1.3 ppm. Furthermore, the mass spectral analysis was unable to find anything near to the expected mass of the amino derivatized C60.

It is possible that the conditions required to hydrolyze the amidine to the amide and free amine would also bring about the cleavage of the C60-N bond. In retrospect, it may have been possible to find conditions that were able to release the amino-C60 compound. Also, it is possible that the approach to mass spec may have been varied and would have indicated the production of the amino adduct; however, this was not attempted.

These investigations began in early 1993 and were presented at the departmental retreat at Asilomar in December of 1993. By late 1994, two papers appeared from the labs of Thomas C. Bruice and Francois Diederich which separately described two aspects of the work discussed in this chapter. Lightstone and Bruice[38] published a communication describing the application of using summations of Boltzmann factors generated from molecular mechanics energies to rationalize intramolecular reactivities in a series of compounds. In an analogous way to the computational methodology I have described, they summed Boltzmann factors for conformations of molecules which could position two groups within a reactive distance of each other. These were then compared to the known rate constants for a series of compounds with varying groups linking the two reactive moieties.

Diederich and co-workers published a paper which described using a tether to deliver a second group to the surface of C60 after an initial group at the other end of the tether has reacted.[11] Their analysis of potential linker molecules did not use any of the computational concepts described here. Having both of these papers appear was somewhat frustrating, in that two fairly fresh concepts which had been synthesized and integrated had been eclipsed by description of each of the ideas separately. A benefit of this, however, was that (in the case of the Diederich paper) there was now an experimental test system with which the computational methodology could be validated or at least supported. The results of this analysis have been written up and submitted as a communication to Journal of the American Chemical Society. This manuscript, which describes the application of the above described computational methodology, to an analysis of a known tether directed regio-specific fullerene synthesis, now follows.

A Computational Strategy For the Design of Regiospecific Syntheses of Fullerene Derivatives

Simon H. Friedman* and George L. Kenyon*

Department of Pharmaceutical Chemistry
University of California, San Francisco
San Francisco, CA 94143

Abstract

We describe a computational methodology to aid in designing regiospecific syntheses of fullerene derivatives. It involves the analysis of tether-directed syntheses, in which two reactive groups are linked with a molecule which directs the placement of the second reactive group after the first has added to the fullerene surface. In our analysis, an ensemble of conformations of the singly reacted fullerene adduct is generated using the technique of systematic conformational searching. The conformations so generated are filtered for their ability to place the second reactive group in a nominally "reactive" position to known reactive sites on the fullerene surface. The energies of the conformations which pass this filtering are generated using molecular mechanics, and then converted into Boltzmann factors, which give a relative probability of observing the given conformation. The sum of the Boltzmann factors of the conformations that pass the filtering test for each potential reactive site gives a prediction of the site that will most likely react with the second reactive group. In this communication, we demonstrate the application of this approach to a specific tether directed synthesis performed by Francois Diederich and co-workers. [Isaacs, L. et. al., Ang. Chem. Int. Ed. **1994**, *33*, 2339-2342]

The ability to add two or more functionalities to the surface of fullerenes with defined geometry is a key element in expanding the potential applications of fullerene derivatives. This is especially true in biochemical systems, where host-guest binding often depends on the accurate positioning and alignment of complementary surface types. In general there are two ways that regiospecificity can be attained in fullerene synthesis : 1) by exploiting relative reactivities of sites, i.e., the reaction at a single site influences the reactivity of the remaining sites, leading to a specific pattern of modification[39] or 2) by tying together two reactive groups with a flexible linker, which then directs the addition of the second group upon reaction of the first[11]. In our pursuit of the synthesis of a diamino C60 compound as a potential HIV-protease inhibitor[27], we have developed a computational strategy for evaluating the ability of different linkers to "deliver" the second reactive group to the fullerene surface in a potentially productive manner, i.e., in the desired position. We elaborate upon this strategy here and demonstrate its application to a known linker-assisted regiospecific fullerene modification accomplished by Diederich and coworkers[11].

