

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Mimicry and analysis of convergent protein interactions

**Permalink**

<https://escholarship.org/uc/item/36q44999>

**Author**

DeLano, Warren Lyford

**Publication Date**

1999

Peer reviewed|Thesis/dissertation

# Mimicry and Analysis of Convergent Protein Interactions

by

Warren Lyford DeLano

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

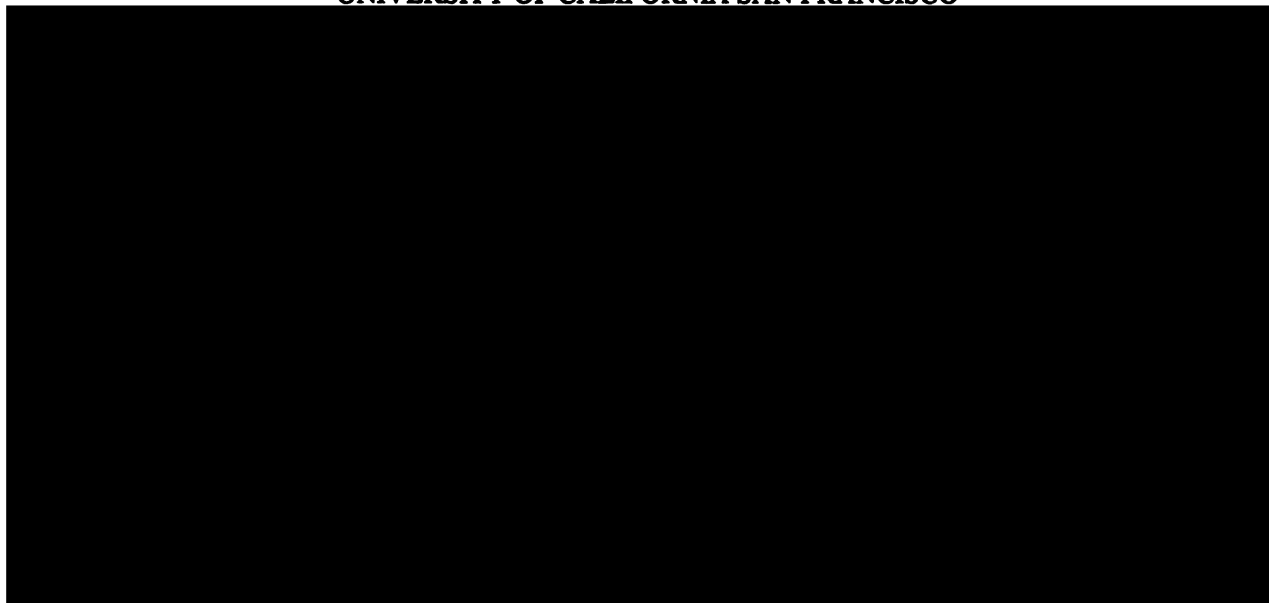
Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA SAN FRANCISCO



Date

University Librarian

Degree Conferred: .....

Copyright ©1999

by

Warren Lyford DeLano

**This thesis is dedicated to my wife  
Elizabeth Susan Pehrson**

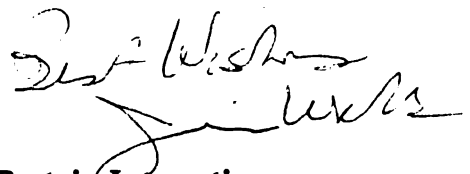
## Preface

I wish to acknowledge my graduate advisor, James Wells, for all of his encouragement, ideas, and support during the past four and a half years. Jim has a contagious sense of enthusiasm and a positive vision that infects everyone around him. The opportunity to work with Jim and his crew at Genentech and Sunesis has been a great honor, and it has provided me with a chance to follow my interests almost without practical constraint.

I also wish to acknowledge two of my coworkers, Andrew Braisted and Brian Cunningham, who together taught me much of what I know in the wet lab. Their remarkable efficiency, “can-do” attitude, and relentless pursuit of results have been particularly inspirational during my graduate career.

Credit must also be given to Julie Ransom, administrator of the Biophysics Graduate Group, who many times helped me to navigate my way through graduate school. My remote locations at Genentech and Sunesis would have much more difficult to handle without her assistance, and I am grateful for it.

Finally, I would like to recognize my undergraduate and post-undergraduate advisor, Axel Brünger, who brought me into the realm of science from the field of computer programming. His influence led me to apply my programming skills in a very interesting and rewarding direction, and it is my experiences working with him over the past decade that are now leading me to follow up a career developing computational systems to address important problems in biology.



## Mimicry and Analysis of Convergent Protein Interactions

by

Warren Lyford DeLano

### Abstract

A diverse assortment of natural proteins target a common site on the Fc fragment of immunoglobulin G, between the C<sub>H2</sub> and C<sub>H3</sub> domains. Alanine scanning of the protein A:Fc binding interface shows that this consensus binding site overlaps an energetic "hot spot." We sought to develop minimal molecules capable of mimicking the hot spot interactions. However, rational development of minimized protein A variants capable of binding tightly at this site proved difficult. Instead, novel peptides engineered to bind Fc using phage display were found to specifically target this consensus binding site with high affinity ( $K_d=15$  nM), in preference to any other site on the Fc surface. The structure of a representative peptide in complex with Fc reveals that these  $\beta$ -hairpin peptides make a number of convergent interactions directly analogous to those found in the Fc binding interfaces of natural proteins. The high affinity for Fc of this 1500 molecular weight peptide demonstrates that protein interfaces can be successfully targeted by smaller molecules capable of mimicking protein interactions using distinct structural scaffolds. Computational analysis reveals that this site may have been targeted by all of these molecules for its high degree of exposed non-polar surface area and for the inherent flexibility of an inter-domain hinge region. Further analysis of this consensus region should help in the identification of potent binding sites on protein surfaces.

## Table of Contents

<b>Copyright</b>	<b>ii</b>
<b>Dedication</b>	<b>iii</b>
<b>Preface</b>	<b>iv</b>
<b>Abstract</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Why Study Molecular Interactions? . . . . .	2
1.1.1 How is Affinity Achieved? . . . . .	3
1.1.2 How is Specificity Achieved? . . . . .	4
1.1.3 How is Cross-Reactivity Achieved? . . . . .	5
1.2 Protein-Protein Interfaces . . . . .	5
1.2.1 Statistical Observations . . . . .	5
1.2.2 Mutagenic Studies . . . . .	6
1.2.3 Additivity . . . . .	7
1.2.4 Energetic Hot Spots . . . . .	8
1.3 Moving Towards Therapeutics . . . . .	9
1.3.1 Therapeutic Proteins . . . . .	10
1.3.2 Therapeutic Small Molecules . . . . .	10
1.4 Focus 1: Targeting the Fc Fragment of Immunoglobulin G . . . . .	12
1.4.1 Minimizing an Fc Binding Domain from Protein A . . . . .	14
1.4.2 Engineering a Novel Fc Binding Peptide . . . . .	15
1.4.3 Understanding Functional Mimicry . . . . .	15
1.5 Focus 2: Development of Tools for Analyzing Protein Interfaces . .	16
1.5.1 Analysis of Surface Properties . . . . .	16
1.5.2 Study and Visualization of Structural Adaptability . . . . .	16
1.6 Conclusion . . . . .	18
<b>2 Dissection and Minimization of a Natural IgG-Fc Binding Domain</b>	<b>19</b>
2.1 Introduction . . . . .	19
2.2 Background: Structure of the Protein A Fc Binding Domain . . . . .	21
2.2.1 Z-Domain, A Functional Analogue of Protein A: B Domain .	22

2.2.2	Protein A:Fc Contact Epitope . . . . .	22
2.3	Alanine Scanning of the Protein A:Fc Interface . . . . .	25
2.3.1	Methods . . . . .	25
2.3.2	Results and Discussion . . . . .	29
2.4	Analysis of Two-helix Variants Based on Protein A . . . . .	33
2.4.1	Methods . . . . .	33
2.4.2	Results and Discussion . . . . .	34
2.5	Analysis of a Single Helix Protein A Mimic . . . . .	37
2.5.1	Methods . . . . .	38
2.5.2	Results and Discussion . . . . .	39
2.6	Conclusion . . . . .	47
2.7	Acknowledgements . . . . .	49
<b>3</b>	<b>Construction of a Non-natural Minimal Fc Binding Domain</b>	<b>50</b>
3.1	Introduction . . . . .	50
3.2	Natural Fc Binding Domains . . . . .	50
3.2.1	Protein A . . . . .	51
3.2.2	Protein G . . . . .	53
3.2.3	Rheumatoid Factor . . . . .	54
3.2.4	Fc Receptor . . . . .	55
3.2.5	Protein H . . . . .	56
3.2.6	Identification of a Consensus Binding Site . . . . .	56
3.3	Initial Isolation of an Fc Binding Peptide . . . . .	58
3.3.1	Methods . . . . .	58
3.3.2	Results and Discussion . . . . .	59
3.4	Peptide Characterization . . . . .	61
3.4.1	Methods . . . . .	61
3.4.2	Results and Discussion . . . . .	62
3.5	Local Optimization . . . . .	65
3.5.1	Methods . . . . .	65
3.5.2	Results and Discussion . . . . .	67
3.6	Global Optimization . . . . .	70
3.6.1	Methods . . . . .	71
3.6.2	Results and Discussion . . . . .	73
3.7	Alanine Scanning . . . . .	74
3.7.1	Methods . . . . .	75
3.7.2	Results and Discussion . . . . .	77
3.8	Further Minimization . . . . .	80
3.8.1	Methods . . . . .	80
3.8.2	Results and Discussion . . . . .	82
3.9	Conclusions . . . . .	84
3.10	Acknowledgements . . . . .	86



<b>4</b>	<b>Crystal Structure of an Engineered Peptide in Complex with Fc</b>	<b>87</b>
4.1	Introduction . . . . .	87
4.2	Methods . . . . .	87
4.2.1	Crystal Growth . . . . .	87
4.2.2	Data Collection and Processing . . . . .	88
4.2.3	Phasing: Molecular Replacement . . . . .	88
4.2.4	Initial Refinement and Building . . . . .	89
4.2.5	Refinement . . . . .	92
4.3	Results and Discussion . . . . .	93
4.3.1	Overall Structure . . . . .	93
4.3.2	Peptide Secondary and Tertiary Structure . . . . .	94
4.3.3	Fc Binding Interactions . . . . .	94
4.3.4	Comparison with Other Fc Binding Interfaces . . . . .	97
4.3.5	Crystal Packing . . . . .	100
4.3.6	Hydration . . . . .	104
4.4	Conclusions . . . . .	105
4.5	Acknowledgements . . . . .	105
<b>5</b>	<b>SiteFinder: A Tool for Comparing Protein Surface Patches</b>	<b>106</b>
5.1	Introduction . . . . .	106
5.2	Software Development . . . . .	107
5.2.1	Algorithm . . . . .	107
5.2.2	Implementation and Testing . . . . .	112
5.3	Application to the Fc Surface . . . . .	113
5.4	Conclusions . . . . .	120
<b>6</b>	<b>RigiMOL: A Tool for Analyzing and Visualizing Structural Changes</b>	<b>122</b>
6.1	Introduction . . . . .	122
6.1.1	Why RigiMOL? . . . . .	124
6.1.2	Background . . . . .	125
6.2	Software Development . . . . .	127
6.2.1	Algorithm . . . . .	127
6.2.2	Implementation . . . . .	132
6.3	Demonstration . . . . .	133
6.4	Applications . . . . .	138
6.4.1	Visualization of Global Conformational Changes in Fc . . .	138
6.4.2	Visualization of Adaptability in the Fc Binding Interface . .	140
6.4.3	Study of Enzyme Mechanisms . . . . .	140
6.4.4	Study of Protein/Nucleic Acid Complexes . . . . .	142
6.5	RigiMOL's Limitations . . . . .	142
6.6	Conclusion . . . . .	143

## 7 Conclusion

7.1 Accomplishments

7.2 What Was Learned

7.2.1 New Concepts

7.2.2 Applications

7.2.3 Future Work

7.3 What Needs to be Done

7.3.1 Present Status

7.3.2 Immediate Needs

7.3.3 Long-Term Needs

7.3.4 Recommendations

7.4 In Closing

## A Theoretical Considerations

A.1 Introduction

A.2 Binding

A.2.1 Equilibrium

A.2.2 Kinetics

A.2.3 Thermodynamics

A.3 Direct Binding

A.3.1 Experimental Methods

A.3.2 Data Analysis

A.3.3 Interpretation

A.4 Competitive Binding

A.4.1 Experimental Methods

A.4.2 Data Analysis

A.4.3 Interpretation

A.4.4 Applications

A.4.5 Summary

## B Laboratory Procedures

B.1 Cleavage

B.2 Preparation

B.3 Titration

B.4 Preparation

B.5 Preparation

B.6 Affinity

B.7 Phage

B.8 ELISA

B.9 Protein

<b>7</b>	<b>Conclusion</b>	<b>144</b>
7.1	Accomplishments . . . . .	144
7.2	What Was Learned? . . . . .	145
7.2.1	Molecular Engineering Requires Robust Technologies . . . . .	145
7.2.2	Attractive Sites Exist on Protein Surfaces . . . . .	146
7.2.3	Function Can Be Transferred Across Molecular Scaffolds . . . . .	147
7.3	What Next? . . . . .	147
7.3.1	Practical Applications of the Fc Binding Peptides . . . . .	147
7.3.2	Implications for Small Molecule Design . . . . .	148
7.3.3	Identification of Potential Binding Sites . . . . .	148
7.3.4	Analysis of Structural Movement in Proteins . . . . .	149
7.4	In Closing . . . . .	149
<b>A</b>	<b>Theoretical Guide to the Practice of Binding Measurements</b>	<b>150</b>
A.1	Introduction . . . . .	150
A.2	Binding Equilibria . . . . .	152
A.2.1	Biomolecular Systems . . . . .	152
A.2.2	Trimolecular Systems . . . . .	157
A.2.3	Trimolecular Systems (Special Case $K = K_B = K_C$ ) . . . . .	161
A.3	Direct Binding Assays: Two Component Systems . . . . .	163
A.3.1	Dose-Response . . . . .	163
A.3.2	Scatchard Analysis . . . . .	167
A.3.3	Direct Fitting of Titration Curves . . . . .	170
A.4	Competition Binding Assays: Three Component Systems . . . . .	172
A.4.1	Assays with Labeled Target ( $EC_{50}$ 's) . . . . .	172
A.4.2	Assays with Labeled Ligand ( $IC_{50}$ 's) . . . . .	181
A.4.3	Fitting Curves from Competition Assays . . . . .	194
A.4.4	Estimating $EC_{50}$ 's and $IC_{50}$ 's from Single Point Data . . . . .	200
A.4.5	Predicting Saturation Levels at Alternate Concentrations . . . . .	201
<b>B</b>	<b>Laboratory Protocols</b>	<b>202</b>
B.1	Cleavage and Purification of Fc' from IgG <sub>1</sub> fusion protein . . . . .	202
B.2	Preparation of M13 Bacteriophage . . . . .	203
B.3	Titration of M13 Bacteriophage . . . . .	204
B.4	Preparation of Single Strand Template DNA . . . . .	205
B.5	Preparation of M13 Bacteriophage Display Libraries . . . . .	206
B.6	Affinity Panning of M13 Bacteriophage Libraries . . . . .	207
B.7	Phage-ELISA Binding Assay ( $EC_{50}$ 's) . . . . .	209
B.8	ELISA Inhibition Binding Assay ( $IC_{50}$ 's) . . . . .	211
B.9	Protein Expression from a Phagemid Vector . . . . .	213

<b>C Contents of the CD-ROM</b>	<b>214</b>
C.1 Introduction . . . . .	214
C.2 Contents . . . . .	214
C.2.1 RigiMOL Movies . . . . .	215
C.2.2 Thesis in PDF Format . . . . .	215
C.2.3 Thesis in Postscript Format . . . . .	215
<b>References</b>	<b>216</b>

## List of Tables

2.1	Z-Domain Side Chain Contacts with Fc . . . . .	24
2.2	Lost Interactions on Helix II . . . . .	44
2.3	Comparison of Z-Domain Variants . . . . .	47
4.1	Crystallization Conditions . . . . .	88
4.2	Data Collection and Processing Statistics . . . . .	89
4.3	Final Refinement Statistics . . . . .	92
4.4	Comparison of Fc Binding Domains . . . . .	100
5.1	Consensus Contacts on Fc . . . . .	114

## List of Figures

1.1	Additivity of Mutations . . . . .	8
1.2	Alanine Scanning Hot Spots . . . . .	9
1.3	Immunoglobulin G Structure . . . . .	12
1.4	Fc Binding Proteins . . . . .	13
2.1	Crystal Structure of the Protein A Fc Binding Domain . . . . .	21
2.2	Protein A:Fc Contact Epitope . . . . .	23
2.3	Z-domain Gene and Protein Sequences . . . . .	25
2.4	Alanine Scan of the Protein A:Fc Interface . . . . .	31
2.5	Hot Spots in the Protein A:Fc Interface . . . . .	32
2.6	Z-Domain Variants . . . . .	33
2.7	Crystal Structure of a Two-Helix Z-domain Variant . . . . .	34
2.8	Alanine Scan of Three Protein A Variants . . . . .	35
2.9	Helix Locking Chemistry . . . . .	37
2.10	Locked-Helix CD Spectrum . . . . .	39
2.11	Crystal Structure of a Single Helix Mimic of Protein A . . . . .	40
2.12	Structural Mimicry . . . . .	41
2.13	Hot Spot Expansion . . . . .	46
3.1	Natural Fc Binding Domains . . . . .	51
3.2	Crystal Structure of an EPO/Peptide Complex . . . . .	52
3.3	Protein A . . . . .	53
3.4	Protein G . . . . .	54
3.5	Rheumatoid Factor . . . . .	55
3.6	Fc Receptor . . . . .	56
3.7	Consensus Binding Site . . . . .	57
3.8	Naive Peptide Library . . . . .	58
3.9	Initial Fc Binding Sequences . . . . .	60
3.10	Peptide Inhibition Curves . . . . .	63
3.11	Circular Dichroism Spectrum of a Selected Peptide . . . . .	64
3.12	Design of Local Optimization Libraries . . . . .	66
3.13	Screening of the Local Optimization Libraries . . . . .	67
3.14	Selected Sequences from Local Optimization . . . . .	68
3.15	Residue Distribution from Local Optimization . . . . .	70
3.16	Design of the Global Optimization Libraries . . . . .	72

3.17	Selected Sequences from Global Optimization . . . . .	74
3.18	Comparison of Protein A and Peptide Hot Spots . . . . .	78
3.19	Alanine Scan of an Fc Binding Peptide . . . . .	79
3.20	Minimized Fc Binding Peptide . . . . .	80
3.21	Fc Binding and pH Dependence of FBP4.1 . . . . .	83
3.22	$K_i$ Determination for FBP4.1 . . . . .	84
3.23	Kinetics of FBP4.1 Binding . . . . .	85
3.24	Steady State FBP4.1 Binding . . . . .	86
4.1	Initial Peptide Density . . . . .	90
4.2	Refinement . . . . .	91
4.3	Structure of the Fc:Peptide Complex . . . . .	93
4.4	Fc:Peptide Binding Interactions . . . . .	95
4.5	Secondary Structure of the Peptide . . . . .	96
4.6	Tertiary Structure of the Peptide . . . . .	96
4.7	Polar Interactions in the Fc:Peptide Interface . . . . .	98
4.8	Footprints on Fc . . . . .	99
4.9	Conserved Fc Interaction Map . . . . .	101
4.10	Conserved Interactions in the Fc Binding Interfaces . . . . .	102
4.11	Crystal Packing Interactions in the Fc:Peptide Complex . . . . .	103
4.12	Hydration of the Fc:Peptide Complex . . . . .	104
5.1	SiteFinder Overview . . . . .	108
5.2	SiteFinder's Solvent Accessible Surface Representation . . . . .	110
5.3	Example Surface Patches . . . . .	113
5.4	Accessibility of the Fc Surface . . . . .	115
5.5	Polarity of the Fc Surface . . . . .	117
5.6	Planarity of the Fc Surface . . . . .	118
5.7	SiteFinder Maps of the Fc Surface . . . . .	119
6.1	Cartesian Versus Conceptual Descriptions of Motions . . . . .	123
6.2	RigiMOL Overview . . . . .	128
6.3	Rigimol Domains . . . . .	130
6.4	Representation of a 3-D Transformation . . . . .	131
6.5	Interpolation Comparison . . . . .	133
6.6	Example RigiMOL Script . . . . .	134
6.7	Example Domain List . . . . .	135
6.8	Example Domain Fit List . . . . .	135
6.9	Example Domain Rotations and Translations . . . . .	136
6.10	Interpolation Demonstration . . . . .	137
6.11	Global Adaptability in the Fc Dimer . . . . .	139
6.12	Adaptability in the Fc Consensus Binding Site . . . . .	141
A.1	Dose Response Measurements . . . . .	165

A.2	Limitations of Dose-Response Measurements . . . . .	165
A.3	Dose-Response as a Function of Ligand Affinity . . . . .	166
A.4	Scatchard Plot . . . . .	169
A.5	An Alternate Scatchard Plot . . . . .	169
A.6	Direct Fitting of Titration Curves . . . . .	171
A.7	Assays with Labeled Target ( $EC_{50}$ 's) . . . . .	172
A.8	Effects of Target Concentration and Ligand Saturation . . . . .	177
A.9	Dependence of the $EC_{50}$ on Ligand Saturation . . . . .	178
A.10	Dependence of the $EC_{50}$ on Ligand Affinity . . . . .	179
A.11	Dependence of the $EC_{50}$ on Ligand Concentration . . . . .	179
A.12	Bottoming Out of an $EC_{50}$ Competition Assay . . . . .	180
A.13	Limitations of $EC_{50}$ Competition Assays . . . . .	181
A.14	Assays with Labeled Ligand ( $IC_{50}$ 's) . . . . .	182
A.15	Effects of Ligand Concentration and Target Saturation . . . . .	187
A.16	Dependence of the $IC_{50}$ on the Fractional Target Saturation . . . . .	187
A.17	Dependence of the $IC_{50}$ on the Ligand Affinity . . . . .	188
A.18	Dependence of the $IC_{50}$ on the Total Target Concentration . . . . .	189
A.19	Bottoming Out of an $IC_{50}$ Competition Assay . . . . .	189
A.20	Limitations of $IC_{50}$ Binding Assays . . . . .	190
A.21	Measurements Under Saturating Conditions: Target Concentration	192
A.22	Measurements Under Saturating Conditions: Target Saturation . .	193



## **Chapter 1**

### **Introduction**

This is an exciting time to be a biologist. In recent decades, scientists have found the keys to unlock many of life's fundamental mysteries through technologies such as DNA sequencing and high resolution structure determination. The process of life need no longer be viewed as magical or somehow incomprehensible, explainable only in terms of gods or mystical legends. Instead, it is apparent that life is a physical process just like any other: it can be disassembled, it can be modified, and it can be understood.

Since the advent of recombinant DNA technology in the 1970's, awesome leaps in understanding have been made. For example, we now have a very good sense of how DNA and associated molecules act to control the cell, and we recognize the significance of many interactions inside the cell. We understand how nucleic acids, proteins, cells, and even entire organisms are replicated. We know how signals are conveyed, and we have grasped how motor proteins generate the motions we experience in our macroscopic world.

However, today's biologists also realize that we only scratching the surface of the true complexity that comprises life. Much of what we now understand in principle, we need to learn in detail. Two aspects of life currently give us the most trouble: (1) the tremendous amount of information we need to be able to

manipulate and interpret, and (2) the colossal interconnectedness that exists inside of living organisms. I believe that much work in the coming decade will be focused on addressing these two problems by developing and applying technologies which can operate on billions of components at once, for purposes of information gathering, analysis, and molecular design. Such technologies will be essential in order to reduce the complexity of life down to useful concepts which can then be rationally applied.

### **1.1 Why Study Molecular Interactions?**

Molecular interactions serve as the medium through which the interconnectedness inside living material is effected. The molecular interaction is the basic mode by which information is transmitted from one component to another. One molecule, A, binds a second molecule, B, and in so doing modulates the behavior of B. The change on B then affects the behavior of molecules C and D, and so on. Indeed, life can be viewed as a massively parallel game of hot potato, as various types of signals are rapidly passed around through a complicated network involving thousands of proteins and signals.

A primary goal of biomedical research is to develop an understanding of life and of diseases processes that will help us to prolong human life and alleviate suffering. With respect to molecular interactions, this means mapping all of the interactions inside of a cell so that we will be able to rationally make changes which will have predictable effects. For a complete mechanistic description of a protein-protein interaction, we need to know why two proteins come together, where they contact one another, how tight that interaction is, and we need to identify the downstream effects of the interaction.

At the heart of a molecular interaction is the set of atoms actually involved in binding. Since many important components in biology are macromolecules (DNA, protein, and carbohydrates), the study of macromolecular interactions is important to developing an understanding of natural process. Also of importance are interactions between a macromolecule and a much smaller molecule (such as calcium or a metabolite), since many signals are propagated through these interactions as well.

One fascinating aspect of living matter is that the closer you look, the more complicated it tends to get. Macromolecular interfaces are no exception. They directly involve hundreds of atoms and can respond to a variety of changes that occur on a protein, sometimes on the other side of a protein. They are dynamic in nature, existing as thermodynamic ensembles of many conformations in both the bound and unbound states. Conceptually, however, there are three identifiable aspects of molecular interfaces that directly relate to their function. These aspects are: (1) affinity, (2) specificity, and (3) cross-reactivity. They are discussed individually below:

### **1.1.1 How is Affinity Achieved?**

All spontaneous events in the universe occur through a decrease in free energy. For protein binding events to occur, they must be energetically favorable. This implies that the unbound state of two molecules must be higher in energy than the bound state. When thinking about any kind of molecular interaction, it is important to remember this and to recognize the driving force for association arises from the difference between these two energies, and not from the bound state in isolation.

Binding energy in non-covalent interactions arises from two fundamental phys-

ical sources: entropy and electrostatics (in the broad sense). The entropic contribution originates primarily from the differential interactions that water has with biological molecules versus other water molecules. The electrostatic component manifests itself in the form of short-range van der Waals interactions as well as a variety of Coulombic effects. Although still a topic of debate, affinity for macromolecular interactions probably originates mostly from the hydrophobic effect, as water molecules try to push less polar solutes out of solution so that they can form more hydrogen bonding interactions without incurring ordering that will lower their entropy.

### **1.1.2 How is Specificity Achieved?**

Although the driving force for macromolecular association arises from the relatively non-specific hydrophobic effect, macromolecules usually bind one other in a very specific and controlled manner. Life would be impossible if all proteins simply bound to one another indiscriminantly. In reality, macromolecules interact with very specific binding partners and do so at precise locations on their surfaces, and these associations are determined by the precise chemical and geometric pairing of interactions on opposing surfaces. For example, if a large region of a protein surface is going to be desolvated, the polar groups in that site will lose their interactions with solvent. This would be a prohibitively expensive proposition (as much as 5 kcal/mole per hydrogen bond, and as much as 70 kcal/mole per formal charge) if those groups were not somehow satisfied in the complex [Sharp, 1990]. Similarly, two macromolecules can not form a tight association if their surfaces are not complementary in shape or able to adapt so as to become complementary upon binding.

### **1.1.3 How is Cross-Reactivity Achieved?**

Affinity and specificity can be carefully balanced to achieve cross-reactive binding. Cross-reactive molecules exhibit specific binding interactions with multiple diverse binding partners using the same contact surface. Examples include cytokines such as human growth hormone [Cunningham, 1990], hetero-oligomeric cytokine receptor subunits (such as gp140) [Wells, 1996], and the receptor binding site on the Fc fragment of IgG [Stone, 1987, Burmeister, 1994a]. What is most amazing is the variety of structures to which these surfaces can bind, but without binding all molecules indiscriminately. Something about these surfaces makes multiple modes of binding possible.

## **1.2 Protein-Protein Interfaces**

Proteins are the most versatile biological molecules found in the cell, and protein-protein interfaces are involved in nearly every aspect of cellular function, from DNA transcription to cell adhesion. In this work, we focus on protein-protein interfaces and interfaces between proteins and peptides.

### **1.2.1 Statistical Observations**

Protein-protein interactions have been studied ever since crystal structures of protein complexes have been available [Chothia, 1975], and considerable effort has been devoted to identifying patterns which might be useful in identifying protein-protein interaction sites [Argos, 1988, Janin, 1990, Janin, 1996]. Unfortunately, universally applicable characteristics which could be used to reliably identify protein-protein recognition sites have not yet been found.

Much of the difficulty stems from the diversity of protein-protein complexes.

Interfaces can range in size from around 500 Å<sup>2</sup> to well over 2000 Å<sup>2</sup> [Jones, 1996]. They may have only a few polar interactions or they may have more than twenty [Conte, 1999]. They vary greatly from being largely hydrophobic to substantially hydrophilic [Tsai, 1997], and all kinds of protein structural elements had been found to be involved in binding [Bernstein, 1977].

The only regular successes which have been obtained in predicting protein recognition sites are for those sites which are associated with recognizable sequence or structure characteristics. Examples include binding sites of proteins with known functions, such as serine proteases [Dodson, 1998], poly-proline SH3 binding domains [Dalgarno, 1998], or coiled coil oligomerization regions [Lupas, 1997] as well as common modules such calcium [Heizmann, 1991] or nucleotide [Bellamacina, 1996] binding motifs.

It remains unclear whether there are any recognizable physical characteristics that will be common to all protein-protein interaction sites, or whether there are so many orthogonal solutions to binding protein surfaces that common properties simply do not exist.

### 1.2.2 Mutagenic Studies

Protein mutagenesis allows us to make a small change on a protein and to observe the effect. With respect to protein interfaces, mutagenic studies have enabled researchers to selective disrupt [Fuh, 1992] or enhance [Lowman, 1993] the natural affinities of binding partners.

In interpreting mutagenesis results, it is essential to remember that the free energy of binding is a system property and is not strictly localized in space or on a single residue. Indeed, all bound and unbound conformations of proteins have the potential to contribute to the binding energy.

RECEIVED  
NOV 27 1997

However, in general most mutations of protein surface residues have little effect on the stability of the unbound state [Sauer, 1996], and so we can usually interpret mutagenesis results in terms of the effect on the bound state. Thus, in a protein interface, a mutated residue that shows a substantial effect on the binding energy usually means one of two things – either that mutated residue has lost a stabilizing role on the complex, or it has introduced additional destabilization into the complex. This can be due to favorable interactions in the original interface that were lost in the mutant, or it can be due to new unfavorable interactions caused by the mutation.

### **1.2.3 Additivity**

A very important observation regarding protein mutations is that the effects of two mutations together will be approximately equal to the sum of the effects of the individual mutations (Figure 1.1) [Sandberg, 1993]. In other words, protein mutations are often additive in terms of their effects on the free energy. This is a remarkable result considering that there is no thermodynamic reason why proteins should necessarily behave that way, and it indicates that multiple perturbations on proteins generally affect the ensembles (and the underlying partition functions) in orthogonal manners. This energetic observation is supported by structural studies of mutant proteins which show that structural changes can also be additive [Skinner, 1996].

Additivity lends credence to the interpretation of mutational effects in terms of the specific interactions that surround individual residues, and to a notion of “localized” energy. Under this conceptual framework, energetically important interactions are those which have large effects when they are lost. Energetically inert interactions have little or no effect when they are eliminated. However, it

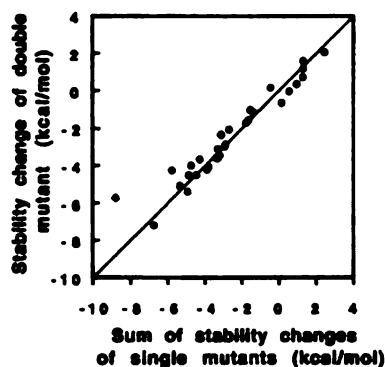


FIG. 1. Additivity of mutational effects on gene V protein stability. The stability change ( $\Delta\Delta G_{2,284}^{\circ}$ ), relative to the WT protein, of gene V protein double mutants is shown on the y axis. The x axis shows the sum of the stability changes, also relative to the WT protein, of the constituent single mutants. Positive values of  $\Delta\Delta G_{2,284}^{\circ}$  indicate proteins with increased stability. The combination of the mutants C33M and I47C is indicated by the diamond ( $\diamond$ ). A line with unit slope is shown for reference.

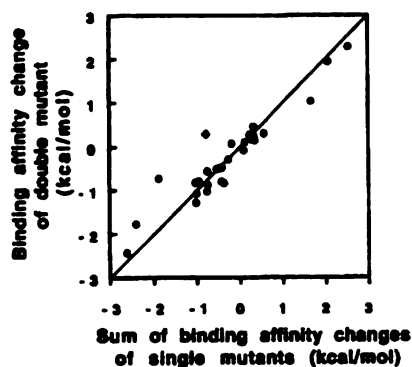


FIG. 2. Additivity of mutational effects on gene V protein DNA binding affinity. Binding affinity change ( $\Delta\Delta G_{2,154}^{\circ}$ ), relative to the WT protein, of gene V protein double mutants is shown on the y axis. The x axis shows the sum of the binding affinity changes, also relative to the WT protein, of the constituent single mutants. A positive value of  $\Delta\Delta G_{2,154}^{\circ}$  indicates enhanced binding to ssDNA relative to WT. The combination of the mutants C33M and I47C is indicated by the diamond ( $\diamond$ ). A line with unit slope is shown for reference.

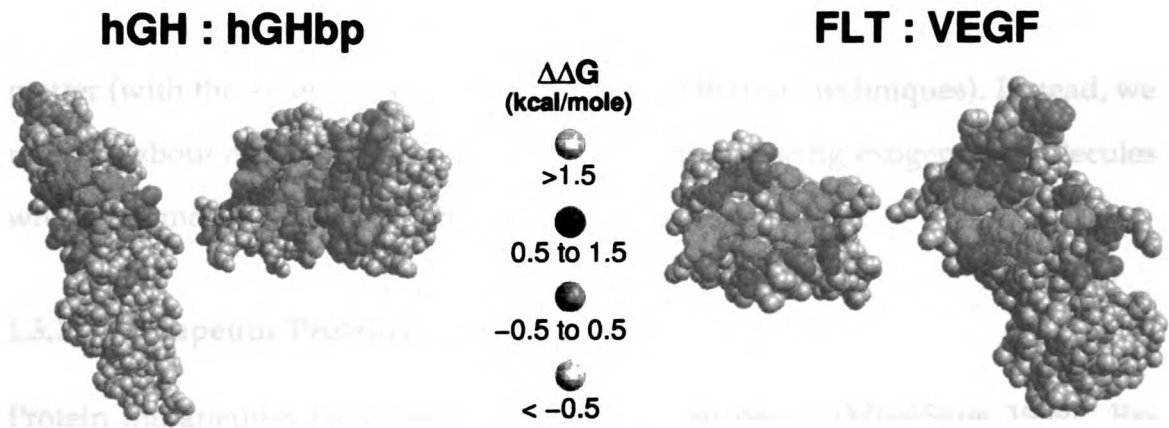
Figure 1.1: **Additivity of Mutations.** These plots (from [Sandberg, 1993]) show that mutational effects can be highly additive. The effects of combined mutations on binding and stability are often very close to the summed effects of the single mutations.

is important to keep in mind that this notion of energy is operational in that it applies to the effects of perturbations and does not directly correspond to the free energy of the system.

#### 1.2.4 Energetic Hot Spots

Systematic mutagenesis techniques have enabled us to locate important binding interactions in a variety of protein binding events. One key result from such work is that only a subset of the contacts in protein interfaces appear to be critical for binding [Cunningham, 1993]. These regions of critical residues are commonly referred to as “energetic hot spots” since in practice, changes made to them have large effects on binding energy. Two examples of hot spots in protein interfaces are shown in Figure 1.2. Other examples of known hot spots are reviewed in [Bogan, 1998].





**Figure 1.2: Alanine Scanning Hot Spots.** Left: Human Growth Hormone interacting with its receptor [Clackson, 1995]. Right: a domain of the FLT receptor interacting with the Vascular Endothelial Growth Factor (VEGF) [Cunningham, 1997]. Residues are color coded to reflect the disruptive effects of alanine substitution.

Experiments have shown that multiple affinity-inert residues can be removed from a protein interface while still preserving the vast majority of the binding energy [Jin, 1994, Pearce, 1996]. This led to a “hot spot” hypothesis which holds that essentially all of the binding energy in a protein interface is concentrated on hot spot residues and that the other parts of protein interface surfaces are only necessary to achieve specificity [Clackson, 1995].

### 1.3 Moving Towards Therapeutics

Since protein interfaces are in widespread use throughout the cell, considerable effort has been devoted to understanding how to modulate protein interactions. Ideally, we would like some way to up or down regulate any protein interaction (include enzyme-substrate interactions) at will. Equipped with such a method, we would be able to cure most ailments including diabetes, cancer, and infections. However, although we can modulate protein interactions in the laboratory

through mutagenesis, currently we have no general way of doing so in living matter (with the exception of some limited gene therapy techniques). Instead, we must go about altering protein interactions by introducing exogenous molecules which can modulate protein activity inside the organism.

### **1.3.1 Therapeutic Proteins**

Protein therapeutics have achieved promising successes [Mire-Sluis, 1999]. Examples include insulin, human growth hormone, erythropoietin, Herceptin, and DNAase. Protein therapeutics have the advantage of being highly specific (and thus less toxic) and are very easy to modify to achieve desired properties (thanks to incredible advances in protein engineering) [Russel, 1999]. The disadvantage of protein therapeutics are that they usually need to be delivered by injection, which can be expensive and inconvenient, and they can only target extracellular components. Until gene therapy becomes a robust technique for delivering proteins to a variety of locations in the body, protein therapeutics will only have limited applications.

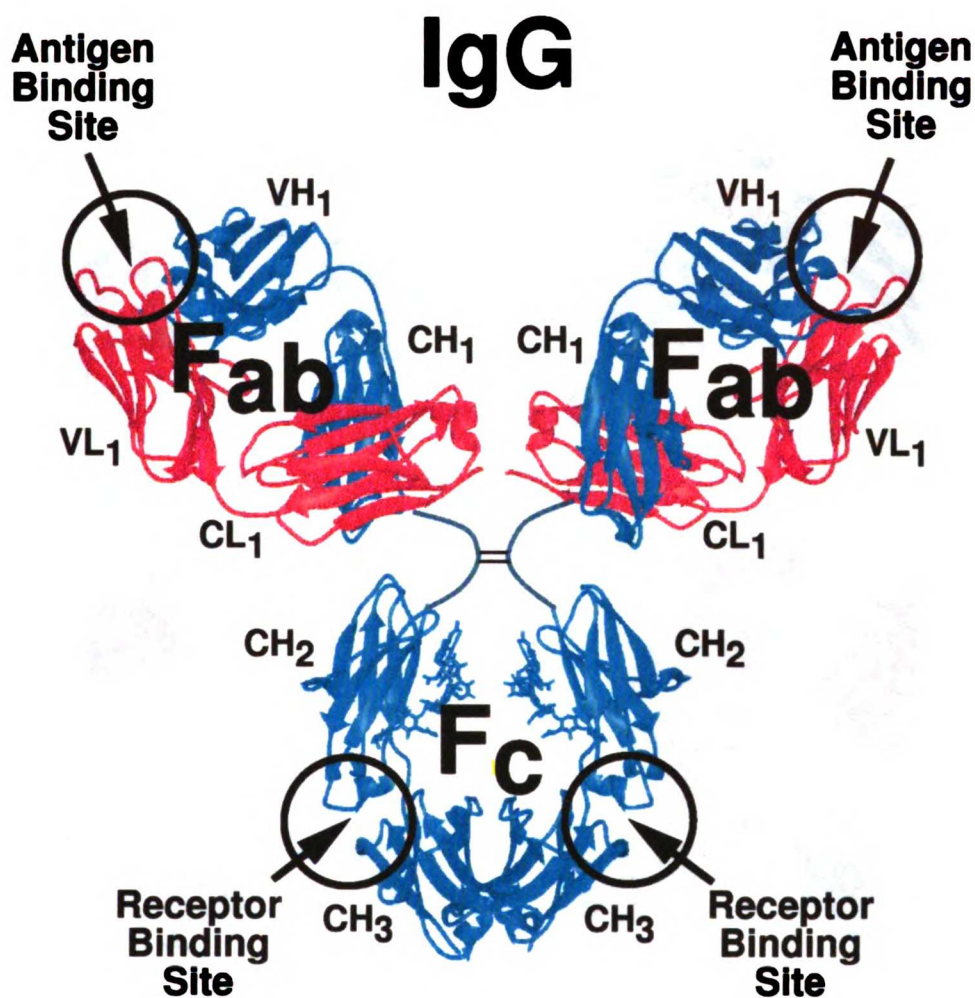
### **1.3.2 Therapeutic Small Molecules**

The gold standard for a therapeutic is the small (<500 m.w.), orally bioavailable drug. Many drugs in use today are based on compounds derived from natural sources such as plants and fungi, and these compounds possessed some therapeutic activity long before they were adopted for use by humans. Furthermore, most drugs target a very particular class of protein interface: the enzyme-substrate interaction. As a result, in therapeutic development we have primarily been limited to improving on substances that evolved naturally and have been only be able to modulate those protein interactions already targeted by such substances.

Ideally, we would like to be able to develop small molecules that target and modulate any kind of protein interaction, but where do we start? Medicinal chemistry techniques are only useful for improving upon existing drugs, and even combinatorial chemistry requires a starting scaffold with some base affinity.

An intriguing observation to arise from the discovery of hot spots on protein surfaces is that these hot spots are comparable in size to small molecule drugs. This led to the suggestion that one might be able to take a protein interface and reduce its functional essence down to a small molecule that mimics the hot spot interactions of the natural protein.

Indeed, promising results have been obtained for continuous binding motifs such as the RGD sequence which was converted into a small molecule with antithrombotic activity [McDowell, 1998a, McDowell, 1998b]. However, similar successes have not been achieved for hot spots involving non-continuous binding motifs such as those commonly found in protein interfaces.

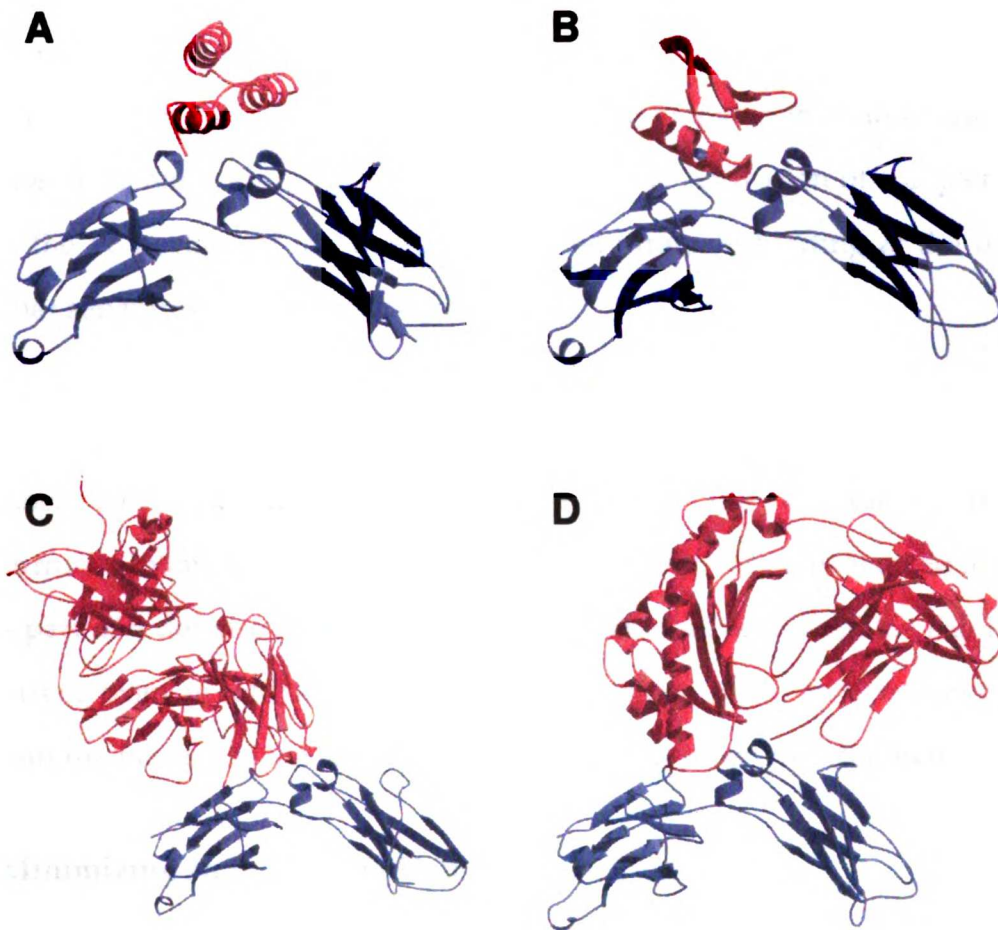


UCSF LIBRARY

Figure 1.3: **Structure of Immunoglobulin G.** This is a schematic model constructed from separate Fab and Fc crystal structures. No structure of the intact dimer yet exists. The fragments and domains are labelled. Arrows and circles indicate the location of the antigen and receptor binding sites. In this work, I focus exclusively on molecules which interact with the receptor binding sites.

#### 1.4 Focus 1: Targeting the Fc Fragment of Immunoglobulin G

The first focus of this thesis is on the development of smaller molecules (1500-2000 M.W.) which can mimic the activity of natural protein binding domains. The target is the receptor binding site on the constant domain (Fc) of Immunoglobulin G (Figure 1.3).



**Figure 1.4: Proteins which bind the Fc fragment of Immunoglobulin G.** One subunit of the Fc dimer is shown in blue, and the binding partners are shown in red. (A) Domain B of protein A [de Vos, 1998]; (B) Domain C2 of protein G [Sauer-Eriksson, 1995]; (C) Rheumatoid factor [Corper, 1997]; (D) Fc receptor [Burmeister, 1994a].

Immunoglobulin G plays a critical role in the immune system. Produced in response to stimulation of B cells by foreign antigens, these homing devices seek out epitopes on the surfaces of foreign proteins and latch on tightly. Because of their bivalent nature, IgG's can form crosslinked networks of antibodies and antigens, particularly symmetric ones. Sufficient concentrations of IgG then activate

the complement cascade, bringing about release of potent defense molecules that will destroy anything nearby.

Although IgG recognizes foreign antigens using loops on the Fab region, the rest of the molecule is also important for full activity. Activation of complement requires the Fc portion of the molecule and may involve binding at the hinge region that separates the Fab and Fc portions.

Fc possesses a remarkably cross-reactive site on the surface of each subunit. At least five natural proteins are known to bind Fc, and the crystal structures of four of these complexes have been solved (Figure 1.4). They are: the Fc binding domains from protein A and protein G, which have independently evolved in bacterial as part of a defense systems; rheumatoid factor, an auto-immune Fab fragment derived from an IgM; and the neonatal Fc receptor, involved in transporting IgG's from the mother to the fetus as part of immune system development.

#### **1.4.1 Minimizing an Fc Binding Domain from Protein A**

First we study the interaction between IgG-Fc and a three-helix binding domain derived from protein A. Systematic mutagenesis is used to map out the functional binding epitope on both sides of the interface and to identify the hot spot residues. Then, through a collaborative project with Andrew Braisted, a minimized two helix mimic of the original domain is created and studied. We find that the hot-spot is preserved intact in the two-helix variant, and with the introduction of a disulfide stabilizer can achieve comparable binding affinity. Next we examine a chemically stabilized single helix mimic which binds Fc but with greatly reduced affinity. A crystal structure (Mark Ultsch) reveals that the helix faithfully reproduces the binding contacts of the full domain, so we explore the likely reasons why the helix is unable to bind with high affinity.

### 1.4.2 Engineering a Novel Fc Binding Peptide

Given the failure of a single helix to bind Fc, we decide to attempt to create a new minimized Fc binding domain from scratch using phage display. This attempt is successful, and we find a 20 amino peptide that binds Fc with  $K_d = 5 \mu\text{M}$  affinity and is competitive with protein A. Two subsequent rounds of optimization give peptides that bind Fc with affinities nearly equal to that of protein A's Fc binding domain. Alanine scanning of the peptide reveals a highly conserved 13-residue sequence that is found to bind tightly ( $K_d=15 \text{ nM}$ ) on its own.

### 1.4.3 Understanding Functional Mimicry

A crystal structure of the minimized 13 residue peptide in complex with Fc shows that it targets the same binding site on Fc as do the natural Fc binding domains. Detailed inspection of the complex reveals a remarkable conservation of Fc binding interactions across the peptide and the natural binding domains. The peptide provides a convincing demonstration that the functional elements in a protein interface can be transferred to and presented from a much smaller structural scaffold, and it suggests that it may be possible to engineer small non-peptidic molecules which do so as well.

## **1.5 Focus 2: Development of Tools for Analyzing Protein Interfaces**

The second focus of this thesis is on development of software to improve our ability to understand protein interfaces. Computer programs were written in order to help elucidate how and why this particular binding site on the surface of Fc was targeted for binding by so many diverse molecules.

### **1.5.1 Analysis of Surface Properties**

The first program, SiteFinder, performs an analysis of properties on a protein surface. Applied to Fc, this program reveals that the consensus binding site shared by all of the natural Fc binding proteins as well as the engineered peptide is part of a large patch of highly accessible, non-polar atoms on the surface of the dimer. Although this approach does not exclusively identify the consensus site as the only patch of this kind, the site is found to be part of one of only two such regions on the protein surface. This circumstantial evidence suggests that the presence of exposed hydrophobic surfaces is an important factor in making the consensus site so cross-reactive.

### **1.5.2 Study and Visualization of Structural Adaptability**

The second program, RigiMOL, enables us to understand and visualize how Fc adapts to interact with the various Fc binding domains. RigiMOL uses a novel rigid body analysis technique to create near-realistic trajectories of protein motion that preserve most internal geometry without resorting to costly and stochastic molecular dynamics computations.

Applied to the Fc system, RigiMOL reveals that the consensus binding site adapts in concert with global shifts in the relative orientations of the  $C_{H2}$  and  $C_{H3}$



## 1.6 Conclusion

Careful application of a robust technique such as phage display can be used to solve molecular design problems even the face of substantial complexity, and its success constrasts strongly against the failure our rational design attempts on the same system.

Our directed evolution of a minimal Fc binding domain demonstrates that smaller molecules are capable of mimicking hot spot binding regions of much larger protein binding domains. Although at about 1500 Daltons, the Fc binding peptide is still about three-fold larger than a small molecule drug, but it shows that high affinity binding can be achieved by presenting important functional groups off of an entirely different structural scaffold.

The consensus binding site on Fc is a potent site for interaction, as evidenced by the fact that four natural proteins and a family of engineered peptides sought it out for binding. Analysis of this region using various computational tools shows that this site is exposed, non-polar, and adaptable. These are all properties which would be expected to assist in making this site good for binding a variety of different proteins, and thus they probably contribute to making this site the preferred locus for binding.

## **Chapter 2**

### **Dissection and Minimization of a Natural IgG-Fc Binding Domain**

#### **2.1 Introduction**

In this chapter, we seek an answer to the question: Given that certain “hot spot” regions of protein interfaces appear to be more important for binding, can we develop minimized binding partners that utilize only the hot spot residues and still bind with high affinity?

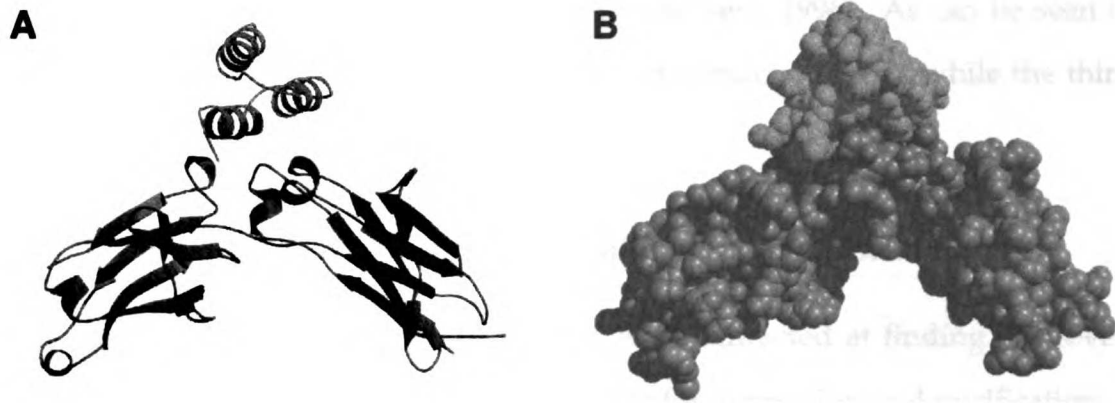
The hot spot hypothesis as originally put forth by Clackson and Wells [Clackson, 1995] holds that virtually all of the binding energy can be attributed to hot spot residues. Under this line of thinking, one would expect that a molecule which faithfully mimicked the hot spot region would bind its target with an affinity near that of the original complete molecule.

A second hypothesis is that of simple additivity with respect to energetic contributions to binding. Under additivity, although the hot spot would still make a major contribution, a molecule which reproduced only the hot spot would be reduced in affinity for its target to an extent equal to the summed contributions of the lost “cool” spot interactions. This affinity reduction could be substantial if there were many such interactions lost.

Yet a third hypothesis is that of cooperative or highly interdependent binding. Under this model, one would expect that affinity for the target of the hot

spot in isolation would be low or undetectable, since other necessary supporting structures would be absent. The O-ring hypothesis proposed by Bogan and Thorn [Bogan, 1998] falls into this category. It holds that the role of surrounding cool regions is to exclude water from the hot spot as opposed to interacting directly with the target, so without an O-ring, the hot-spot would not exist.

In order to determine which of the three above hypothesis is most applicable to the interaction between Protein A (Z-Domain) and IgG-Fc, we adopted a stepwise minimization strategy which involved shrinking the three-helix binding domain down to a single helix that presents most of the hot-spot residues and is stabilized in a helical conformation by a non-natural chemical linker.



**Figure 2.1: Crystal Structure of the Protein A Fc Binding Domain.** 2.5 Å crystal structure ( $R=22.5\%$ ,  $R_{free}=30.4\%$ ) of the complex between Fc (blue) and Protein A: B domain (red) [de Vos, 1998]. Ribbon (A) and space-filling (B) representations are shown. Note that in the original structure from [Deisenhofer, 1981], the third helix in Protein A appeared unfolded, possibly due to crystal packing interactions. Independent circular dichroism and NMR measurements confirm that the helix remains folded when the B domain interacts with Fc in solution [Gouda, 1992, Jenderberg, 1996].

## 2.2 Background: Structure of the Protein A Fc Binding Domain

Protein A is a multidomain protein derived from *Staphylococcus aureus* which has the natural function of binding G class immunoglobulins in the host organism [Langone, 1982]. Each of the five domains adopts a three-helix bundle conformation, and they apparently evolved through gene duplication events [Uhlén, 1984, Guss, 1990]. The domains have diverged to take on differential immunoglobulin binding activities, but the interaction with the Fc fragment of IgG<sub>1</sub> is best characterized. The second Ig binding domain in the Protein A gene is referred to as B domain, and it is the domain most relevant to this study.

Domain B of Protein A binds the Fc fragment of IgG with an affinity of around  $K_d = 10$  nM [Bottomley, 1994]. The domain exists as a three-helix bundle protein

alone in solution and in complex with Fc [Jenderberg, 1996]. As can be seen in Figure 2.1, two of the three helices make direct contacts with Fc while the third helix plays primarily a stabilizing role.

### **2.2.1 Z-Domain, A Functional Analogue of Protein A: B Domain**

Early research on the B domain of protein A was directed at finding improved variants which would be good affinity handles for expression and purification of exogenous proteins in *E. coli*. A double mutant of domain B (A2V,G29A) was designed to be resistant to hydroxylamine and cyanogen bromide cleavage [Nilsson, 1987]. This variant, referred to as Z-domain, is the most intensely studied derivative of Protein A and can be expressed at high levels in *E. coli* as well as being displayed on the surface of M13 bacteriophage [Nord, 1995].

The mutations in domain B applied to form Z-domain have no significant effect on the Fc binding activity of the protein [Bottomley, 1994, Jenderberg, 1995]. Thus, the conclusions derived regarding the Fc binding of Z-domain will be equally applicable to domain B and may also be applicable some of the other Fc binding domains in Protein A.

### **2.2.2 Protein A:Fc Contact Epitope**

Two structures of the complex between Fc and the B domain from Protein A are shown in Figure 2.2. While there are some minor differences between them, they are largely in agreement with respect to the location and nature of contacts on Fc and Protein A. The structures are derived from different crystal forms, so it is not surprising that the  $C_{H2}$  and  $C_{H3}$  domains have shifted relative to one another. Superposition of core atoms in  $C_{H3}$  gives a  $4^\circ$  rotation and 1 Å translation of core atoms in the  $C_{H2}$  domain. Although many of the differences between the two

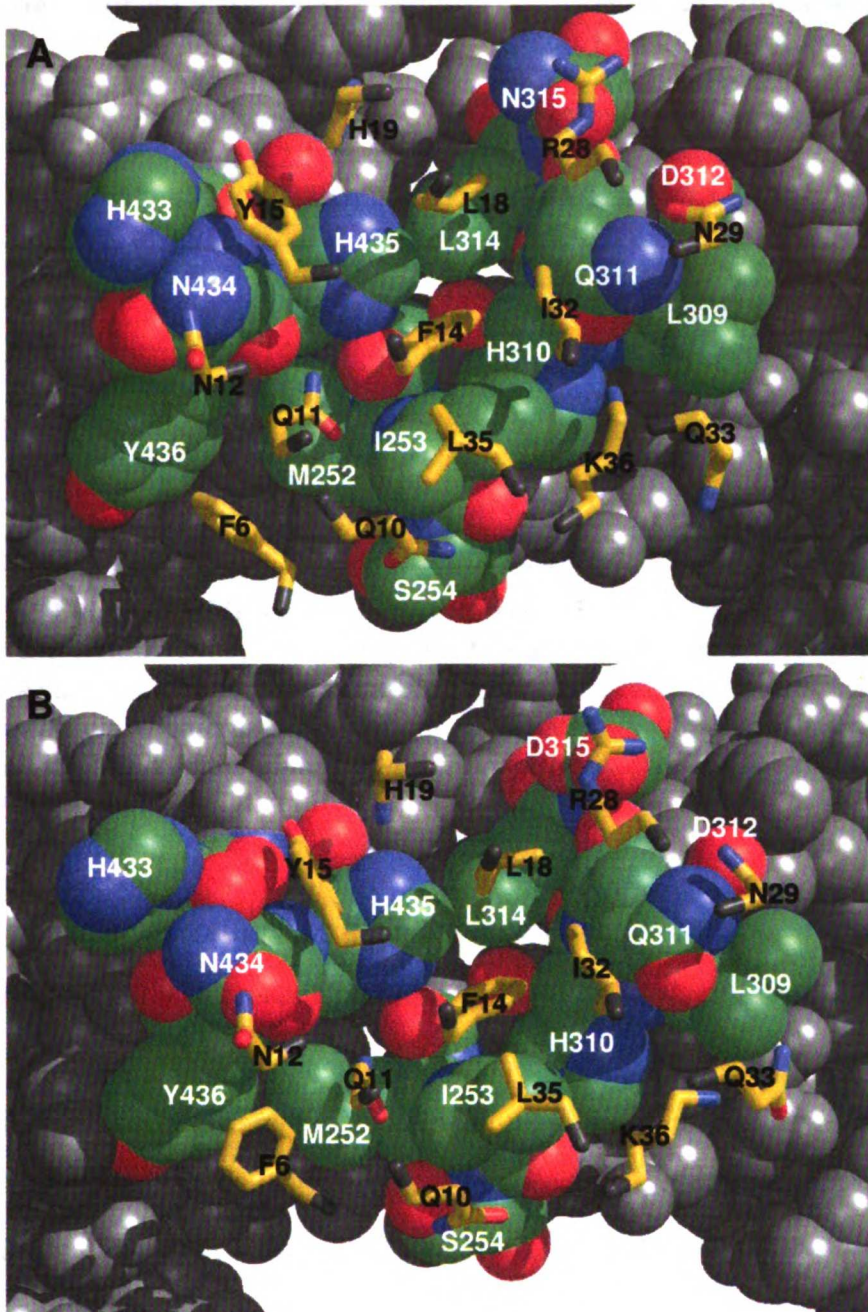


Figure 2.2: **Protein A:Fc Contact Epitope** These two structures are from (A) [de Vos, 1998] and (B) [Deisenhofer, 1981]. Nitrogen and oxygen atoms are shown in blue and red respectively. Carbon and sulfur atoms are shown in green on Fc and yellow on Protein A (B-Domain). Only interfacial atoms are colored. Residues on Fc and Protein A are labeled in white and black respectively. Hydrogens are not shown.

<b>Residue</b>	<b>Buried Surface Area Å<sup>2</sup></b>	<b>Principal Contacts</b>
Phe 6	99.7 (51.6)	Met 252, Tyr 436
Gln 10	39.2 (39.2)	Ile 253, Ser 254*
Gln 11	70.2 (69.6)	Met 253, Ile 253*, Asn434*, His 435
Asn 12	21.2 (12.1)	Asn 434*
Phe 14	71.9 (67.3)	Ile 253, His 310, Gln 311, His 435
Tyr 15	93.1 (88.3)	His 433, Asn 434, His 435
Leu 18	60.8 (60.4)	Gln 311, Leu 314, His 435
His 19	9.8 (8.8)	His 435
Arg 28	14.6 (14.6)	Gln 311, Asn 315
Asn 29	47.9 (44.1)	Gln 311*
Ile 32	24.0 (22.0)	Ile 253, Gln 311
Gln 33	15.0 (14.3)	Leu 309, Gln 311
Leu 35	2.5 (2.4)	Ile253
Lys 36	59.9 (57.9)	Ile253, Leu 309, His 310*, Gln 311*

**Table 2.1: Z-Domain Side chain Contacts with Fc.** The buried surface area is shown for the whole residue and, in parenthesis, for side chain atoms beyond the C<sub>β</sub> atom. Asterices (\*) indicate possible hydrogen bonding interactions. Buried surfaces areas were calculated from the [Deisenhofer, 1981] structure. Interactions were tabulated from visual inspection of both the [Deisenhofer, 1981] and [de Vos, 1998] structures.

structures are likely to reflect errors in the structure determination process, some of them may actually represent genuine accommodation of this conformational change in Fc.

Overall, the interface between Protein A and Fc is largely non-polar. There appear to be only three to six hydrogen bonds formed between Protein A and Fc, and no salt bridges. The total area buried in the interface is about 1360 Å<sup>2</sup> with about 635 Å<sup>2</sup> buried on Protein A and 725 Å<sup>2</sup> buried on Fc. Table 2.1 contains a detailed summary of the contacts that Protein A makes with Fc, along with information about how much surface area on each of the residues is buried in the complex.

```

      GCC GTA GAC AAC AAA TTC AAC AAA GAA CAA
1   A  V  D  N  K  F  N  K  E  Q
      CAA AAC GCG TTC TAT GAG ATC TTA CAT TTA
11  Q  N  A  F  Y  E  I  L  H  L
      CCT AAC CTG AAT GAG GAG CAG CGT AAT GCC
21  P  N  L  N  E  E  Q  R  N  A
      TTT ATC CAA AGT TTA AAA GAT GAC CCA AGC
31  F  I  Q  S  L  K  D  D  P  S
      CAA AGC GCT AAC CTT TTA GCA GAA GCT AAA
41  Q  S  A  N  L  L  A  E  A  K
      AAG CTA AAT GAT GCT CAG GCG CCT AAG CAG
51  K  L  N  D  A  Q  A  P  K  Q

```

**Figure 2.3: Z-domain Gene and Protein Sequences.** This sequence contains the natural Protein A B-domain gene sequence (Protein A 120 to 177) with two point mutations, A2V and G29A, a combination commonly referred to as Z-domain [Nilsson, 1987].

### 2.3 Alanine Scanning of the Protein A:Fc Interface

The first step in the development of minimized Z-domain variants was to map out the important binding interactions in this interface using alanine scanning mutagenesis. Both sides of the interface were alanine scanned using M13 bacteriophage as an expression and display vehicle for the mutants.

#### 2.3.1 Methods

##### Z-domain Vector Construction

Z-domain was displayed on the surface of M13 bacteriophage as an N-terminal fusion to the second extracellular domain to the gene III protein. Human growth hormone (hGH) in the monovalent phage display vector pGHam-g3 [Lowman, 1991] was replaced with the 180 nucleotide gene sequence (Figure 2.3) which contains the 58-residue Z-domain gene [Nilsson, 1987] flanked by an alanine on the N-terminus and a glutamine on the C-terminus. In this construct, pABZD, the



gene was immediately preceded by the 23 residue stII secretion signal [Lee, 1983], and immediately followed by residue 249 of the mature M13 gIII protein (which includes the SGGGSGSG linker sequence). Expression of this fusion protein was under control of the AP promoter. Also prepared was vector pWZD which is identical to pABZD except that the glutamine codon 60 in Figure 2.3 is replaced by an amber (TAG) stop codon.

Numbering of the residues in this construct are consistent with [Gouda, 1992], where the first residue in the sequence is residue number 1 (as shown). Sequences of the constructs were confirmed by dideoxy sequencing [Sanger, 1977]

### **Fc Vector Construction**

The gene for residues 230-448 of IgG<sub>1</sub> (the Fc fragment) was cloned out of the pEK3 Fc expression vector [Carter, 1996] and into pGHam-g3 in place of hGH to form vector pW0437. The gene was immediately followed by a leucine codon (CTC), a suppressible (TAG) stop codon, and residue 249 of the mature M13 gIII protein. Site directed mutagenesis [Kunkel, 1987] was then applied using the oligonucleotide T ACA AAT GCC TAT GCT GCA GTC ACA TGC CCC ACC GTG CCA GCA CCT GAA CTC to add residues 224-229 onto the N-terminus of the Fc in order to include two cysteine residues that would be capable of forming intramolecular disulfide bridges. Thus, Fc expressed from this vector would have the potential of forming intramolecular disulfide bonds in a manner resembling those found in Fc' fragment obtained from papain cleavage of intact IgG<sub>1</sub>. Also introduced was an N-terminal alanine-valine sequence in order to facilitate cleavage of the signal peptide.

Due to the amber (TAG) stop codon at the Fc/gene III junction, expression of this vector in amber suppressor *E. coli* will produce both free Fc and Fc-gene

III fusion protein, and thus permit formation of Fc dimer on the surface of M13 bacteriophage.

### **Verification of Covalent Fc Dimerization on Phage**

Expression of disulfide-linked Fc on the surface of M13 bacteriophage was confirmed by Western blot [Towbin, 1987]. Briefly, M13 bacteriophage produced using the pW0437 and pW0438 vectors were prepared and purified as described in Protocol B.2, and then 2  $\mu$ L aliquots of  $10^{13}$  p.f.u./mL bacteriophage were denatured and loaded onto a 10% Tris-glycine polyacrylamide gels under both reducing and non-reducing conditions. The protein was then electrophoretically transferred to nitrocellulose and probed using an  $\alpha$ -Fc polyclonal goat antibody followed by an  $\alpha$ -goat alkaline phosphatase conjugate. Detection was accomplished by chemiluminescence.

Binding of pW0438 Fc displaying phage to immobilized Z-domain was confirmed by phage ELISA, as described in Protocol B.7, with Z-domain immobilized at 5  $\mu$ g/mL concentration.

### **Mutagenesis**

All interface residues in the complex structure of Protein A (B-domain) with Fc [Deisenhofer, 1981] were selected for alanine scanning. A residue was defined as being in the interface if at least one of its side chain atoms was within 5 Å of an atom on opposing molecule, and a script written for the computer program X-PLOR [Brünger, 1992] was used to identify residues fitting this criterion. As a result, residues 6, 10, 11, 12, 14, 15, 18, 19, 28, 29, 32, 33, 35, and 36 on Z-domain and residues 252, 253, 254, 309, 310, 311, 312, 314, 315, 434, 435, and 436 on Fc were selected for alanine substitution.

Mutagenesis was carried out by applying the method of [Kunkel, 1987] using sets of 39-mer mutagenic oligonucleotides. For each residue, the alanine codon with maximum similarity to the original sequence was chosen for the mutation. All mutants were verified by dideoxy sequencing [Sanger, 1977].

## **Material**

Z-domain was obtained from expression of vector pWZD in the non-amber suppressor *E. coli* strain 27C7 using Protocol B.9. Supernatant from a 15' 12k spin of the resuspended freeze-thaw mixture was first purified by affinity chromatography on IgG Sepharose (Pharmacia), and then by reverse phase HPLC on a preparative C-18 column. Purity was assessed at >99% by analytical reverse phase HPLC and electrospray mass spectrometry.

Purified Fc' fragment was obtained from Papain cleavage of CD4-IgG fusion protein [Capon, 1989] and purified by column chromatography as described in Protocol B.1. Fc purity was assessed at >95% by SDS-PAGE.

## **Protein A Mutant Binding Assays**

ELISAs were performed as described in Protocol B.7 for each mutant along with one nonmutant control per plate. Fc was immobilized at a coating concentration of 10  $\mu\text{g}/\text{mL}$ , and 0.1% BSA (Sigma) was used as a blocking agent.

Assays were performed in duplicate at uniform saturation levels of 25%, a condition where the ratio of  $EC_{50}$  measurements for two molecules will closely track the ratio of the dissociation constants (derived in Section A.4.1).

For each variant, twelve-point competition curves were recorded with Fc concentrations ranging from 9  $\mu\text{M}$  down to 50 pM in three-fold serial dilutions. Curves were fit by computer to a 3-parameter sigmoidal curve and used to determine

midpoint concentrations on the competition curves.

### **Fc Mutant Binding Assays**

ELISAs were performed as described in Protocol B.7 for each mutant along with at least one unmutated control per plate. Z-domain was immobilized on the surface of 96 well format MaxiSorp plates (Nunc) at a coating concentration of 5  $\mu\text{g}/\text{mL}$ .

Assays were also performed in duplicate and at uniform saturation levels of 25%. For each variant, twelve-point competition curves were recorded with Protein A concentrations ranging from 10  $\mu\text{M}$  down to 56 pM in three fold serial dilutions. Curves were fit by computer to a 3-parameter sigmoidal curve and used to determine midpoint concentrations on the competition curves.

### **2.3.2 Results and Discussion**

#### **Fc Display and Function Confirmed on Bacteriophage**

Development of the Western blot revealed a 70 kD species present in the pW0438 phage that was absent from pW0437 phage. This is consistent with the molecular weight of the dimer formed between one chain of Fc/gIII fusion and one chain of Fc (expected molecular weight, 67 kD). pW0437 showed only a single band at around 45 kD. Under reducing conditions, pW0437 and pW0438 showed an identical banding pattern with a strong signal at approximately 50 kD, consistent with reduced monomeric Fc/gene III fusion protein. Therefore, pW0438 phage were concluded to be displaying covalent Fc dimer on their surface.

The affinity of Fc displayed on phage for binding Z-domain in solution was measured at  $EC_{50} = 25 \text{ nM}$ , only slightly lower than the  $EC_{50} = 5 \text{ nM}$  binding affinity exhibited by Z-domain phage for binding Fc in solution. Thus, Fc is func-

tionally displayed on the surface of bacteriophage and still exhibits high affinity binding to Z-domain.

### Alanine Scanning

The results from the alanine scan of the Protein A:Fc interface are shown in Figure 2.4. Only about 40% of the residues mutated have large effects upon binding when they are replaced by alanine. Foremost among these is Phe14, which has greater than a 3.5 kcal/mole effect upon binding when mutated. Next in importance come Glu11 and Ile32, which each disrupt binding by about 2.5 kcal/mole, followed by Leu18 and His19 which have approximately 1.5 kcal/mole effects.

Similarly, on the complementary side of the interface, only 25% of the residues have large impacts upon affinity when replaced with alanine. Ile253 has the largest effect with 3 kcal/mole of disruption, followed by His435 at 2.5 kcal/mole and His310 with a 2 kcal/mole drop in affinity.

These results suggest that there is a hot spot of binding affinity in the Protein A:Fc system that resembles hot spots found in other protein interfaces [Bogan,1988]. When the effects of these mutations are painted onto the structure of the Z-domain: Fc complex (Figure 2.5), we see that these residues form a roughly complementary patch in the center of the interface.

Figure 2.4 also shows the amount of surface area buried on each of the side chains that would be lost upon mutation to alanine. As has been seen with other protein interfaces, the correlation between burial and sensitivity to alanine substitution is fairly weak. Many heavily buried residues have little effect on binding when they are removed, and several very sensitive residues account for very little buried surface area in the interface.

Four out of the five hot spot residues on Z-domain (those with  $\Delta G_{mut} > 1.5$

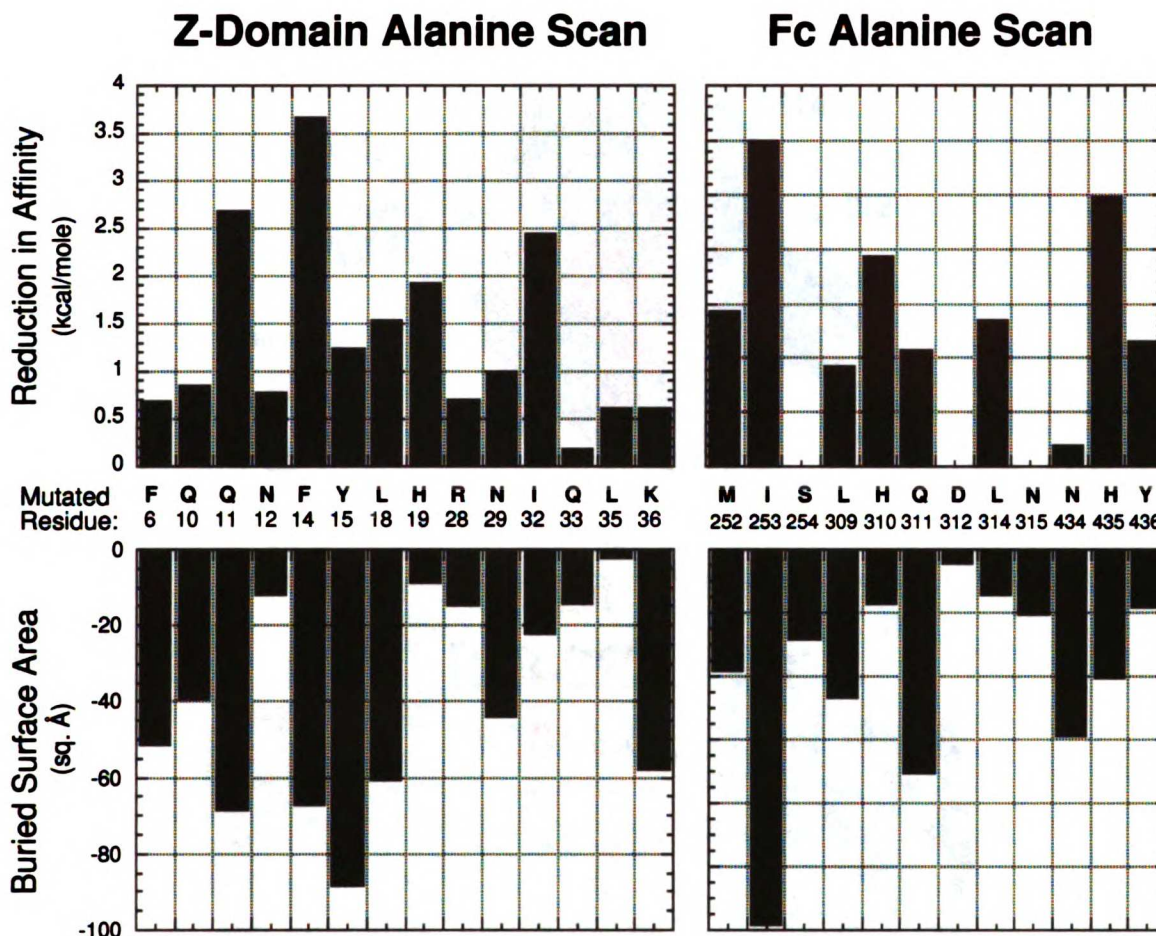
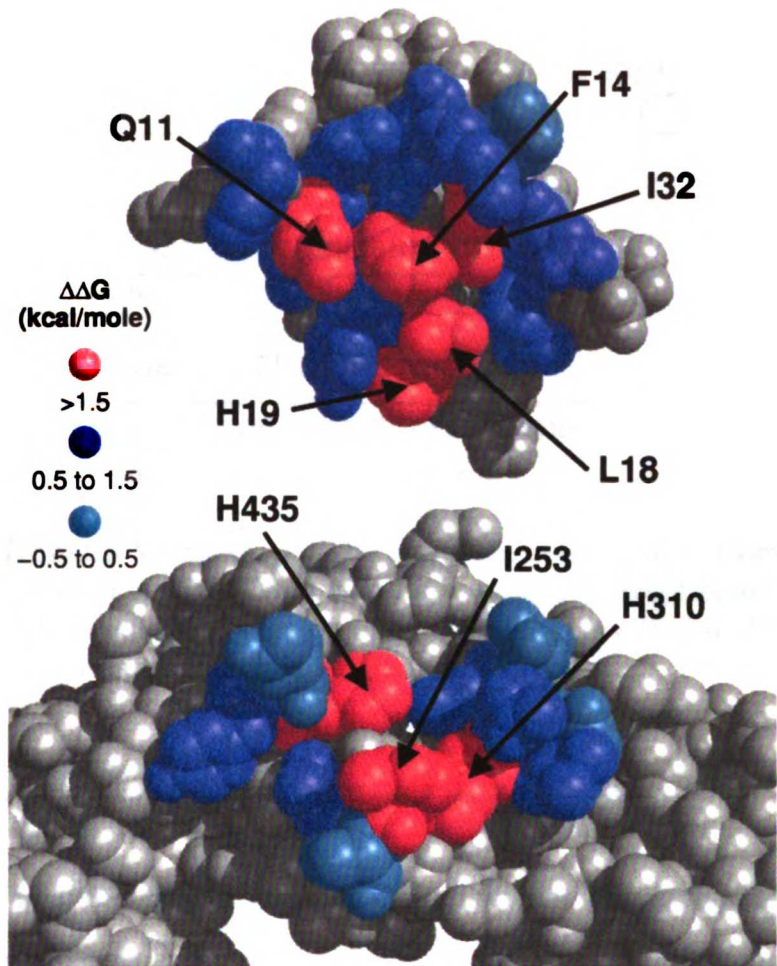


Figure 2.4: **Alanine Scan of the Protein A:Fc Interface.** The top row shows the reduction in affinity resulting from alanine substitution of the indicated residues on Protein A, Z-domain (left) or on Fc (right). The bottom row shows how much buried surface area would be lost in the complex due to removal of atoms beyond the  $C_{\beta}$  atom.

kcal/mole), are located on Helix 1 (see Figures 2.1 and 2.5), suggesting that a single helix might ultimately be capable of binding to and mimicking the energetically important Fc binding interactions.

A natural intermediate along the pathway of design from three-helix Z-domain to a one-helix mimic is a two-helix version. In the next section, we examine two such variants and explore how their functional binding interactions differ from

1100711DDADA



**Figure 2.5: Hot Spots in the Protein A:Fc Interface.** Hot spot residues (red) on either side of the interface have large effects on affinity when mutated to alanine. Four out of five such residues on Protein A, Z-domain are located on the first helix of the bundle.

that of Z-domain.

1100E LIBRARY

(A)

(B)

(C)

Figure  
derivative  
displacement  
that has

2.4 A

Develop

Brainte

of these

2.6. Z

holds

of Z39

(Z38) a

pattern

2.4.1

The vec

1996] ar



### (A) 3 Helix

AVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDQAQPK



### (B) 2 Helix

AVAQSFNMQQORRFYEALHDPNLNEEQRNAKIKSIRDD



### (C) 2 Helix (disulfide constrained)

AVAQSFNMQCORRFYEALHDPNLNEEQRNAKIKSIRDDC



Figure 2.6: **Z-Domain Variants.** (A) The complete three-helix Z-domain sequence derived from protein A. (B) Z38, a two-helix sequence engineered using phage display. (C) Z39C, a modified version of Z38 with a stabilizing disulfide bridge that holds the two helices together.

## 2.4 Analysis of Two-helix Variants Based on Protein A

Development of two-helix variants of Z-domain was accomplished by Andrew Braisted in his directed evolution of "Mini-Z" [Braisted, 1996]. Sequences of two of these variants, Z38 and Z39C, can be seen juxtaposed with Z-domain in Figure 2.6. Z39C is stabilized by a disulfide bridge between residues 10 and 39 which holds the two helices together. A crystal structure of a slightly truncated version of Z39C can be seen in Figure 2.7. Here we alanine scan both the unconstrained (Z38) and constrained (Z39C) variants of Z-domain in order to assess whether the pattern of functionally important interactions in the interface has changed.

### 2.4.1 Methods

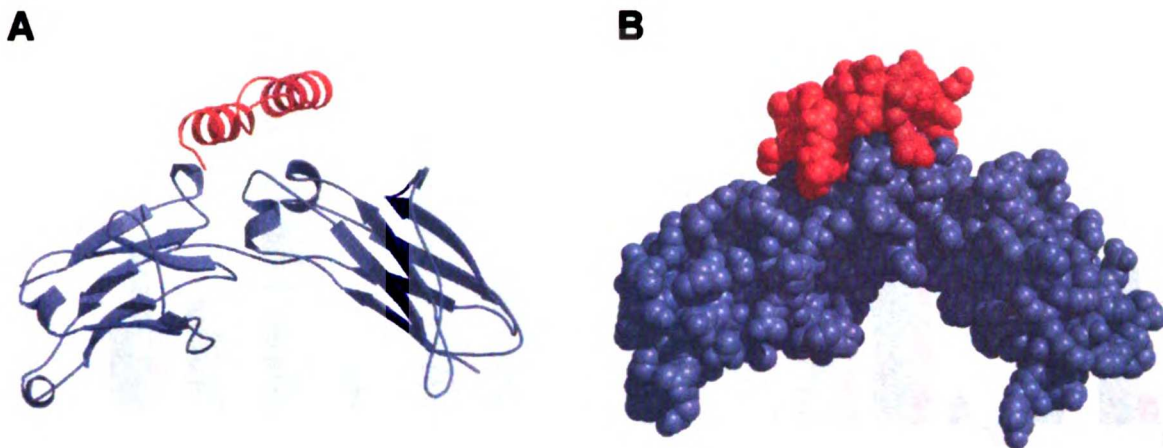
The vectors pABZ38 and pABZ39C were obtained from Andrew Braisted [Braisted, 1996] and were analogous in composition to pWZD from Section 2.3.1. Alanine



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100



**Figure 2.7: Crystal Structure of a Two-Helix Z-domain Variant.** This is a 1.95 Å crystal structure of Z34C [Ultsch, 1999] (red), a truncated variant of Z39C which lacks four N-terminal residues, in complex with Fc (blue). Ribbon (A) and space-filling (B) representations are shown. The two helices bind in virtually an identical manner to the parent three-helix protein A domain.

mutants of pABZ38 and pABZ39C at the indicated positions were also prepared as described, and measurements of their binding affinities was performed in an identical fashion using the unmutated variants as controls.

#### 2.4.2 Results and Discussion

The results from the comparative alanine scan of the three variants in shown in Figure 2.8. As with Z-domain, Gln11, Phe14, Leu18, His19, and Ile32 are among the most important residues for binding in the two-helix variants, indicating that the hot spot has been preserved essentially intact.

However, there are some notable differences in the minimized molecules. Tyr15 and Leu18 appear to be ten-fold more important for binding in the two-helix variants than in Z-domain. One possible explanation for this could be that the side chains of these residues have taken on additional structural roles in the minimized proteins that were not necessary or important in Z-domain.

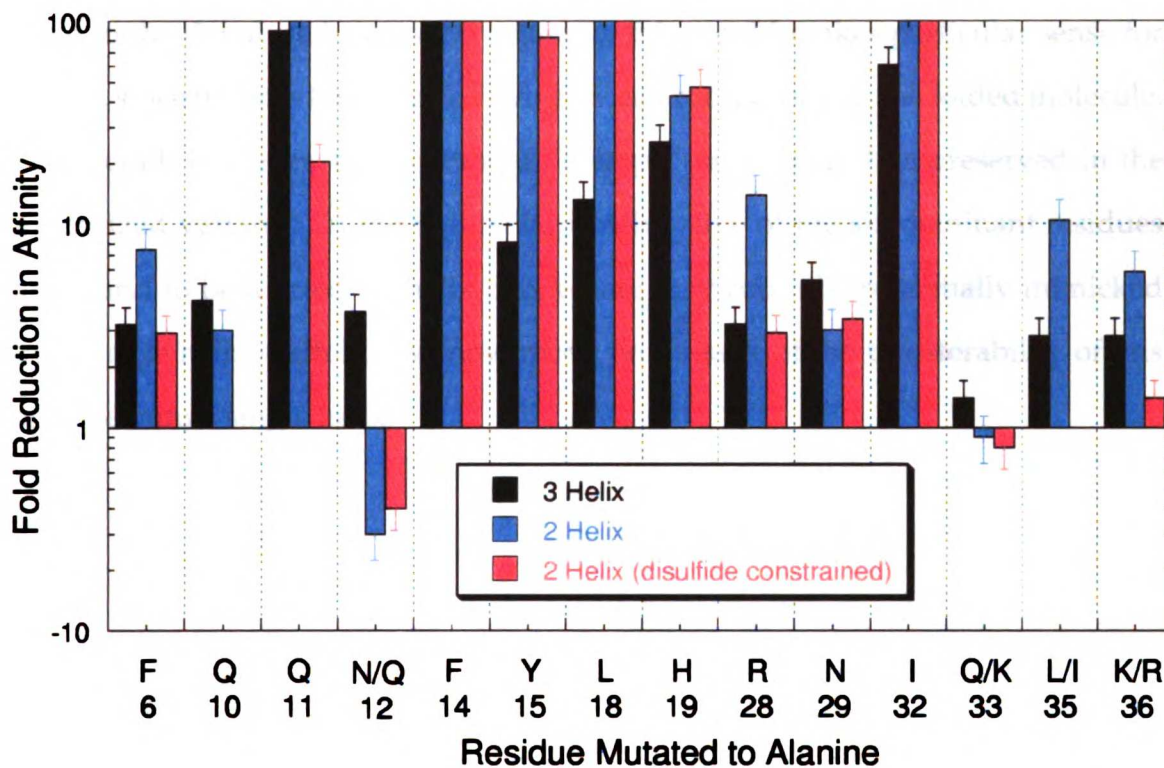


Figure 2.8: **Alanine Scan of Three Protein A Variants.** Shown are the effects of alanine mutations on the binding affinity for the variants: Z-domain (black), Z38 (blue), and Z39C (red). Positions that were mutated in Z38 and Z39C relative to Z-domain are labelled with the original residue on the left and the mutated residue on the right.

Some differences also exist between Z38 and Z39C, which differ with respect to the presence of a stabilizing disulfide bridge. Several of the residues in the unconstrained Z38 variant, Phe6, Arg28, and Lys36, are more sensitive to mutation than in Z39C. Thermal melts of Z38 and Z39C monitored by circular dichroism show that Z38, but not Z39C, is partially unfolded at the room temperature [Braisted, 1996]. Thus, an energetic penalty for folding must be paid by some molecules upon binding, and the binding affinity will be sensitive to this folding penalty. The differential effects of these mutants on affinity in the various contexts could therefore be explained by effects on the stability of the folded proteins rel-



ative to the denatured states. Structurally, this would make particular sense for Arg28, since it packs between the two helices near the loop in the folded molecule.

Overall, however, the functional "epitope" on Protein A is preserved in the minimized variants. In the two-helix proteins, five of the six dominant residues are found to be located on helix one, where they could be potentially mimicked by a single helix scaffold. We now move on to the issue of the transferability of this epitope onto a single helix.

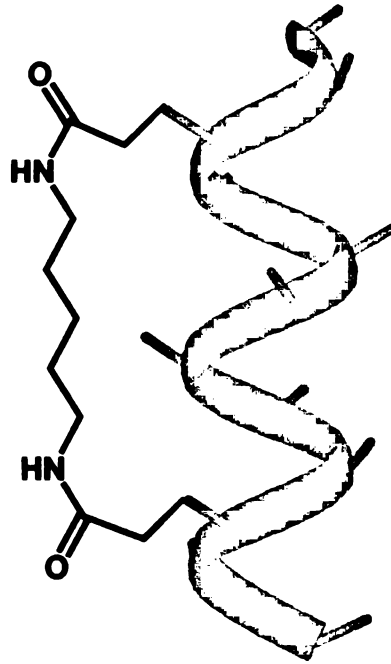


Figure 2.9: **Helix Locking Chemistry.** This chemical “lock” is capable of stabilizing the conformation of isolated helices in solution by linking side chains of  $i$  and  $i + 7$  residues [Phelan, 1997]. It can be thought of as two glutamine side chains linked together by a five carbon alkane chain.

## 2.5 Analysis of a Single Helix Protein A Mimic

Because a single isolated  $\alpha$ -helix is not normally folded or stable in solution, various chemical methods have been developed which can enforce a helix conformation [Jackson, 1991, Pavone, 1992, Bracken, 1994, Phelan, 1997]. Here we apply an  $i$  to  $i + 7$  side chain linkage strategy (Figure 2.9) to create a single  $\alpha$ -helix mimic of Protein A.

## 2.5.1 Methods

### Locked-Helix Synthesis

The synthesis of locked helices is published [Phelan, 1997] and not covered here. Locked-helix sequences were designed using Z34C as a template (Figure 2.7).

Briefly, the peptide sequence of LH1, N-succinyl-FNMXQRRFYZALH-NH<sub>2</sub>, where X is gamma-L-glutamyl allyl ester and Z is gamma-L-glutamyl (N-allyloxycarbonyl pentylamine) was synthesized by solid phase on Wang resin. The allyl and allyloxycarbonyl protecting groups were selectively removed using palladium catalyst, and then the free amine and carboxylate were cyclized using standard peptide coupling reagents. The peptide was then simultaneously deprotected and cleaved from the resin. Purification was carried out by reverse phase and confirmed by electrospray mass spectrometry. A related peptide, LH2, was also prepared. It is identical to LH1 except for containing alanines in place of the two arginines.

### CD Spectra

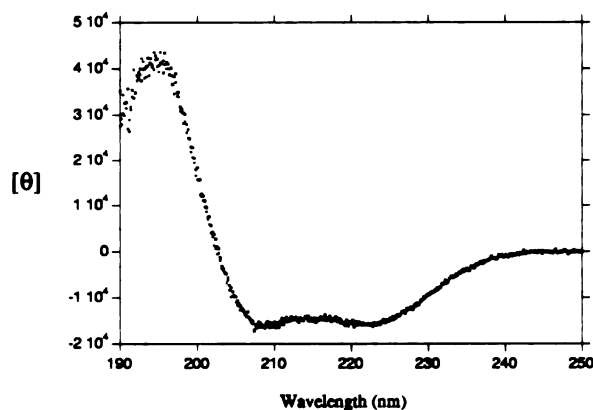
CD spectra were acquired on an Aviv 60DS spectrometer (Aviv Associates) over the range 190-250 nm (0.2 nm intervals) in a cell with a path length of 0.1 cm at 25° C. Samples were at a 0.1 mg/mL concentrations of peptide in 100 mM NaCl and 10 mM Tris HCl pH 7.2.

### Binding Assays

Inhibition of biotinylated Z-domain binding to immobilized Fc by peptide was measured as described in Protocol B.8. Fc was immobilized at 10 μg/mL, and biotin:Z-domain conjugate was bound at a subsaturating level. 0.1% BSA (Sigma)







**Figure 2.10: Locked-Helix CD Spectrum.** This is the circular dichroism spectrum of LH2 alone in solution, which shows the characteristic double minima at 209 and 222 nm indicative of a helical peptide.

was used as a blocking agent and detection was carried out using avidin-HRP conjugate. Competition data were obtained for 8 two-fold serial dilutions of peptide starting at 1 mM.