The strategy is to find the bond on the C60 that is most likely to be approached by the second reactive group as the linker sweeps through its possible conformations. The computational approach is summarized as follows: 1) Construct a model of the singly reacted species. 2) Generate the entire ensemble of conformers available to the flexible linker that joins the first reactant (now bonded to the C60) and the second reactant. This is done using the technique of systematic conformational searching where rotatable bonds are driven by some incremental change. 3) Filter this large ensemble for the conformers that place the second reactive equivalent near the surface of the fullerene in a productive manner. 4) For each potential reactive bond on the C60 surface, query whether or not a given conformer positions the second reactive equivalent closely enough and in the correct orientation to react with the bond. 5) Finally, determine the molecular mechanics energy of the conformations that satisfy condition 4, and, from these values and the apparent global conformational energy minimum, determine the Boltzmann factor for each of these reactive conformations. By examining the sum of these Boltzmann factors for each potential reactive bond, the probability of finding the second reactive equivalent in a reactive position is assessed and

then used to determine the likely position of addition. A similar approach has also been pursued by Bruice and coworkers to help understand rates of intramolecular esterifications[38,40].

We have applied this methodology to analyze a synthesis performed by Diederich and coworkers[11] which utilizes the approach of tying two reactive equivalents to a flexible linker. They utilized the species shown in Figure 4.10, which is the product of an initial addition across a 6-6 bond forming a cyclopropyl link. Following this first addition, the diene at the end of the tether adds to a 6-6 bond on the surface in a Diels-Alder fashion. These authors designed the linker they used based on the relative energy of the possible products, a reasonable approach that we will comment on below.

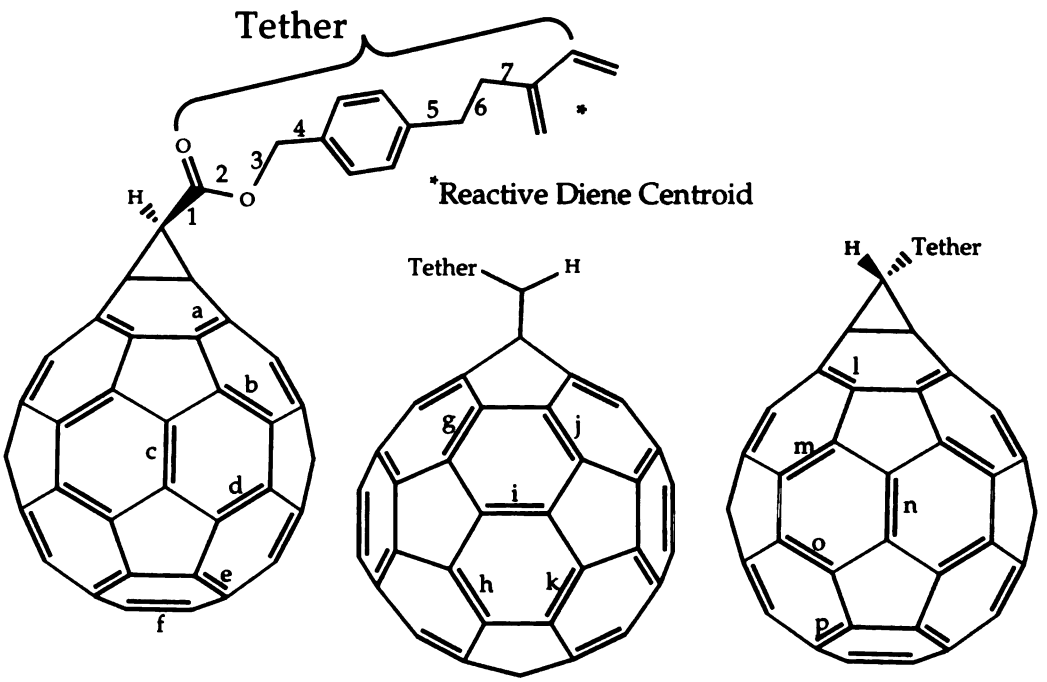


Figure 4.10 Schematic showing tether linked diene and the 15 unique 6-6 bond targets (a-p) and the 7 bonds driven in the search