### X-ray Crystallography

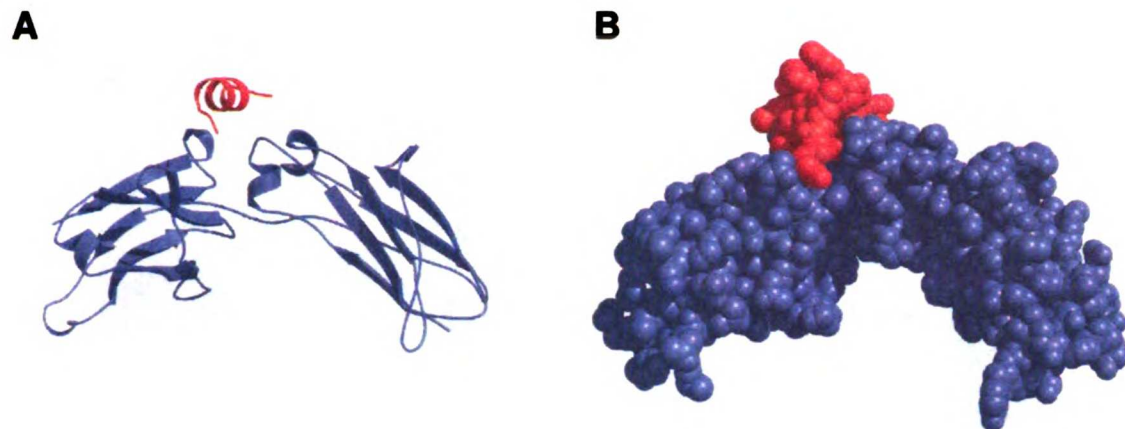
The crystal structure of LH1 in complex with Fc is primarily the work of Mark Ultsch [Ultsch, 1999] and will not be described in detail here. Briefly, the asymmetric unit in the crystal consists of a single locked-helix and a single subunit of the Fc dimer. The 1.75 Å complex was solved by molecular replacement and refined to an R-factor of 21.9% and an R-free of 28.6%.

### 2.5.2 Results and Discussion

Circular dichroism spectra of the locked helices confirms that they adopt helical conformations in solution (Figure 2.10), suggesting that the peptide should not need to pay an energetic penalty for folding prior to binding Fc.

However, in competition binding assays with Z-domain, the binding affinity

Accepted for publication  
 10/15/2011

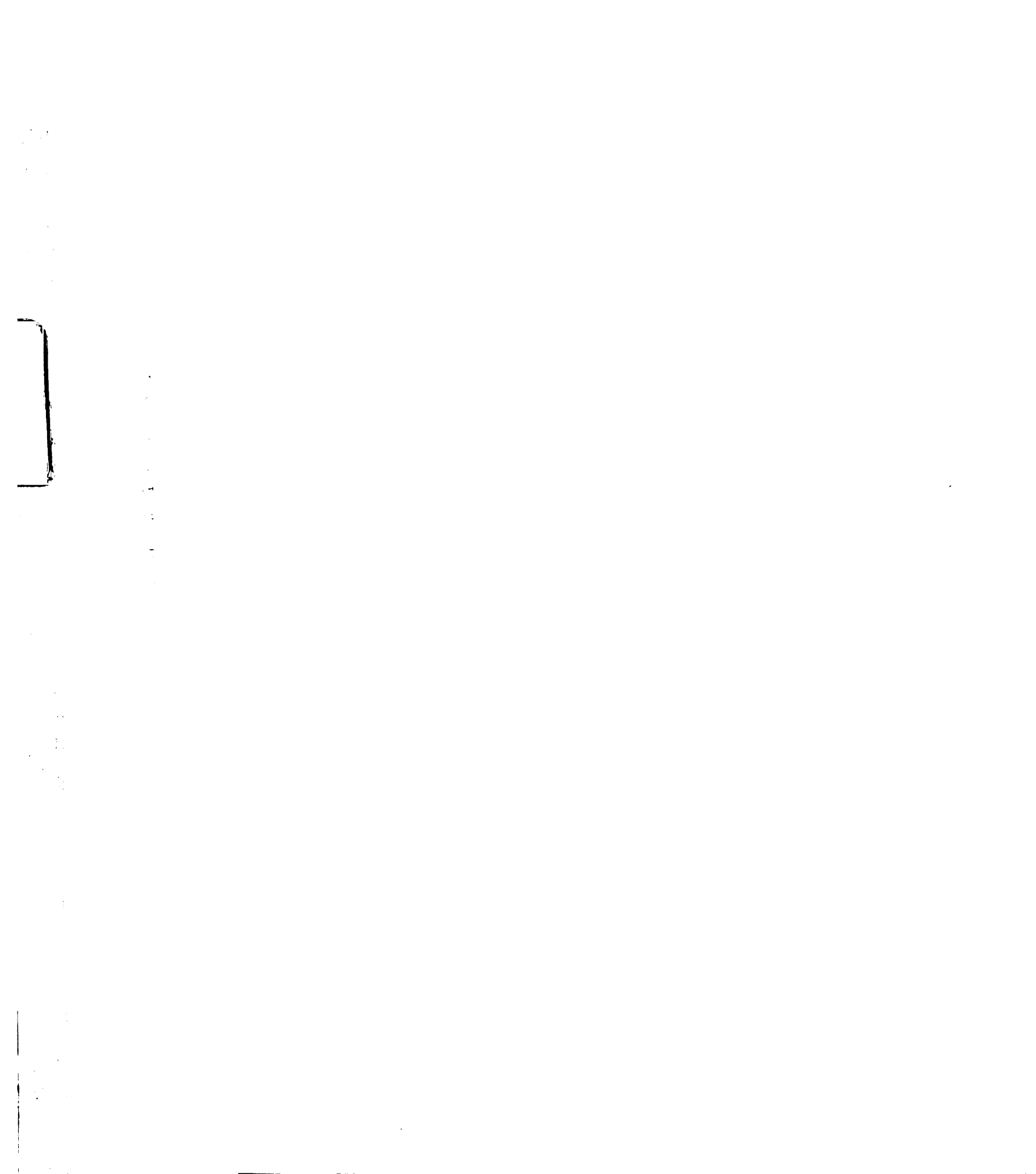


**Figure 2.11: Crystal Structure of a Single Helix Mimic of Protein A.** Shown is the 1.75 Å crystal structure of LH1 in complex with Fc [Ultsch, 1999] in ribbon (A) and space-filling (B) representations. The single helix (red) binds Fc (blue) in the same position as helix 1 in the structures of two-helix and three-helix variants complexed with Fc.

of LH1 for Fc appears to be substantially reduced. The  $IC_{50}$  for LH1 is about 250  $\mu M$ , well over four orders of magnitude lower in affinity. This is equivalent to loss of about 6 kcal/mole of binding energy, a tremendous difference when compared to the two-helix variant, which retained most of the binding affinity.

The crystal structure of the locked-helix peptide bound to Fc (Figure 2.11) confirms that the constrained peptide binds Fc, and that it does so in the same manner as Z-domain and the two-helix variants. The detailed interactions of the LH1 with Fc can be compared to those of Z-domain in Figure 2.12. All of the hot spot residues on the helix are contacting Fc, and all except His19 occupy identical conformations. His19 has been repositioned apparently due to unraveling of the helix. The other non-hot-spot side chains which make contacts with Fc are preserved intact.

Given that we have eliminated 75% of Z-domain and 40% of the side chain contacts in creating LH1, it is natural to ask the question: what percentage of the



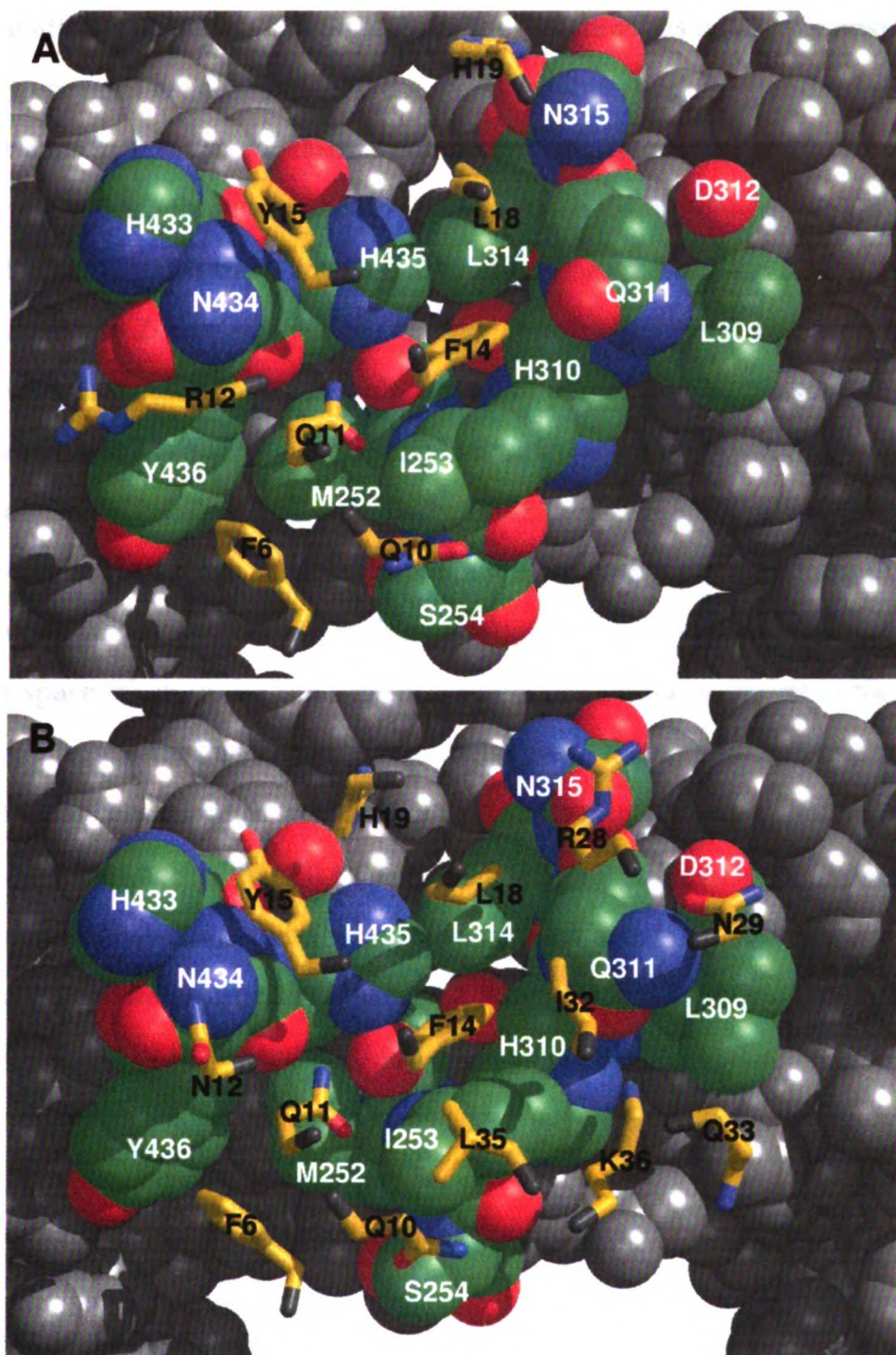


Figure 2.12: **Structural Mimicry.** The crystal structure of the Fc binding interface of (A) the locked-helix (LH1) and (B) Protein A, B-domain. The hot spot residues (Gln11, Phe14, Leu18, and His19) in the single helix interact with Fc much as they do in their natural context. Only His19 shows any substantial perturbation when presented on the locked-helix scaffold.



binding affinity remains? Unfortunately, the question has no simple answer since binding energies are all relative to the standard state, and the standard state itself is arbitrary. In this case, Z-domain binds Fc with an affinity of 10 nM, about 11 kcal/mole tighter than a molecule with “zero” binding affinity would bind, so about half of *that* arbitrary amount of binding energy has been lost in the single helix relative to Z-domain.

A more relevant question would be: what percentage of the specific binding affinity remains? We can derive a simplistic answer to this question if we think in terms of the probability that a similarly sized molecule with no specific binding affinity would be bound to Fc in the exact site where the peptide normally binds.

How do we define bound? In this case, we will define it as occupying the volume of space taken up by the ligand  $V_l$  when it is bound to target at the specific binding site. Assuming that all ligands distribute themselves randomly in solution, at what concentration will half of the binding sites be occupied with ligand? It will be when half of the all such sites in the solution are filled. Under dilute target conditions, this will be when 50% of the solution volume is composed of ligand. We can determine the molarity of such a solution if we know the partial specific volume of the ligand ( $\bar{v}_l$ ) in liters per mole.

$$\frac{V_{ligands}}{V_{total}} = \frac{1}{2} \quad (2.1)$$

$$molarity = \frac{V_{ligands}/\bar{v}_l}{V_{total}} = \frac{1}{2\bar{v}_l} \quad (2.2)$$

Proteins have partial specific volumes in solution in the range  $7.2 \times 10^{-4}$  to  $7.5 \times 10^{-4}$  L/g [Creighton, 1993]. For a 2000 molecular weight peptide such as LH1, that would imply a partial specific volume of about 1.4 liters per mole. Thus, if the peptide was utterly non-specific in its binding, then it would be 50% bound to

target at a concentration of about 0.33 M. Coincidentally, this is fairly close to the standard state concentration of 1 M. In comparison, a non-specific ligand the size of Z-domain (6.7 kD) would be half-bound at about 0.1 M concentration.

We can use these simplistic volumetric estimates of non-specific binding affinities to estimate the percentage of specific binding affinity remaining in the locked-helix peptide. Z-domain's  $K_d$  of 10 nM is  $10^7$ -fold lower than 0.1 M, which is equivalent to 9.8 kcal/mole of specific binding energy ( $RT \ln 10 = 1.4$  kcal/mole). LH1's  $K_d$  of 250  $\mu$ M is  $1.3 \times 10^3$ -fold lower than 0.33 M, giving 4.4 kcal/mole of specific binding energy. Thus it would appear that only about 45% of Z-domain's specific binding energy is retained by the locked-helix peptides.

Of course, this oversimplified analysis ignores other sources of non-specific binding such as general hydrophobic interactions, and it ignores non-ideality of solutions, rotational considerations, binding at other sites, and the fact that peptides and proteins are not usually soluble to such high concentrations. However, it does provide a conceptual framework for thinking about specific binding energies without worrying about standard state concentrations, and it places an upper limit on how much material would be needed to achieve truly nonspecific binding in the absence of other considerations.

According to the hot spot hypothesis, most or all of the binding energy should be due to a few critical interactions. Here we appear to have successfully mimicked most of those interactions, yet a huge fraction of the binding affinity has been lost. One of the hot spot residues on Z-domain, Ile32, is indeed missing in the locked-helix. However, that side chain accounts for only 2.4 kcal/mole of binding energy, which is less than half of the observed observed reduction in affinity.

If we add up the apparent importance of the remaining interactions that Z-domain makes with Fc (Table 2.2), we find that all of the lost interactions on the





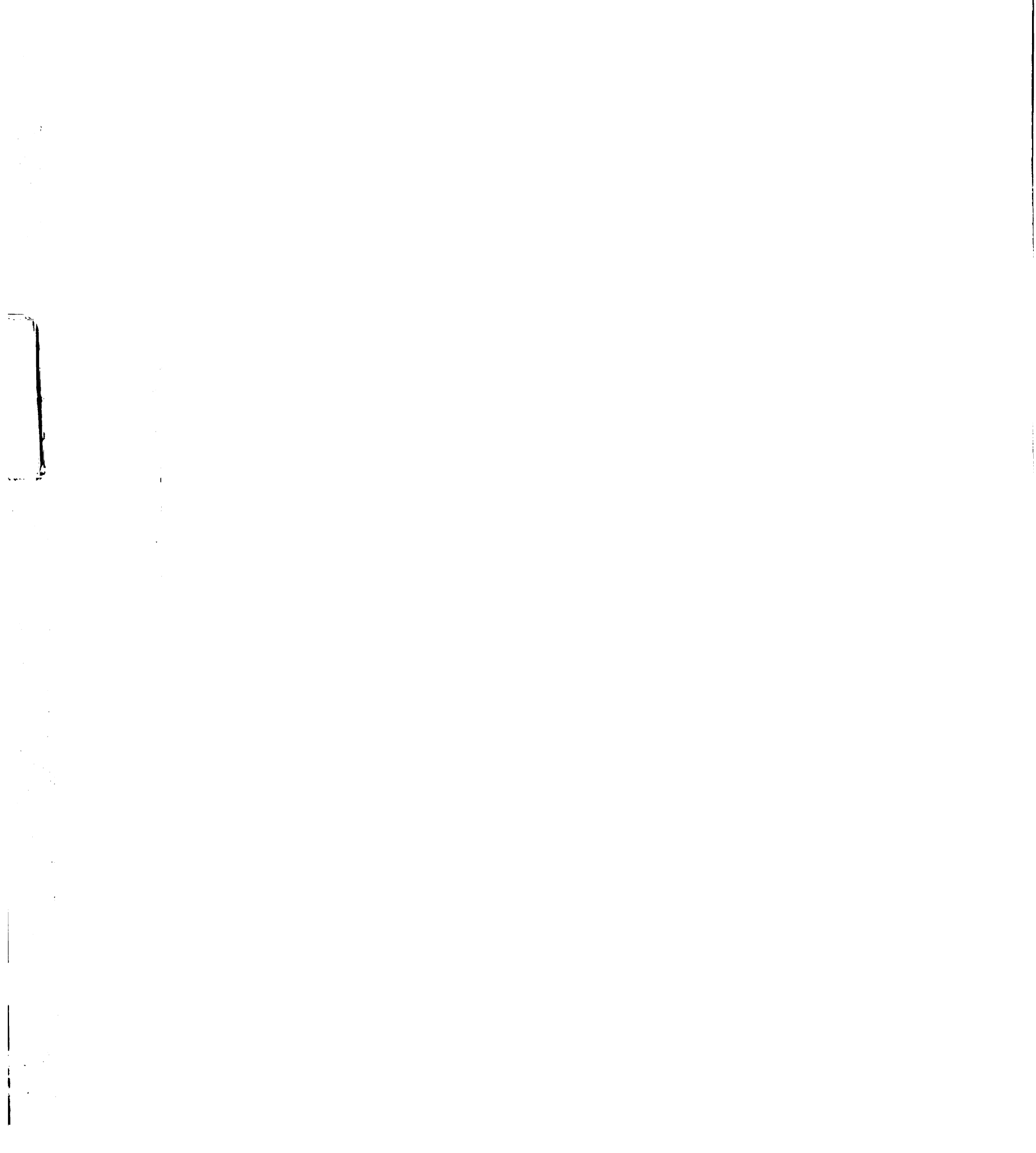
Alanine Substitution	Energy Lost (kcal/mole)
Arg28	0.7
Asn29	1.0
Ile32	2.4
Gln33	0.2
Lys35	0.6
Arg36	0.6
<b>Total</b>	<b>5.5</b>

**Table 2.2: Lost Interactions on Helix II.** From alanine scanning mutagenesis, we can estimate the cumulative effect of lost interactions found in the natural protein A:Fc interface, but which are absent in the locked-helix:Fc interface.

second helix would appear to account for 5.5 kcal/mole of binding energy. That is very much in agreement with the 6 kcal/mole that one would compute from a  $2.5 \times 10^4$  fold drop in binding affinity. This result suggests that binding energies are exhibiting additive behavior with respect to removal of contacts on the second helix.

The displacement of His19 on the locked-helix relative to its position in the three other structures may also be important. An alanine substitution of His19 results in a 1.9 kcal/mole drop in binding affinity, so the predicted combined effect of removing Ile32 and His19 would be 4.3 kcal/mole, or about 75% of the observed drop in affinity. However, given that His19 is still interacting with Fc, it may still be making energetically favorable interactions, and so the effect of this displacement remains unknown. Although the importance of His19 could in theory be measured by an alanine substitution of His19 on the locked-helix, the solubility limit of these peptides makes accurate assessment of the importance of this interaction nearly impossible.

Are there any other differences which could explain the observed drop in affinity? A comparison of the three Z-domain variants for which structures are known

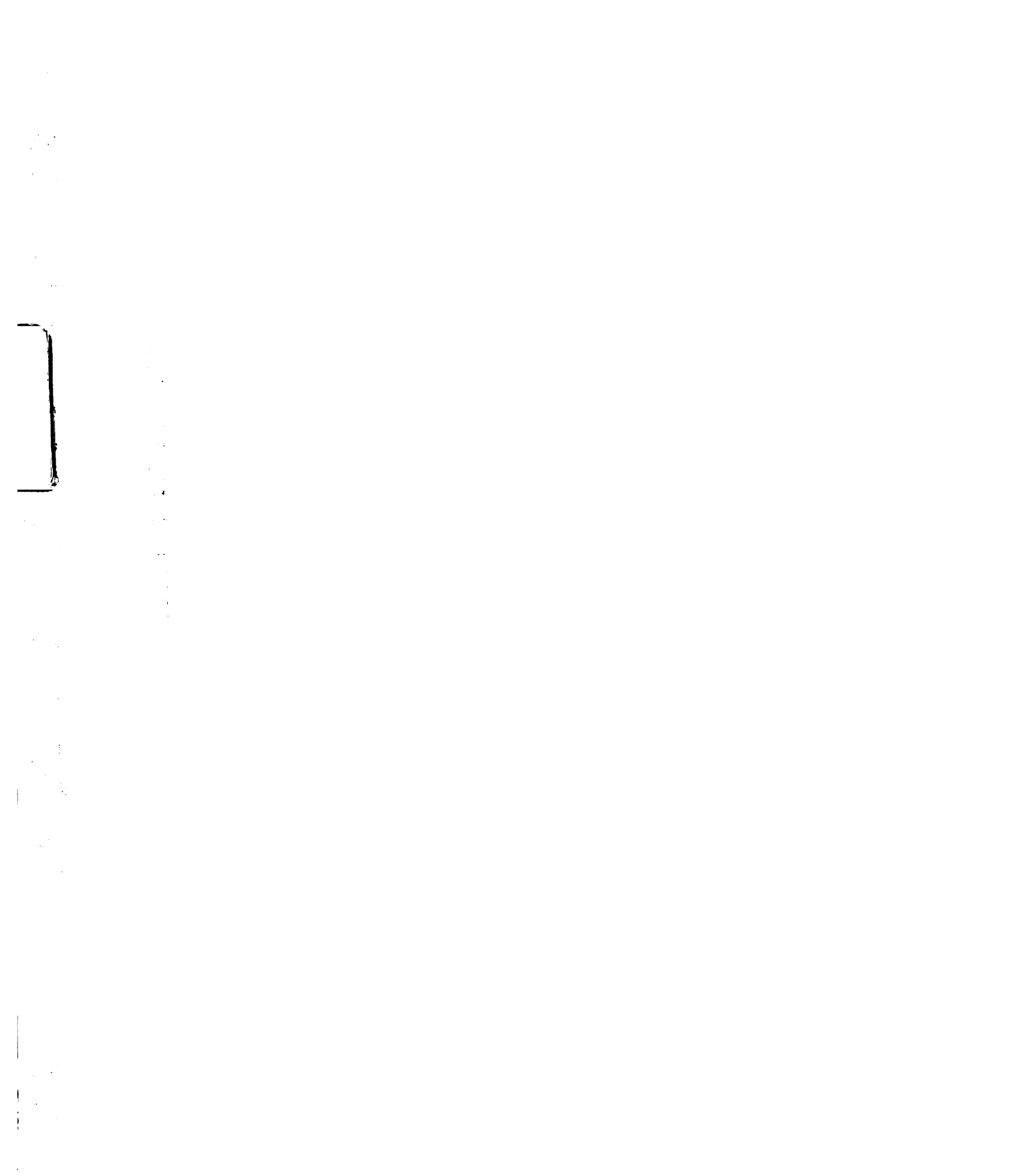


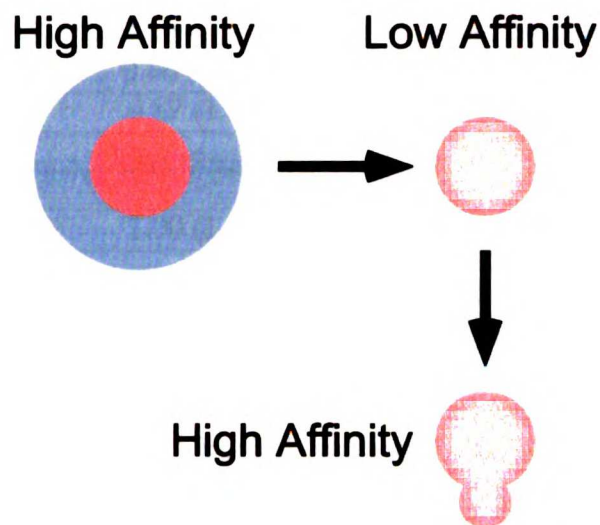
is shown in Table 2.3. In terms of overall R.M.S.D. geometric variations, the three variants are about as similar to one another as the two crystal structures of B-domain are to one another, so no substantial global conformation changes can be detected. Surprisingly however, the amount of surface area buried in the complex for the single helix is much greater than one would expect for the single helix alone.

Where does the 210 Å<sup>2</sup> of additional buried surface come from in the locked helix complex, when compared to helix 1 of the Z34C complex? Calculation of buried surface area on a per-residue basis along with visual inspection of the structure reveals that it arises from several sources. First, the N-terminal succinylate on LH1 interacts with the 380's loop on the C<sub>H3</sub> domain Fc to bury an additional 80 Å<sup>2</sup>. Next, subtle improvements in the packing between the helix and Fc along the length of the molecule give rise to another 80 Å<sup>2</sup> of buried surface area. Apparently, the absence of a second helix allows the remaining helix to form a tighter association with Fc than it does in Z34C. Finally, the repositioning of His19 buries an additional 50 Å<sup>2</sup> of surface area.

In total, the single locked-helix complex buries a remarkable 86% of the surface area buried in the Z34C complex, despite having lost over half of the specific binding energy (as defined above). Thus, the nature of the interactions appears to be more important for affinity than simply the raw amount of surface area buried.

In order to achieve high affinity binding with a minimized molecule, it may not be sufficient to simply mimic a hot spot. It may also be necessary to incorporate additional high potency hot-spot-like interactions which can counteract the combined effects of the lost low affinity interactions. This idea of hot-spot expansion (Figure 2.13) may be difficult to achieve in a helical context, since a fixed helical scaffold makes it relatively difficult to add new interactions except by adding a





**Figure 2.13: Hot Spot Expansion.** Protein interfaces typically consist of a few high energy interactions (red) and many lower energy interactions (blue). A minimized molecule may be able to mimic those high energy interactions but bind with a low relative affinity because of the numerous low energy interactions that will be absent. For a minimized molecule to achieve an affinity comparable to the parent, it may be necessary to recruit additional high affinity interactions into the minimized binding domain and thus expand the hot spot

turn onto either end of the helix. Also, the main-chain hydrogen bond donors on the helix are all satisfied and generally unavailable for interaction with other groups in the protein interface. Less regular structural motifs, such as loops and partial  $\beta$ -sheet structures, can incorporate additional groups in a more flexible fashion, and these motif can also use some of their main chain hydrogen bond donors and acceptors to directly participate in the interaction.

KXHXQ77 1500



Protein/Peptide	R.M.S.D. vs. B-Domain	Buried Surface Area	Affinity
B-Domain	0.86 (1.61)*	1310	10 nM
Z34C	0.87 (1.32)	1370	10 nM
Z34C Helix 1	-	970	-
Z34C Helix 2	-	510	-
Locked-Helix	1.00 (1.44)	1180	250 $\mu$ M

Table 2.3: **Comparison of Z-Domain Variants.** R.M.S.D. figures are shown for the  $C_{\alpha}$  backbone, and for all non-hydrogen atoms (in parenthesis) between the indicated structure and that of Protein A B-domain [de Vos, 1998]. (\*) the R.M.S.D. value for B-domain is computed between the [de Vos, 1998] and [Deisenhofer, 1981] structures. Buried surface areas were calculated for non-hydrogen atoms using CNS [Brünger, 1998] with a solvent probe radius of 1.4 Å.

## 2.6 Conclusion

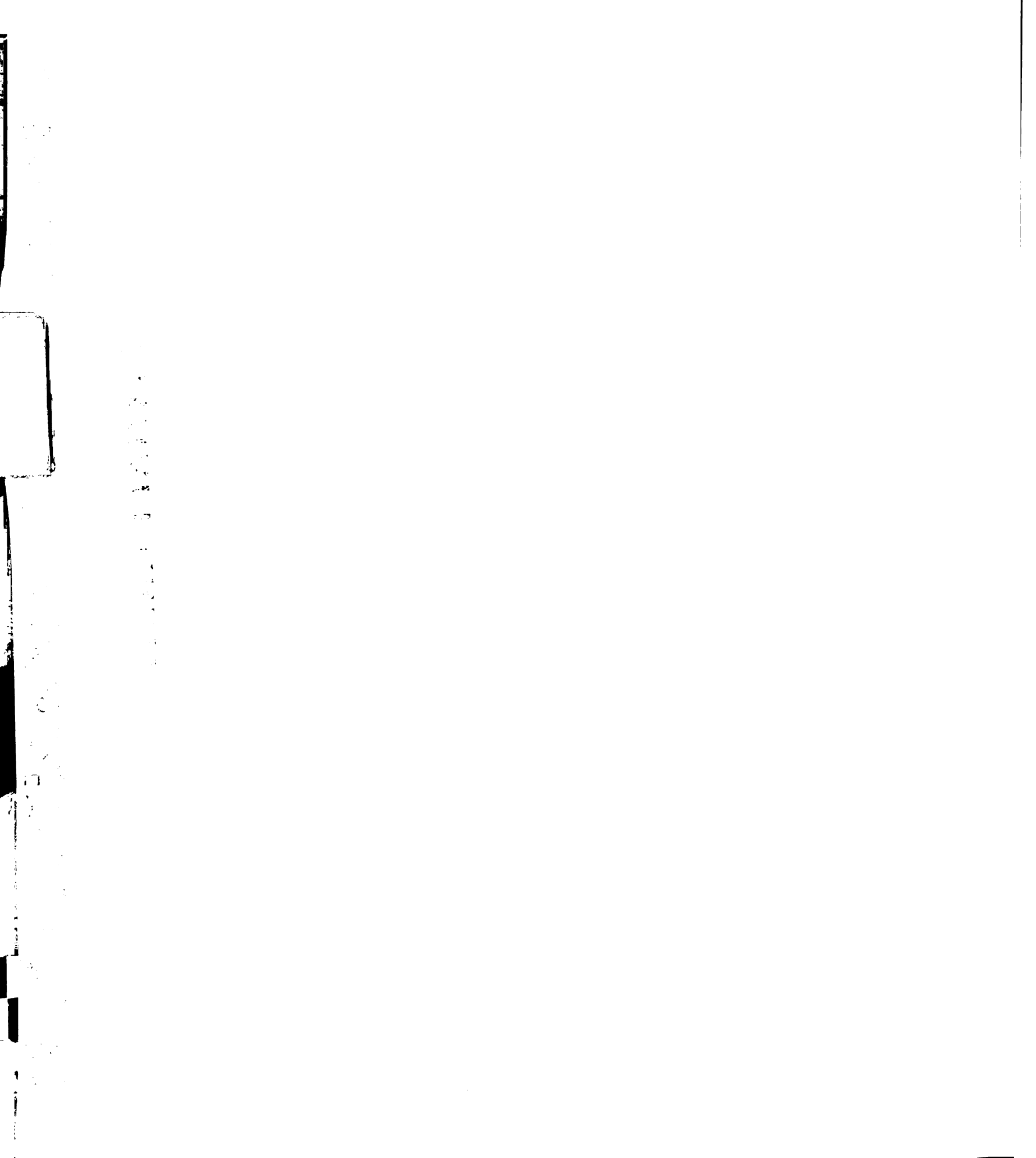
In the Protein A/Fc system, it appears that mimicry of hot spot residues alone is not sufficient to achieve high affinity binding. Although we have successfully generated a mimic of the first helix of the Protein A Fc binding domain, this molecule binds with an affinity over four orders of magnitude lower than the full binding domain. The crystal structure of the molecule demonstrates that it binds Fc in the expected manner and that the hot spot residues interact with Fc. Remarkably, the observed drop in affinity matches what would be expected from the alanine scanning results on the second helix, since all of those interactions are absent in the locked-helix. Thus, it appears that the interactions on the second helix make an additive contribution to the binding energy.

In order to develop a hot spot mimic of Protein A that binds with high affinity, it will apparently be necessary to find additional high affinity interactions with the Fc that will compensate for the many low affinity interactions that have been lost. *A priori* design of such interactions is beyond our current capabilities given our limited understanding of energetics in molecular recognition. In order to insure



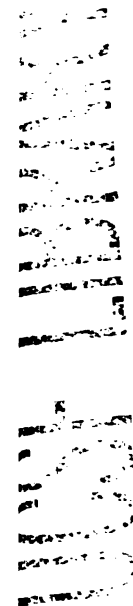
success, it would probably be necessary to resort to some form of combinatorial synthesis and high throughput screening method. However, the inflexible and regular nature of the helical scaffold makes it a sub-optimal platform from which to conduct such studies where the goal is to seek out and explore a variety of new interactions.

11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100



## 2.7 Acknowledgements

The work described in this chapter reflects a collaborative project between Andrew Braisted, Mark Ultsch and myself. I performed the mutagenesis and binding assays, Andrew carried out the peptide synthesis, and Mark performed the crystallography. The analysis is entirely my own work product.





## **Chapter 3**

### **Construction of a Non-natural Minimal Fc Binding Domain**

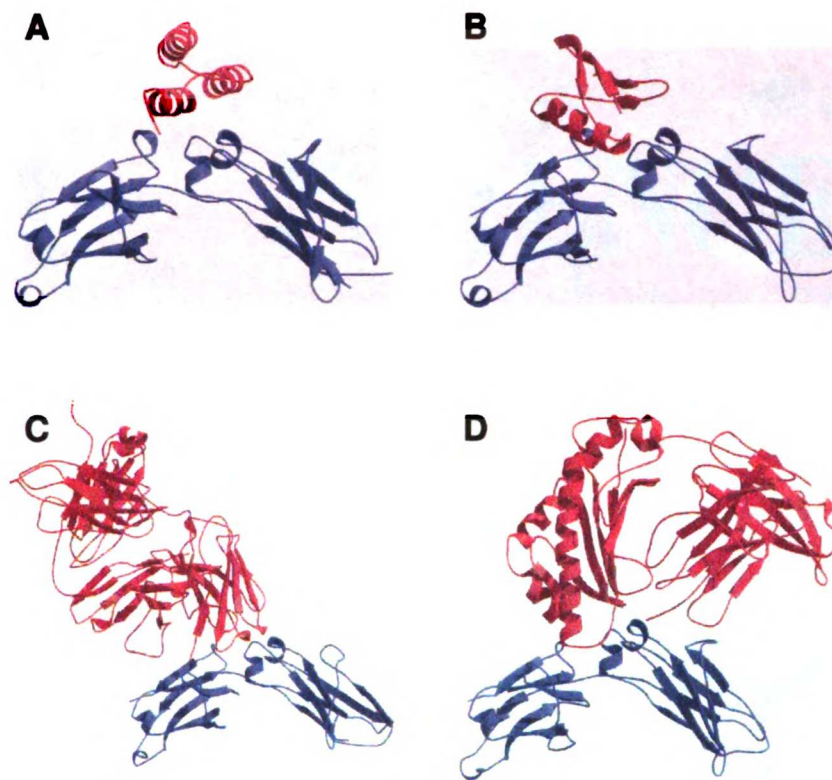
#### **3.1 Introduction**

Attempts to develop a minimized variant of Z-Domain resulted in short helical peptides which could mimic the binding interactions of Z-domain but not the high affinity binding. In nature, a variety of structural motifs have evolved for binding Fc which utilize helices, loops, beta sheet structures or combinations of the above (Figure 3.1). One possible reason for our lack of success may be that an isolated helical epitope is not the most efficient structural scaffold to use for developing the smallest Fc binding molecules. By working only with helical molecules, we were unable to explore other potentially fruitful Fc binding motifs.

Recent successes in the development of novel peptides that can mimic the binding activities of natural protein hormones suggest that alternate solutions to binding at protein interfaces can be found relatively easily (Figure 3.2) [Kay, 1998].

#### **3.2 Natural Fc Binding Domains**

The remarkable cross-reactivity of the receptor binding site on Fc gives us a unique opportunity to investigate and develop our understanding of protein-protein interactions and the phenomenon of cross reactive binding. We briefly review the



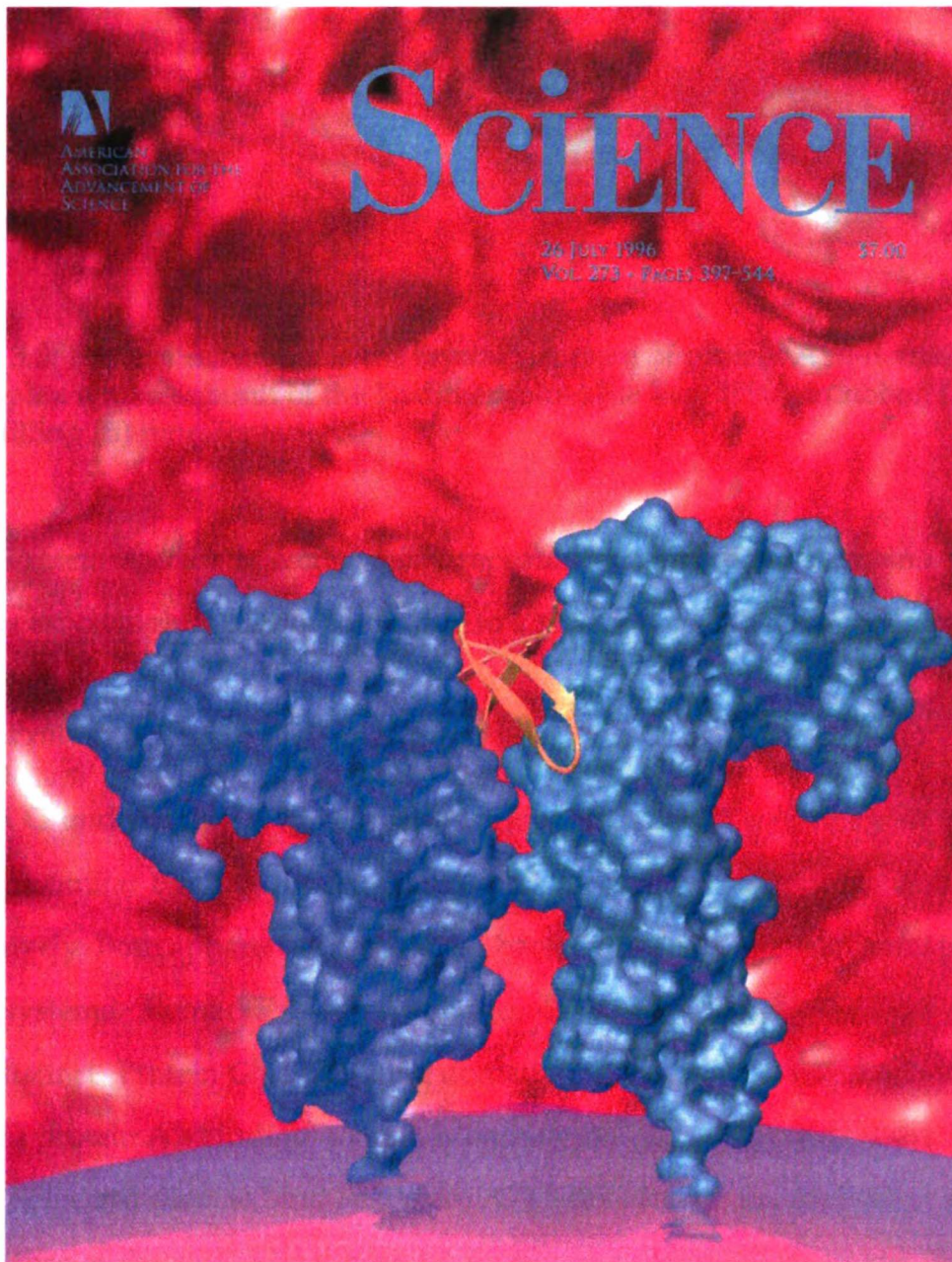
**Figure 3.1: Natural Fc Binding Domains.** One subunit of the Fc dimer is shown in blue, and the binding partners are shown in red. (A) Domain B1 Protein A [de Vos, 1998] (B) Domain C2 of Protein G [Sauer-Eriksson, 1995] (C) Rheumatoid Factor [Corper, 1997] (D) Fc Receptor [Burmeister, 1994a].

four natural Fc binding domains below before moving on to the process of evolving an Fc binding domain from scratch.

### 3.2.1 Protein A

The Fc binding domain from protein A is shown in Figure 3.3. This is an all-helical binding domain that was used as a starting point for design of minimized one or two helix mimics, as described in Chapter 2.

It binds Fc with an affinity of about 10 nM for the individual domain [Karlsson,

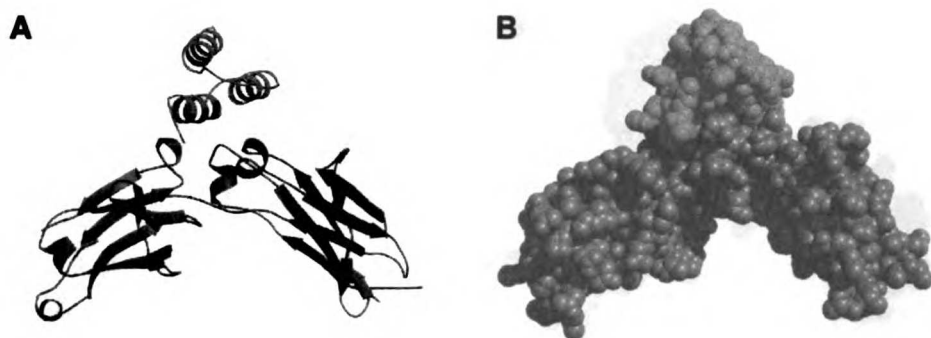


AMNH Q7 150M

**Figure 3.2: Crystal Structure of an EPOR/Peptide Complex.** This crystal structure of an engineered dimeric peptide in complex with the erythropoietin receptor (EPOR) [Livnah, 1996] demonstrated in a dramatic fashion that peptides can mimic the binding activities of larger protein domains.







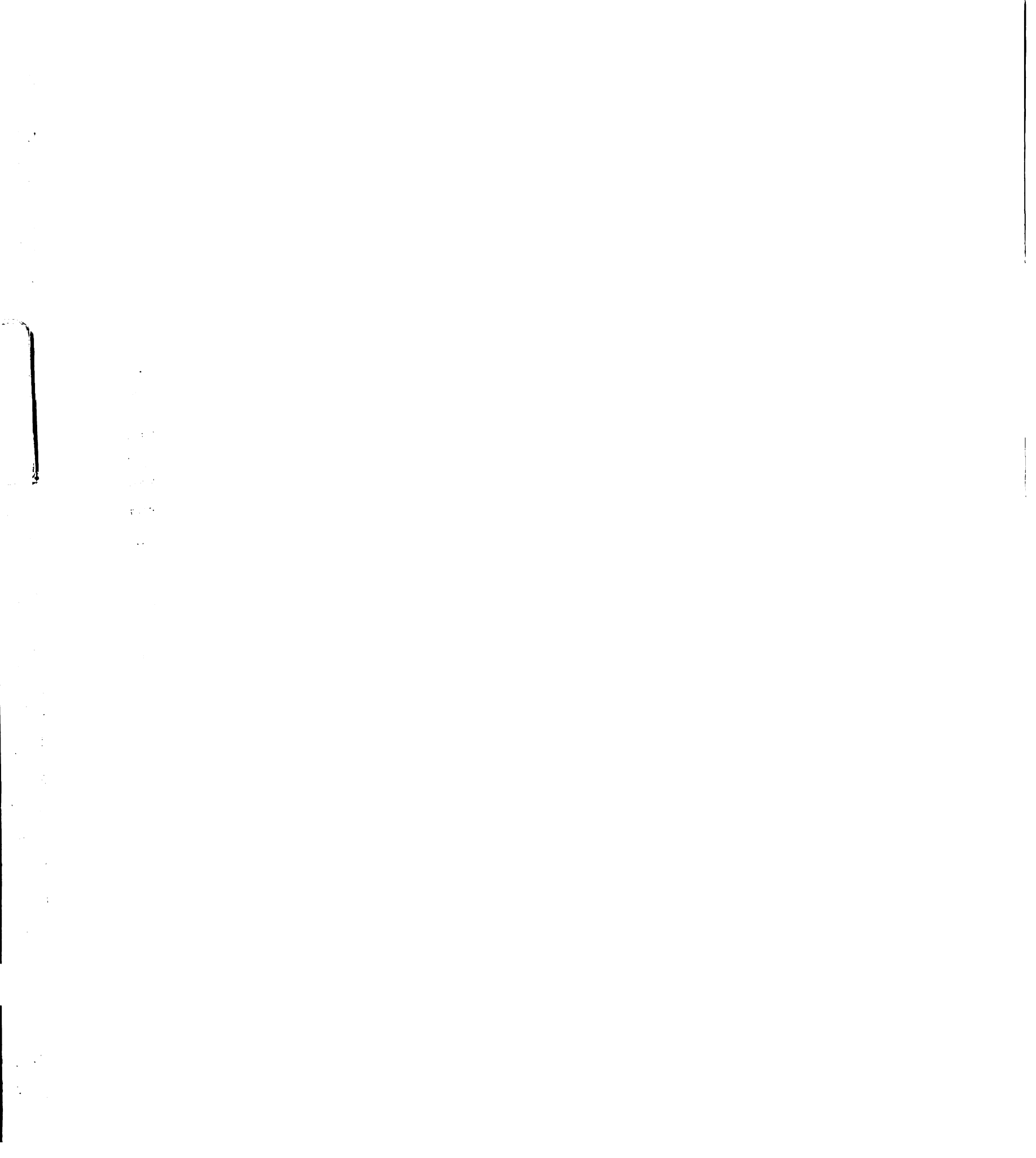
**Figure 3.3: Protein A.** Shown is the crystal structure of protein A, B-domain (red) in complex with a subunit of the Fc dimer (blue) [de Vos, 1998], in ribbon (A) and space-filling (B) representations.

1995]. Protein A actually contains five repeated variants of this domain and thus exhibits much higher affinity for Fc due to a polyvalent mode of interaction.

### 3.2.2 Protein G

Protein G and protein A have evolved independently in a convergent manner to adopt similar binding functions in *Streptococcus* G148 and *Staphylococcus aureus* respectively. Both apparently assist these bacteria in evading the mammalian immune systems. Remarkably, the structures of their Fc binding domains are entirely distinct. Protein G's Fc binding domain (Figure 3.4) is a mixed  $\alpha/\beta$  protein, whereas protein A's binding domain is entirely helical, and protein G uses a mixture of helix and sheet to bind Fc. Protein G's helix binds in an orientation distinct from either of the two helices used by protein A to bind Fc. However, both domains exhibit similar affinities for Fc of around 10 nM [Fahnestock, 1990], and they are approximately the same size ( $\approx 6.5$  kD).