We modeled the monoadduct using the program SYBYL[9,41]. A systematic conformational search was initiated by driving each of the seven rotatable bonds with a 30° resolution (a theoretical limit of 3.6×10^7 conformers). Each of the conformers was analyzed to determine the distance between two centroids. Centroid 1 was at the center of mass of the two reacting diene carbons. Centroid 2 was at the center of mass of the C60 sphere. As each conformation was generated, it was discarded if the centroid 1-

centroid 2 distance was greater than 8Å, i.e., distant from the surface of the fullerene. This filter resulted in a reduction to ~1.2 x 10^4 potential reactive conformations.

Each of these remaining conformations was further tested for ability to position the diene carbons within a reactive distance to possible reactive C60 bonds. This process is simplified because of the known preference in C60 for Diels-Alder reactions to modify at 6-6 bonds. Therefore, only 6-6 bonds were analyzed. For each 6-6 bond, we filtered the potential reactive conformations based on the ability of the conformation to position each of the two reacting diene carbons within a specific distance to the reacting 6-6 bond carbons. In order to assess the effect of this distance on the results of the simulation, we repeated the filtering using a range of distance cutoffs, from 4.0 to 3.4 Å. In addition to the distance filter, each conformation was also filtered for the angle of the diene relative to the 6-6 bond. This criteria was set at $\leq 30°$[42]. Because the diene is unsymmetrical, this meant that for each 6-6 bond there were two possible final products related through a 180° rotation of the diene at the time of reaction, designated with subscripts a and b in Table 1[43]. A simplifying factor was the presence of a plane of symmetry in the monoadduct which roughly halved the number of 6-6 bonds necessary to be examined. The final number of filters performed was 31, two for each of the 6-6 bonds except for the one directly opposite to the cyclopropyl group, where symmetry causes the two possible diene adducts to be the same.

The conformations that passed the filter for each bond were noted, and their molecular mechanics energy was converted into a Boltzmann factor using the expression $p=e^{-[(E_c-E_g)/RT]}$, where E_c is the energy of the conformation and E_g is the lowest energy conformation found in the complete search of 3.6 x 10^7 conformers. This expression represents a quantity that is proportional to the probability of observing a specific energy state, in this case a potentially reactive conformation. The results of this summation are listed in Table 4.1.

6-6 Bond	4.0Å	3.8Å	3.6Å
$\mathbf{b_a}$	7.2×10^{-2}	3.0×10^{-2}	7.1×10^{-9}
$\mathbf{b_b}$	3.9×10^{-5}	9.0×10^{-13}	
$\mathbf{c_a}$	1.4	3.1×10^{-2}	1.1×10^{-2}
$\mathbf{d_a}$	1.2×10^{-2}		
$\mathbf{d_b}$	4.8×10^{-2}		
$\mathbf{g_a}$	9.1×10^{-5}	2.3×10^{-5}	
$\mathbf{i_a}$	7.9×10^{-9}	7.7×10^{-9}	7.7×10^{-9}
$\mathbf{i_b}$	3.3×10^{-8}	3.3×10^{-8}	
$\mathbf{j_a}$	4.9×10^{-7}	4.8×10^{-8}	
$\mathbf{l_b}$	8.3×10^{-12}		
$\mathbf{m_a}$	3.0×10^{-8}	3.3×10^{-9}	4.0×10^{-10}
$\mathbf{m_b}$	3.6×10^{-8}	3.6×10^{-8}	3.6×10^{-8}

Table 4.1. Summation of Boltzmann factors for productive conformations (satisfying angle and distance cut-offs) at different reactive atom distance cut-offs. No conformations were observed at the 3.4Å level. Subscripts a and b for each bond identifier indicate one of two possible orientations of the diene relative to the 6-6 bond[43].The bonds not listed had no conformations which satisfied the filtering criteria.