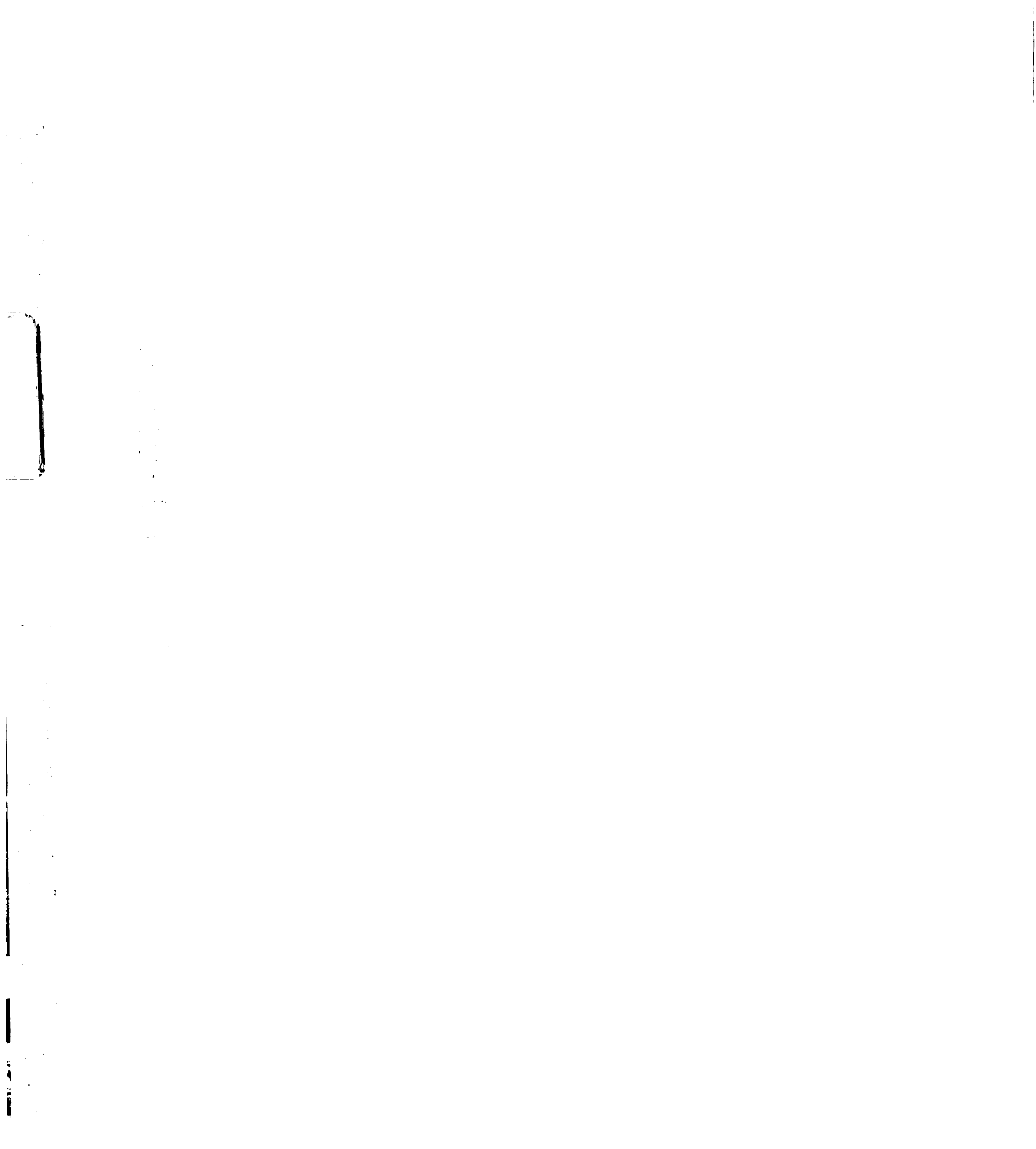
10/10/2017  
 10/10/2017  
 10/10/2017

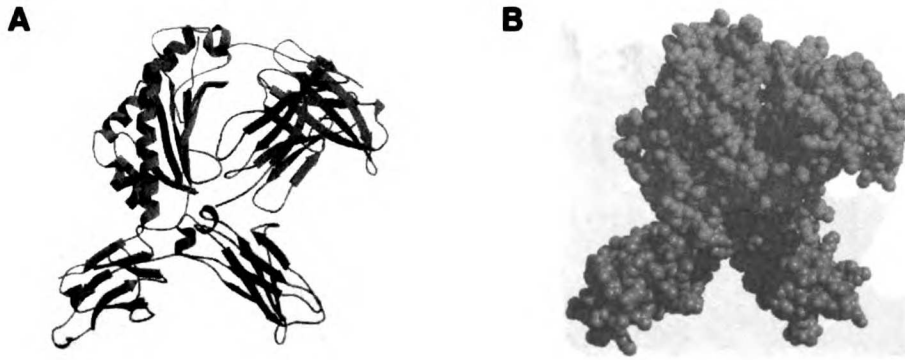












**Figure 3.6: Fc Receptor.** Shown is the crystal structure of neonatal Fc receptor (red) in complex with a subunit of the Fc dimer (blue) [Burmeister, 1994a], in ribbon (A) and space-filling (B) representations.

between the domains.

### 3.2.5 Protein H

Protein H is a protein from *Streptococcus pyogenes* which shows Fc binding activity that is competitive with protein A and Protein G [Frick, 1992], and thus it probably binds in the same region. Its structure is not yet known and it exhibits no sequence identity with protein A or any of the other Fc binding proteins. CD and sequence analysis suggests that the Fc binding portion of the molecule adopts a dimeric coiled-coil conformation in solution that would be distinct from three-helix bundle of protein A [Nilson, 1995]. Thus, Protein H may represent another unique Fc binding motif.

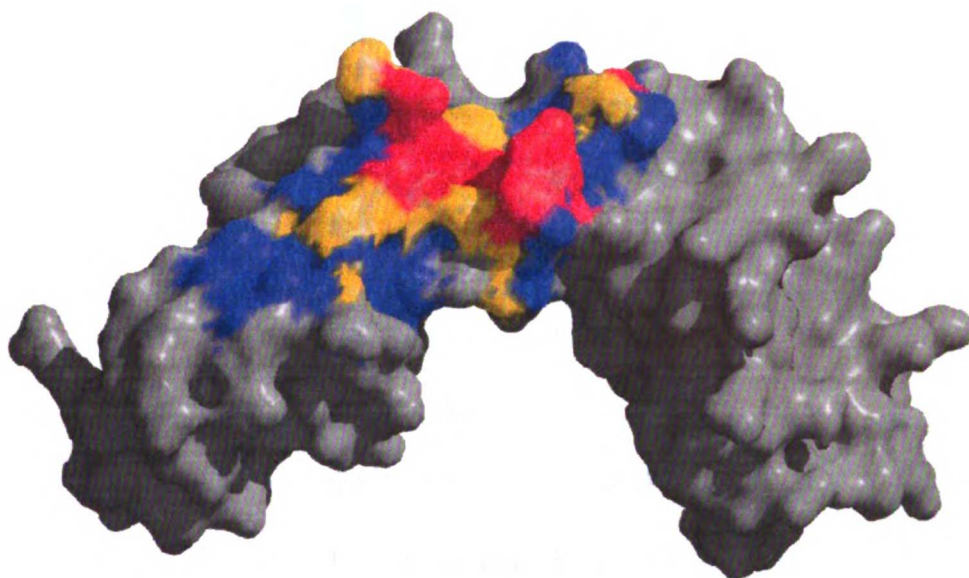
### 3.2.6 Identification of a Consensus Binding Site

Figure 3.1 shows that these proteins bind to an overlapping region on Fc. For the higher resolution structures (protein A, protein G, and rheumatoid Factor), the sets of atoms in each interface are known and it is possible to construct a map of

UNIVERSITY OF  
 CALIFORNIA  
 LIBRARY  
 STANFORD







**Figure 3.7: Consensus Binding Site.** Shown is a map of the overlap between binding sites of the natural Fc binding proteins whose Fc-complexes have been solved to high resolution (protein A, protein G, and rheumatoid Factor). Blue regions participate in a least one interface, yellow in two, and red regions are found all three interfaces.

the intersections of these sets (Figure 3.7). We find that there are approximately 35 atoms distributed over half a dozen amino acid side chains and adjacent main chain atoms found in all three Fc binding interfaces. These proteins are apparently recognizing a common feature on the surface of Fc.

We wondered whether naive peptides selected to bind Fc would also interact preferentially at this site. If binding to this region on Fc was selected in the natural domains for biological reasons alone, then peptides selected to bind Fc might bind anywhere on the surface of the dimer. If however, some intrinsic property of this site makes it a particularly good location for binding, then we would expect Fc binding peptides to cluster in this region as well.

F  
o  
X  
9  
3  
V  
s  
n  
b  
3  
C  
t  
s  
t  
a

$X_7CX_4CX_7$	$5.3 \times 10^8$
$X_7CX_5CX_6$	$5.6 \times 10^8$
$X_6CX_6CX_6$	$5.0 \times 10^8$
$X_6CX_7CX_5$	$6.3 \times 10^8$
$X_5CX_8CX_5$	$4.5 \times 10^8$
$X_4CX_9CX_5^*$	$1.9 \times 10^8$
$X_4CX_{10}CX_4$	$2.1 \times 10^8$
Total Diversity	$3.1 \times 10^9$

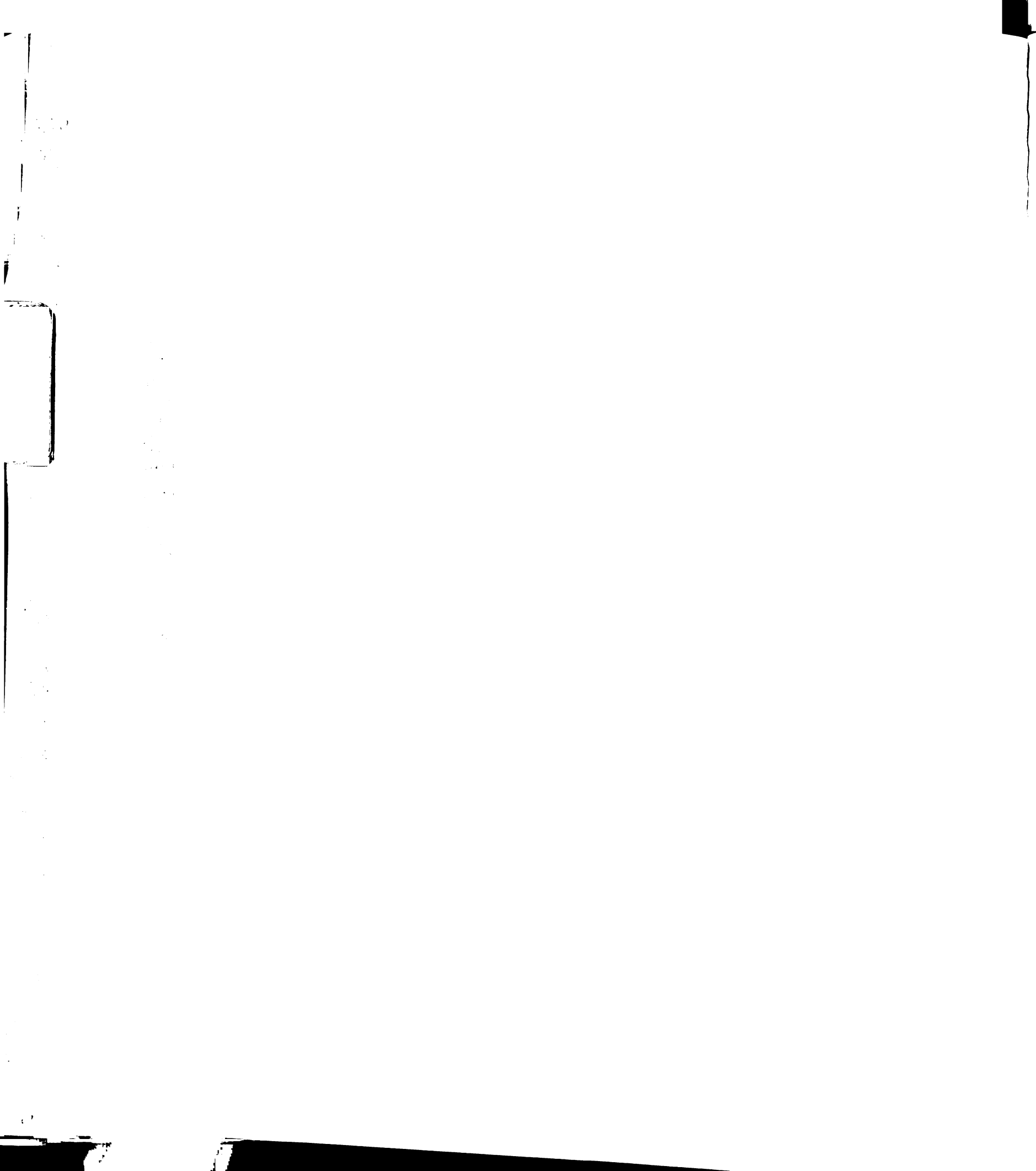
Figure 3.8: **Naive Peptide Library.** This library was designed to contain a variety of disulfide constrained peptide sequences with potential for binding proteins. X represents any random amino acid generated using an NNS DNA codon sequence. (\*) This library also contained sequences of the format  $X_5CX_9CX_4$ .

### 3.3 Initial Isolation of an Fc Binding Peptide

We used polyvalent M13 bacteriophage display of a random pool of peptides to select members which specifically bound Fc. This technique has achieved some remarkable successes during the past few years in finding novel peptides which bind to proteins [Arza, 1998, Koivunen, 1999].

#### 3.3.1 Methods

Construction of a the polyvalent M13 bacteriophage library of peptides fused to the gene VIII protein is described in [Lowman, 1998], and their composition is shown in Figure 3.8. The library construct contained an stII secretion signal peptide [Lee, 1983], the twenty residue peptide library ( $X_iCX_jX_k$  where X is a random amino acid from an NNS codon,  $i + j + k = 18$ , and  $j = 4 - 10$ ), a GGGSGGG



linker, and the M13 gene VIII starting at the first residue of the mature protein.

Selections were performed as described in [Lowman, 1998] and in Protocol B.6, with IgG<sub>1</sub>-Fc' immobilized as the target protein at (5 µg/mL). IgG<sub>1</sub>-Fc' was obtained from papain cleavage of CD4-IgG immunoadhesion followed by column chromatography as described in Protocol B.1. Fc purity was assessed at >95% by SDS-PAGE.

High levels of specific binding of library phage to immobilized were observed after only three rounds of selection. Up to 500-fold more phage were found to bind Fc coated wells compared to those coated only with blocking protein.

Clones were sequenced and the following two gene sequences obtained:

AAG GAG GCG AGC TGC TCC TAC TGG CTC GGC  
GAG CTG GTC TGG TGC GTC GCC GGG GTG GAG (FBP1.1)

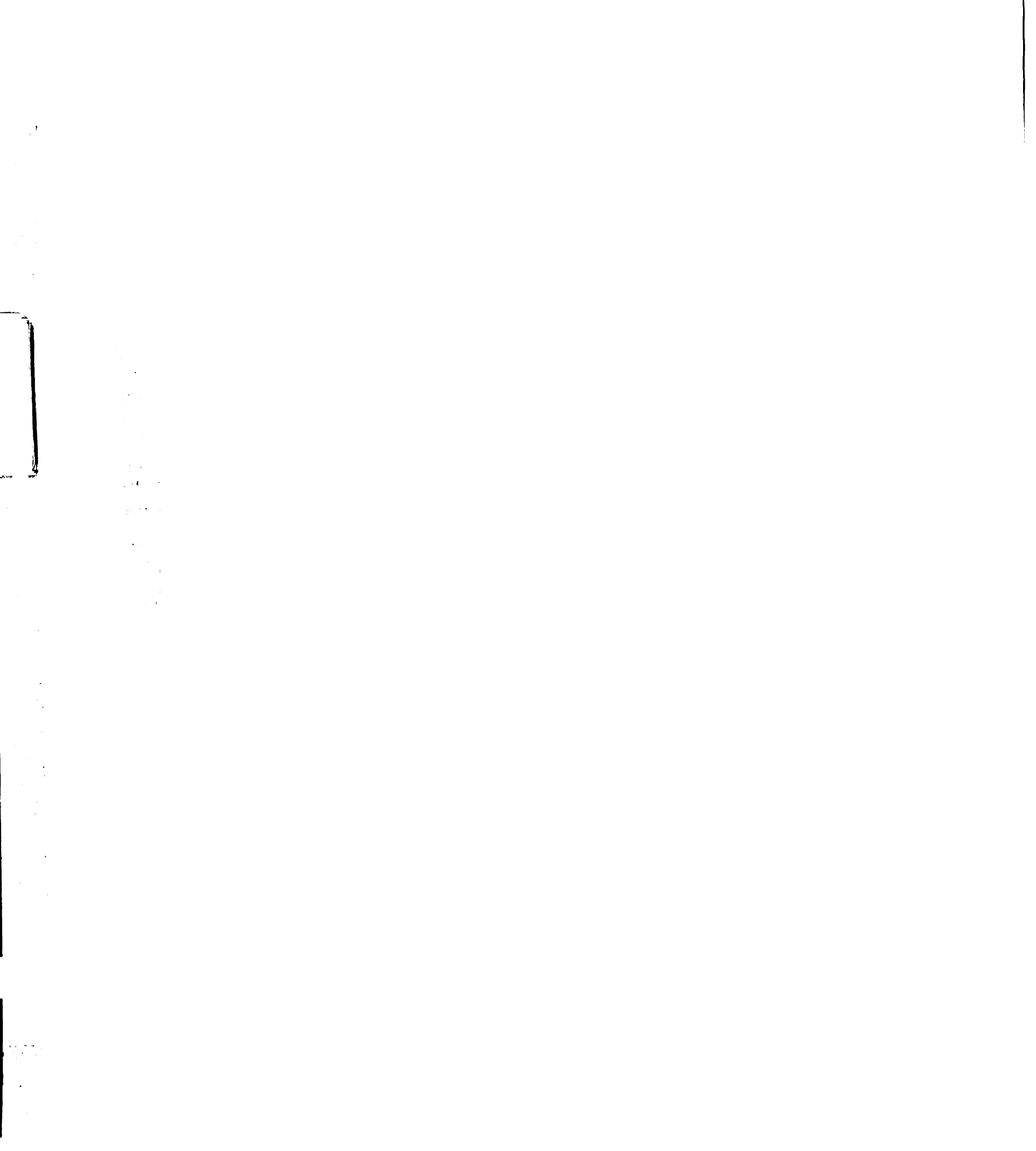
and

GAG ACG CAG AGG TGC ACC TGG CAC ATG GGC  
GAG CTG GTG TGG TGC GAG AGG GAG CAC AAC (FBP1.2)

Several repetitions of the selection experiment showed similar selection patterns but did not result in any additional sequences.

### 3.3.2 Results and Discussion

From the initial selections for binding against Fc, two and only two peptides sequences were obtained, FBP1.1 and FBP1.2 (Figure 3.9). Both peptides were of the X<sub>4</sub>CX<sub>9</sub>CX<sub>5</sub> format and shared seven residues of identity including the two cysteines. The other five residues constituted a contiguous GELVW sequence found in the C-terminus of the loop.



**FBP1.1**  
**K E A S C S Y W L G E L V W C V A G V E**  
**FBP1.2**  
**E T Q R C T W H M G E L V W C E R E H N**

**Figure 3.9: Initial Fc Binding Sequences.** These were the only Fc binding sequences isolated from the initial naive peptide library. They shared the cysteine positioning and a GELVW sequence near the N-terminus of the loop.

The fact that two and only two such Fc binding sequences were selected is an intriguing result. From an infinitely large library of peptides, we would expect to select a family of Fc binding peptide with varying affinities for Fc. Since we found only two sequences, any other Fc binding peptides in the starting pool must have had affinities substantially weaker than these peptides.

The critical elements in the peptide sequence appear to be the five residue GELVW sequence along with the cysteine spacing. There are  $20^5=3.2 \times 10^6$  different five residue sequences, but  $32^5=3.4 \times 10^7$  different DNA sequences are needed to exhaustively sample those peptide sequences when using NNS codons. The fraction of the naive peptide library that contained nine randomized residues between the two cysteines consisted of  $1.9 \times 10^8$  sequences. Thus, a about a five-fold excess of diversity was present, assuming that the positioning of the GELVW sequence was critical, that sampling was perfectly even, and that nothing else in the peptide mattered for or against binding.

In light of this analysis, perhaps it is not so surprising that only two peptides were selected. It also suggests that this experiment was just barely tractable given the size of our libraries. If six residues were absolutely required in order to bind Fc along with precise cysteine positioning, probably no Fc binding sequences would have been found.





### **3.4 Peptide Characterization**

The selected peptides showed substantial binding to Fc when displayed polyvalently on phage, but analysis in that format was difficult because of multiple site binding. To accurately obtain physical information it was necessary to both synthesize the peptide and transfer its gene to a monovalent bacteriophage display context [Lowman, 1991].

Characterization consisted of measuring the binding affinity of the peptide, observing and measuring competition with protein A Z-domain, and recording the CD spectrum of the peptide. Attempts to obtain structural information by NMR were unsuccessful due to aggregation of the peptide at high concentration.

#### **3.4.1 Methods**

##### **Peptide Synthesis**

The FBP1.1 and FBP1.2 peptides were synthesized with a C-terminal amide by solid phase peptide synthesis on Wang resin, and purified by reverse phase HPLC on a C-18 column. Cyclization was carried out using iodine/acetic acid and confirmed by electrospray mass spectrometry.

##### **Competition Binding Experiments**

Competitive Fc binding experiments between protein A Z-domain and the FBP1.1 and FBP1.2 peptides were performed as described in Protocol B.8. Competitions were carried out using immobilized Fc and biotinylated Z-Domain as well as in an inverse fashion, with immobilized Z-Domain and biotinylated Fc. The saturation level of biotinylated label on the immobilized protein was varied to enable extrapolation of the  $IC_{50}$  values down to zero label concentration in order to obtain  $K_i$



values.

### **Measurement of Circular Dichroism Spectra**

Circular dichroism (CD) spectra were acquired on an Aviv 60DS spectrometer (Aviv Associates) over the range 190-250 nM (0.2 nM intervals) in a cell with a path length of 0.1 cm at 25° C. Samples were at a 0.1 mg/ml concentrations of peptide in 100 mM NaCl and 10 mM Tris HCl pH 7.2.

### **Transfer to a Monovalent Phage Display Format**

The gene sequence of FBP1.1 was transferred to a monovalent phagemid vector by cassette mutagenesis to create vector pW0803. The resulting construct contained the stII signal sequence, the peptide KEASCSYWL GELVWCVAVE, a GGGPGGG linker, and the M13 gene III protein starting at residue 253.

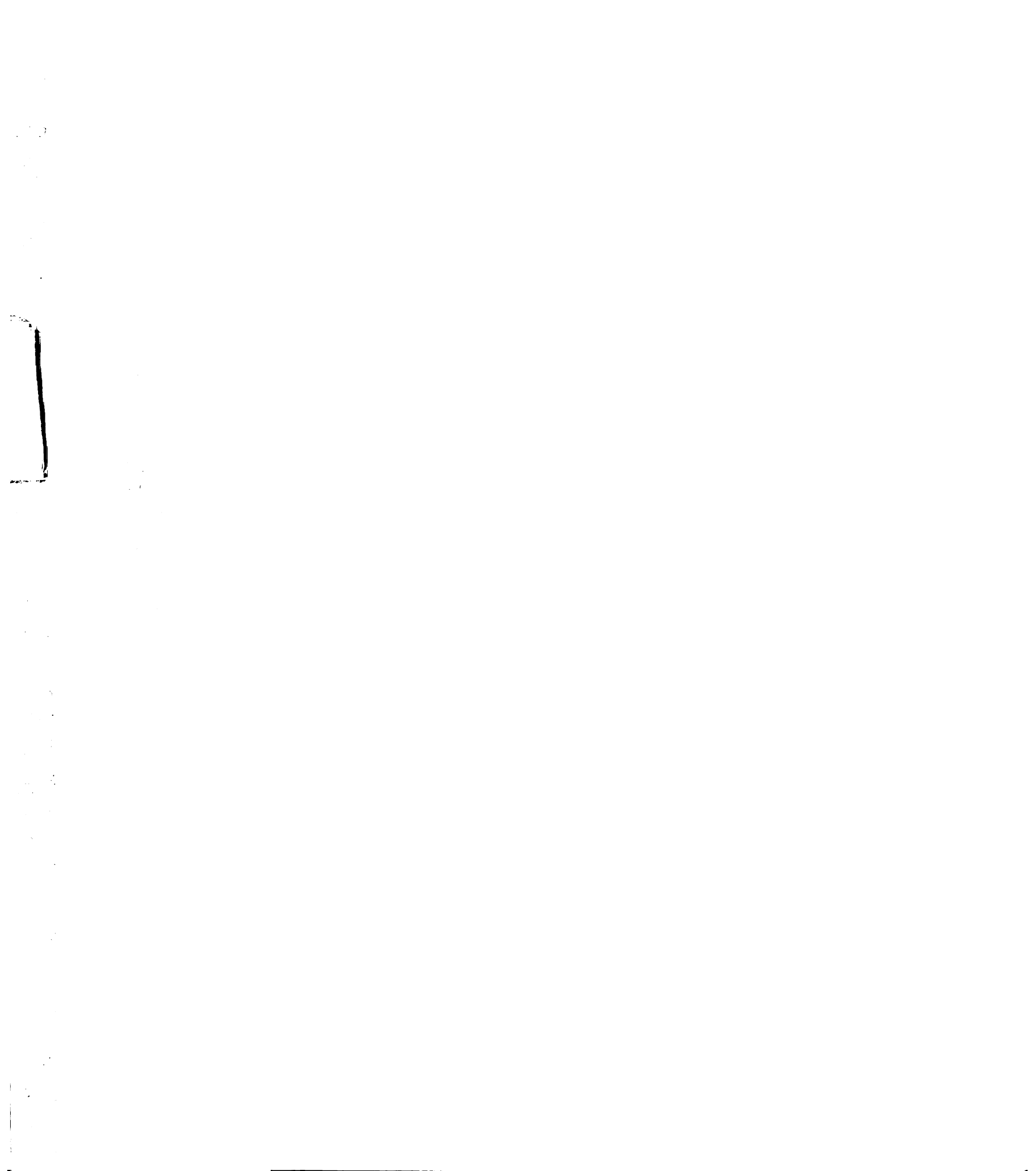
### **Binding Affinity Measurements on Phage**

Phage ELISA binding assays were performed on pW0803 using Protocol B.7. Z-domain phage produced from pABZD were used as a control.

### **3.4.2 Results and Discussion**

FBP1.1 and FBP1.2 were found to be fully competitive with Z-domain for binding to Fc, with  $K_i$ 's near 5  $\mu$ M. One such competition experiment is shown in Figure 3.10. This provided a strong indication that the peptide bound Fc in a region overlapping the protein A binding site on Fc, and it was consistent with possible interaction between the peptide and the consensus binding region on Fc.

Interestingly, the CD spectrum of the peptide in solution (Figure 3.11) showed



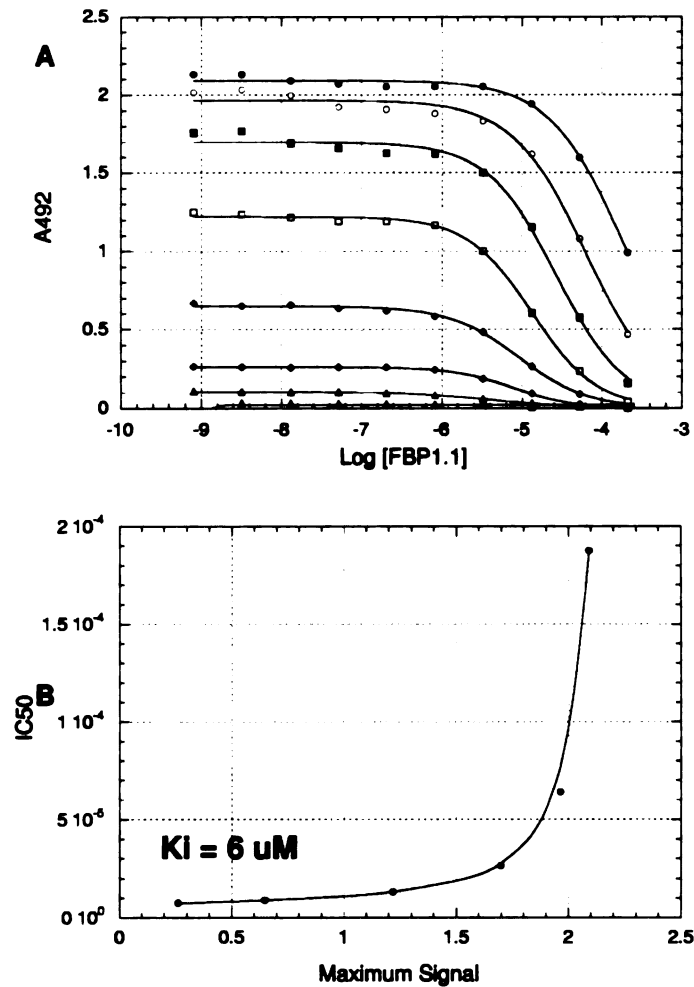


Figure 3.10: **Peptide Inhibition Curves.** (A) Competition binding measurements of the FBP1.1 peptide's ability to inhibit binding of biotinylated Fc to immobilized Z-domain were performed at a variety of biotin-Fc concentrations. (B) The  $IC_{50}$  values were extrapolated to zero label concentration in order to obtain a  $K_i$  for the peptide.

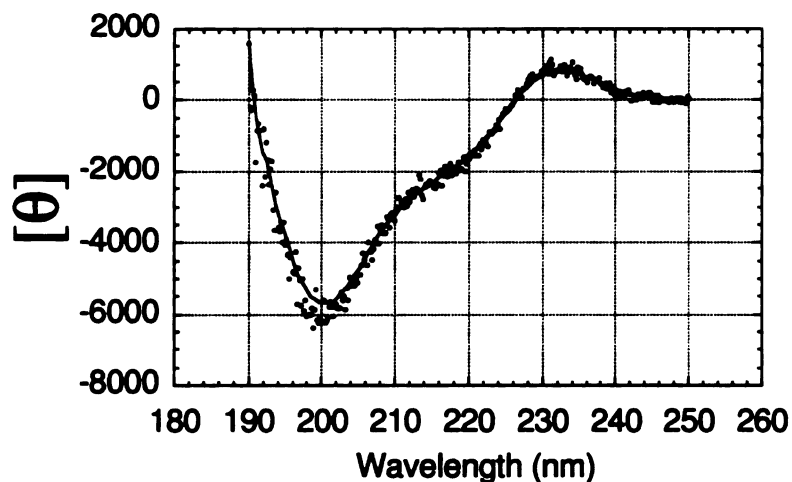
that the peptide did not adopt a stable helical structure like that of protein A, Z-domain or of the locked-helix peptides based on Z-domain. Instead, the CD spectrum suggested that the peptide was either disordered in solution or might adopt some kind of extended  $\beta$  structure.

The selected gene for peptide FBP1.1 was transferred from a polyvalent phage display format onto a monovalent phage display format (pW0803). Phage pro-



11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

Fig  
sp  
the  
is  
  
du  
to  
in:  
an



**Figure 3.11: Circular Dichroism Spectrum of a Selected Peptide.** Shown in the spectrum of FBP1.1 alone in solution. The spectrum is dramatically different from that of the locked-helix peptides (Figure 2.10), and suggests that its conformation is distinct and not helical.

duced using pW0803 showed an  $EC_{50}$  of  $2 \mu M$  for binding immobilized Fc, close to the  $5 \mu M$  inhibition constant for antagonization of the Fc/protein A Z-Domain interaction. The peptide's  $EC_{50}$  is about 400-fold weaker than that of Z-domain in an equivalent binding assay.





### 3.5 Local Optimization

While an affinity near  $5 \mu\text{M}$  is quite good in comparison to the locked-helix peptides, the manner in which the peptide had been selected suggested that further optimization of the peptide affinity would indeed be possible. Because the original library contained only  $3.1 \times 10^9$  combinations out of a possible  $2.6 \times 10^{20}$ , the chance was very slight that all positions would have their best Fc binding combinations present in the starting library, and so the two selected sequences were expected to be sub-optimal for binding. Many of the residues in the 20-mer peptide were likely to still have potential for improvement.

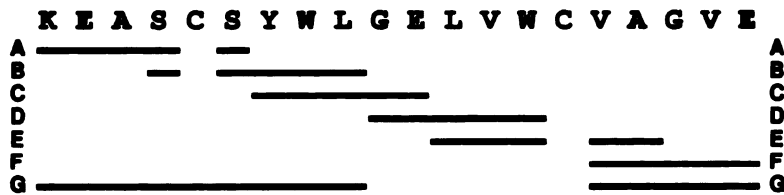
Because we had essentially no information about the probable structure of the Fc binding peptide, I adopted an exhaustive scanning library strategy for optimization. In this approach, a family of libraries were prepared in which the diversity was focused on a short contiguous stretch of residues.

However, because the cysteine positioning was specifically selected from a library of peptides with varied cysteine spacings, it was assumed to have already been optimized, and so the cysteine positions were not changed in any of the subsequent optimization work.

#### 3.5.1 Methods

##### Library Design and Preparation

Libraries were designed as shown in Figure 3.12 using NNS codons in the randomized positions. In order to prevent the parent pW0803 sequence from dominating the selection, library template DNA was generated from pW0806, a variant of pW0803 in which positions 12-15 in the peptide have been mutagenized into two stop codons and a single base deletion (CTG GTC TGG became TGA TGA



**Figure 3.12: Design of Local Optimization Libraries.** From top to bottom, the solid bars represent the regions mutated in the local optimization libraries (A through G). These regions were mutated in a nearly-exhaustive fashion using NNS codons.

GG).

All mutagenesis was carried out by the method of [Kunkel, 1987] using oligos with 18 bases of identity flanking the regions of mutation, and the libraries A-G were prepared according to Protocol B.5. The final library sizes were (A)  $1.1 \times 10^8$  (B)  $1.6 \times 10^8$  (C)  $1.4 \times 10^8$  (D)  $6.2 \times 10^7$  (E)  $9.4 \times 10^7$  (F)  $1.2 \times 10^8$  and (G)  $1.5 \times 10^8$ .

### Selection

Six rounds of selection were performed against Fc using the method described in Protocol B.6. Measurements of specific and non-specific binding were made by quantitating the phage bound to Fc coated wells against that of phage bound to wells without Fc. Consistent increases in the specific binding of phage to Fc were observed in all seven libraries (Figure 3.13).

After the sixth round of selection, at least 18 clones from each library were sequenced. Sequences represented more than once in the population of sequenced clones were then selected for further characterization in a phage ELISA binding assays. The Fc binding activity of fourteen clones was measuring using Protocol B.7.

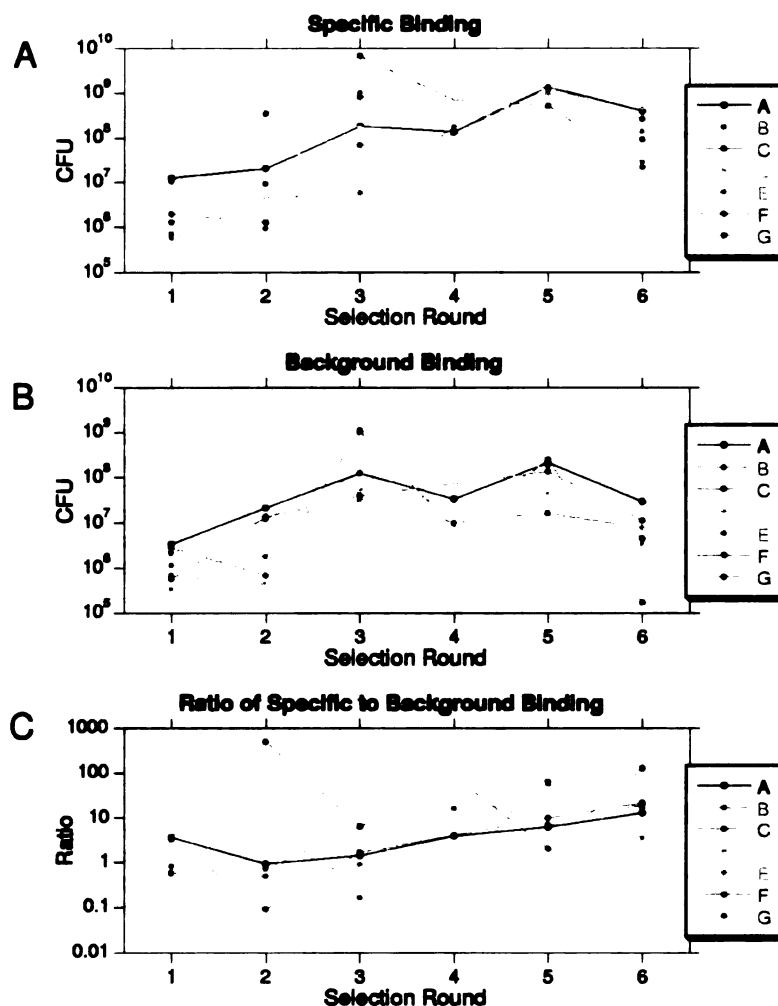


Figure 3.13: **Screening of the Local Optimization Libraries.** Across each of the libraries, (A) shows the amount of phage (in colony forming units) eluted from Fc-coated wells, and (B) shows the amount eluted from wells without Fc. The ratio of these figures in (C) shows how the phage populations were enriched with Fc binding sequences.

### 3.5.2 Results and Discussion

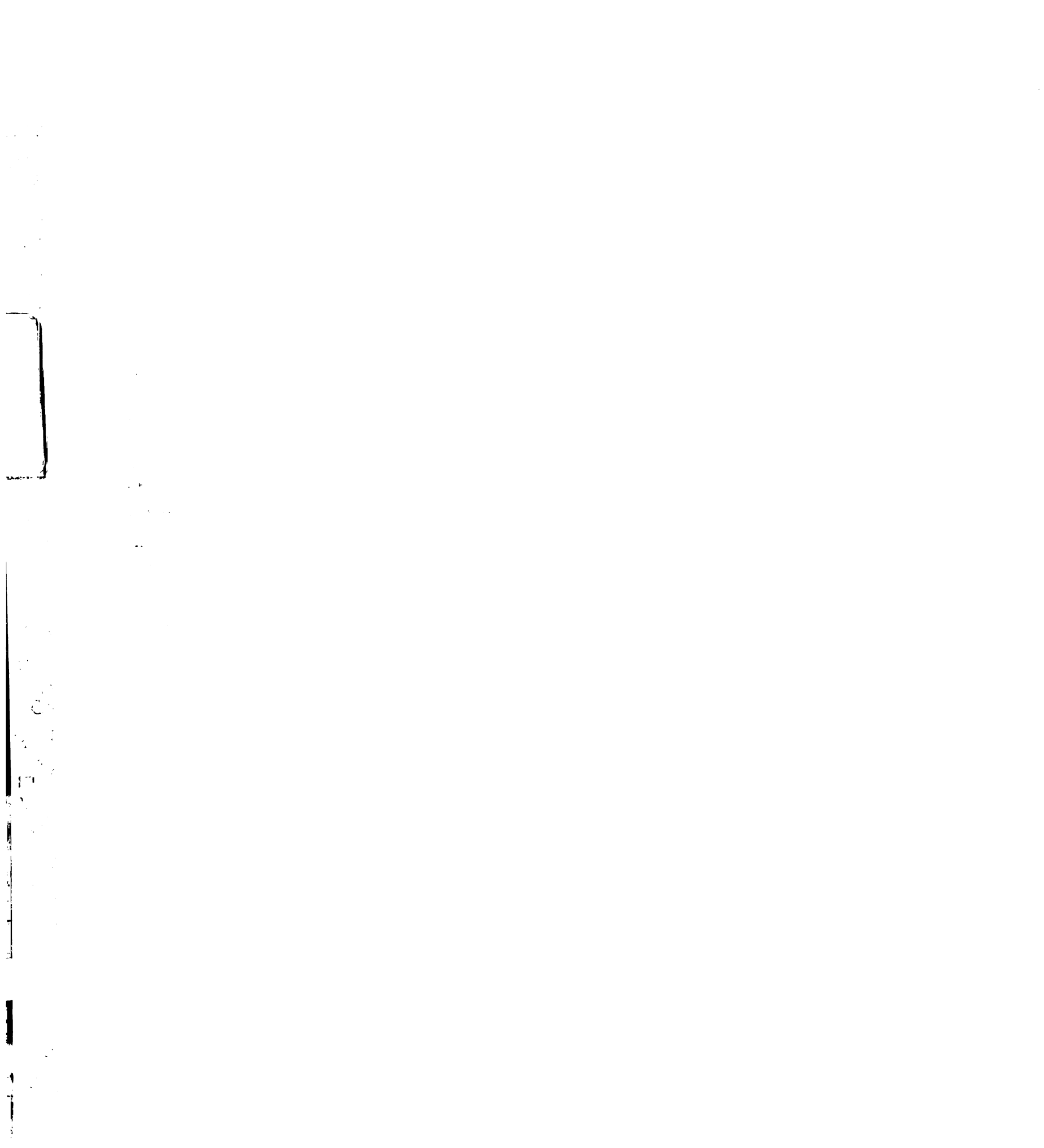
The increases observed in the specific binding of library phage particles to immobilized Fc (Figure 3.13) indicated that the populations of phage in the libraries were being enriched for molecules that bound tightly to Fc.

Sequencing of a sample of the library populations after the sixth round of selec-



Sequence	Occurrence	EC50 (nM)
ESED <b>CS</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	1	
EKED <b>CS</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	1	
EDPD <b>CS</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	1	
EFAD <b>CS</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	1	
NADD <b>CS</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	1	
SETT <b>CS</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	1	
AWKT <b>CQ</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	1	
DLAD <b>CS</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> R <b>V</b> E <b>G</b>	2	94
KEAD <b>CA</b> W <b>H</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	14	138
KEAE <b>CS</b> Y <b>H</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	1	
KEARC <b>W</b> Y <b>H</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> D <b>P</b> E <b>E</b>	1	809
KEASC <b>S</b> Y <b>H</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	10	416
KEASC <b>S</b> W <b>H</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> A <b>G</b> V <b>E</b>	1	225
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>T</b> E <b>G</b> V <b>E</b>	17	818
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>D</b> D <b>G</b> V <b>E</b>	9	838
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> E <b>G</b> V <b>E</b>	5	743
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> P <b>G</b> V <b>E</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> D <b>K</b> S <b>G</b> V <b>E</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>D</b> N <b>G</b> V <b>E</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>D</b> T <b>L</b> T <b>E</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> P <b>G</b> V <b>E</b>	2	734
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> G <b>V</b> E <b>G</b>	3	776
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> A <b>G</b> V <b>E</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>D</b> T <b>F</b> D <b>E</b>	9	301
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>D</b> G <b>L</b> D <b>E</b>	3	328
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>V</b> G <b>L</b> D <b>E</b>	2	278
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>E</b> D <b>T</b> L <b>E</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>E</b> D <b>T</b> M <b>E</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>E</b> D <b>M</b> M <b>E</b>	1	
WVED <b>CS</b> W <b>H</b> M <b>G</b> E <b>L</b> V <b>W</b> C <b>D</b> G <b>G</b> E <b>F</b>	14	139
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>D</b> W <b>M</b> N <b>G</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>D</b> D <b>T</b> P <b>V</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>D</b> D <b>Y</b> G <b>E</b>	1	
KEASC <b>S</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> D <b>L</b> W <b>E</b>	1	
WRGG <b>CS</b> W <b>H</b> M <b>G</b> E <b>L</b> V <b>W</b> C <b>E</b> H <b>D</b> M <b>E</b>	1	
AVSK <b>CS</b> F <b>F</b> M <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> D <b>V</b> M <b>N</b>	1	
NQVS <b>CS</b> Y <b>S</b> R <b>G</b> E <b>L</b> V <b>W</b> C <b>S</b> K <b>Q</b> S <b>Q</b>	1	
GRME <b>CA</b> W <b>H</b> Q <b>G</b> E <b>L</b> V <b>W</b> C <b>T</b> P <b>T</b> L <b>E</b>	1	
GTME <b>CS</b> W <b>H</b> Q <b>G</b> E <b>L</b> V <b>W</b> C <b>T</b> P <b>T</b> L <b>A</b>	1	
EMRD <b>CS</b> W <b>H</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>A</b> H <b>M</b> E <b>G</b>	9	
GSWE <b>CA</b> Y <b>H</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>E</b> T <b>G</b> S <b>G</b>	1	
VAEP <b>CA</b> Y <b>H</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>E</b> V <b>L</b> K <b>G</b>	1	
KEAM <b>CS</b> Y <b>W</b> L <b>G</b> E <b>L</b> V <b>W</b> C <b>E</b> S <b>D</b> M <b>E</b>	1	

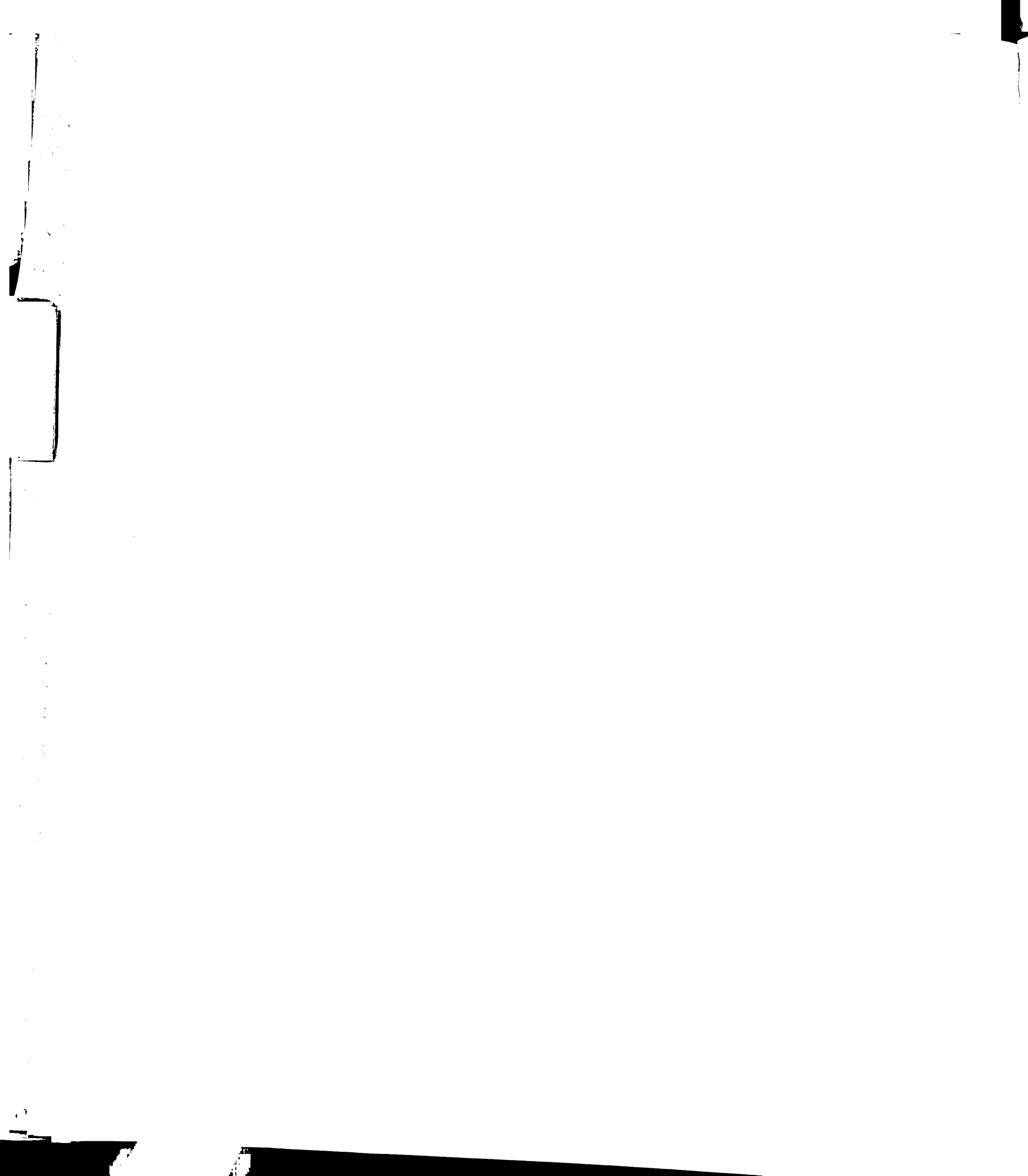
Figure 3.14: Selected Sequences from Local Optimization. Shown are the sequences obtained from the local optimization libraries. Residues that differ from the initial Fc binding sequence (FBP1.1) are shown in red. Also shown is the occurrence of these sequences in the pool of sequenced clones along with the Fc binding affinities (EC<sub>50</sub>'s) for most sequences present more than once in the population.



tion showed that the libraries were consensing towards particular sequences (Figure 3.14). Since all peptides in each library are in competition with one another, it would be expected that peptides with higher affinity would be the sequences most represented in the library. When those sequences were then run through Fc binding assays, it was found that they indeed had increased binding affinities for Fc.

The highest affinity peptides from the local optimization libraries bound Fc with  $EC_{50}$ 's near 100 nM. This is about ten-fold tighter than the starting sequence FBP1.1 which bound with an affinity of 1  $\mu$ M ( $EC_{50}$ ) in the same binding assay. A variety of different changes were selected in the peptide that could give similar improvements of binding affinity, so it appears that there are many ways to achieve better binding with this peptide.

Various regions of the molecule exhibited different levels of consensus, as quantified in Figure 3.15. For example, despite the fact that portions of the GELVW sequence were randomized in three libraries (C-E), only a single clone was found that contained any mutations in this region of the peptide. Thus, residues in this region must be highly sensitive to mutation and are likely play a critical function in the molecule. In comparison, residues in the N-terminal portion of the loop showed moderate consensus with preference for certain residues at each position, and the N and C termini of the peptide showed very little consensus. In the next section, we apply these consensus patterns in a global fashion to create fully optimized Fc binding peptides.





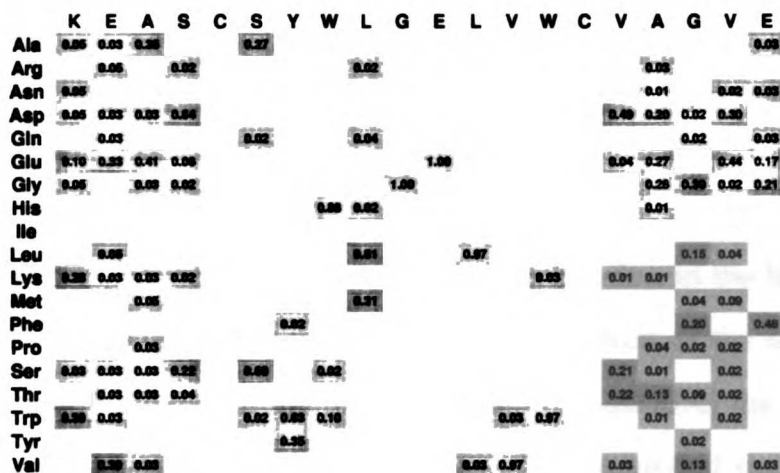


Figure 3.15: **Residue Distribution from Local Optimization.** Shown is the frequency of occurrence for specific residue types at each position in the peptide. Residues in gray boxes were present in less than 10% of clones, those in blue were selected in 10%-75%, and those in red were highly conserved, occurring in >75% of selected sequences. These frequencies were only tabulated for positions which had been randomized.

### 3.6 Global Optimization

While the local optimization libraries produced a series of changes which individually gave up to a hundred-fold increase in affinity, these optimizations were selected independently, and many of these affinity improvements were likely to be incompatible with one another. In order to find out which combinations of improvements were compatible, a secondary global round of optimization was necessary.

In this global optimization, the combinatorial nature of the genetic code was exploited using an original software program to design several libraries with preferred residues distributed throughout the peptide gene.

121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200



### **3.6.1 Methods**

#### **Library Design**

Three libraries were designed based on selection patterns observed in the local optimization libraries (Figures 3.14 and 3.15). For each codon in the library, a combination of bases was used that would produce a mixture of the desired amino acids as well as any other combinations that might result from the genetic code. The first library, A, consisted of those residues which exhibited strong conservation ( $\geq 10\%$ ) in sequenced clones. Library B contained those residues which were found in a minimum of two clones ( $\geq 2\%$ ), and library C contained residues seen at least once in any clone.

The composition of the libraries is shown in Figure 3.16. The potential diversity of library A is low enough that it is possible to exhaustively sample almost the entire library in a single selection experiment. Library B and especially library C are too large to be exhaustively sampled. The actual oligo design was carried out using a custom computer program (DESIGN) written for this purpose and which generated the output in Figure 3.16.

#### **Library Preparation**

Considerable difficulty was encountered during attempts to prepare these libraries using Kunkel mutagenesis, so it was necessary to fall back on a less efficient cassette mutagenesis approach. Oligonucleotides containing the randomized bases were extended and made double stranded by standard fill in and ligation reactions with additional oligonucleotides so as to include flanking XbaI and ApaI sites corresponding to those in pW0806. The vector and the insert were then double digested, purified, and ligated together. DNA from this ligation reaction

A		DRG	GWA	GMA	RRC	TGC	KCT	TRS	CAC	MTG	GGC	GAG	CTG	GTC	TGG	TGC	RVC	RVM	BKC	GAS	KDW
1.	DRG:		E	G	K			R	W	O	(6)										Cumulative Size: 6.00e+00
2.	GWA:		E						V		(2)										1.20e+01
3.	GMA:	A	E								(2)										2.40e+01
4.	RRC:	D	G			N	S				(4)										9.60e+01
5.	TGC:	C									(1)										9.60e+01
6.	KCT:	A						S			(2)										1.92e+02
7.	TRS:	C							W	Y	(4)										7.68e+02
8.	CAC:		H								(1)										7.68e+02
9.	MTG:				L	M					(2)										1.54e+03
10.	GGC:		G								(1)										1.54e+03
11.	GAG:	E									(1)										1.54e+03
12.	CTG:				L						(1)										1.54e+03
13.	GTC:								V		(1)										1.54e+03
14.	TGG:								W		(1)										1.54e+03
15.	TGC:	C									(1)										1.54e+03
16.	RVC:	A	D	G		N	ST				(6)										9.22e+03
17.	RVM:	A	DE	G	K	N	RST				(12)										1.11e+05
18.	BKC:	C	FG	L	R	V					(6)										6.64e+05
19.	GAS:	DE									(2)										1.33e+06
20.	KDW:	CDEFG	L	V	YO						(12)										1.59e+07

B		DRS	VWG	SVG	RRC	TGC	KCC	TRS	YRS	MTG	GGC	GAG	CTG	GTC	TGG	TGC	RNC	VVS	NBS	GWS	KDM
1.	DRS:	C	D	E	G	K	N	RS	W	Y	(12)										Cumulative Size: 1.20e+01
2.	VWG:	E				K	L	M	Q	V	(6)										7.20e+01
3.	SVG:	A	E	G		P	Q	R	S		(6)										4.32e+02
4.	RRC:	D	G			N	S				(4)										1.73e+03
5.	TGC:	C									(1)										1.73e+03
6.	KCC:	A				S					(2)										3.46e+03
7.	TRS:	C						W	Y	(4)											1.38e+04
8.	YRS:	C	H			Q	R	W	Y	(8)											1.11e+05
9.	MTG:				L	M					(2)										2.21e+05
10.	GGC:		G								(1)										2.21e+05
11.	GAG:	E									(1)										2.21e+05
12.	CTG:				L						(1)										2.21e+05
13.	GTC:							V			(1)										2.21e+05
14.	TGG:							W			(1)										2.21e+05
15.	TGC:	C									(1)										2.21e+05
16.	RNC:	A	D	G	I	N	STV				(8)										1.77e+06
17.	VVS:	A	DE	GH	K	NPQRST					(18)										3.19e+07
18.	NBS:	AC	FG	I	LM	P	RSTVW				(24)										7.64e+08
19.	GWS:	DE				V					(4)										3.06e+09
20.	KDM:	CDEFG	L	V	YO						(12)										3.67e+10

C		DNS	NNS	NNS	VNS	TGC	BVG	TDS	HRS	MDS	GGC	GAG	STC	KKG	WRG	TGC	RNM	NNS	NNS	NNS	NNM
1.	DNS:	ACDEFG	I	K	L	M	N	R	S	T	V	W	Y	(24)							Cumulative Size: 2.40e+01
2.	NNS:	ACDEFGH	I	K	L	M	N	P	Q	R	S	T	V	W	Y	(32)					7.68e+02
3.	NNS:	ACDEFGH	I	K	L	M	N	P	Q	R	S	T	V	W	Y	(32)					2.46e+04
4.	VNS:	A	DE	G	H	I	K	L	M	N	P	Q	R	S	T	V	(24)				5.90e+05
5.	TGC:	C									(1)										5.90e+05
6.	BVG:	A	E	G		P	Q	R	S	W	O	(9)									5.31e+06
7.	TDS:	C	F		L			W	Y	(6)											3.19e+07
8.	HRS:	C	H	K	N	Q	R	S	W	Y	(12)										3.82e+08
9.	MDS:		H	I	K	L	M	N	Q	R	S	(12)									4.59e+09
10.	GGC:		G								(1)										4.59e+09
11.	GAG:	E									(1)										4.59e+09
12.	STC:				L			V			(2)										9.17e+09
13.	KKG:		G	L		V	W				(4)										3.67e+10
14.	WRG:			K	R	W	O				(4)										1.47e+11
15.	TGC:	C									(1)										1.47e+11
16.	RNM:	A	DE	G	I	K	N	R	S	T	V	(16)									2.35e+12
17.	NNS:	ACDEFGH	I	K	L	M	N	P	Q	R	S	T	V	W	Y	(32)					7.51e+13
18.	NNS:	ACDEFGH	I	K	L	M	N	P	Q	R	S	T	V	W	Y	(32)					2.40e+15
19.	NNS:	ACDEFGH	I	K	L	M	N	P	Q	R	S	T	V	W	Y	(32)					7.69e+16
20.	NNM:	ACDEFGH	I	K	L	N	P	Q	R	S	T	V	Y	(32)							2.46e+18

**Figure 3.16: Design of the Global Optimization Libraries.** An original computer program (DESIGN) was used to generate oligonucleotide sequences for the global optimization libraries (A-C), which incorporated only specific bases at each position. The specific mixtures of bases will give the combinations of amino acid residues selected in the local optimization libraries: (A) in >10% of clones, (B) in >2% of clones, or (C) in any clone. The IUPAC codes for base mixtures are: **B=C,G,T**; **D=A,G,T**; **H=A,C,T**; **K=G,T**; **M=A,C**; **N=A,C,G,T**; **R=A,G**; **S=G,T**; **V=A,C,G**; **W=A,T**; **Y=C,T**.



was then purified and used to generate libraries as described in Protocol B.5, step 5. In the end, libraries of the following sizes were obtained: (A)  $3.9 \times 10^7$  (B)  $6.3 \times 10^7$  and (C)  $4.6 \times 10^7$ .

## **Selection**

Again, six rounds of selection were performed against Fc using the method described in Protocol B.6. Measurements of specific and non-specific binding were made by quantitating the phage bound to wells with immobilized Fc to the phage bound to wells without Fc. High levels in the specific binding of phage to Fc were observed in all three libraries (Figure 3.13). In the final round of selection, the specific binding reached as high as 5000-fold.

After the final round of selection, at least 18 clones from each library were sequenced, and most of these were further characterized in phage ELISA binding assays (Protocol B.7).

### **3.6.2 Results and Discussion**

Very high levels of specific binding (up to 5000-fold) were obtained during the selections of the global optimization libraries. By using the selection patterns from local optimization in a combinatorial fashion, it appears that we were able to evolve peptides with even higher affinities for Fc.

Phage ELISA binding assays on many of the selected sequences demonstrated that the high level of specific binding observed during selection translated into higher binding affinities for Fc. As shown in Figure 3.17, several of the selected sequences showed  $EC_{50}$  affinities below 10 nM, which represents a 100-fold improvement over the starting sequences and a 10-fold improvement over peptides selected from the local optimization libraries.



1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

	Sequence	Occurrence	EC50 (nM)
<b>A</b>	KVADCANHMGELVWCTEVEG	1	23
	GEEDCSYHLGELVMCTELDD	1	69
	GVADCANHMGELVWCTERED	1	
	GEEDCANHMGELVWCSGGDF	1	100
	WEADCANHMGELVWCTKVEE	1	7
	GEADCSYHLGELVWCHDFEE	1	156
	WVDCAYHLGELVWCSTFEE	1	9
	WVEDCANHMGELVWCTKVDE	1	70
	READCANHMGELVWCSEIDL	2	47
	KEASCAYHLGELVWCDAFDV	1	77
	KVASCANHMGELVWCDGLDG	1	
	GEADCANHMGELVWCTKVEE	1	38
	GEASCAYHLGELVWCDEGGG	1	386
	KVEDCAYHLGELVWCTEGDE	1	63
	<b>B</b>	KEPDCSWHLGELVMCTPMIEV	1
KEADCANHMGELVWCSEIEG		1	66
EQADCANHMGELVWCTPMVF		1	8
KEPDCSWHLGELVWCTPIEV		1	15
GEPPCANHMGELVWCTPMVF		12	7
GEQDCSYHMGELVWCTTFVGG		1	210
GVRMCAYHLGELVWCTPMIEF		1	10
KVADCANHMGELVWCSEIEV		1	44
GEADCANHMGELVWCTPMDL		2	
GEQDCSWHLGELVWCTPMIEV		1	
GEQDCSYHLGELVWCSDMEL		1	
<b>C</b>		EVADCSWHLGELVWCTEIEF	1
	GEEDCANHMGELVWCTDVED	1	52
	EVEDCAYHLGELVWCSDIEG	1	82
	WEEDCANHMGELVWCSEIEFDE	1	44
	KEASCANHMGELVWCSEIEE	1	130

Figure 3.17: **Selected Sequences from Global Optimization.** Shown are the sequences selected in the global optimization libraries along with their occurrences and Fc binding affinities ( $EC_{50}$ 's, where available).

### 3.7 Alanine Scanning

Given that the peptide competed with protein A for binding to Fc, it was expected that the two molecules would bind to an overlapping site. However, we wondered whether the functional binding epitopes of the two molecules would be similar or distinct.

Peptide FBP2.1 was the highest affinity peptide ( $EC_{50}=94$  nM) available after local optimization, so it was chosen for use as the binding partner in an alanine scan of the Fc surface. It had the sequence:





D L A D C S W H L G E L V W C S R V E G

Since no structural information was available about the interaction of the peptide and Fc, the same set of alanine scanning mutants prepared for the Z-domain alanine scan was used in scanning the Fc:peptide interaction on the Fc side of the interface.

After global optimization, the highest affinity peptide was FBP3.1, which has the sequence:

W E A D C A W H L G E L V W C T K V E E

This peptide was selected for alanine scanning in order to determine which regions would be most important in binding Fc.

### **3.7.1 Methods**

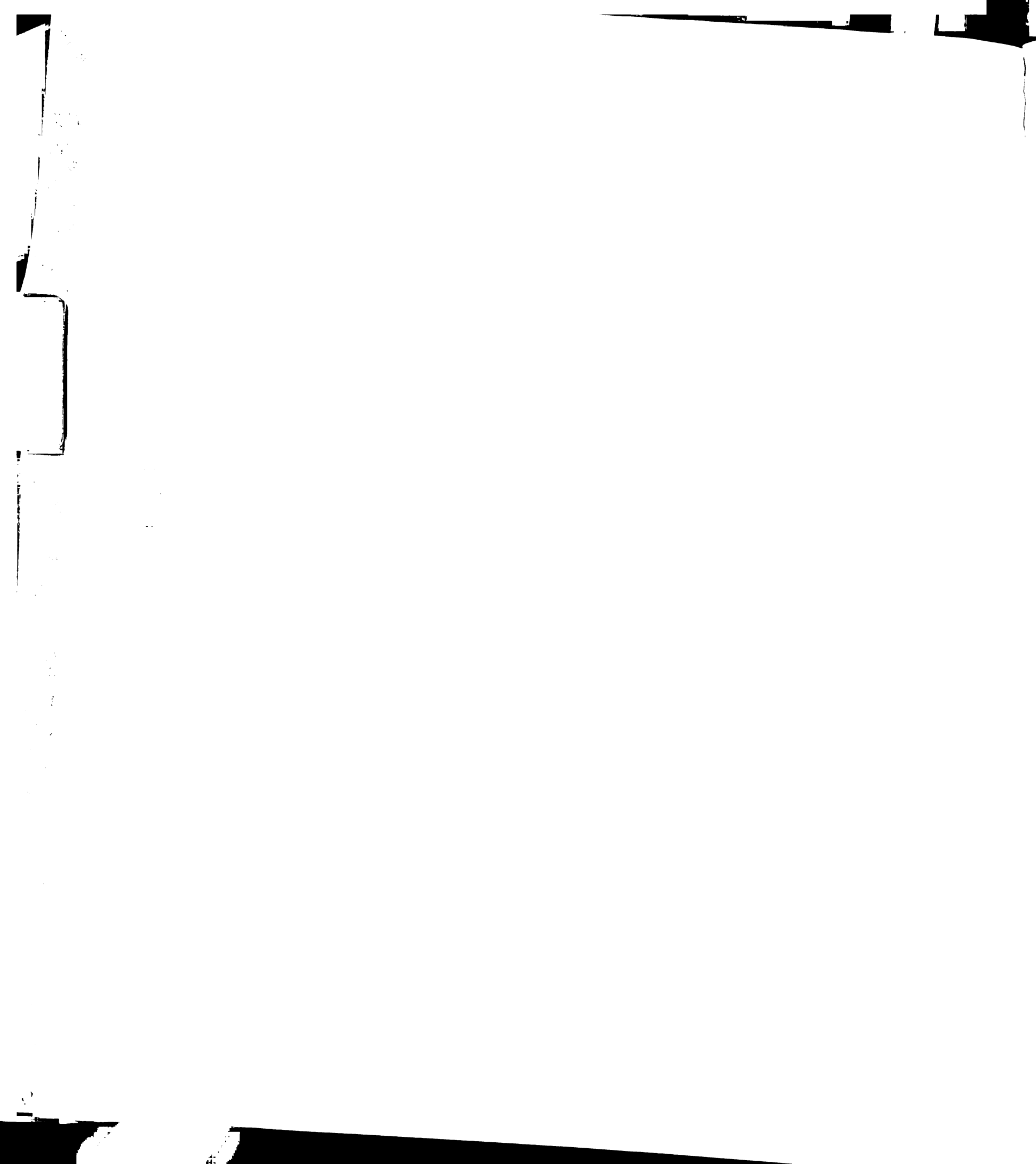
#### **Peptide Synthesis**

FBP2.1 was synthesized with a C-terminal amide by solid phase peptide synthesis on Wang resin, purified by reverse phase HPLC on a C-18 column. Cyclization using iodine/acetic acid was confirmed by electrospray mass spectrometry.

A portion of FBP2.1 was then biotinylated at the sole amine group in the peptide using EZ-link sulfo-NHS-LC-LC-biotin (Pierce) at a 1:1 ratio in 50 mM carbonate buffer at pH 6.0. Completion (>90%) of this reaction was monitored by analytical HPLC and then the reaction was quenched with excess ethanolamine.

#### **Mutagenesis**

Mutants of Fc in a phage display context were prepared as described in Section 2.3.



Single mutants of vector pW0910 containing the FBP3.1 peptide were prepared using mutagenesis [Kunkel, 1987]. Conservative alanine substitutions were made at each residue except for the two cysteines, Ala3, Ala5, and Gly10, which were replaced with proline, alanine, and serine respectively. Mutagenic oligonucleotides were prepared with 18 bases of identity flanking the mutated residue. All mutants were confirmed by dideoxy sequencing [Sanger, 1977]. The last residue of the peptide (Glu20) was not successfully mutated.

### **Binding Assays, Fc Mutants**

ELISAs were performed as described in Protocol B.7 for each mutant along with one nonmutant control per plate, with the following modifications. Avidin was coated onto the plates at 5  $\mu\text{g}/\text{mL}$  overnight. Plates were washed 5X with PT, and blocked 1 hr with CBB. Biotinylated FBP2.1 at 20 nM concentration was incubated on the avidin coated plates for 1 hour, then the plates were washed 10X with PT. The remainder of the assay was performed as described. Peptide concentrations during the competition phase of the assay ranged from 20  $\mu\text{M}$  down to 11 pM over twelve wells.

### **Binding Assays, Peptide Mutants**

ELISAs were performed as described in Protocol B.7 for each mutant along with one nonmutant control per plate. Fc was immobilized at a coating concentration of 10  $\mu\text{g}/\text{mL}$ .

Assays were performed in duplicate at uniform saturation levels of 25%, a condition where the ratio of  $\text{EC}_{50}$  measurements for two molecules will closely track that ratio of the the dissociation constants (described in Section A.4.1).



For each variant, twelve-point competition curves were recorded with solution Fc concentrations ranging from 2.7  $\mu$ M down to 15 pM in three-fold serial dilutions. Curves were fit by computer to a 3-parameter sigmoidal curve and used to determine midpoint concentrations on the competition curves, and ratios were computed for the mutants relative to the unmutated control peptides.

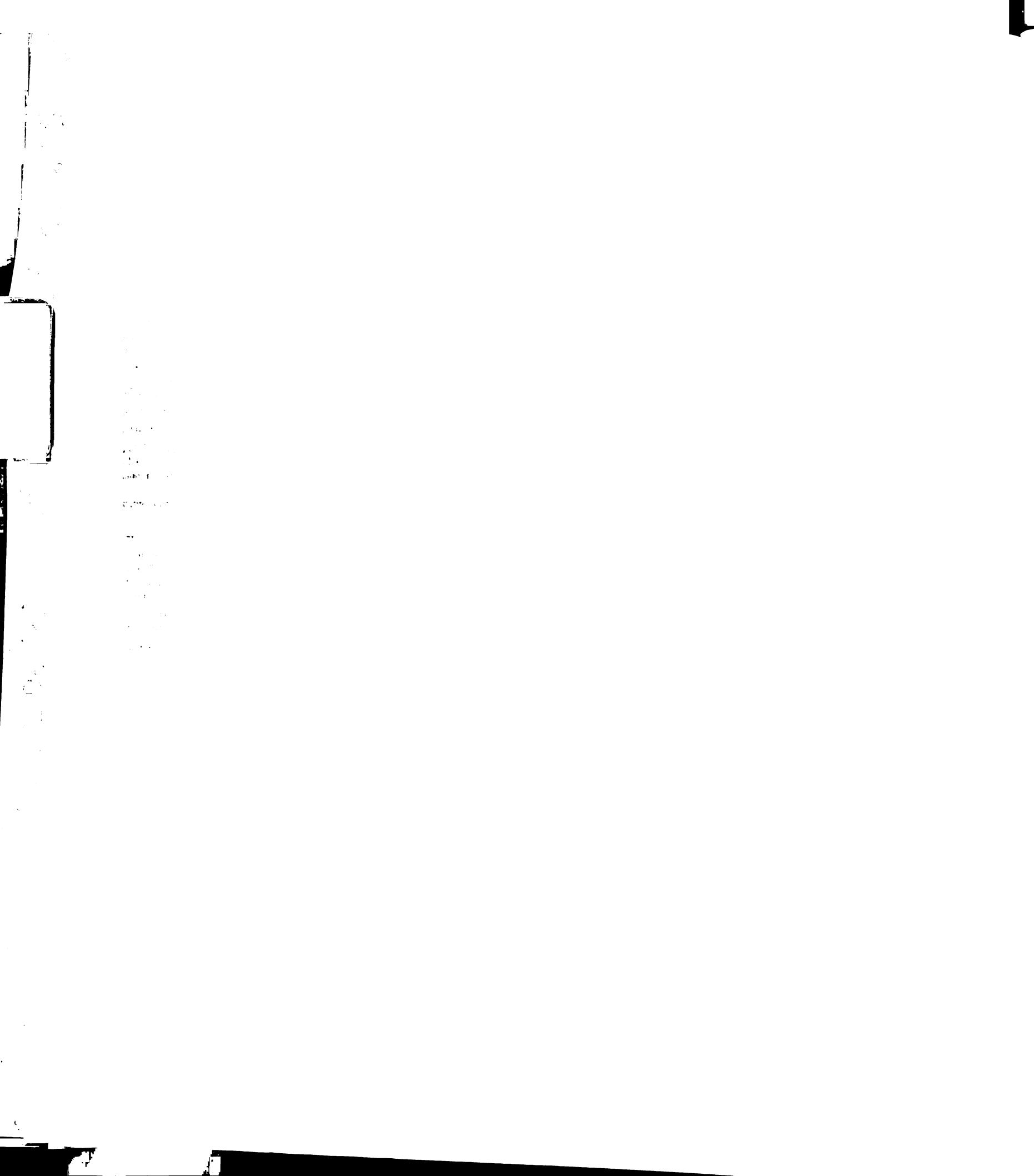
### **3.7.2 Results and Discussion**

#### **Fc Alanine Scan**

A comparison of alanine scans of Fc binding to Protein A, Z-domain or to the FBP2.1 peptide (Figure 3.18) shows that the patterns of molecular recognition in the two systems are distinct. Although both binding partners are sensitive to mutation of His310 and His435, the peptide is not greatly affected by mutation of Ile253. Instead, the peptide is quite sensitive to mutation of Asn434 or Tyr436. Thus, although FBP2.1 and Protein A have overlapping binding sites on Fc, the relative importance of the various interactions in that region are different. This difference further suggests that the peptides bind Fc in a manner distinct from that of Protein A.

#### **Peptide Alanine Scan**

The alanine scanning results on peptide FBP3.1 are shown in Figure 3.19. Clearly, the most important parts of the peptide for binding Fc are located within the loop region of the peptide. In particular, Leu12, Val13, and Trp14 near the end of the loop are tremendously important. Removal of any one of these three residues results in about a 4 kcal/mole drop in binding affinity. Also of significance are residues Trp7 and His8, residues that showed strong consensus (69% and 88%



### A. Protein A: Z-Domain

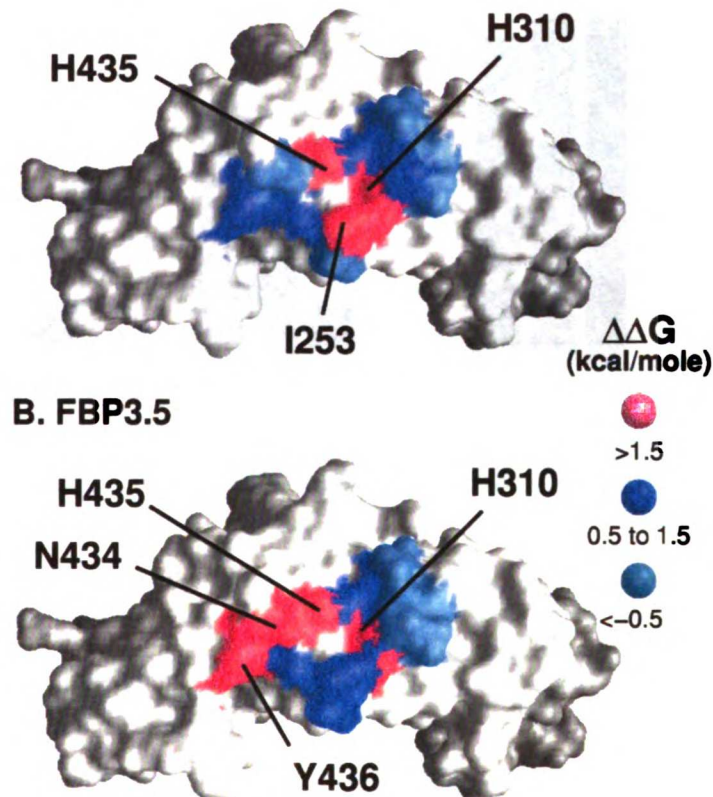
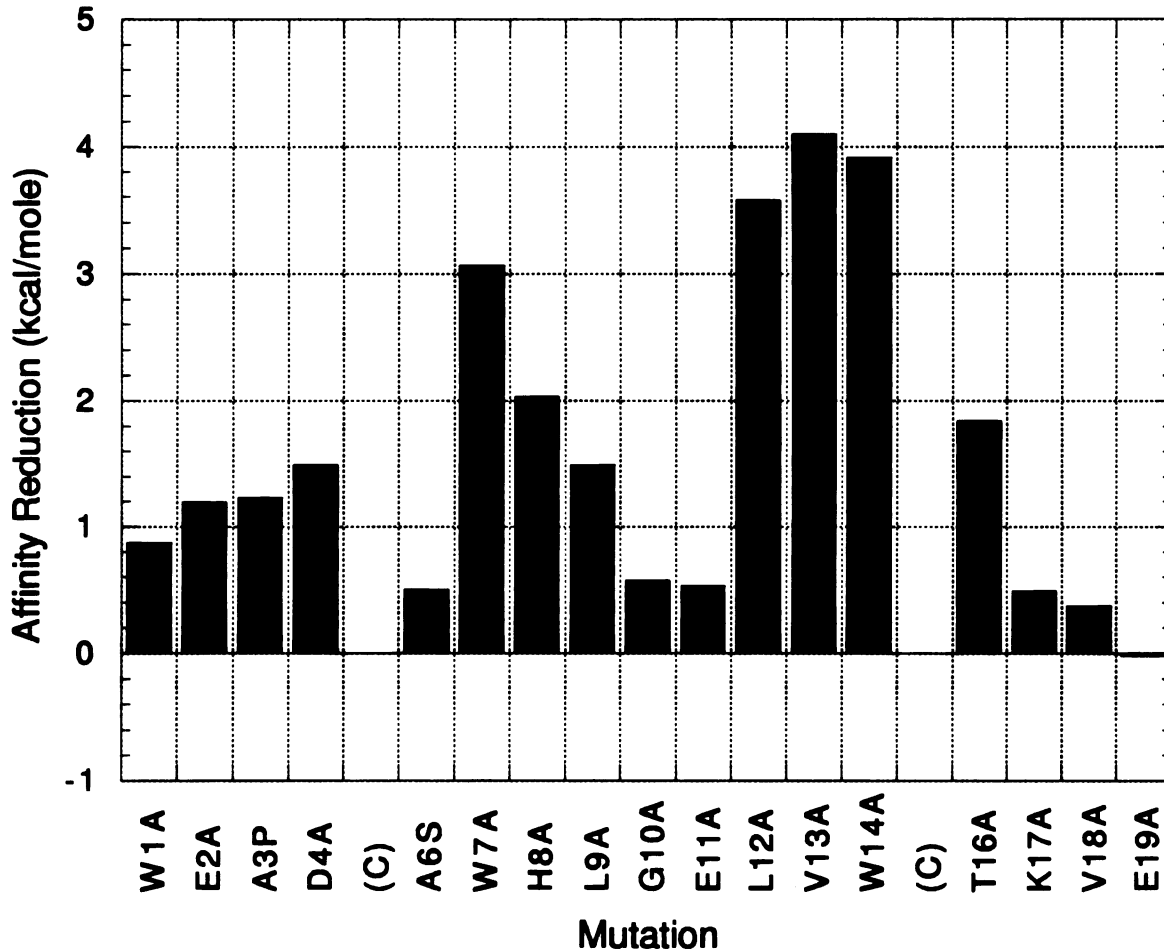


Figure 3.18: **Comparison of Protein A and Peptide Hot Spots.** Shown is a comparison of the alanine scanning results on Fc for binding to (A) Protein A, Z-domain or to (B) a selected Fc binding peptide. The mutated residues are color-coded according to their disruptive effect when replaced with alanine.

respectively) in the local optimization libraries.

An apparent correlation exists between the importance of a residue as measured by alanine scanning (Figure 3.19) and the extent to which it was conserved during selection (Figure 3.15). However, the reverse is not uniformly true. Both Gly10 and Glu11 were highly conserved during selection, but substitution of either of these residues with alanine has little effect on the affinity. One explanation





**Figure 3.19: Alanine scan of an Fc Binding Peptide.** Shown is an alanine scan of the Fc binding peptide, FBP3.1, residues 1-19. Note, Ala3 and Ala6 were instead substituted with proline and serine respectively, which were both residues found at these positions in some selected clones.

for such a puzzling result is that these residues confer some other form of selective advantage which increases the fitness of these peptides relative to others in the population without affecting the ultimate affinity for Fc. Such advantages might include superior expression, better display, faster folding, or perhaps improved resistance to proteolysis.



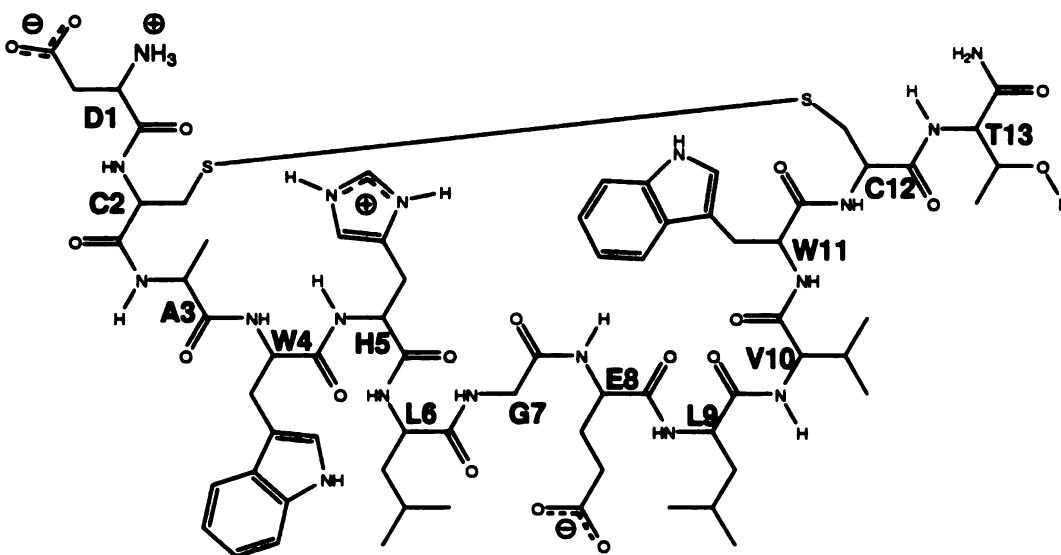


Figure 3.20: **Minimized Fc Binding Peptide.** Selection patterns from the global optimization libraries combined with alanine scanning results on peptide FBP3.1 suggested that this 13-residue peptide (FBP4.1) might be capable of binding Fc with high affinity.

### 3.8 Further Minimization

The alanine scan of FBP3.1 demonstrated that most of the important binding determinants were located inside or immediately adjacent to the the loop region. Therefore, it seemed quite possible that an even smaller version of the Fc binding peptide would be capable of binding Fc with high affinity.

#### 3.8.1 Methods

##### Peptide Design

Residues 4-16 of FBP3.1 were highly representative of the selection patterns seen in the global optimization libraries (Figure 3.17) and were therefore selected as a design for FBP4.1, a minimized 13 residue Fc-binding peptide (Figure 3.20). As with the other peptides, charge at the C-terminus of the peptide was neutralized



by designing the peptide with a C-terminal amide in order to better mimic the parent molecule.

### **Peptide Synthesis**

FBP4.1 was synthesized with a C-terminal amide by solid phase peptide synthesis on Wang resin, and purified by reverse phase HPLC on a C-18 column. Cyclization using iodine/acetic acid was confirmed by electrospray mass spectrometry.

### **Competition Binding Assays**

FBP4.1 peptide and protein A: Z-domain were used in competition binding assays as described in Protocol B.8. Z-domain was coated at 10  $\mu\text{g}/\text{mL}$  concentration, and biotinylated Fc was used in solution as the label at about a 25% saturation level. Assays were performed at both pH 7.2 in PBS and at pH 6.0 in 10 mM MES, 100 mM NaCl. Competitor concentrations were varied from 1  $\mu\text{M}$  down to 50 pM in three-fold dilutions.

### **$K_i$ Determination of FBP4.1**

A series of competition binding assays (Protocol B.8) with FBP4.1 were performed against biotinylated Fc binding to immobilized Z-domain. Z-domain was coated onto plates at 10  $\mu\text{g}/\text{mL}$  concentration and the concentration of biotinylated Fc was varied to achieve saturation binding levels between 5 and 80% . FBP4.1 concentration was varied between 1.25  $\mu\text{M}$  and 560 pM in 2-fold dilutions. The  $\text{IC}_{50}$  values obtained were extrapolated down to zero label concentration in order to compute a  $K_i$  for the peptide.



## Surface Plasmon Resonance Binding Assays

Fc was immobilized on an CM-5 chip using standard protocols (Pharmacia). All experiments were performed using a running buffer of 25 mM MES pH 6.0 with 0.05% tween-20. Kinetic and steady state binding of the FBP4.1 peptide to Fc was measured at concentrations from 6.1  $\mu$ M down to 3 nM in two-fold dilutions.

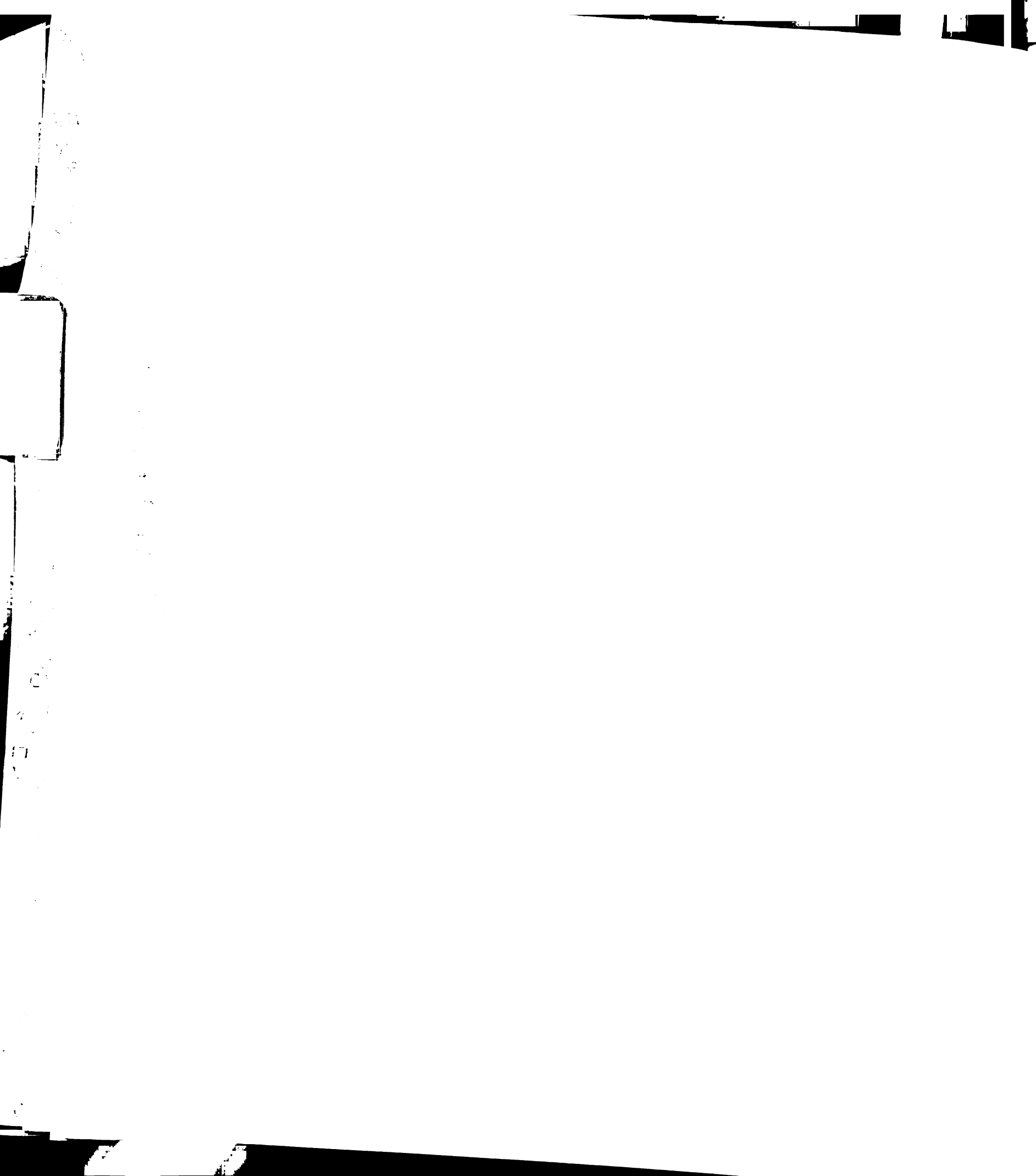
### 3.8.2 Results and Discussion

The minimized peptide (Figure 3.20) was indeed found to compete for binding to Fc. Interestingly, the pH dependence of peptide binding appeared to be nearly opposite that of Protein A:Z-domain. Figure 3.21 shows that while Z-domain loses about half an order of magnitude in binding affinity on going from pH 7.2 down to pH 6.0, the Fc peptide gains in affinity to a comparable extent. This opposite differential behavior is another clue that the binding interactions of the peptide are distinct from that of Z-domain.

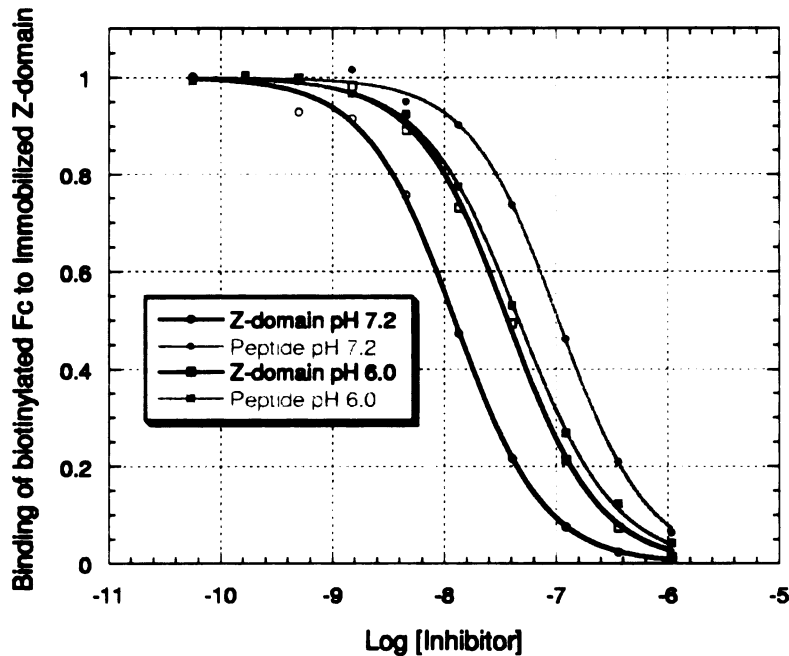
We find the binding affinity of the peptide for Fc at its optimum pH of 6.0 is quite high. The data shown in Figure 3.22 give a  $K_i$  of 24 nM for the peptide inhibiting the Protein A, Z-domain:Fc interaction. That result is backed up by kinetic binding studies of the peptide (Figure 3.23) which give  $k_{on} = 1.6 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_{off} = 2.6 \times 10^{-2} \text{ s}^{-1}$ , and  $K_d = 16 \text{ nM}$ .

The near-steady state association levels from Figure 3.23 were used to compute a steady-state binding curve (Figure 3.24). Under steady state conditions, the peptide appears to have a  $K_d = 14 \text{ nM}$  binding affinity for Fc.

Thus, a peptide of comparable size to the locked-helix Protein A variants was created which bound Fc with an affinity comparable to that of Protein A. How is this binding accomplished? The lack of sequence homology between Protein A







**Figure 3.21: Fc Binding and pH Dependence of FBP4.1.** The FBP4.1 peptide was able to inhibit the interaction between protein A, Z-domain and biotinylated Fc with an  $IC_{50}$  below 100 nM (red) at pH 7.2. This inhibition was increased at pH 6.0, down to an  $IC_{50}$  of about 25 nM. Interestingly, Z-domain (black) exhibits an opposite pH dependence.

and the selected peptides, and the alanine scanning results on the Fc side of the interface all suggest that the peptide is interacting with Fc a fashion distinct from Protein A. What is the structure of this peptide? Where does the peptide bind Fc? Does it mimic one of the other natural Fc binding domains or does it employ a novel Fc recognition motif? The answers to these and other pressing questions will be addressed in the next chapter as we examine the crystal structure of one such peptide in complex with Fc.