The aim of this simulation is to find which of the possible reactive 6-6 bonds is most likely to be near the reactive diene in a productive orientation. The $\mathbf{c_a}$ isomer is the sole product experimentally observed by Diederich and coworkers[11]. These simulation results show that, for the three limiting cutoff distances, the sole experimentally observed product, ($\mathbf{c_a}$), is the product which has the highest "contact probability" from among the 31 possible products, although for the intermediate cut-off value of 3.8Å, the experimentally unobserved $\mathbf{b_a}$ isomer has close to the same Boltzmann summation as the $\mathbf{c_a}$ isomer. The ability of this relatively simple analysis to rapidly indicate likely targets of addition suggests that it may be useful in designing other regiospecific fullerene syntheses, especially as greater sophistication of the models used is introduced (i.e., higher resolution searches and inclusion of quantum mechanics to analyze "productive" conformations).

A significant factor that has not been included in the simulation is the intrinsic reactivity of each 6-6 bond. There is evidence that there is considerable variation in subsequent 6-6 bond reactivity following an initial 6-6 bond addition[44]. As other information about intrinsic site reactivity becomes available, either from theory or experiment, it can be included in our computation by scaling the final summed Boltzmann factors. Also, further experimental results can feed back into determining what are optimal filtering distances and angles.

Diederich and coworkers approached the tether design problem by analyzing the theoretical energy of the possible reaction products produced by a given tether. The tether that they chose produced a modeled product (the $\mathbf{c_a}$ isomer final reaction product) that was 5 kcal/mol lower in energy than the $\mathbf{b_a}$ product (i.e., the product that would be formed by reaction of the bond **b**), as determined using semiempirical calculations. This is a reasonable approach insofar as the product of the reaction fairly represents the structure and energy of the activated complex. The methodology we have presented in this communication provides an alternate approach to the solution of the problem of predicting the sites of intramolecular reactivity by analyzing the relative populations of potentially reactive conformations. It is our hope that this approach to the problem of designing regiospecific fullerene syntheses may prove a practical tool as well as a route to understanding and modeling reaction dynamics.

Acknowledgments

This research was supported by USPHS Grant GM39552. We thank Prof. Leonard Peller for useful discussions.

Chapter 5

Summary and Conclusions

This thesis describes some of the first applications of fullerenes to biological investigations. I was not motivated to do this work, however, by the wish to introduce fullerenes into the biological world. Instead my motivation was based on a recognition that C60 fullerene was an ideal starting point specifically for the design of HIV protease inhibitors: This is so because, at the risk of harping on a point, the C60 fullerene is highly complementary to the active site of the HIV protease, both sterically and chemically.

Although my original motivation was to solve a specific problem, I have come to view fullerenes as potentially generally useful in the realm of drug design. I believe that fullerenes may be useful, generic, conformationally-restricted pharmacophore scaffolds. In other words, they can act as platforms on which the groups necessary for binding to a target site may be rigidly positioned and oriented.

In the case of the HIV protease design work, the analysis of ligand-protein interaction was simplified by the conformationally restricted nature of the fullerene sphere. Because the central sphere of C60 is totally conformationally restricted, when a substituent is added to the fullerene surface, it is limited in the orientations that it can adopt. The effect of this is two-fold. First of all, modeling the solution conformation of the derivative is simplified. Secondly, there is no conformational entropy loss upon binding due to the freezing out of bond rotations. A common theme in ligand design is the rigidification of flexible molecules to increase binding affinity by reducing conformational entropy changes upon binding.