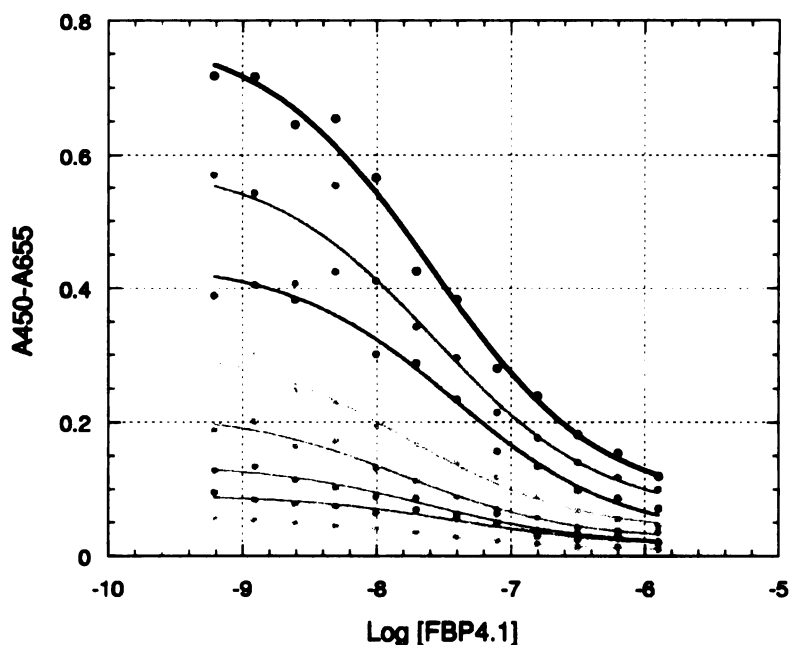
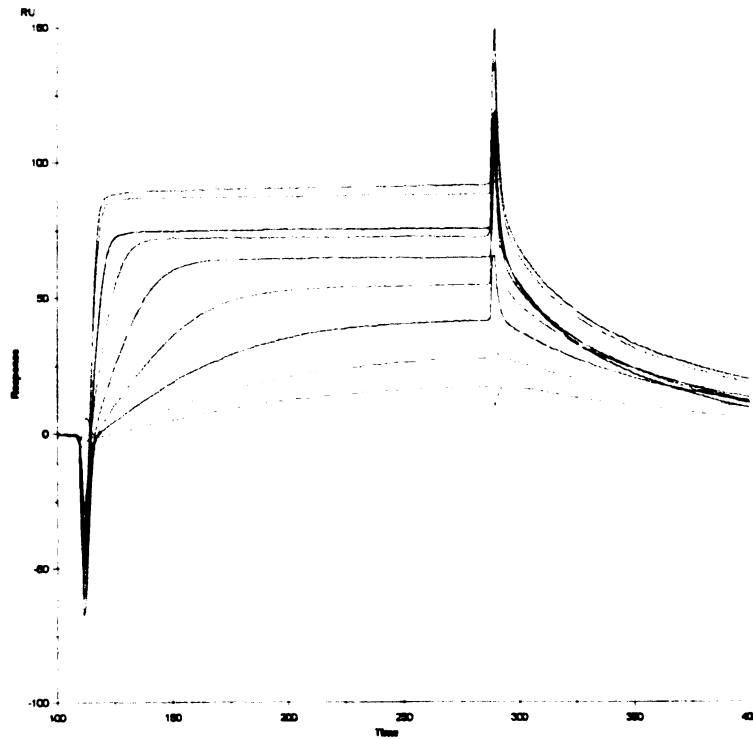


Figure 3.22:  $K_i$  Determination for FBP4.1. Competition binding measurements of the FBP4.1 peptide's ability to inhibit binding of biotinylated Fc to immobilized Z-domain were performed at a variety of biotin-Fc concentrations. A  $K_i$  of 24 nM was obtained from extrapolation.

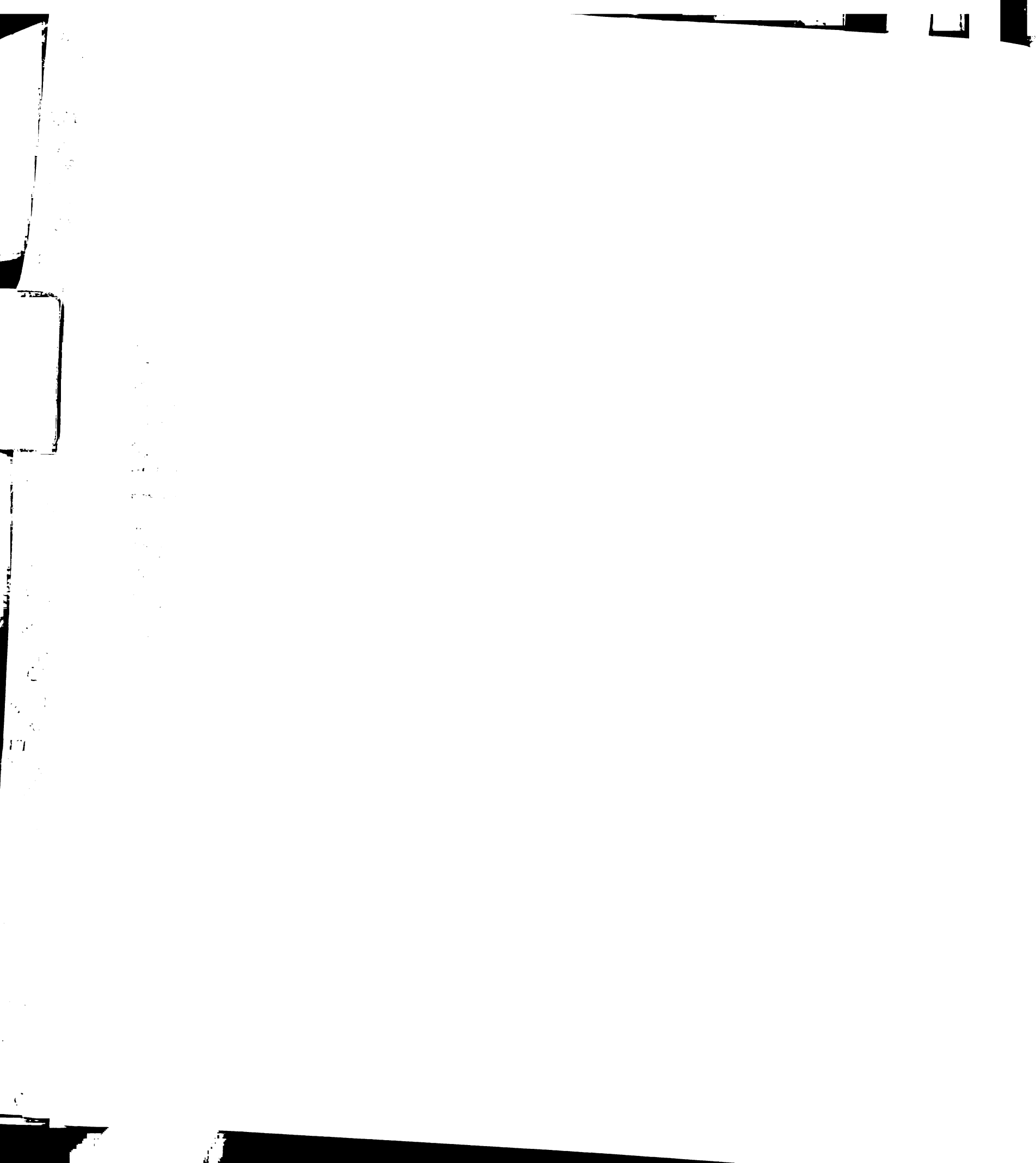
### 3.9 Conclusions

From a naive population of random disulfide constrained peptides displayed on M13 bacteriophage, members were selected which bound Fc. These peptides were found to be competitive with Protein A with a  $K_i$  of 5  $\mu\text{M}$ . Two rounds of optimization focused on local and then global improvements gave peptides which bound Fc with affinities improved over 200-fold. Alanine scanning of the Fc surface binding to peptide showed that the nature of the peptide:Fc interaction was distinct from the interaction with protein A, Z-domain. Consensus sequence patterns and alanine scanning of the peptide sequences suggest that a 13-residue minimized peptide might also bind Fc with high affinity. This peptide was synthesized and indeed found to bind Fc tightly to Fc ( $K_i=24$  nM,  $K_d=14-16$  nM) in



**Figure 3.23: Kinetics of FBP4.1 Binding.** Kinetic measurements of FBP4.1 binding to immobilized Fc were recording by surface plasmon resonance over a range of concentrations. From these curves, rate constants of  $k_{on} = 1.6 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$  and  $k_{off} = 2.6 \times 10^{-2} \text{ s}^{-1}$  were computed.

multiple types of binding assays. Thus, this minimized peptide can bind Fc with an affinity comparable to that of natural binding domains over four times larger in size.



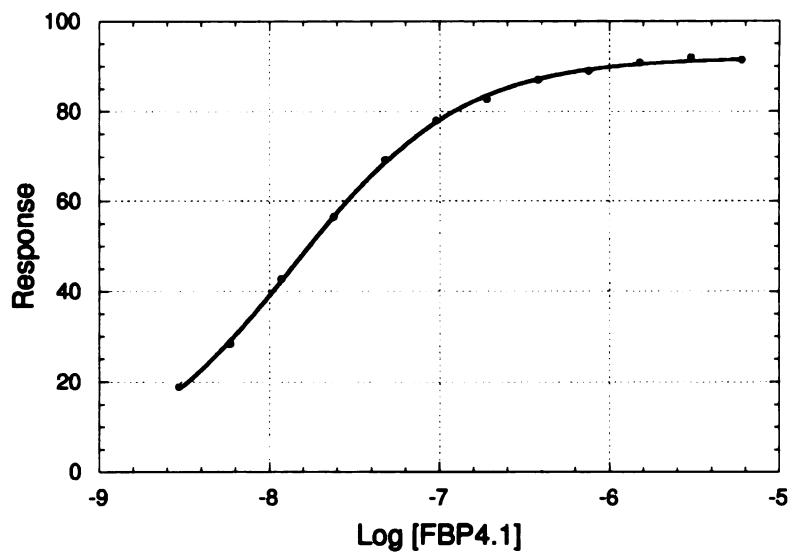


Figure 3.24: **Steady State FBP4.1 Binding.** Measurements of bound fraction as a function of steady state FBP4.1 peptide concentrations were used to compute a dissociation constant of  $K_d = 14$  nM.

### 3.10 Acknowledgements

Brian Cunningham, Henry Lowman, Jeffrey Tong, Mark Dennis for assistance and discussions pertaining to peptide selection. Shane Atwell for help with molecular biology and library creation. Nick Skelton for NMR.



## **Chapter 4**

### **Crystal Structure of an Engineered Peptide in Complex with Fc**

#### **4.1 Introduction**

The 13-residue Fc binding peptide FBP4.1 formed crystals in complex with Fc that diffracted to 2.6 Å, making it possible to solve the structure of the peptide in complex with Fc. The peptide is shown to have a unique structure, which makes it a new addition to the already diverse family of Fc binding proteins with known structure. Remarkably, the peptide is also found to target the same region on Fc as the natural protein domains, suggesting that the intrinsic properties of that site appear to be attracting molecules to bind there. A survey of the Fc binding interactions of all the Fc binding domains reveals that there are actually a number of common features presented off of the diverse scaffolds.

#### **4.2 Methods**

##### **4.2.1 Crystal Growth**

The Hampton-50 factorial screening kit (Hampton Research) was used to identify crystallization conditions. Promising results were obtained in acetate buffer at pH 5-6 with PEG and isopropanol as precipitants, which were particularly convenient since they eliminated the need for addition cryosolvents during low temperature





Faint, illegible text or markings, possibly bleed-through from the reverse side of the page.

Component	Concentration
Sodium Acetate pH 6.0	100 mM
Polyethylene Glycol 4K	20%
Isopropanol	20%
IgG <sub>1</sub> -Fc	100 $\mu$ M
DCAWHLGELVWCT-NH <sub>2</sub>	150 $\mu$ M

**Table 4.1: Crystallization Conditions.** Shown are the optimal conditions for crystallization of the Fc/FBP4.1 complex. They were optimized with respect to pH, buffer, precipitate concentrations, and protein concentrations.

data collection. The optimized crystal conditions are shown in Table 4.1.

Crystals were grown using standard procedures in “sitting drops” with 0.5-1 ml reservoir buffers in 24-well culture plates sealed with cover slip and vacuum grease. Best results were obtained by mixing the reservoir solution in 1:3 to 3:1 ratios with protein drops. Drops sizes were 2-10  $\mu$ l. Crystal plates appeared within several days and grew to up to 0.3 mm over several weeks.

#### 4.2.2 Data Collection and Processing

Crystals were frozen in liquid nitrogen and evaluated on a 300 mm MAR-Research image plate system with a Rigaku RU2000 rotating anode generator. Once crystals with reasonable diffraction ( $\approx 3.2$  Å) were obtained, a full data set was collected at the Stanford Synchrotron Radiation Laboratory. The data collection statistics are presented in Table 4.2

#### 4.2.3 Phasing: Molecular Replacement

Phasing of this structure was achieved rapidly since the structure of the Fc subunit was known. The AMoRe program [Navaza, 1994] was applied because of its speed and convenience. It successfully found two solutions with correlation



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

<b>Source</b>	Stanford Synchrotron Radiation Laboratory Beam 7.1, $\lambda = 0.908$
<b>Space Group</b>	$P2_1$
<b>Unit Cell</b>	$a = 67.55 \text{ \AA}, b = 60.83 \text{ \AA},$ $c = 68.17 \text{ \AA}, \beta = 103.87^\circ$
<b>Resolution</b>	25 to 2.7 $\text{\AA}$
<b><math>R_{\text{merge}}</math></b>	8.3%
<b>Completeness</b>	99.9%
<b>Total Reflections</b>	60,434
<b>Unique Reflections</b>	14,847
<b>Redundancy</b>	4.1
<b>Average <math>I/\sigma_I</math></b>	11.3

Table 4.2: Data Collection and Processing Statistics.

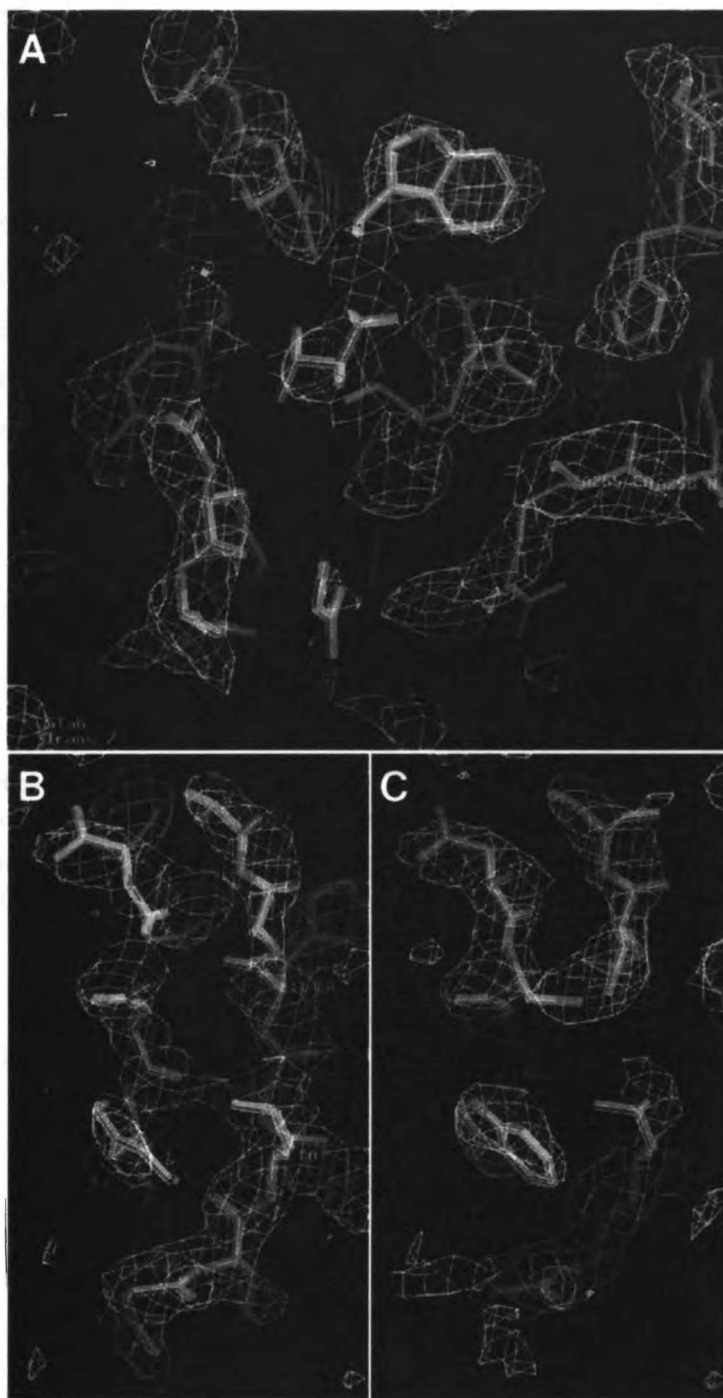
coefficients that exceeded the nearest non-equivalent solutions by over 30%. The search model was based on a 1.9  $\text{\AA}$  structure of IgG-Fc [Ultsch, 1999] which was in turn based on Deisenhofer's structure, and the rotation search was performed in an 85  $\text{\AA}$  cube using reflections from 10.0 to 5.0  $\text{\AA}$  and Patterson vectors from 4.0 to 25.0  $\text{\AA}$ . The following rotation and translation solutions were used to build an initial model of the Fc dimer.

Rotation	Translation			C-Coeff.	R-factor
203.09 78.47 194.90	0.1467	0.0000	0.1047	34.7	47.8
146.62 97.86 15.64	0.5611	0.8730	0.8927	26.0	50.8
Together				37.8	46.7

#### 4.2.4 Initial Refinement and Building

The initial model of the Fc dimer was subjected to rigid body refinement using X-PLOR [Brünger, 1992] with reflections down to 3.2  $\text{\AA}$ . Both the R and the  $R_{\text{free}}$  dropped from around 50% down to about 38%. At this point, a map was calcu-





**Figure 4.1: Initial Peptide Density.** Shown is a  $\sigma_a$  weighted electron density map [Read, 1986, Kleywegt, 1996] calculated using phases from the Fc dimer alone after rigid body minimization against the observed structure factors. Peptide atoms from the final structure (C-yellow, N-blue, and O-red) are shown to for comparison with the initial electron density. Frames A-C show slabs of increasing distance from the Fc surface. Fc atoms are shown in green (note: sulfur atoms on the peptide in (C) are also green).



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

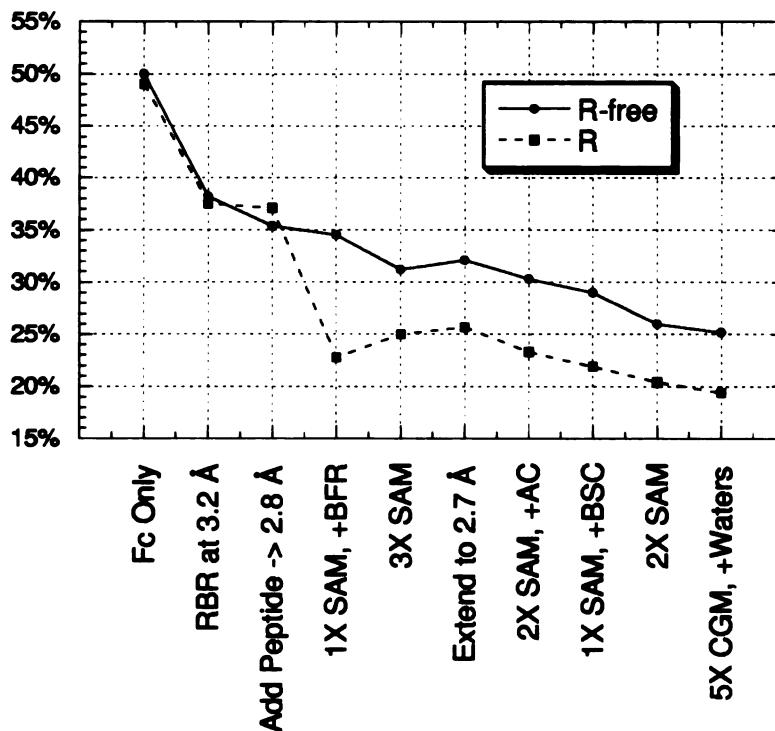


Figure 4.2: **Refinement.** Abbreviations: RBR = Rigid Body Refinement; SAM = Simulated Annealing Minimization followed by Conjugate Gradient Minimization; BFR = Individual Atomic B-factor Refinement. AC = Overall Anisotropic Correction; BSC = Bulk solvent correction; CGM = Conjugate Gradient Minimization.

lated and electron density was observed for the peptide near the  $C_{H2} - C_{H3}$  region of Fc. The density (Figure 4.1) was of such good quality that it was possible to build an entire model of the peptide including all of the side chain atoms. It was also necessary to rebuild several residues in the middle of the 290's loop which participated in crystal contacts near the peptide. After including X-ray data down to 2.8 Å and applying a round of positional refinement, the R-factor dropped to about 35%.





<b>Resolution</b>	20 to 2.7 Å
<b>Unique Reflections (<math>F &gt; 0.1\sigma</math>)</b>	14,266
<b>Bulk Solvent Correction</b>	0.26 e, 10.0 Å <sup>2</sup>
<b>R.M.S.D. in Bonds</b>	0.007 Å
<b>R.M.S.D. in Angles</b>	1.64°
<b>Average B-factor</b>	26.6 Å <sup>2</sup>
<b>R.M.S.D. in B-factor for Bonded Atoms</b>	2.1 Å <sup>2</sup>
<b>R</b>	19.4%
<b>R<sub>free</sub></b>	25.2%
<b>Number of Atoms</b>	4863
<b>Number of Waters</b>	55

Table 4.3: Final Refinement Statistics.

#### 4.2.5 Refinement

The refinement was carried out as shown in Figure 4.2. NCS restraints were applied throughout the refinement to all Fc atoms distant ( $>10$  Å) from non-equivalent crystal contacts. NCS restraints were applied to the peptide initially, but were removed for the rest of the refinement due to the presence of non-equivalent crystal contacts.

After several rounds of refinement, resolution was extended down to 2.7 Å, and then an overall anisotropic and a bulk solvent correction were applied. Once the structure appeared to have converged, 55 waters were added using the criterion of at least a  $1\sigma$  peak in (Fo-Fc) difference density along with two plausible hydrogen bonding partners in the vicinity. The structure was complete after several more rounds of energy minimization. Refinement statistics are shown in Table 4.3.



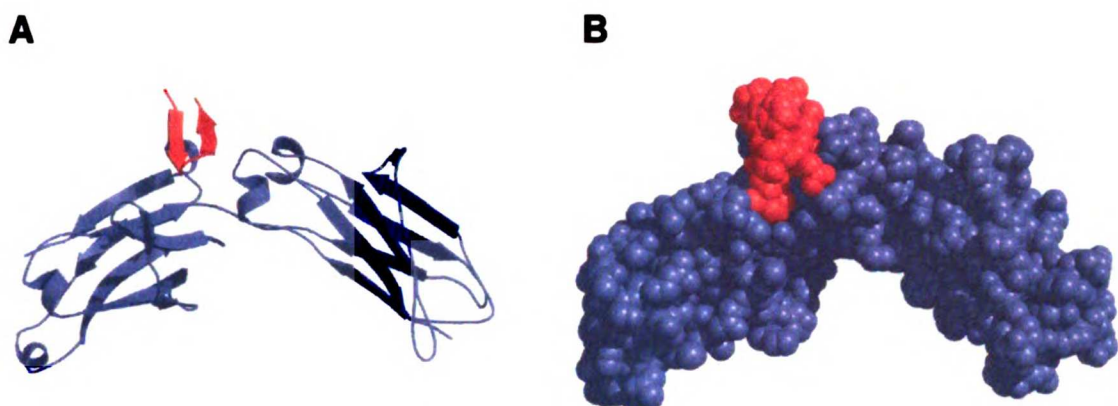


Figure 4.3: **Structure of the Fc:Peptide Complex.** Shown is the 2.5 Å crystal structure of FBP4.1 (red) in complex (blue) in ribbon (A) and space-filling (B) representations.

### 4.3 Results and Discussion

The final  $R$  and  $R_{free}$  for the refinement were 19.4% and 25.2% respectively, and the structure is shown in Figure 4.3. Note the remarkable similarity in size between the minimized Fc binding peptide and the single helix variant of Protein A (Figure 2.11), which binds in nearly the same location but is about four orders of magnitude weaker in affinity.

#### 4.3.1 Overall Structure

FBP4.1 adopts a  $\beta$ -hairpin structure (Figure 4.4) and nestles down in a groove on Fc formed between the 430's and 250's loops. The peptide binds Fc in a monomeric fashion, in contrast to the EPO mimetic, which binds as a dimer [Livnah, 1996].

The structure of this  $\beta$  hairpin is also distinct from that of the EPO mimetic in that it has nine residues separating the two cysteine residues instead of eight. Although both Fc binding and EPO receptor binding peptides exhibit superficial



Faint, illegible text or markings, possibly bleed-through from the reverse side of the page.

similarity in adopting bulged hairpin conformations, and in showing similar consensus binding sequences (GELVW in the Fc peptide, GPLTW in the EPO peptides [Wrighton, 1996]), the ways in which they bind their targets are completely distinct. In the EPO mimics, much of the peptide is devoted to forming a dimerization interface with the other peptide, and only part of one side of each hairpin actually contacts the EPO receptor. In comparison, one full face FBP4.1 is able to interact with Fc, and it can bury about about 50% more surface area than one of the EPO mimics. The Fc binding peptide also shows no resemblance to other phage selected peptides evolved to bind VEGF [Weismann, 1998] or to a polysaccharide binding antibody [Young, 1997].

#### **4.3.2 Peptide Secondary and Tertiary Structure**

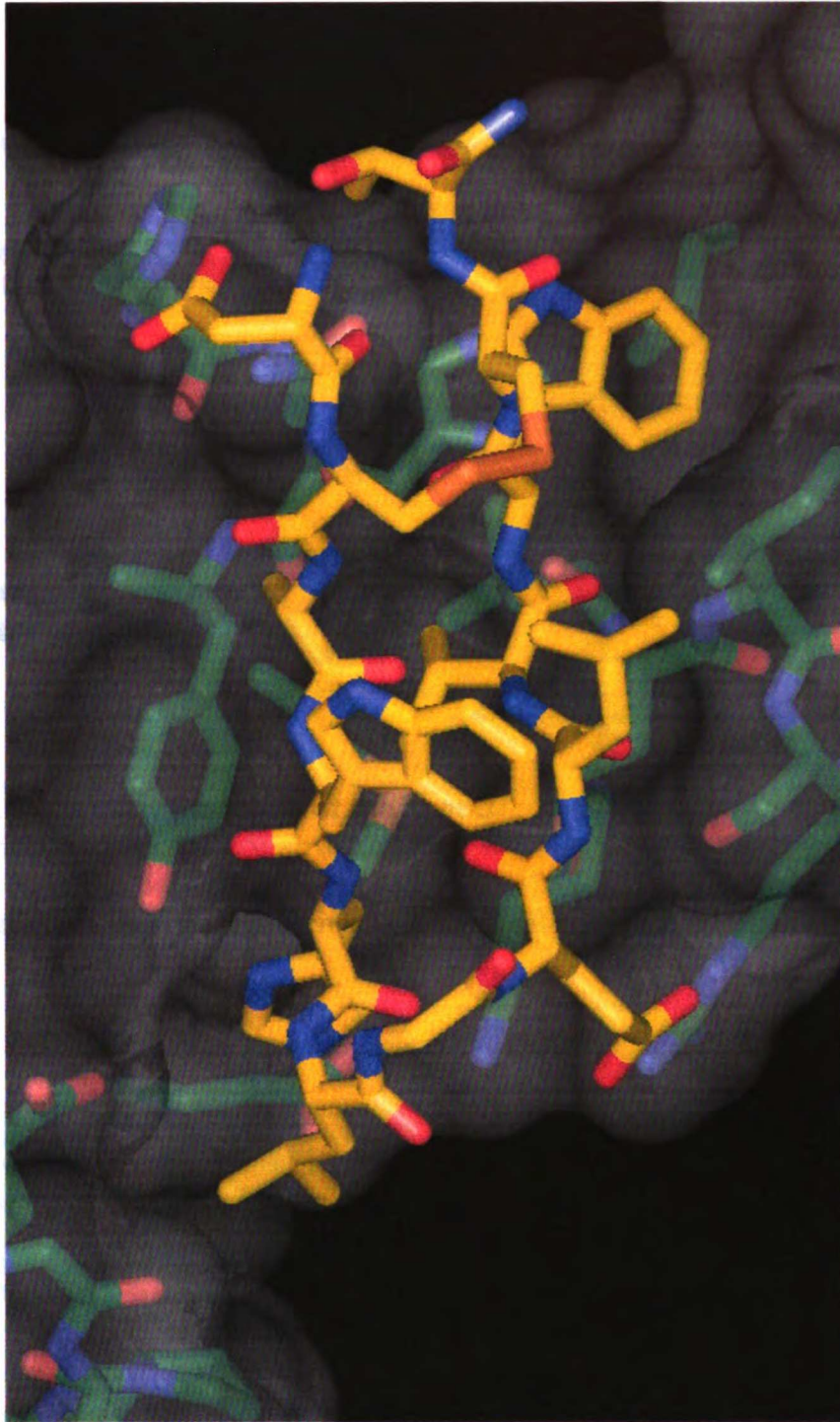
A close look at the secondary structure of the peptide (Figure 4.5) reveals that the regular alternating hydrogen bonding pattern of the  $\beta$ -hairpin is disrupted at Val10 in the peptide, which forms a  $\beta$ -bulge in the hairpin. This  $\beta$ -bulge structure helps project the leucine side chain back towards the disulfide bridge and enables effective packing against Trp5 (Figure 4.6). These two residues along with the disulfide bridge form a miniature hydrophobic core on the backside of the peptide. The sensitivity of peptide affinity ( $\Delta G=4$  kcal/mole) to alanine substitution at this position, and the extent to which both residues were conserved during selection strongly hints that this core may be particularly important in stabilizing the active conformation of the peptide.

#### **4.3.3 Fc Binding Interactions**

A schematic diagram of the peptide's polar Fc binding interactions is shown in Figure 4.7. There were a total of eight potential hydrogen bonding interactions in-



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100



**Figure 4.4: Fc:Peptide Binding Interactions.** This close-up view of the complex between FBP4.1 (DCAWHLGELVWCT-NH<sub>2</sub>, yellow) and Fc (green) illustrates how many of the peptide side chains are able to interact directly with the Fc surface, including the critical residues Trp11 (upper right), Val10 (center), as well as salt bridge forming residues such as Asp1 (top left), His5 (lower left), and Asp8 (lower right).





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

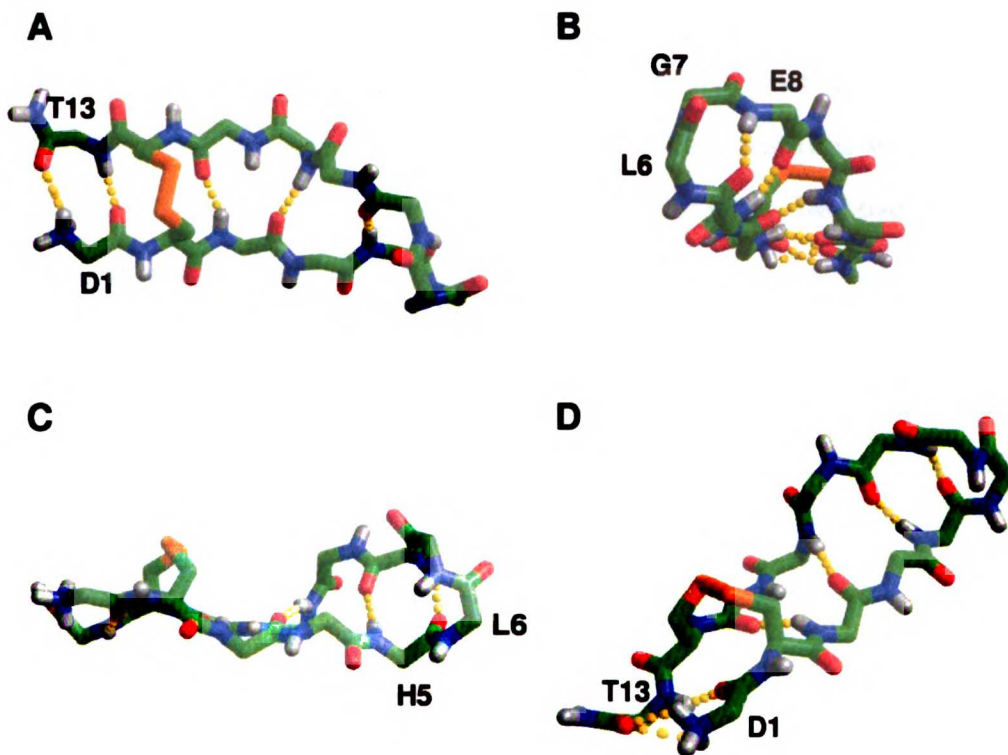


Figure 4.5: **Secondary Structure of the Peptide.** (A-D) Four views of the backbone conformation and secondary structure in FBP4.1. Note how the regular  $\beta$ -sheet hydrogen bonding pattern is disrupted at Val10 causing a bulge to form in the peptide structure. A type II  $\beta$ -turn at Leu6 and Gly7 brings the chain back around.

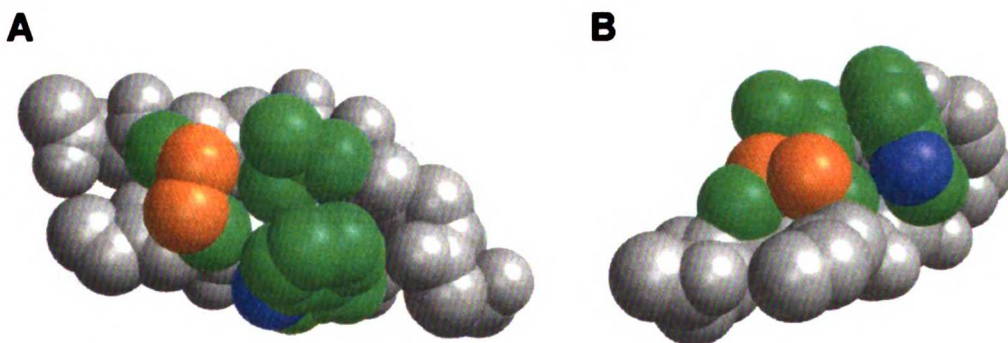


Figure 4.6: **Tertiary Structure of the Peptide.** The bulge conformation at residue 10 in the peptide causes Leu9 to point up along the backside of the peptide, enabling the side chain to pack against the disulfide bridge and Trp4. Both Leu9 and Trp4 were highly conserved throughout the selections, apparently for their structural roles.



10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

cluding four involved in intermolecular salt bridges. The pH dependent binding of the peptide is explained by the two histidines which participate in the interaction (His5 on the peptide, and His433 on Fc). These residues would be more charged at pH 6.0 than at with pH 7.2, thus accounting for the observed  $\approx 4$  fold improvement in binding affinity at the lower pH.

There are also a number of non-polar interactions between the peptide and residues on Fc. Most notable are Val10 and Trp11 which pack tightly against the Fc surface. Their critical contributions to binding are reflected by their high sensitivity to alanine substitution ( $\Delta G \approx 4$  kcal/mole), and the extreme degree to which they were conserved throughout the selection process.

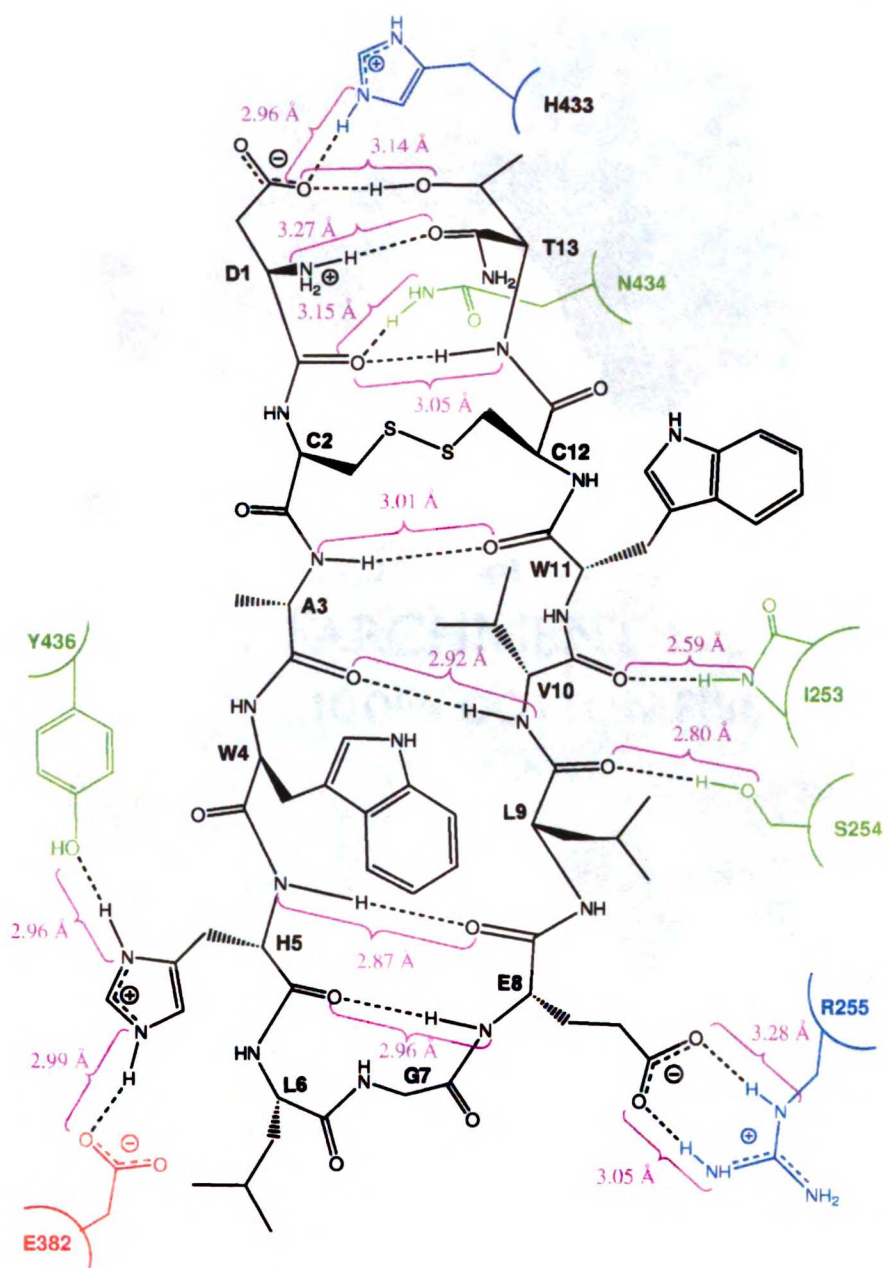
#### **4.3.4 Comparison with Other Fc Binding Interfaces**

The substantial overlap between the consensus binding site of the natural Fc binding proteins and of FBP4.1 is shown in Figure 4.8. The peptide almost entirely covers the consensus binding regions, suggesting that it too has targeted this particular site for some reason.

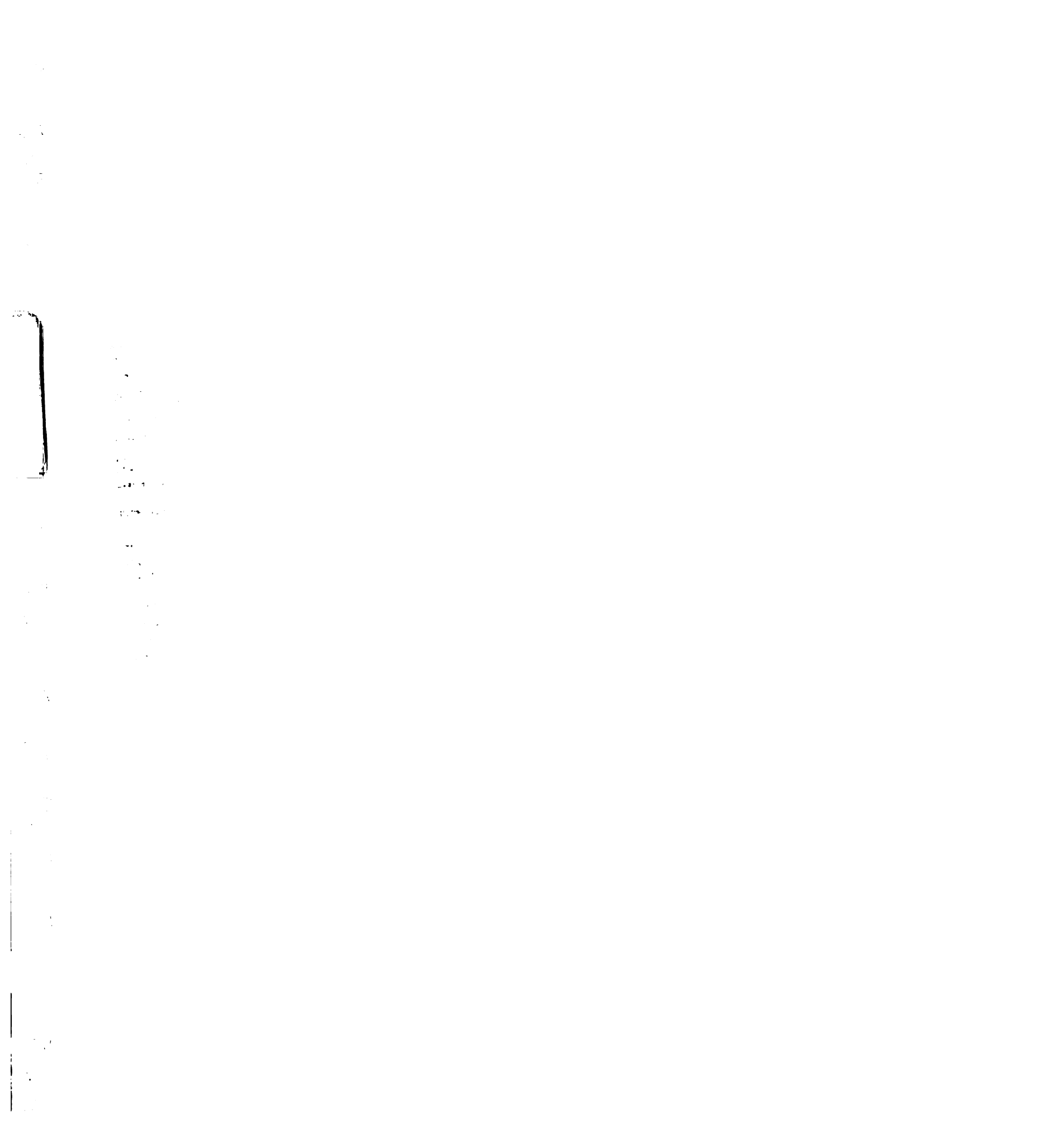
A summary of the statistics for FBP4.1 and the other Fc binding domains with known complex structures is presented in Table 4.4. Remarkably, the 13 residue FBP4.1 peptide is comparable to the natural binding domains in terms of buried surface area, hydrogen bonding, salt bridge formation, and affinity, even though it is less than one quarter of their size.

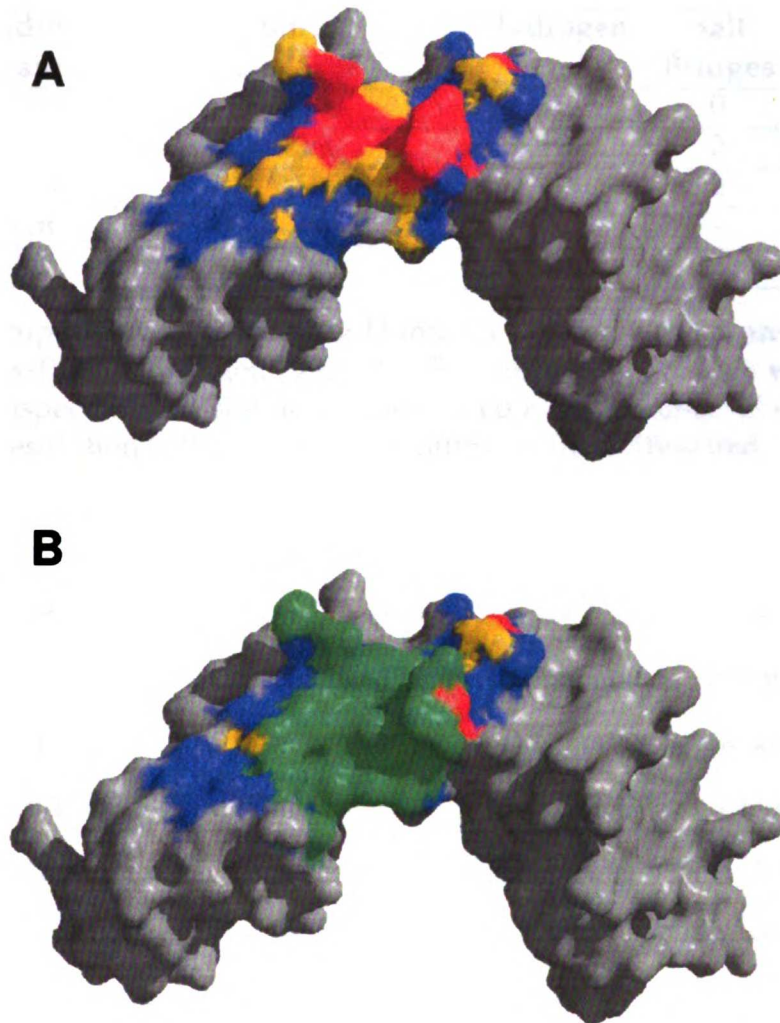
If we juxtapose the various Fc binding interfaces from Protein A, Protein G, Rheumatoid Factor, and FBP4.1, some remarkable patterns appear. Figure 4.9 shows a composite map of the conserved interactions found in the consensus binding site on Fc, and Figure 4.10 shows those interactions in detail. For example, the phenyl ring of Phe14 in Protein A (Figure 4.10E:3) occupies the same





**Figure 4.7: Polar Interactions in the Fc:Peptide Interface.** Intramolecular hydrogen bonding interactions in the FBP4.1 (black) are shown. Also shown are interactions with positively (blue) and negatively (red) charged groups on Fc, as well as interactions with neutral polar groups (green). Up to eight hydrogen bonds appear to be made between the peptide and Fc, including those involved in intermolecular salt bridges.