These two points, important for the design work with the HIVP, also make the fullerene a generally interesting starting point for drug design. A major avenue in drug design is based on pharmacophores, i.e., the spatial arrangement of key chemical features required of a compound to bind to a site. Following the discovery of a pharmacophore for a given site, one can then design molecules that can present the important features of the pharmacophore with the correct relative geometry. *Fullerenes can*

potentially fill the role of a conformationally rigid scaffold upon which the elements of pharmacophores can be placed. Carbon-to-carbon distances vary from 1.4 to 7Å within C60, a range which is typical of many drug molecules. These distances can of course be increased through the actual substituent attached to the fullerene surface. Angles of two or more substituents relative to each other can also be modulated by the positioning on the surface, as well as the nature of the substituent. Another way to modulate the substituent distances and angles is through using higher fullerenes such as C70, C84, etc.

Fullerene derivatives are roughly the right size to be drug candidates. An analysis of Indinavir and Retinovir, two clinically used HIV protease inhibitors, shows each has approximately 500Å^2 of surface area, with ~15% of this being heteroatomic. This corresponds closely with the analyzed fullerene derivatives, which range in surface from 400-600 Å^2 and heteroatomic percentages from 1-12%.

A range of studies of the interaction of fullerenes in biological systems has not turned up major signs of acute toxicity or carcinogenicity. These issues will have to be fully examined in the future, before introducing fullerene based pharmaceuticals into human use. Because only a fraction of the total surface in typical fullerene pharmaceutical applications will be typically used for presentation of binding elements, it may be possible to modify surface reactivity, toxicity and carcinogenicity by introducing modifications onto non-binding portions of the surface, for example to alter the redox behavior of the molecule.

The key to accessing these advantages of the fullerene core will be the development of regiospecific synthetic methodology so that the key groups of the pharmacophore can be properly positioned on the fullerene surface. The strategy described in Chapter 4 begins to address this issue. It is my hope that what started as the solutions to a specific problem may prove useful avenues to pursue in response to a variety of molecular design problems.

References

(1) Wood, R. W.; Collier, A. Acquired Immunodeficiency Syndrome Infectious Disease Clinics of North America **1987**, *1*, 145-163.

(2) Hirsch, M. S.; D'Aquila, R. T. Therapy For Human Immunodeficiency Virus Infection New England Journal of Medicine **1993**, *328*, 1686-95.

(3) Debouck, C. The HIV-1 protease as a therapeutic target for AIDS Aids Research and Human Retroviruses **1992**, *8*, 153-64.

(4) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease Science **1989**, *245*, 616-21.

(5) Fersht, A. *Enzyme Structure and Mechanism*; W.H. Freeman: New York, 1985.

(6) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking With Grid-Based Energy Evaluation Journal of Computational Chemistry **1992**, *13*, 505-524.

(7) Gschwend, D. A.; Kuntz, I. D. Orientational Sampling and Rigid-Body Minimization in Molecular Docking Revisited - On-The-Fly Optimization and Degeneracy Removal Journal of Computer-Aided Molecular Design **1996**, *10*, 123-132.

(8) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. The MIDAS Display System Journal of Molecular Graphics **1988**, *6*, 13-27.

(9) Molecular Modeling System SYBYL, Version 6.2, Tripos Associates, Inc.

(10) Kuntz, I. D. Structure Based Strategies For Drug Design and Discovery Science **1992**, *257*, 1078-82.

(11) Isaacs, L.; Haldimann, R. F.; Diederich, F. Tether-Directed Remote Functionalization of Buckminsterfullerene-Regiospecific Hexaadduct Formation Angewandte Chemie-International Edition in English **1994**, *33*, 2339-2342.

(12) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A., H.; Teller, E. Equation of State Calculations by Fast Computing Machines Journal of Chemical Physics **1953**, *21*, 1087-92.

(13) Woolley, D. W. Antimetabolites Annals of the New York Academy of Sciences **1950**, *52*, 1197-1378.

(14) Liu, S. Z.; Ju, Y. J.; Kappes, M. M.; Ibers, J. A. The Structure Of The C60 Molecule - X-Ray Crystal Structure Of A Twin At 110-K Science **1991**, *254*, 408-410.

(15) Tunon, I.; Silla, E.; Pascual-Ahuir, J. L. Molecular Surface Area and Hydrophobic Effect Prot. Eng. **1992**, *5*, 715-716.