**Figure 4.8: Footprints on Fc.** (A) The consensus binding site for the natural Fc binding domains. Blue regions participate in a least one interface, yellow in two, and red regions are found all three interfaces. (B) is the same as (A) but with the FBP4.1 binding site superimposed in green. Remarkably, the peptide virtually covers the regions of medium to high consensus.

position as the indole ring of Trp11 of Fc-III (Figure 4.10B:3). The indole nitrogen of Trp43 in Protein G (Figure 4.10C:2) makes the same H-bond with Asn433 on Fc as does the main-chain amide from Thr13 of Fc-III (Figure 4.10B:2). Tyr 98H and Asp31H (Figure 4.10D:6,9) from the Rheumatoid Factor heavy chain make identical hydrophobic and polar interactions, respectively, as Val10 and Glu8 on Fc-III





<b>Fc Binding Domain</b>	<b>Buried Surface Area ( Å<sup>2</sup>)</b>	<b>Hydrogen Bonds</b>	<b>Salt Bridges</b>	<b>Affinity (K<sub>d</sub>)</b>
Protein A	1310	4	0	10 nM
Protein G	1350	10	2	7 nM
Rheumatoid Factor	1520	8	2	≈ 10 μM
Fc Receptor	≈1600	-	-	15 nM
FBP4.1	1180	8	3	25 nM

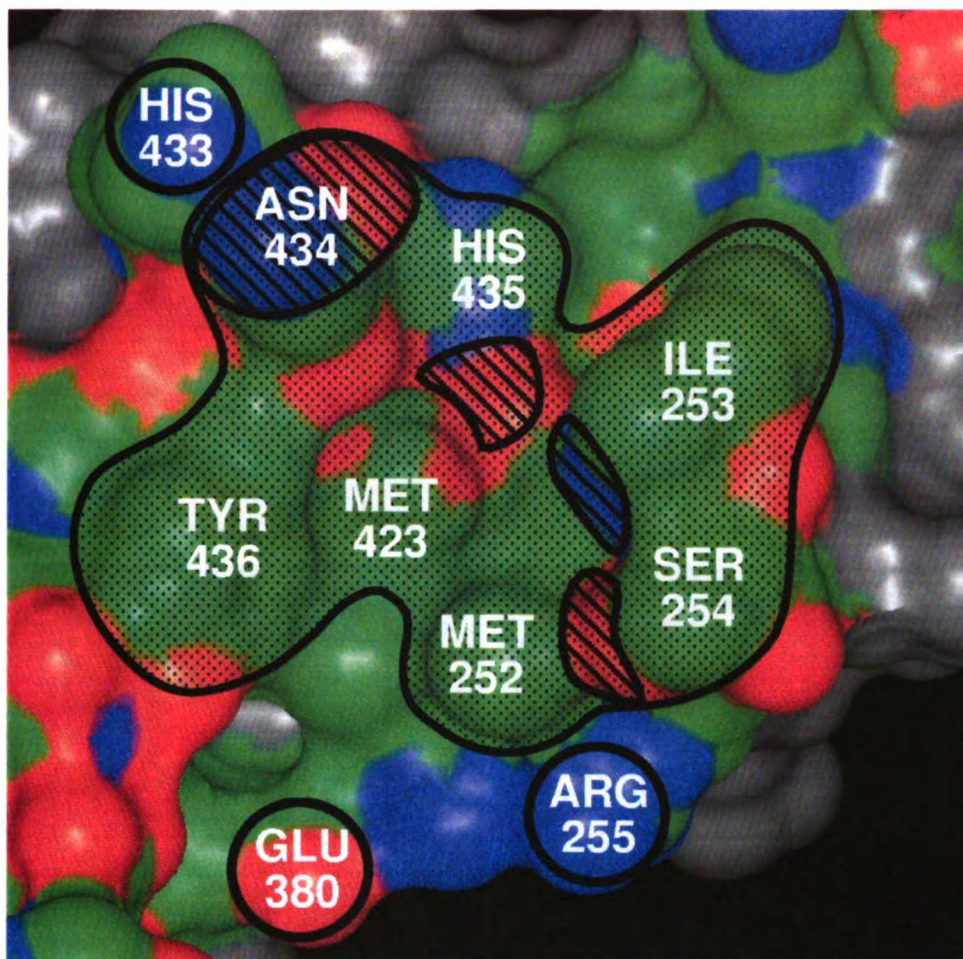
**Table 4.4: Comparison of Fc Binding Domains.** Buried surface areas were computed using X-PLOR [Brünger, 1992]. Hydrogen bonding counts were obtained from visual inspection and should be considered *estimates* since all of these structures have a resolution > 2.5 Å. Literature affinities from: [Braisted, 1996] (Protein A), [Fahnestock, 1990] (Protein G), [Corper, 1997] (Rheumatoid Factor).

(Figure 4.10B:6,9), and Lys28 from Protein G (Figure 4.10C:8) makes the same salt-bridge that His5 does from the Fc-III (Figure 4.10B:8). In a striking example of repeated convergent evolution at the atomic level, the backbone amide of Val10 in Fc-III (Figure 4.10B:5), Glu27 in Protein G (Figure 4.10C:5), the backbone amide of Tyr98H in Rheumatoid Factor (Figure 4.10D:5), and Gln11 in Protein A (Figure 4.10E:5) all make the same buried hydrogen bond with the backbone amide proton of Ile253 on Fc.

#### 4.3.5 Crystal Packing

As indicated above, dimerization of the peptide does not appear to occur when it interacts with Fc. The minimum separation between peptides in the crystal lattice is about 20 Å. There are two kinds of packing interactions in the crystal lattice: those that are pseudo-symmetric with respect to the Fc dimer (Figure 4.11A) and those that are not (Figure 4.11B). The bound peptide is located at a crevice that exists between the two crystal packing layers, and thus the peptides have both equivalent and non-equivalent crystal contacts.

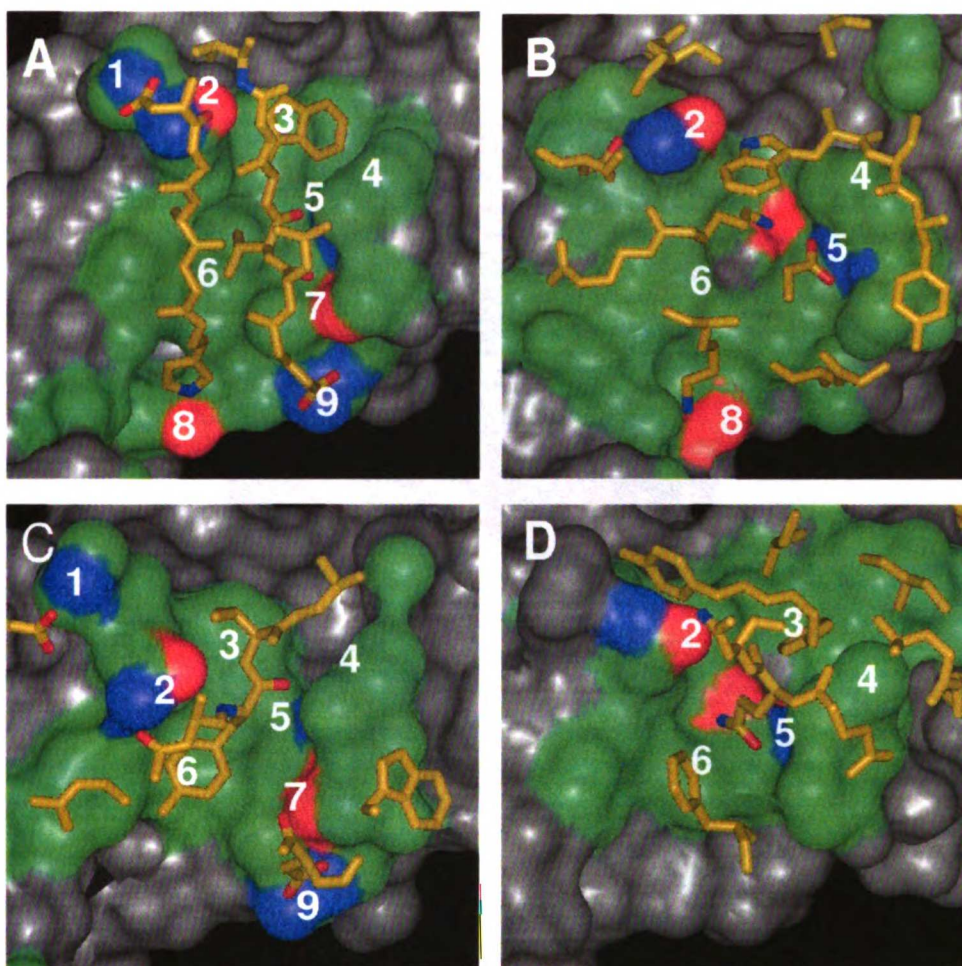
The pseudo-symmetric contact is with the 290's loop. Tyr295 actually packs



**Figure 4.9: Conserved Fc Interaction Map.** Topological distribution of conserved interactions in the consensus binding site on Fc. The predominantly hydrophobic consensus region is shaded. Hydrogen bonding sites are shown with diagonal lines and salt bridging locations are denoted by open circles. Nitrogen and oxygen atoms are colored blue or red respectively, and carbon and sulfur atoms are colored green.

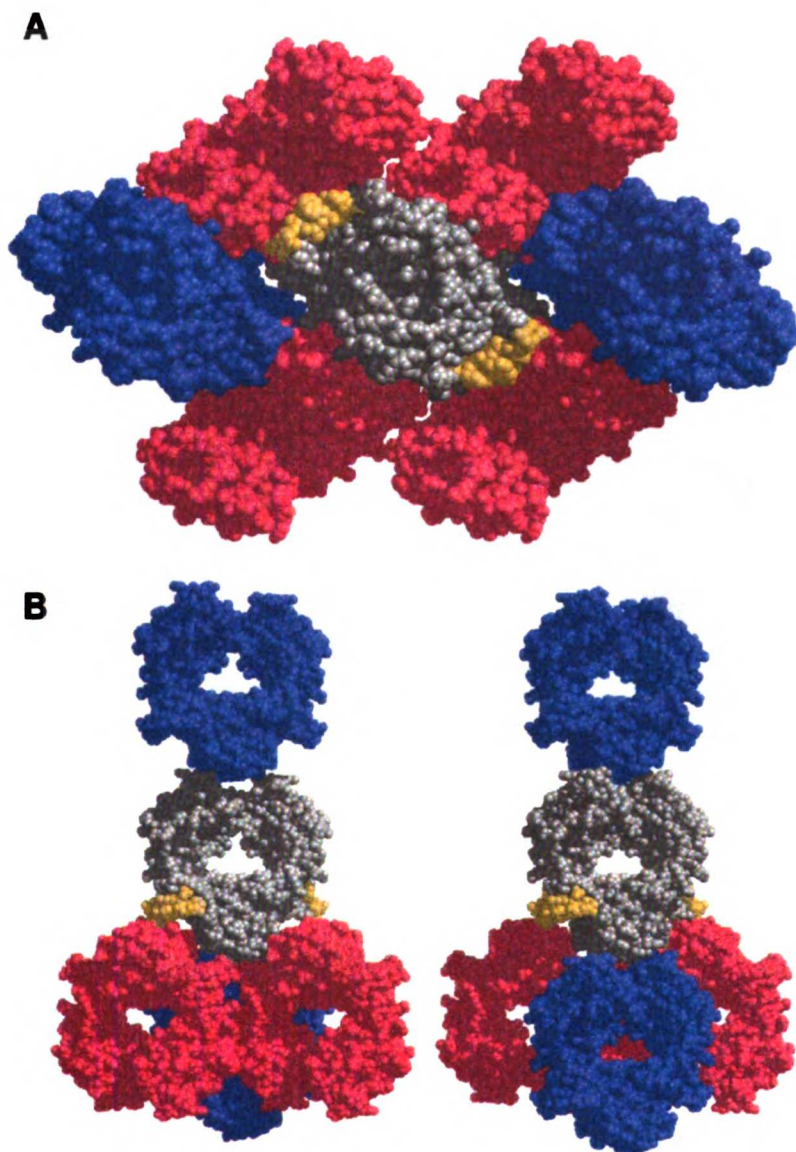
against Trp11 in FBP4.1. The asymmetric contacts involve Asp1 of the peptide and the C-terminus of subunit A on Fc and residues on the short 350's helix.

Although some of the full length peptides were found to crystallize in complex with Fc, they adopted an alternate crystal form which did not diffract well (typically only to about 5 Å). The presence of packing interactions between the



1  
2  
3  
4  
5  
6  
7  
8  
9

**Figure 4.10: Conserved Interactions in the Fc Binding Interfaces.** Comparison of the Fc binding interactions of (A) the selected peptide, FBP4.1 (DCAWHLGELVWCT-NH<sub>2</sub>), (B) domain C2 from Protein G, (C) rheumatoid factor, and (D) domain B of Protein A. Numbers indicate the following conserved interactions: (1) salt-bridges with His433, (2) hydrogen bonding to Asn434, (3) hydrophobic packing onto His435, (4) burial of the hydrophobic "knob" formed by Ile253 and Ser254, (5) hydrogen bonding to main chain (N-H) of Ile253, (6) hydrophobic packing onto Met252 and Tyr436, (7) hydrogen bonding to Ser254, (8) salt-bridges with Glu380, and (9) salt-bridges with Arg255. For clarity, only interfacial atoms are shown, and only nitrogen and oxygen atoms involved in conserved polar interactions are colored blue or red respectively. The remaining contact atoms are colored yellow and green. The dynamic adaptability of this site can be viewed in a movie on the CD-ROM (Section C.2.1).



**Figure 4.11: Crystal Packing Interactions in the Fc:Peptide Complex.** The asymmetric unit in the Fc:FBP4.1 complex consists of one dimer of Fc (gray) and two peptides (yellow). The crystal packing interactions can be broken down into (A) those interactions which are pseudo-symmetric with respect to the Fc dimer, and (B) those which are not. Both types of interactions occur near the peptide binding site.

1. The first part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

2. The second part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

3. The third part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

4. The fourth part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

5. The fifth part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

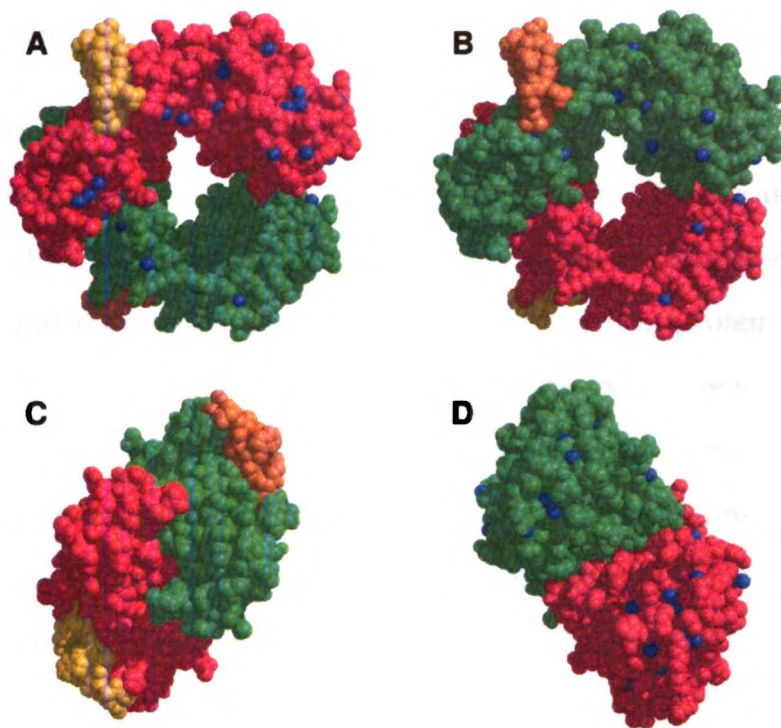
6. The sixth part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

7. The seventh part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

8. The eighth part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

9. The ninth part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

10. The tenth part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".



**Figure 4.12: Hydration of the Fc:Peptide Complex.** Fifty-five waters (blue) were placed in the Fc:FBP4.1 structure, but none of them were found to participate in any Fc:peptide binding interactions.

N-terminus of the 13 residue peptide perhaps explains why the longer peptides were unable to generate good crystals of this form.

#### 4.3.6 Hydration

Although evidence for 55 tightly bound waters was seen in the crystal structure, no water mediated interactions were found to exist between the peptide and Fc. The close packing of the peptide onto the Fc surface would appear to exclude the possibility of waters participating in this interface. Bound waters were distributed fairly randomly over the structure (Figure 4.12), but none were found within proximity of the peptide.

#### **4.4 Conclusions**

The crystal structure of the FBP4.1 peptide bound to Fc reveals a novel Fc binding motif not present in any of the natural Fc binding proteins. The peptide targets the consensus binding site identified from overlapping the binding sites of the natural domains, and it occupies the same region as helix 1 on protein A. However, instead of adopting a helical scaffold resembling the locked-helices, the peptide forms a  $\beta$ -hairpin with a bulged residue that assists in the formation of a small hydrophobic core on the backside of the peptide. From this novel scaffold, the peptide presents a family of Fc binding interactions which are closely related to those in the natural Fc binding domains. Comparison of all the Fc recognition domains gives a consensus picture of the binding site along with a map of the preferred locations for interactions. The geometric constraints in such a map could have utility in engineering even smaller binding domains to associate with the consensus binding site.

#### **4.5 Acknowledgements**

Mark Ultsch and Chris Weismann for assistance in the experimental and computational areas of crystallography, respectively. Bart de Vos for mentorship in structural biology.





## Chapter 5

### SiteFinder: A Tool for Comparing Protein Surface Patches

#### 5.1 Introduction

The extraordinary convergent binding behavior on Fc led us to wonder whether this “consensus” binding region was selected for binding due to its intrinsic physical and chemical properties in addition to whatever biological function binding Fc may have. To address this question, we performed an *in silico* patch-based analysis of the Fc surface.

The idea of using patch analysis to identify protein interface binding sites was originally suggested by Jones and Richardson in their surveys of protein interactions [Jones, 1996]. In their implementation, patches were based on a central residue and surrounding nearest neighbor residues. Only a small number of patches (<100) were generated and scored by various criteria.

Here we extend that work in two ways, (1) by increasing the size of the patch set four orders of magnitude, and (2) by using a more flexible atom-based patch definition.

Our computational approach to identifying binding sites on Fc is designed to mirror the phage selection experiment performed in the laboratory. Just as we screened our billion-member peptide library against Fc, we screen our million-member patch library against a set of characteristics likely be involved in binding.

Since we have empirical results which demonstrate that the consensus binding site is the preferring binding site on the molecule, we can compare this binding site against the patch distributions in order to learn what distinguishes it.

## **5.2 Software Development**

SiteFinder was developed in the C programming language on a Linux workstation using the standard development tools. The MidasPlus package [Ferrin, 1988] was used extensively to visualize SiteFinder output, and various scripts for the CNS program [Brünger, 1998] were created to test and reformat SiteFinder results.

### **5.2.1 Algorithm**

A schematic overview of how the sitefinder program works is shown in Figure 5.1. The program starts with a solvent accessible surface, generates a large set of patches, scores those patches according to selected criteria, and then outputs that information as lists, or alternatively, as average ranks mapped across the protein surface.

#### **Surface Area Calculation**

For convenience and speed, SiteFinder uses a discrete approximation of protein surfaces rather than an analytical one. The surface representation employed by sitefinder is similar that generated by the DMS molecule surface algorithm from MIDAS [Ferrin, 1988] but it generates solvent accessible surfaces instead of molecular surfaces. Solvent accessible surfaces exist one water radius (1.4 Å) farther from van der Waals surface of the protein.

A spherical set of points for each atom was derived from a dodecahedron.

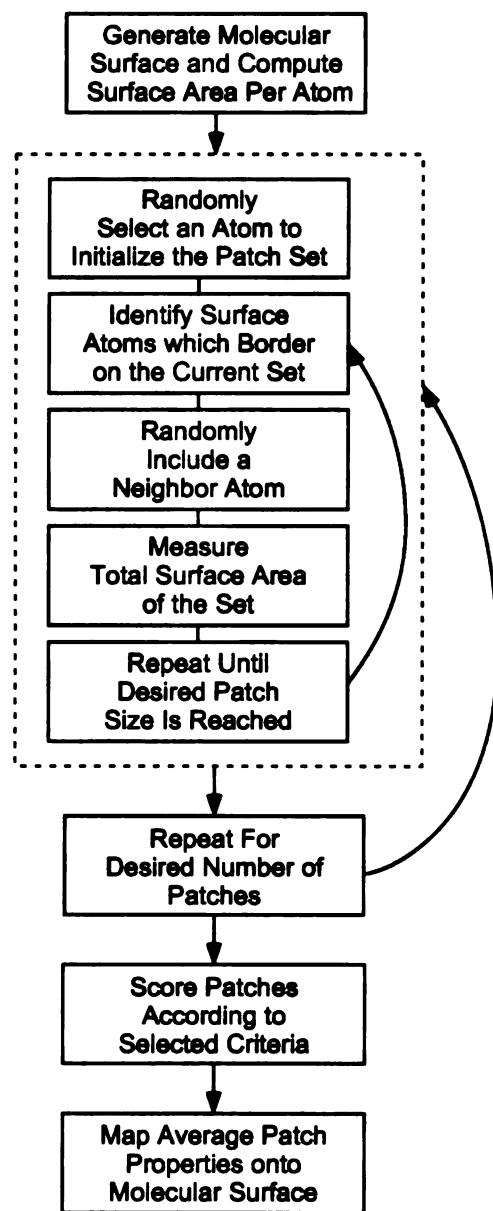


Figure 5.1: **SiteFinder Overview.** The SiteFinder program operates in the above manner to generate and score populations of up to  $10^6$  surface patches on a protein surface.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the specific procedures and protocols that must be followed to ensure that all records are properly maintained and updated. It details the roles and responsibilities of various staff members in this process.

3. The third part of the document provides a summary of the key findings and recommendations from the audit. It highlights areas where improvements are needed and offers practical suggestions for addressing these issues.

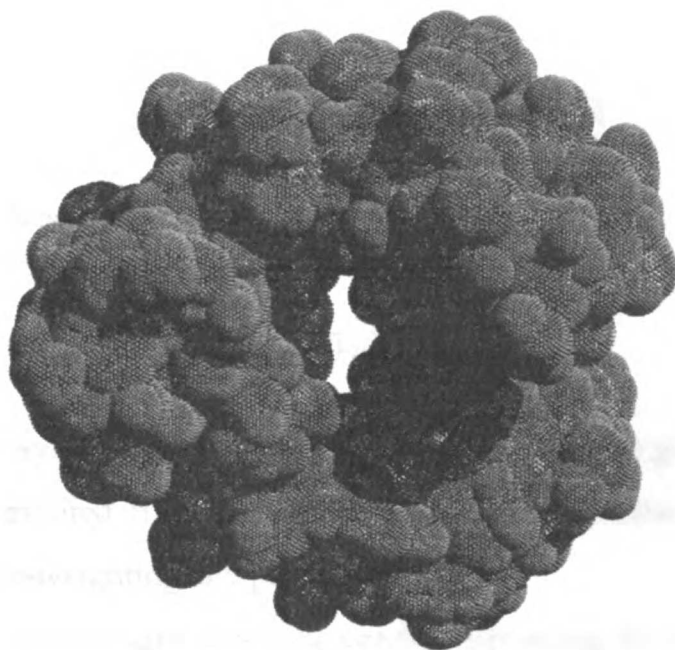
Each equilateral triangle in the dodecahedron is subdivided into four smaller triangles whose vertices are pushed out to the surface of the sphere. This subdivision process is repeated until 1280 triangles are obtained. This results in a highly regular sampling of the spherical surface. The area associated with each point is the sum of one third of the areas attributed to each triangle that shares the vertex. The area of each spherical triangle can be computed using the formula  $area = R^2(\alpha + \beta + \gamma - \pi)$  where  $R$  is the radius of the sphere, and  $\alpha, \beta, \gamma$  are the angles in radians [Bartsch, 1974]. Because it is impossible to tessellate a spherical surface with a completely regular array of >20 points, the amount of surface area attributed to each point varies within a small range.

The set of all points over all atoms in the molecule is then assembled and filtered for intersection with the solvent accessible surface of other atoms. Points that are within the solvent accessible radius are eliminated. What remains is the solvent accessible surface representation shown in Figure 5.2.

Also computed is a set of solvent accessible surface points for each atom in the presence of only those atoms that are covalently bonded to it and within 2.5 Å (or within 3.0 Å for main chain atoms in adjacent residues). This gives a measure of the potential solvent accessible surface that each atom is capable of achieving in the absence of any secondary or tertiary packing interactions, given the unavoidable obstruction from covalent neighbors. For example, using these cutoffs, the maximum solvent accessible surface of the  $C_\epsilon$  atom in a phenylalanine is the solvent accessible surface area of that atom, minus the fraction of surface occluded by the  $C_{\gamma 1}, C_{\gamma 2}, C_{\delta 1},$  and  $C_{\delta 2}$  atoms in the ring. Similarly, the maximum accessible surface area of a  $C_\alpha$  atom in an alanine is its own accessible surface area minus that occluded by  $C_\beta, N, C, O, C^{i-1},$  and  $O^{i-1}$ .

1. The first part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

2. The second part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".



**Figure 5.2: SiteFinder's Solvent Accessible Surface Representation.** SiteFinder uses a point-based molecular surface representation of a protein's solvent accessible surface. Following generation, each point has associated with it an  $(x,y,z)$  coordinate, a surface area, and a  $(x,y,z)$  surface normal.

### **Patch Set Generation**

Patches are generated in a random fashion using a weighting scheme which insures that atoms are sampled evenly across all sets of patches. To generate a patch, an atom is first chosen at random using the current weights ( $W_i$ 's). For each subsequent atom, a neighboring atom is randomly chosen using the weighting formula:

$$weight(i) = count(neighbors\ of\ i\ already\ in\ set)^2 * W_i \quad (5.1)$$



1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The list includes names such as Mr. J. H. Smith, Mr. J. B. Jones, and Mr. W. C. Brown.

2. The second part of the document is a list of the names and addresses of the members of the committee who were present at the meeting. The names are listed in alphabetical order, and the addresses are listed below each name. The list includes names such as Mr. J. H. Smith, Mr. J. B. Jones, and Mr. W. C. Brown.

3. The third part of the document is a list of the names and addresses of the members of the committee who were absent from the meeting. The names are listed in alphabetical order, and the addresses are listed below each name. The list includes names such as Mr. J. H. Smith, Mr. J. B. Jones, and Mr. W. C. Brown.

$W_i$  is the overall weight for a given atom computed from the formula:

$$W_i = \frac{B_i}{1 + \text{count}(\text{usage of atom } i \text{ in all sets})} \quad (5.2)$$

and where  $B_i$  is a base weighting to correct for adjacency bias:

$$B_i = \frac{1}{\text{count}(\text{neighbors of atom } i)} \quad (5.3)$$

This weighting system favors selection of patches which are globular in shape and which are distributed in an even fashion over the entire surface without over-weighting or under-weighting any particular region.

The solvent accessible surface for the patch is computed by summing up the surface areas on each atom, and a patch is considered complete when the total surface area exceeds the specified threshold. About a million patches can be generated in an hour on a machine with 128 MB of RAM.

### Calculation of Patch Properties

After all patches have been generated, they are analyzed according to a series of properties. The following criteria are available:

where:

$I$  is the set of atoms in a patch.

$A_i$  is the area of atom  $i$ .

$A_T = \sum_{i \in I} A_i$  is the total area of a patch.

**Charge Density** =  $\sum_{i \in I} |\text{charge}_i| / A_T$

1. The first part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

2. The second part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

**Polarity** =  $\sum_j A_j/A_T$  :  $j \in \{\mathbf{N},\mathbf{O}\}$  in  $I$  and  $|charge_j| > 0.25$

**Accessibility** =  $A_T/\sum_{i \in I} M_i$  :  $M_i$  is the maximum potential exposed surface area for each atom in the absence of packing interactions.

**Planarity** =  $|\sum_k \mathbf{v}_k|/\text{count}(\mathbf{v}_k)$  :  $\mathbf{v}_k$  are all surface normals on all atoms in the patch.

**Area Per Atom** =  $A_T/\text{count}(I)$

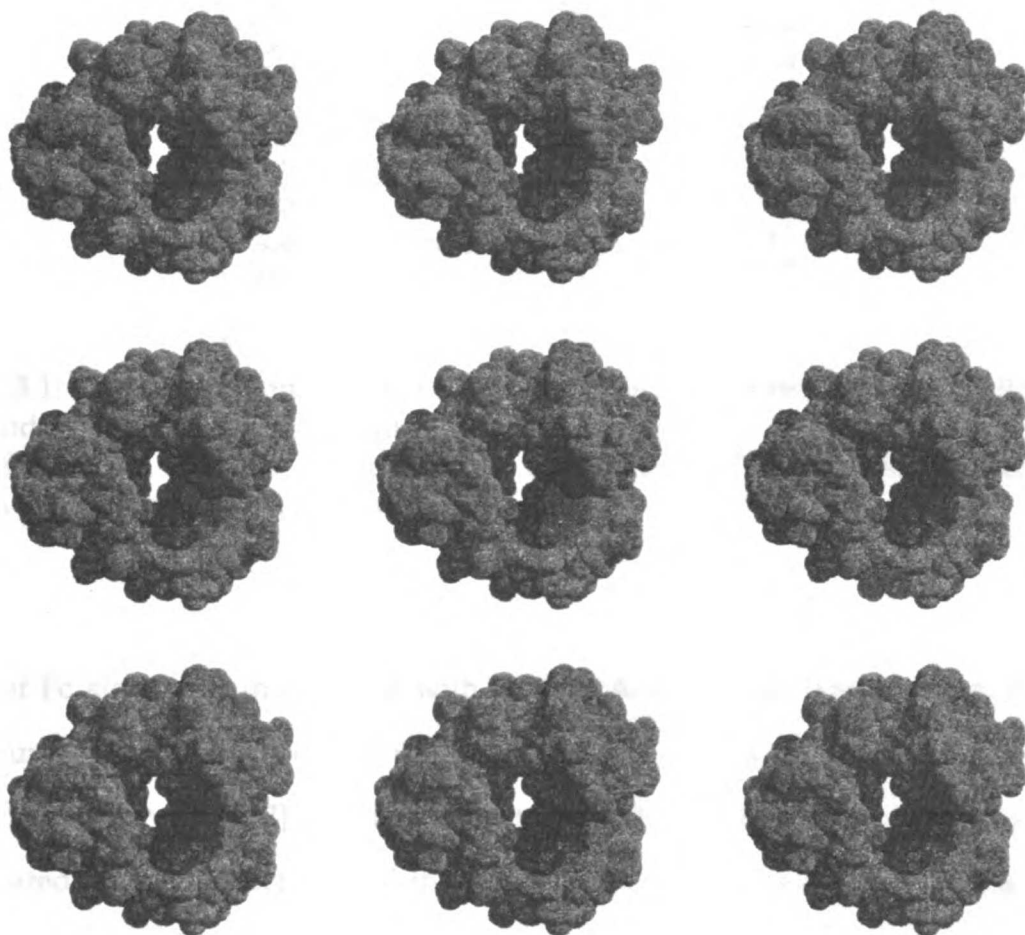
**Non-polar Area Per Atom** =  $\sum_j A_j/\text{count}(I)$  :  $j \in I$  where  $|charge_j| < 0.25$

## 5.2.2 Implementation and Testing

Approximately 6000 lines of C code were required to implement this system, of which 50% were written specifically for this project. The remaining code was assembled from other original programs written by myself over the past five years. About 75% is devoted to the parser, vector libraries, error checking, and utilities, and so the algorithm itself only consists of about 1500 lines of code.

Extensive testing was performed along the way to insure that SiteFinder produced accurate surface area statistics and that it generated reasonable and diverse patch sets. Specific numeric results output by SiteFinder were verified with CNS scripts whenever possible using the built-in surface area measurement tools. Some example surface patches on Fc are shown in Figure 5.3.

The probability weighting mechanism was introduced in order to obtain even coverage of the final surface over all the patches. Over the Fc surface, each accessible atom participates in  $2.5 \pm 0.24\%$  of the patches, as would be expected since the ratio of the patch size to the total surface of the molecule is 0.024. Thus for every million Fc surface patches, each atom participates in 25,000 different patches on average.



**Figure 5.3: Example Surface Patches.** Shown in red are nine example  $525 \text{ \AA}^2$  surface patches generated on the Fc surface. The probability weighting algorithm insures that the Fc surface is sampled evenly over all regions.

### 5.3 Application to the Fc Surface

The strict consensus binding site consisted of solvent accessible atoms that were present in each of the Fc binding interfaces, and those atoms are shown in Table 5.1. On average, this set of atoms accounts for  $525 \text{ \AA}^2$  of solvent accessible surface on the surface of Fc.

2.5 million  $525 \text{ \AA}^2$  surface patches were generated over five Fc crystal struc-

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for consistent and reliable data collection processes to support informed decision-making.

3. The third part of the document focuses on the role of technology in data management and analysis. It discusses how modern software solutions can streamline data collection, storage, and reporting, thereby improving efficiency and accuracy.

4. The fourth part of the document addresses the challenges associated with data management, such as data quality, security, and privacy. It provides strategies to mitigate these risks and ensure that data is used responsibly and ethically.

5. The fifth part of the document concludes by summarizing the key findings and recommendations. It stresses the importance of ongoing monitoring and evaluation to ensure that data management practices remain effective and up-to-date.

<b>Residue</b>	<b>Consensus Contacts</b>
Leu251	$O$
Met252	$C_{\alpha}, C_{\beta}, C_{\gamma}, S_{\delta}, C_{\epsilon}$
Ile253	$N, C_{\beta}, C_{\gamma 1}, C_{\gamma 2}, C_{\delta 1}$
Ser254	$N, C_{\alpha}, C_{\beta}, O_{\gamma}$
His433	$C_{\epsilon}, O$
Asn434	$N, C_{\alpha}, C_{\beta}, C_{\gamma}, O_{\delta 1}, N_{\delta 2}, C, O$
His435	$C_{\beta}, C_{\gamma}, N_{\delta 1}$
Tyr436	$C_{\beta}, C_{\gamma}, C_{\delta 1}, C_{\delta 2}, C_{\epsilon 1}, C_{\epsilon 2}, C_{\zeta}$

**Table 5.1: Consensus Contacts on Fc.** These are the surface atoms found in all four Fc binding interfaces resolved at high resolution (protein A, protein G, rheumatoid factor, and FBP4.1). Together, these atoms account for an average of about 525 Å<sup>2</sup> of surface area on an Fc dimer subunit.

tures. The structures used were the dimer Fc binding domains from the structure of Fc alone and in complex with Protein A: B-domain [Deisenhofer, 1981], Fc bound to protein G: domain C2 [Sauer-Eriksson, 1995], Fc bound to rheumatoid factor [Corper, 1997], and Fc bound to FBP4.1. 500,000 surface patches were generated on each structure and then the results were combined. By using five structures instead of just one, it was hoped that the effects of biases or errors in the individual structures would be reduced. However, because the patch analysis depends more on the chemical nature of surface atoms than on their exact conformations, measurements made on any of the individual Fc structures were found to be nearly equivalent to results computed over all structures.

The distribution of solvent accessibilities of surface patches on Fc is shown in Figure 5.4, along with that of the consensus binding site. Accessibilities are computed both in terms of solvent accessible surface area per atom and solvent accessible surface fraction. The latter measurement is superior because not all covalent structures present in protein side chains can bury the same amount of surface area per atom. For instance, the solvent accessible surface area of a  $C_{\gamma}$  atom in a tyro-

1. The first part of the document is a list of names and addresses of the members of the committee.

2. The second part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of the chairperson.

3. The third part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of the secretary.

4. The fourth part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of the treasurer.



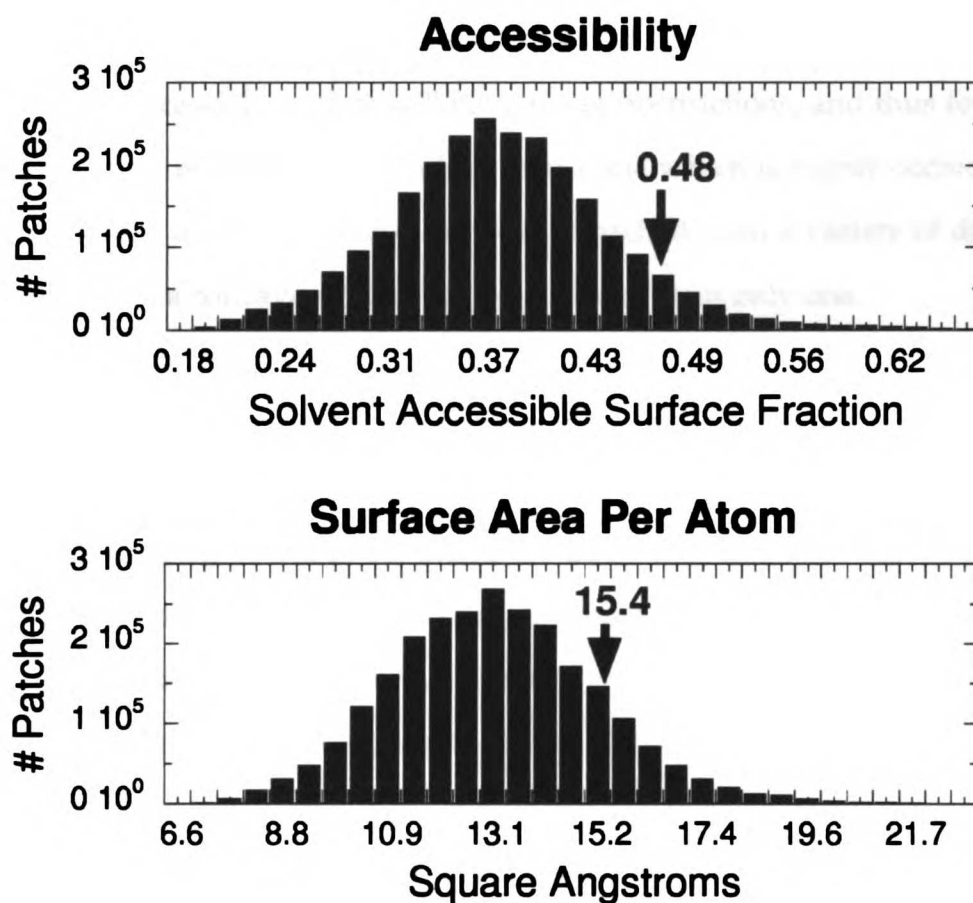


Figure 5.4: **Accessibility of the Fc Surface.** Histograms showing (A) the solvent accessible surface fraction and (B) the accessible surface area per atom over the population of patches generated on the Fc surface. Arrows indicate where the consensus binding site falls in the distributions.

sine is inherently more occluded than the oxygen of the hydroxyl group, due to the higher valency. In order to obtain an atomic metric for surface area which takes into account the covalent context of an atom, a separate measurement is made of how much surface area each atom has in the absence of non-covalent packing interactions, and then this number is used to normalize the actual amount of accessible surface to obtain solvent accessible surface fraction.

As shown in Figure 5.4A, the consensus binding site is more solvent exposed than the vast majority of surface patches on the Fc surface. The exposed nature

1. The first part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

2. The second part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

3. The third part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".

of this site may make it possible for a larger variety of structures to interact here, since a highly accessible surface will have fewer obstructions, and thus fewer inherent geometric restrictions in binding than a site which is highly occluded. In geometric terms, a convex surface can be approached from a variety of different directions, while a concave surface can be explored from only one.

Also, note that the solvent accessible surface area per atom (Figure 5.4B) in the consensus region is closer to the average value than the fractional accessibility because many of the atoms in the consensus site are part of ring systems (His and Tyr side chains), in which the atoms are inherently more obstructed by their covalently bound neighbors. The accessible nature of the consensus binding site would not be as apparent if we only measured the solvent accessible surface area per atom without taking into account the background environment of the different atoms.

The polarity of surface patches on Fc is shown in Figure 5.5, along with that of the consensus binding site. Three different measurements of polarity were made: the fraction of the surface capable of hydrogen bonding, the non-polar surface area per atom, and the charge density. In each case, we find that the consensus binding site is of lower polarity, and is thus more hydrophobic than most other patches on the protein surface. Given that polar interactions are thought to be the primary determinants of specificity in protein interactions, it would seem logical that a cross-reactive binding site would be less specific since it must interact with a variety of different binding partners.

The planarity of surface patches on Fc is shown in Figure 5.6. Here find that even though the consensus binding site is located at a cleft between the C<sub>H2</sub> and C<sub>H3</sub> domains, it is of only average planarity. This metric is therefore not useful for distinguishing the consensus binding site from the rest of the IgG-Fc surface.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for consistent and reliable data collection processes to support informed decision-making.

3. The third part of the document focuses on the role of technology in data management and analysis. It discusses how modern software solutions can streamline data collection, storage, and reporting, thereby improving efficiency and accuracy.

4. The fourth part of the document addresses the challenges associated with data management, such as data quality, security, and privacy. It provides strategies to mitigate these risks and ensure that data is used responsibly and ethically.

5. The fifth part of the document concludes by summarizing the key findings and recommendations. It stresses the importance of ongoing monitoring and evaluation to ensure that data management practices remain effective and aligned with the organization's goals.

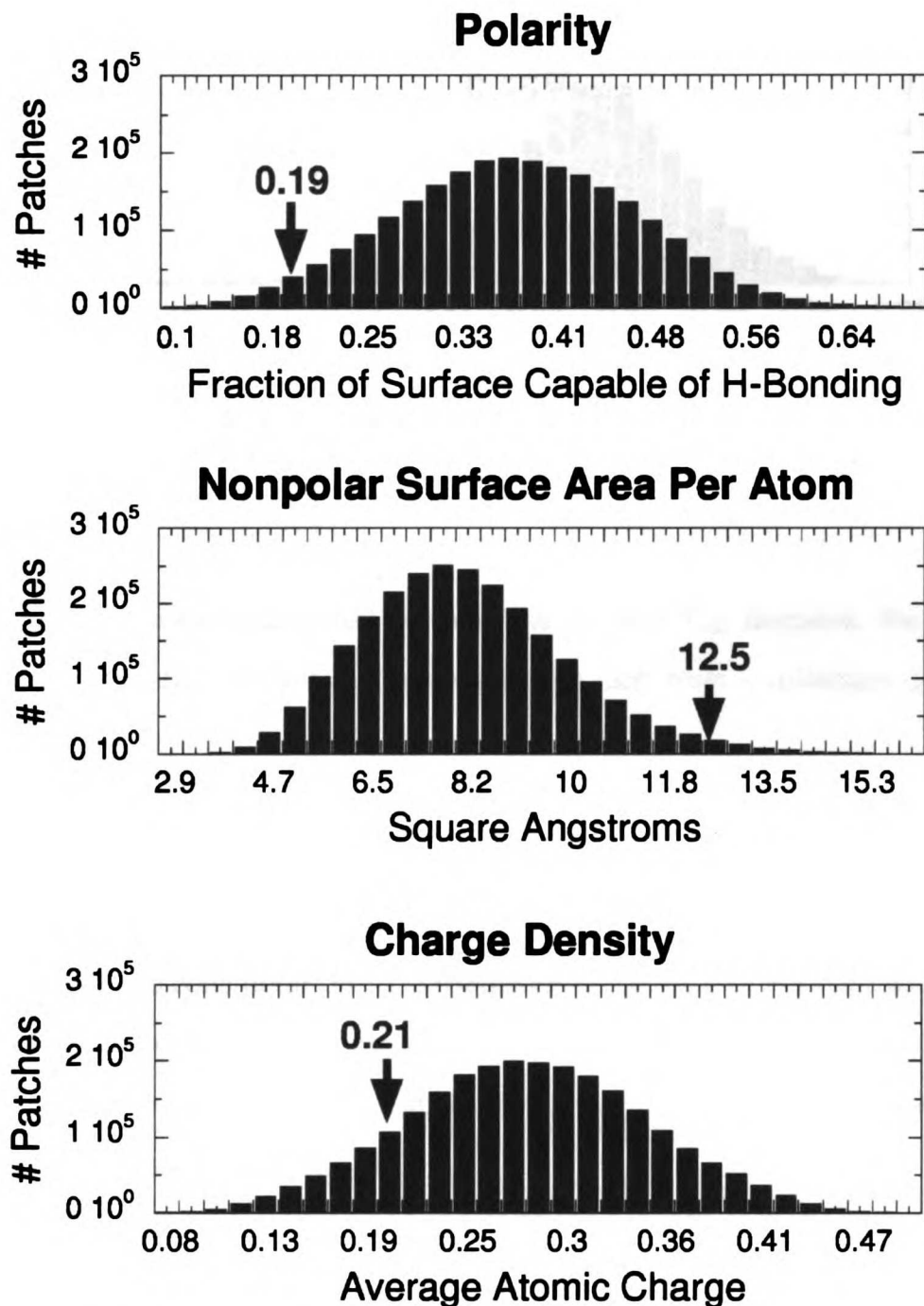


Figure 5.5: **Polarity of the Fc Surface.** Histograms showing (A) the polarity, (B) the non-polar surface area per atom, and (C) the charge density over the population of patches generated on the Fc surface. Arrows indicate where the consensus binding site falls in the distributions.

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

2. Next, it is important to gather relevant information and resources. This can include research, consulting experts, or reviewing existing data.

3. Once the information is gathered, the next step is to analyze it and identify the key factors that influence the outcome. This often involves breaking down the problem into smaller, more manageable parts.

4. After analysis, a plan or strategy should be developed. This plan should outline the steps that need to be taken to solve the problem or answer the question.

5. The final step is to implement the plan and monitor the progress. This involves carrying out the tasks outlined in the plan and adjusting as needed based on the results.

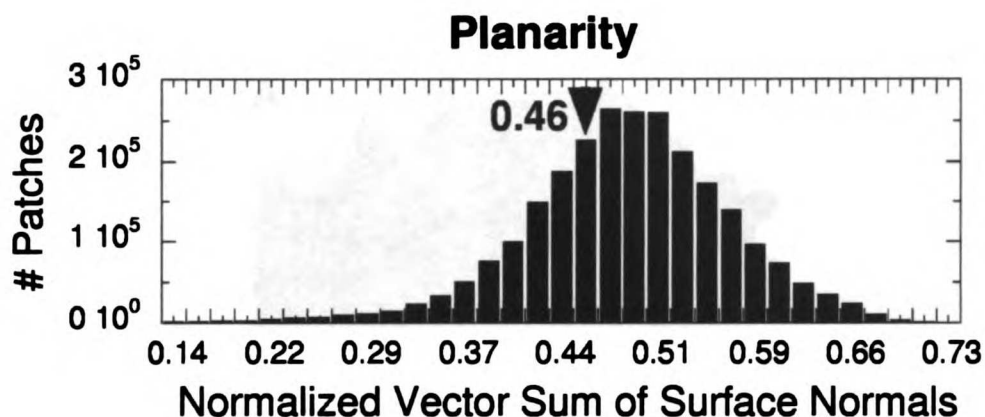


Figure 5.6: **Planarity of the Fc Surface.** Histogram showing (A) the planarity population of patches generated on the Fc surface. Arrows indicate where the consensus binding site falls in the distribution.