(16) Novotny, J.; Bruccoleri, R. E.; Saul, F. A. On the attribution of binding energy in antigen-body complexes McPC 603, D1.3, and HyHEL-5. Bioc. **1989**, *28*, 4735-49.

(17) Sijbesma, R.; Srdanov, G.; Wudl, F.; Castoro, J. A.; Wilkins, C.; Friedman, S. H.; DeCamp, D. L.; Kenyon, G. Synthesis of a Fullerene Derivative for the Inhibition of HIV Enzymes Journal of the American Chemical Society **1993**, *115*, 6510-6512.

(18) DesJarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Furth, P. S.; Alvarez, J. C.; Ortiz de Montellano, P. R.; DeCamp, D. L.; Babe, L. M.; Craik, C. S. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease Proc.Natl.Acad.Sci. USA **1990**, *87*, 6644-6648.

(19) Thaisrivongs, S.; Tomasselli, A. G.; Moon, J. B.; Hui, J.; McQuade, T. J.; Turner, S. R.; Strohbach, J. W.; Howe, W. J.; Tarpley, W. G.; Heinrikson, R. L. Inhibitors of the protease from human immunodeficiency virus: design and modeling of a compound containing a dihydroxyethylene isostere insert with high binding affinity and effective antiviral activity. Journal of Medicinal Chemistry **1991**, *34*, 2344-2356.

(20) Ikeda, S.; Ashley, J. A.; Wirsching, P.; Janda, K. D. HIV protease inhibitor J. Am. Chem. Soc. **1992**, *114*, 7604-7606.

(21) DeCamp, D. L.; Babe, L. M.; Salto, R.; Lucich, J. L.; Koo, M. S.; Kahl, S. B.; Craik, C. S. Specific inhibition of HIV-1 protease by bornonated porphyrins J. Med. Chem. **1992**, *35*, 3426-3428.

(22) Woon, T. C.; Brinkworth, R. I.; Fairlie, D. P. Inhibition of HIV-1 Proteinase By Metal Ions Int. J. Biochem. **1992**, *24*, 911-914.

(23) Zhang, Z. Y.; Reardon, I. M.; Hui, J. O.; O'Connell, K. L.; Poorman, R. A.; Tomasselli, A. G.; Heinrikson, R. L. Zinc Inhibition of Renin and the Protease from Human Immunodeficiency Virus Type 1 Biochem. **1991**, *30*, 8717-8721.

(24) Santi, D. V.; Pena, V. A. "amine salt linkage" J. Med. Chem. **1973**, *16*, 273-280.

(25) Hirsch, A.; Li, Q. Y.; Wudl, F. Globe trotting hydrogens on the surface of the fullerene compound C60H6(N(CH2CH2)2O)6. Angewandte Chemie-International Edition in English **1991**, *30*, 1309-1310.

(26) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. Reconciling the Magnitude of the Microscopic and Macroscopic Hydrophobic Effects Science **1991**, *252*, 106-109.

(27) Friedman, S. H.; DeCamp, D. L.; Sijbesma, R. P.; Srdanov, G.; Wudl, F.; Kenyon, G. L. Inhibition of the HIV-1 Protease by Fullerene Derivatives- Model Building Studies and Experimental Verification Journal of the American Chemical Society **1993**, *115*, 6506-6509.

(28) Michael Connolly, University of California, San Francisco

(29) Several conformation of this compound were originally used for docking in selecting it as a synthetic target, all of them chair conformations. The major experimentally observed conformation is a boat however. The hydrophobic desolvation indicated is the value of desolvation for this boat structure, a value that is somewhat smaller than the originally determined value, but still a large enough improvement over compound **1** to warrant testing.

(30) Ganapathi, P. S.; Friedman, S. H.; Kenyon, G. L.; Rubin, Y. Sequential Double-Micahel Additions of Dienolates With C-C60 - Rapid Access to Sterically Congested Buckminsterfullerene Derivatives With Defined Stereochemistry Journal of Organic Chemistry **1995**, *60*, 2954-2955.

(31) Bingel, C. Cyclopropanation of Fullerenes Chemische Berichte **1993**, *126*, 1957-1959.

(32) Rose, J. R.; Salto, R.; Craik, C. S. Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. Journal of Biological Chemistry **1993**, *268*, 11939-45.

(33) An, Y. Z.; Anderson, J. L.; Rubin, Y. Synthesis of Alpha-Amino Acid Derivatives of C60 From 1,9-(4-Hydroxycyclohexano)-Buckminsterfullerene. Journal of Organic Chemistry **1993**, *58*, 4799-4801.

(34) Balch, A. L.; Lee, J. W.; Olmstead, M. M. Structure of a Product of Double Addition to C70-[C70(IR(CO)CL(PPHME2)2)2].3C6H6. Angewandte Chemie-International Edition in English **1992**, *31*, 1356-1358.

(35) Darke, P. L.; Jordan, S. P.; Hall, D. L.; Zugay, J. A.; Shafer, J. A.; Kuo, L. C. Dissociation and Association of the HIV-1 Protease Dimer Subunits: Equilibria and Rates Biochemistry **1994**, *33*, 98-105.

(36) Jordan, S. P.; Zugay, J.; Darke, P. L.; Kuo, L. C. Activity and Dimerization of Human Immunodeficiency Virus Protease as a Function of Solvent Composition and Enzyme Concentration Journal of Biological Chemistry **1992**, *267*, 20028-20032.

(37) Kuzmic, P.; Peranteau, A. G.; Garcia-Echeverria, C.; Rich, D. H. Mechanical Effects on the Kinetics of the HIV Proteinase Deactivation Biochemical and Biophysical Research Communications **1996**, *221*, 313-317.

(38) Lightstone, F. C.; Bruice, T. C. Geminal-Dialkyl Substitution, Intramolecular Reactions, and Enzyme Efficiency Journal of the American Chemical Society **1994**, *116*, 10789-10790.

(39) Hirsch, A.; Lamparth, I.; Karfunkel, H. R. Fullerene Chemistry in Three Dimensions-Isolation of Seven Regioisomeric Bisadducts and Chiral Trisadducts of C-60 and Di (Ethoxycarbonyl)Methylene Angewandete Chemie-International Edition in English **1994**, *33*, 437-438.

(40) Lightstone, F. C.; Bruice, T. C. Ground State Conformations and Entropic and Enthalpic Factors in the Efficiency of Intramolecular and Enzymatic Reactions .1. Cyclic Anhydride Formation by Substituted Glutarates, Succinate, ... Journal of the American Chemical Society **1996**, *118*, 2595-2605.

(41) Molecular mechanics energies were generated using the Tripos force field, Gasteiger-Hückel partial charges and a distance dependent dielectric.

(42) The angle constraint is based on a pseudo-dihedral formed between the reactive diene carbon, the centroid connecting the two diene reactive carbons, the centroid connecting the two reactive 6-6 bond carbons, and the reactive 6-6 bond carbon. This dihedral is limited to plus or minus 30°

(43) The designations of subscripts a and b for each labeled 6-6 bond are to indicate the manner in which the diene adds to it. The a isomer is the one in which the reactive diene carbon that is adjacent to the carbon which is

attached to the tether, is within the cut-off distance from the carbon in the 6-6 bond which is closest through space to the cyclopropyl base (or to the carbonyl carbon in the tether in cases where the two carbons in the 6-6 bond are equidistant from the cyclopropyl base.)

(44) Hirsch, A.; Lamparth, I.; Grosser, T.; Karfunkel, H. R. Regiochemistry of Multiple Additions to the Fullerene Core- Synthesis of a T-H-Symmetric Hexakisadduct of C-60 with Bis (Ethoxycarbonyl) Methylene Journal of the American Chemical Society **1994**, *116*, 9385-9386.