Although the site exists at a cleft between the  $C_{H2}$  and  $C_{H3}$  domains, the nature this planarity metric can not distinguish a large cleft from a collection of small clefts.

Given that the consensus binding site is more accessible and less polar than most patches on the Fc surface, we can combine these two properties together in order to attempt a *post-hoc* "prediction" of where molecules would bind if driven by only these two considerations. We tried combining sites with a high accessible surface fraction with those having a low fraction of the surface capable of hydrogen bonding. The results of this experiment is shown in Figure 5.7 juxtaposed with the consensus binding site on Fc.

We find that only two large regions of the Fc surface are identified as being substantially non-polar (Figure 5.7). The first is a region which includes the consensus binding site at the  $C_{H2}$ - $C_{H3}$  domain hinge and continues over a portion of the  $C_{H2}$  domain. The second is a patch on the inside of  $C_{H3}$  near the interior homodimerization region of the domain. A smaller non-polar region is identified near the tip of the  $C_{H2}$  domain.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that proper record-keeping is essential for the integrity of the financial system and for the ability to detect and prevent fraud.

2. The second part of the document outlines the specific requirements for record-keeping, including the need for clear, legible entries and the requirement to retain records for a minimum of seven years. It also discusses the importance of regular audits and the role of internal controls in ensuring the accuracy of the records.



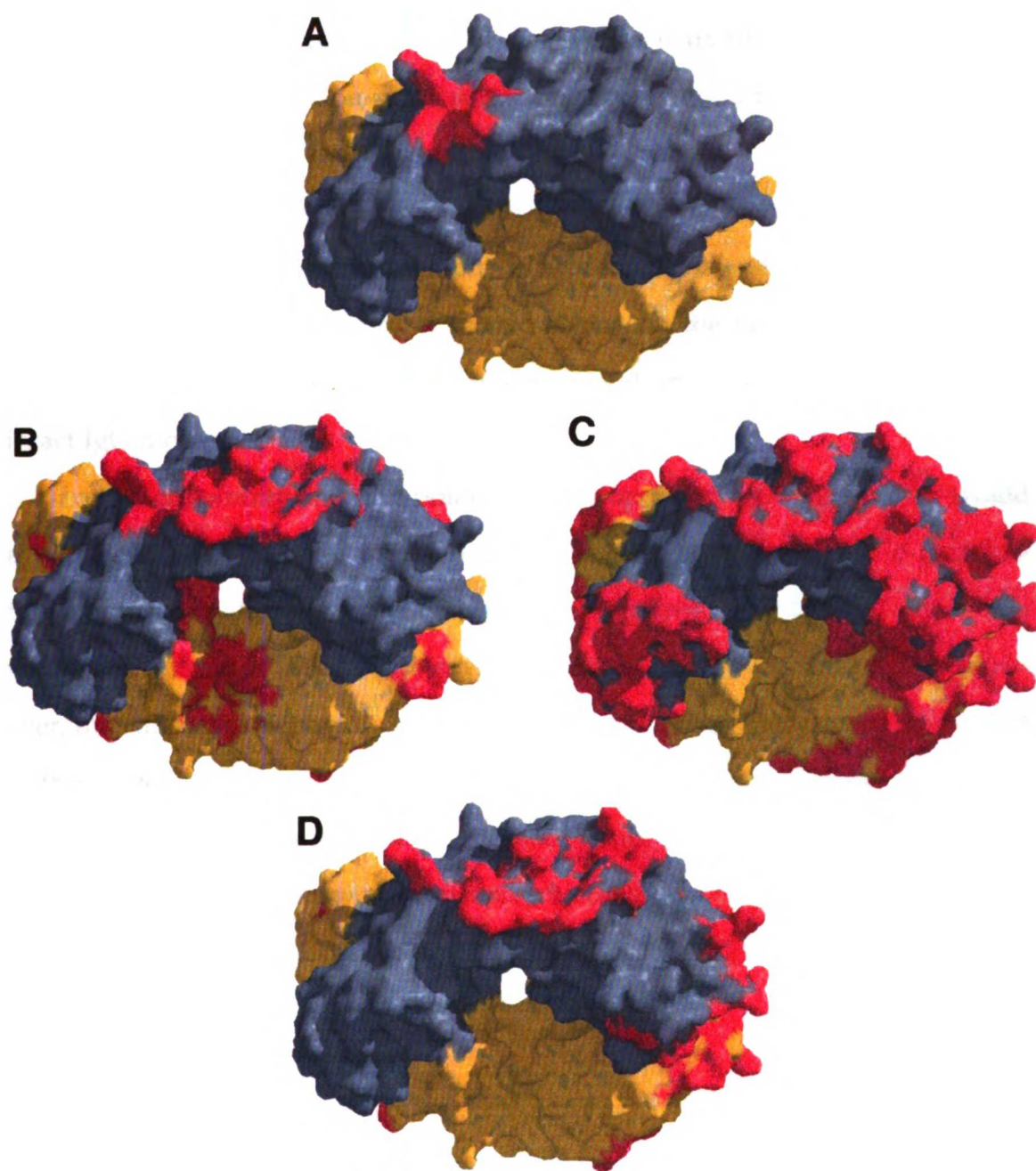


Figure 5.7: **SiteFinder Maps of the Fc Surface.** (A) the consensus binding site on Fc, (B) regions of the Fc surface with an average polarity ranking of  $<0.40$  (C) regions of the Fc surface with an accessibility ranking of  $>0.44$ , and (D) regions with an average combined ranking  $[((1 - \text{polarity}) + \text{accessibility})/2]$  of  $>0.55$ .



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

In terms of accessibility, there are a number of exposed surfaces on the Fc dimer with accessibilities comparable to that of the consensus site (Figure 5.7C). However, the large patch of non-polar surface area near the interior of the Fc dimer does not include any of these accessible surface patches.

By combining the two metrics together, we can eliminate much of the Fc surface (Figure 5.7D) from consideration. What remains is one region that contains the consensus binding site and a second region on the tip of the C<sub>H2</sub> domains which corresponds remarkably to where the C<sub>H1</sub> domain would be attached in an intact IgG molecule.

If only accessibility and non-polarity were important in binding Fc, we would expect Fc binding domains to have evolved to interact with all parts of the identified exposed hydrophobic region. Indeed, both Protein A (Figure 3.3) and the neonatal Fc receptor (Figure 3.6) cover additional portions of this region. However, the consensus behavior tells us that molecules specifically seek out the part of this region which includes the domain hinge. What distinguishes this site in particular? Perhaps the location at a domain hinge is itself important. In the next chapter, we will try to answer the question by looking at how the Fc dimer adapts in the various complexes, and how that adaptability affects this hinge region in particular.

## 5.4 Conclusions

Surface patch analysis performed on a collection of Fc crystal structures reveals that the consensus binding site is more accessible and less polar than most surfaces on the Fc dimer. By combining these two properties into a search criterion, it proved possible to eliminate most of the Fc surface from consideration, leaving



only two regions remaining on the Fc surface. One of these regions is located on the tip of the C<sub>H2</sub> domain, where the C<sub>H1</sub> domain would be located in the context of an intact IgG molecule. The other region contains the consensus binding site, but also includes additional surface on the C<sub>H2</sub> domain. Since natural proteins and selected peptides target the subset of the region which includes the domain hinge, we conclude that some additional properties of this area must be responsible for making it the preferred interaction site.



## Chapter 6

### RigiMOL: A Tool for Analyzing and Visualizing Structural Changes

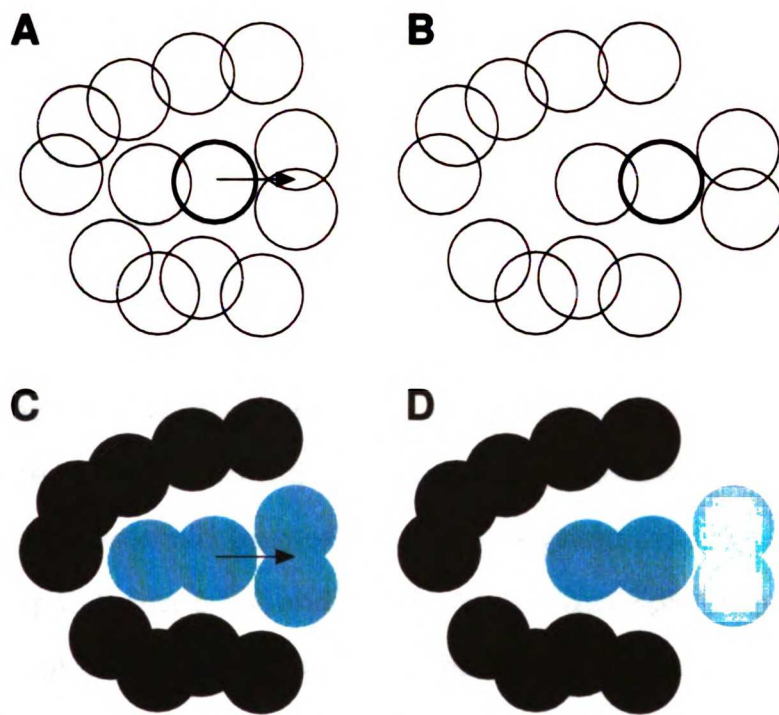
Difficulties in understanding the adaptable nature of the consensus binding site on the Fc led me to conclude that additional methods were needed to assist us in understanding and visualizing how proteins change and adapt to various conditions. A survey of the literature and available software did not turn up anything like what I envisioned: a program that could apply an objective, reference-frame independent analysis to two or more conformations of a protein structure and reduce all of the atomic motions down into a list of concepts that could be readily understood by the human mind. Therefore, I decided to develop my own solution to this problem. RigiMOL (pronounced "RIDGE-ih-mall") is the result of this effort.

#### 6.1 Introduction

Protein structures consist of many thousands of atoms that are interconnected by a variety of covalent and non-covalent interactions. For the most part, these interactions act over a very short range. For example, covalent bond distances between carbon, hydrogen, nitrogen, oxygen, and sulfur are all  $<2.0 \text{ \AA}$  in length. Strong hydrogen bonds involve close contacts ( $<3.0 \text{ \AA}$ ) between the donor and acceptor atoms. Van der Waals interactions fall off very rapidly as atoms become separated







**Figure 6.1: Cartesian Versus Conceptual Descriptions of Motion.** Conceptually, changes in protein structure can be viewed as the individual motions of many small atoms (circles in A and B), or alternative, as the relative motions of associated groups of atoms (blue and black groups in C and D). Although equivalent, the latter approach is simpler and much easier to comprehend.

by more than the length of their combined van der Waals radii, and  $\pi$ -stacking interactions involving aromatic side chains also require adjacency of atoms. Even potentially long range electrostatic interactions often involve contacts between oppositely charged groups that are only 3.0-5.0 Å apart.

Because most interactions act over a short range, one would expect adjacent atoms to often exhibit correlated motion. In other words, if an atom, A, is displaced by an angstrom, the atoms to which it is bonded will need to move as well, and so will the atoms A will be displacing. Those atoms not interacting with A would remain fixed (Figure 6.1, A and B). Imagine that A is surrounded by a neighborhood of a dozen atoms. What would be the best was to describe the



displacement process?

On the one hand, we could describe the displacement by recording the start and end coordinates of all atoms. This would be a laborious process that would provide little conceptual insight, but it would encapsulate all the information in an exact fashion. This is essentially what we do when we solve multiple crystal structures of a protein adopting different conformations. On the other hand, we could break the set of atoms down into two groups: group 1, the set of atoms that moves with A, and group 2, the set of atoms that does not move with A. The displacement process could then be described as group 1 moves one angstrom relative to group 2 (Figure 6.1, C and D). This simple description provides a better understanding of what actually occurred than would a list of coordinates and relative displacements.

Of course it is quite easy for the human mind to analyze the motion of a handful of atoms and provide a conceptual description of a process such as that shown in Figure 6.1. However protein structures consist of thousands of atoms. It is far beyond the capability of the human mind to *simultaneously* analyze all atoms and draw meaningful conceptual conclusions. However, a computer can analyze many thousands of parameters at once and distill the information down into something interpretable.

### 6.1.1 Why RigiMOL?

RigiMOL reduces the motions of many thousands of atoms down to conceptual motions of relatively few objects, referred to as domains. These domains consist of groups of atoms that show correlated motion. Once RigiMOL has identified domains in proteins, it can generate three dimensional movies which morph proteins from one conformation to another. This enables a human being to readily

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full, including the street name, number, and city. The list includes names such as Mr. J. H. Smith, Mr. W. B. Jones, and Mr. C. D. Brown, among others.

2. The second part of the document is a list of the names and addresses of the members of the committee who were present at the meeting. This list is also in alphabetical order and includes names such as Mr. J. H. Smith, Mr. W. B. Jones, and Mr. C. D. Brown, among others.

grasp motions in very complicated systems with minimal effort.

In order for a human being to achieve a conceptual understanding of complicated protein motions in the absence of a tool like RigiMOL, it is necessary to reduce the complexity of the system being analyzed to the point where a human can simultaneously focus on all of the components being compared. This could mean visualizing the motion of the  $C_{\alpha}$  atoms only in order to look for global conformational changes, or it could mean focusing on a small region of the protein. However, this simplification process can often introduce biases into the analyses since it is the human being that decides what to look at and what to compare. By necessity, the reference frame and comparison set are subjective.

In contrast, a program like RigiMOL is objective. It performs its analysis in a reference-frame independent manner by using internal coordinates of the structure. It also operates on the entire structure at once when it makes comparisons. Complete manual analysis of two or more protein structure can take hours or even days. In comparison, RigiMOL does its job in about five minutes for a 3000 atom molecule. A structural biologist equipped with RigiMOL can thus spend much more time on interpreting motions in protein structures instead of just trying to understand what the motions are.

As an added benefit, RigiMOL's 3-D morph sequences can be used to create highly effective animations graphics that clear illustrating how proteins change and adapt.

### **6.1.2 Background**

The dynamic nature of protein structures have led many research groups to investigate the nature, source, and effects of these motions [Gerstein, 1998]. One of the earliest examples of a large scale conformational change was that of the haemag-

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full, including the street name, city, and state.

2. The second part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of chairman and vice-chairman. The names are listed in alphabetical order, and the addresses are given in full, including the street name, city, and state.

glutinin membrane glycoprotein which brings about membrane fusion through coil helix transition resulting from a pH shift [Carr, 1993]. However, many much more subtle but important conformational changes occur in nature. For example, the highly allosteric enzyme glycogen phosphorylase undergoes a variety of structural changes in response binding of glucose-1-phosphate, glycogen, and AMP, and in response to phosphorylation [Browner, 1992].

Several research groups have attempted to develop algorithms for automated identification of protein domains. Visual inspection of difference maps is perhaps the simplest of these approaches [Rossman, 1974], and this can be augmented through use of clustering [Crippen, 1978]. More sophisticated approaches involve mathematical tools such as normal mode analysis [Levitt, 1985].

More recently, software for carrying out these types of analyses has started to appear. Nichols, et al. have developed an approach based on difference distance matrices which operates on sets of fixed residues groups as the "atomic" unit in its domain search [Nichols, 1995]. Wriggers and Schulten have developed an approach that operates only on  $\alpha$ -carbon coordinates and uses an adaptive least-squares fitting method to identify domains [Wriggers, 1997]. This method also locates effective rotation axes which can be visualized on the static structures. Most recently, Gerstein and Krebs have developed a web server for visualizing protein motions which employs a molecular dynamics based algorithm to generate trajectories between two conformations [Gerstein, 1999].

RigiMOL improves on the above work in the following ways: (1) it includes motions of all non-hydrogen atoms, (2) it uses both distance difference matrix and R.M.S.D. fitting criteria, (3) it provides both analytical results and dynamic trajectories, (4) it uses a deterministic algorithm, (5) it is highly automated, and (6) it is not restricted in its application to proteins only.

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full, including the street name and the city.

2. The second part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of chairman and vice-chairman.

3. The third part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of secretary and treasurer.



## 6.2 Software Development

### 6.2.1 Algorithm

An overview of how RigiMOL works is shown in Figure 6.2. It requires two or more structure files in PDB format. The residue numbers and chain identifiers in the structures must be aligned in order for RigiMOL to properly match them.

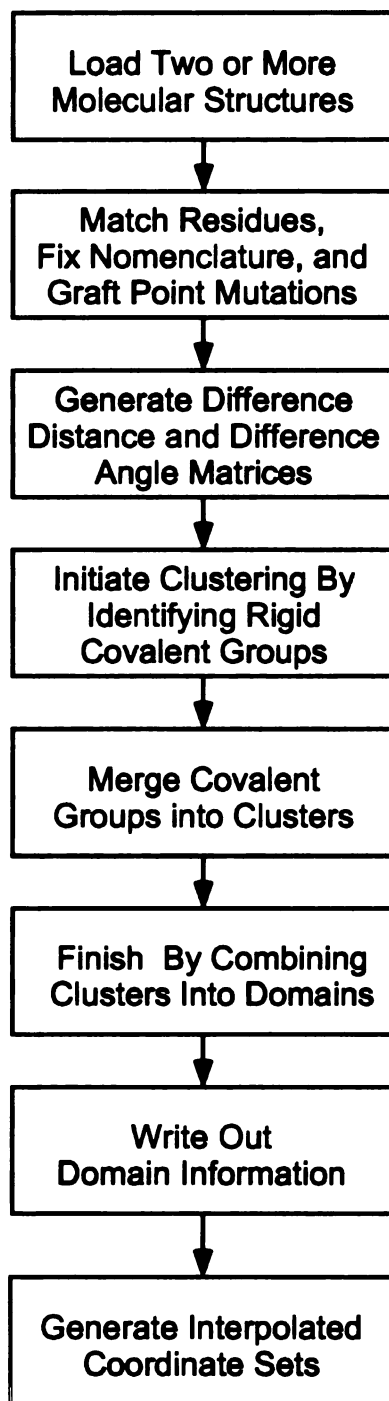
#### Matching

The matching process requires sorting of the structures, and then a careful geometric examination of the pseudo-chirality of equivalent atoms in the structure. This problem arises because although atoms such as  $C_{\gamma 1}$  and  $C_{\gamma 2}$  in a residue such as phenylalanine are chemically equivalent, they have different atom labels and different coordinates in the PDB file. Crystallographers do not generally adopt any consistent convention in labeling these atoms, so RigiMOL must make sure that all the structures it is going to compare are consistent with one another. RigiMOL uses the first PDB file as a template and then interchanges atoms in Phe, Leu, Val, Asp, Glu, Phe, Arg, and Tyr as required in order to obtain structures with matching pseudo-chirality.

For visualization purposes, RigiMOL also has the capability of morphing any of the natural 20 amino side chains into another natural side chain. If this option is activated, matching of atoms in such residues then would be carried out at this point. Alternatively, RigiMOL can simply just match the backbone atoms, deleting the unmatched side chain atoms.

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The names are: Mr. J. H. Smith, Mr. J. B. Jones, Mr. W. C. Brown, Mr. T. E. White, Mr. R. L. Green, Mr. S. D. Black, Mr. M. A. Gray, Mr. K. P. Blue, Mr. N. O. Red, Mr. H. I. Purple, Mr. G. F. Yellow, Mr. J. K. Orange, Mr. L. M. Silver, Mr. P. Q. Bronze, Mr. R. S. Gold, Mr. T. U. Platinum, Mr. V. W. Iron, Mr. X. Y. Copper, Mr. Z. A. Lead, Mr. B. C. Tin, Mr. D. E. Zinc, Mr. F. G. Nickel, Mr. H. I. Cobalt, Mr. J. K. Manganese, Mr. L. M. Magnesium, Mr. N. O. Calcium, Mr. P. Q. Sodium, Mr. R. S. Potassium, Mr. T. U. Lithium, Mr. V. W. Barium, Mr. X. Y. Strontium, Mr. Z. A. Rubidium, Mr. B. C. Cesium, Mr. D. E. Francium, Mr. F. G. Actinium, Mr. H. I. Thorium, Mr. J. K. Uranium, Mr. L. M. Plutonium, Mr. N. O. Americium, Mr. P. Q. Curium, Mr. R. S. Berkelium, Mr. T. U. Californium, Mr. V. W. Einsteinium, Mr. X. Y. Fermium, Mr. Z. A. Mendelevium, Mr. B. C. Nobelium, Mr. D. E. Lawrencium, Mr. F. G. Rutherfordium, Mr. H. I. Dubnium, Mr. J. K. Seaborgium, Mr. L. M. Bohrium, Mr. N. O. Hassium, Mr. P. Q. Tennessine, Mr. R. S. Oganesson.

2. The second part of the document is a list of the names and addresses of the members of the committee who were present at the meeting. The names are listed in alphabetical order, and the addresses are listed below each name. The names are: Mr. J. H. Smith, Mr. J. B. Jones, Mr. W. C. Brown, Mr. T. E. White, Mr. R. L. Green, Mr. S. D. Black, Mr. M. A. Gray, Mr. K. P. Blue, Mr. N. O. Red, Mr. H. I. Purple, Mr. G. F. Yellow, Mr. J. K. Orange, Mr. L. M. Silver, Mr. P. Q. Bronze, Mr. R. S. Gold, Mr. T. U. Platinum, Mr. V. W. Iron, Mr. X. Y. Copper, Mr. Z. A. Lead, Mr. B. C. Tin, Mr. D. E. Zinc, Mr. F. G. Nickel, Mr. H. I. Cobalt, Mr. J. K. Manganese, Mr. L. M. Magnesium, Mr. N. O. Calcium, Mr. P. Q. Sodium, Mr. R. S. Potassium, Mr. T. U. Lithium, Mr. V. W. Barium, Mr. X. Y. Strontium, Mr. Z. A. Rubidium, Mr. B. C. Cesium, Mr. D. E. Francium, Mr. F. G. Actinium, Mr. H. I. Thorium, Mr. J. K. Uranium, Mr. L. M. Plutonium, Mr. N. O. Americium, Mr. P. Q. Curium, Mr. R. S. Berkelium, Mr. T. U. Californium, Mr. V. W. Einsteinium, Mr. X. Y. Fermium, Mr. Z. A. Mendelevium, Mr. B. C. Nobelium, Mr. D. E. Lawrencium, Mr. F. G. Rutherfordium, Mr. H. I. Dubnium, Mr. J. K. Seaborgium, Mr. L. M. Bohrium, Mr. N. O. Hassium, Mr. P. Q. Tennessine, Mr. R. S. Oganesson.



**Figure 6.2: RigiMOL Overview.** RigiMOL converts three dimensional coordinate files into lists of concepts and interpolated coordinate files which can be used to generate dynamic movies of structural motions.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for consistent data collection practices and the use of advanced analytical techniques to derive meaningful insights from the data.

3. The third part of the document focuses on the role of technology in data management and analysis. It discusses how modern software solutions can streamline data collection, storage, and analysis processes, thereby improving efficiency and accuracy.

4. The fourth part of the document addresses the challenges associated with data management, such as data quality, security, and privacy. It provides strategies to mitigate these risks and ensure that the data remains reliable and secure throughout its lifecycle.

5. The fifth part of the document concludes by summarizing the key findings and recommendations. It stresses the importance of ongoing monitoring and evaluation to ensure that the data management processes remain effective and aligned with the organization's goals.

## **Measuring**

Throughout the entire domain identification process, RigiMOL operates in internal coordinates. These coordinates are based on the geometrical relationships of the chemical bonds. Following atom matching, two square matrices are generated of dimension equal to the number of bonds in the structure. The first is filled with the inter-bond distances, and the second with the inter-bond angles. These large matrices are used throughout the clustering process to measure similarity.

## **Clustering Algorithm**

Over the first two levels, RigiMOL's clustering algorithm operates on the assumption that elements of maximum similarity should be clustered together, provided that they are within appropriate proximity of one another. Differences between elements are sorted based on their differences, and then the closest pair is merged together. Then the next pair is considered, and so on.

However, when considering the merger of an element into an existing group, a cutoff threshold is applied. The element under consideration can not join the group if that element is too distant from an existing element in the group. Instead, it will either need to join another group, or remain isolated. The same rule is applied to merger of two existing groups, with one exception. If an element in one group can be transferred over to another group to give lower total deviations within both groups, then an element transfer will occur.

At the first level of clustering, only covalently bonded atoms are clustered together into groups. This stage is designed to locate structures such as aromatic rings and find other inflexible elements of protein structure (such as peptide bonds). By grouping these rigid elements together, the complexity of the problem



1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for consistent and reliable data collection processes to support informed decision-making.

3. The third part of the document focuses on the role of technology in modern data management. It discusses how advanced software solutions can streamline data collection, storage, and analysis, leading to more efficient and accurate results.

4. The fourth part of the document addresses the challenges associated with data security and privacy. It stresses the importance of implementing robust security measures to protect sensitive information from unauthorized access and breaches.

5. The fifth part of the document concludes by summarizing the key findings and recommendations. It reiterates the importance of a data-driven approach and encourages the organization to continue investing in data management capabilities to stay competitive in the market.

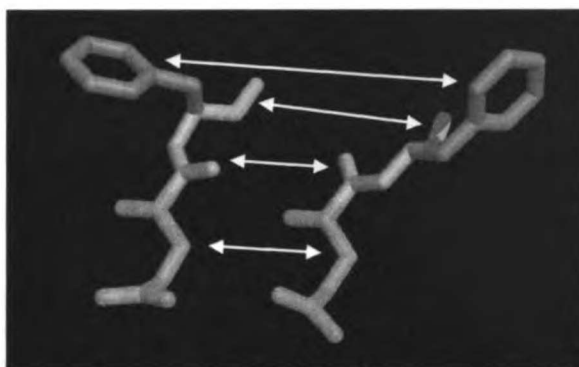


Figure 6.3: **RigiMOL Domains.** RigiMOL defines a domain as a collection of atoms that move together. The domains shown here in different colors are small and were generated using fairly stringent parameters. Larger domains (up to the size of complete protein folding unit) can be obtained through use of less stringent criteria.

is significantly reduced.

At the next level, covalent groups are merged together to form clusters. Clusters can be disconnected with respect to covalent structure, but must contain elements close enough to satisfy the neighbor cutoff criterion.

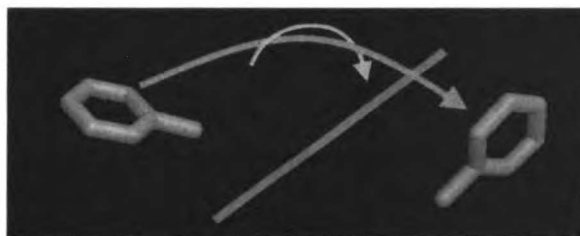
At the third level of clustering, a different approach is taken which involves a real-space R.M.S.D. comparison. In this case, the user specifies a minimum level of divergence (in ratios of Å R.M.S.D.) which two domains must exhibit to qualify as unique domains. Clusters which do not meet this threshold are merged together. Example domains are shown in Figure 6.3.

Throughout all levels of clustering, a shape cutoff is applied. This dimensionless quantity relates the number of bonds in a domain to the size of a sphere required to contain those bonds. Low shape cutoffs ( $\approx 25$ ) give much more spherical domains, while larger cutoffs (up to 500) permit much more extended structures. The primary function of this parameter is to prevent formation of unrealistically elongated groups and clusters at the first two levels of clustering.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for consistent and reliable data collection processes to ensure the validity of the results.





**Figure 6.4: Representation of a 3-D Transformation.** RigiMOL moves domains through space by rotating them about the characteristic axis (blue) of the 3-D transformation while coupling that movement to a screw rotation (yellow) about the path of motion (red).

### Fitting

Although the entire domain identification process was performed in internal coordinates, visualization of protein domains requires some kind of real-space superposition of the structures. The user can specify whether RigiMOL should simply superimpose the  $\alpha$ -carbon backbone or superimpose a particular identified domain.

### Interpolation

Once a superposition has been chosen, RigiMOL can then generate real-space trajectories of the structure moving from one conformation to another. For reasons of simplicity, determinism, and smoothness, this process is accomplished using distance geometry exclusively rather than molecular dynamics.

In order to convert one structure into another, each domain identified by RigiMOL must undergo a three dimensional transformation. Except for a few degenerate cases (rare), this transformation can be described as a rotation about a fixed axis combined with a screw rotation along the direction of motion (Figure 6.4).

By iterating along this transformation in a stepwise fashion, it is possible to

1. The first step is to identify the problem or goal. This involves understanding the current situation, the resources available, and the desired outcome. It is important to be clear and specific about what you want to achieve.

2. Next, you need to develop a plan. This involves breaking down the goal into smaller, manageable tasks and determining the order in which they should be completed. It is also important to consider potential obstacles and how to overcome them.

3. Once you have a plan, you need to execute it. This involves putting your plan into action and monitoring your progress. It is important to stay focused and motivated throughout the process.

4. Finally, you need to evaluate your results. This involves comparing your actual performance against your goals and identifying areas for improvement. It is important to be honest and objective in your evaluation.

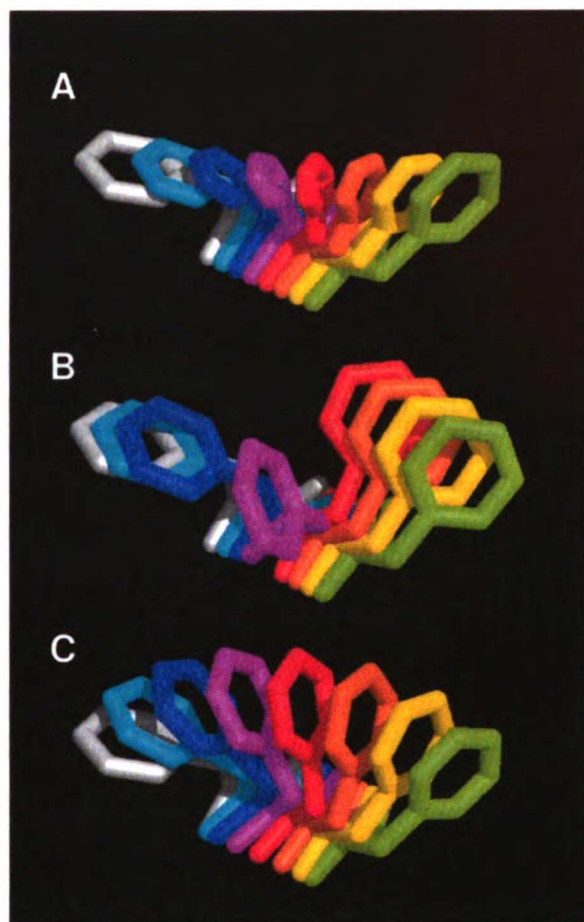
generate a trajectory which interpolates between the two structures. For most simple motions of side chains, the axis of motion coincides fairly well with an actual bond axis in a structure, resulting in a very natural looking transformation. The relative geometry of atoms within domains remain largely unchanged (by definition), so their motion is quite plausible. Thus, it is possible to obtain realistic looking trajectories of protein motion without doing a single step of molecular dynamics!

Some refinement to the initial transformations is applied in order to minimize distortion of bond lengths and steric clashes. In truth, the energies of the intermediate structures are still likely to be quite high, but the purpose of the software is to illustrate differences between states, not to generate energetically plausible pathways between them.

Figure 6.5 shows a comparison between RigiMOL's approach to interpolation and two other approaches. RigiMOL produces smoother trajectories than restrained molecular dynamics while still preserving much of the covalent geometry that would be distorted by a smooth linear interpolation.

### **6.2.2 Implementation**

RigiMOL was written in the C programming language on Linux workstations using the standard development tools. The 12,000 lines of source code are roughly modular and share a number of routines in common lineage with SiteFinder. The program is controlled using a script language, and reads and writes standard format PDB files as well as command scripts for MIDAS which aid in visualization of motions.



**Figure 6.5: Interpolation Comparison.** Three approaches to interpolated motion are compared. A linear interpolation (A) is smooth but distorts the covalent geometry of the structure, in this case compressing the aromatic ring to a third of its true size half-way through the interpolation. (B) molecular dynamics approaches introduce random variability into the interpolation and tend to hop discretely between low energy states. Such interpolations appear jerky and the extraneous motion is distracting. (C) RigiMOL generates a smooth path between two states while preserving the internal geometry of the domains in motion.

### 6.3 Demonstration

Here we demonstrate the application of RigiMOL to a small test system. The two structures being compared are crystal structures of two  $V_L$  chains of an Fab from [Muller, 1998], which are unremarkable except for the fact that their elbow angles

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for consistent and reliable data collection processes to support informed decision-making.

3. The third part of the document focuses on the role of technology in enhancing data management and analysis. It discusses how modern software solutions can streamline data collection, storage, and reporting, thereby improving efficiency and accuracy.

4. The fourth part of the document addresses the challenges associated with data management, such as data quality, security, and privacy. It provides strategies to mitigate these risks and ensure that data is used responsibly and ethically.

5. The fifth part of the document concludes by summarizing the key findings and recommendations. It stresses the importance of ongoing monitoring and evaluation to ensure that data management practices remain effective and aligned with the organization's goals.

```

load
  pdb filename v11.pdb end
  pdb filename v12.pdb end
end

reference
  bond_cutoff = 0.4
  neighbor_cutoff = 2.9
end

compare end

analyze
  group_distance_cutoff = 0.25 ! Distance (Angstroms)
  group_angle_cutoff = 30.00 ! Angle (Degrees) 0-180
  group_shape_cutoff = 25.00 ! Shape Ratio (25-500)

  cluster_distance_cutoff = 1.50 ! Distance (Angstroms)
  cluster_angle_cutoff = 45.00 ! Angle (Degrees) 0-180
  cluster_shape_cutoff = 100.00 ! Shape Ratio (25-500)

  domain_similarity_cutoff = 1.25 ! Ratio > 1.0
  domain_size_cutoff = 20.0 ! Atom Count
  domain_shape_cutoff = 500.00 ! Shape Ratio (25-500)

end

fit mode = domain domain = 1 end

write
  analysis
    prefix = m/v1
    min_domain_size = 2
    num_colors = 36
  end

  fits
    min_domain_size = 10
  end

  interpolation
    prefix = m/int
    steps = 30
  end

end

stop

```

Figure 6.6: **Example RigiMOL Script.** RigiMOL is controlled using a script language similar to other molecular modeling and refinement programs. Various parameters can be adjusted to suit the particular application.

differ by over 30 degrees.

The RigiMOL input file used in analyzing these two structures is shown in Figure 6.6. Because these structure are relatively small (only 1652 atoms) the entire analysis and interpolation process takes only a couple minutes on a Pentium II class machine.

In addition to providing a number of MIDAS scripts as output to assist in vi-

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for consistent and reliable data collection processes to support informed decision-making.

Domain	Size	Shape-Ratio	Residues
1	734	69.4	1-101
2	683	105.1	108,110-155,157-164,171-203,205-213
3	33	68.2	106-111,140
4	33	40.8	106,166-171
5	27	92.9	101-106
6	16	20.2	154-157
7	17	83.6	164-167
8	15	75.7	2-3
9	11	4.6	203-204
10	7	2.6	83
11	6	6.2	105
12	7	47.1	28
13	5	2.3	142
14	5	13.4	103
15	5	2.0	165
16	5	2.5	123
17	5	2.1	24
18	5	2.2	18
19	5	2.4	166
20	5	2.4	143
21	5	12.4	188
22	4	0.8	170
23	4	0.9	26
24	4	0.9	93-94
25	4	0.9	152
26	4	1.3	54
27	4	7.3	108
28	4	1.1	109
29	4	7.4	42
30	4	0.8	211
31	3	1.8	204-205
32	3	3.7	183
33	3	4.2	126
34	3	2.9	129
35	3	3.5	24
36	3	2.5	172
37	3	4.3	169
38	3	4.1	107
39	3	3.5	18
40	3	2.1	160

... (continues)

Figure 6.7: **Example Domain List.** RigiMOL outputs a list of domains which includes the atom count, shape ratio, and a list of residues whose atoms participate in the domain.

RMSD Statistics and relative RMSD of domain in column after domain in row.											
Overall Fit	= 3.06		1	2	3	4	5	6	7	8	9
Domain 1	Fit = 0.37			37.2	13.7	19.9	1.3	38.9	15.0	1.7	38.8
Domain 2	Fit = 0.34		29.2		4.0	4.7	8.8	1.1	2.4	39.5	3.5
Domain 3	Fit = 0.44		28.2	12.3		3.2	9.4	14.2	3.4	36.0	8.1
Domain 4	Fit = 0.38		40.0	11.8	4.2		15.7	12.1	2.5	51.9	14.9
Domain 5	Fit = 0.26		4.2	45.7	16.4	24.1		47.5	17.9	5.4	48.0
Domain 6	Fit = 0.23		36.0	5.1	13.3	15.6	15.5		10.5	52.8	9.2
Domain 7	Fit = 0.34		41.1	12.8	2.6	2.6	15.1	15.2		53.7	11.1
Domain 8	Fit = 0.41		7.6	23.4	13.3	14.2	7.3	24.7	11.7		29.2
Domain 9	Fit = 0.31		21.6	15.0	6.5	14.6	11.7	15.7	14.2	33.7	

Figure 6.8: **Domain Fit List.** RigiMOL outputs a matrix of domain fits for domains of a specified size. For each column, the R.M.S.D. for atoms within the indicated domain are shown first. Then, while this domain is superimposed, R.M.S.D. values are calculated for atoms in the other domains. For example, if atoms in domain 1 are superimposed in the two structures being compared, atoms in domain 2 have an average R.M.S.D. 37 times higher than that of atoms inside domain 1.



Rotations and Translations					
Domain	RMSD	Axis	Angle	Translation	Distance
1	0.37	( 0.00, 0.00, 0.00)	0.00	( 0.00, 0.00, 0.00)	0.00
2	0.34	(-0.81, 0.17, 0.44)	31.87	( 3.97, 11.87, 2.30)	12.73
3	0.44	(-0.70, 0.06, 0.69)	39.98	( 1.75, 3.50, 1.44)	4.17
4	0.38	(-0.87, 0.04, 0.48)	44.41	( 4.00, 0.25, 5.42)	6.74
5	0.26	(-0.48, 0.45, 0.54)	3.58	(-0.05, -0.02, 0.26)	0.26
6	0.23	(-0.89, 0.11, 0.38)	29.23	( 1.62, 13.70, -3.17)	14.15
7	0.34	( 0.71, -0.18, -0.56)	-41.41	( 2.56, 2.02, 3.76)	4.97
8	0.41	(-0.05, 0.04, 1.00)	9.93	(-0.06, 0.21, 0.17)	0.27
9	0.31	(-0.45, 0.17, 0.87)	27.63	( 3.48, 13.70, 0.98)	14.17
10	0.03	(-0.33, 0.33, -0.74)	-137.76	(-4.41, 0.33, 2.08)	4.88
11	0.48	(-0.53, 0.89, 0.13)	76.73	( 1.57, 1.22, -1.57)	2.53
12	0.42	( 0.68, -0.37, -0.45)	33.92	( 0.43, -0.04, 0.24)	0.49
13	0.05	( 0.48, 0.22, 0.58)	-90.91	( 1.99, 2.58, -0.52)	3.30
14	0.18	(-0.82, 0.39, -0.32)	86.92	(-0.65, -0.86, 0.19)	1.09
15	0.11	( 0.43, 0.69, -0.42)	-43.03	( 1.11, 0.97, 2.23)	2.67
16	0.33	(-0.48, 0.90, -0.18)	-79.77	( 7.23, 10.18, 4.84)	13.39
17	0.04	(-0.12, 0.91, 0.30)	92.94	( 0.55, 0.57, -1.58)	1.77
18	0.15	( 0.24, -0.48, 0.99)	-153.22	( 0.82, 1.23, 0.47)	1.55
19	0.25	( 0.14, 0.46, 0.69)	-19.67	( 1.62, 1.55, 2.51)	3.37
20	0.24	(-0.78, 0.40, -0.46)	88.81	( 0.70, 6.71, -1.33)	6.88
21	0.60	(-0.21, 0.94, -0.19)	21.84	( 3.65, 18.53, -0.06)	18.89
22	0.03	( 0.97, 0.08, -0.27)	-88.34	( 4.80, 3.38, 6.78)	8.97
23	0.03	(-0.51, 0.28, 0.54)	75.38	( 0.18, -0.35, 0.01)	0.40
24	0.04	( 0.77, -0.48, -0.18)	35.62	(-0.22, -0.10, 0.06)	0.25
25	0.05	(-0.39, 0.24, 0.88)	54.01	( 4.21, 18.84, -0.92)	19.33
26	0.05	(-0.34, 1.05, -0.92)	-174.90	(-0.25, 0.10, 0.07)	0.28
27	0.73	(-0.89, 0.58, 0.27)	134.71	( 2.88, 2.78, 2.74)	4.85
28	0.06	(-0.20, 0.81, -0.37)	38.48	( 2.86, 4.79, 3.04)	6.36
29	0.05	(-0.54, -0.20, 0.81)	39.66	( 1.16, -0.81, 0.55)	1.52
30	0.08	(-0.48, 0.13, -0.68)	-125.56	( 5.90, 18.17, 3.60)	19.44
31	0.15	(-0.25, 0.52, 0.53)	75.72	( 4.20, 14.22, 1.38)	14.89
32	0.01	(-0.06, 0.86, -0.34)	-37.71	( 4.76, 13.96, 2.66)	14.99
33	0.01	( 0.10, 0.88, 0.32)	-67.65	( 6.87, 11.36, 7.00)	15.01
34	0.01	( 0.69, -0.41, 0.58)	-86.66	( 3.97, 10.71, 1.44)	11.51
35	0.03	( 0.35, -0.88, -0.13)	-166.39	( 0.17, -0.05, -0.29)	0.34

**Figure 6.9: Example Domain Rotations and Translations.** After superimposing two structure according to a specified criterion, RigiMOL outputs a list of the rotations and translations required to superimpose domains on one another. These are the rotations and translations used as a basis for generating interpolated coordinate sets.

ualizing domains found in the structures, RigiMOL writes out statistics which can help in understanding and describing the domains and their motions. First, RigiMOL outputs a list of the domains in the structure, their size, their shape ratio, and which residues contain atoms in the various domains (Figure 6.7). Next, for domains down to a specified size cutoff, RigiMOL outputs R.M.S.D. statistics for atoms within each of the domains, and R.M.S.D. statistics for each of the other domains while one domain is being superimposed (Figure 6.8). This chart is analogous to a highly distilled difference distance matrix. Finally, RigiMOL outputs a list of the domains along with their rotation and translation statistics under the conditions of the current superposition (Figure 6.9).

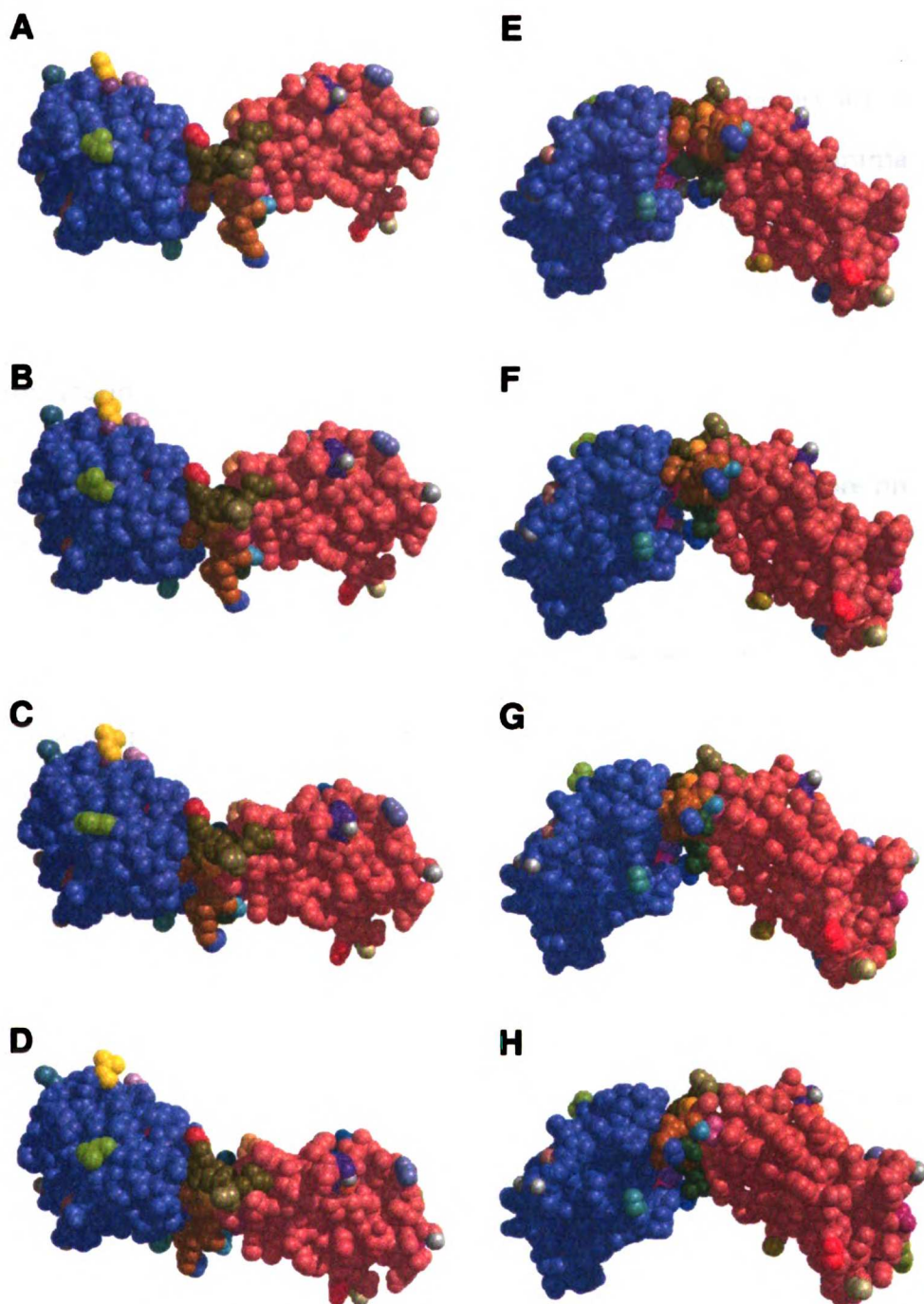


Figure 6.10: **Interpolation Demonstration.** Shown is an example set of structures of an Fab  $V_L$  chain from a RigiMOL interpolation. Atoms within a given domain are colored uniformly. For example, the large set of blue atoms is domain 1, and the large set of light red atoms is domain 2. (A-D) are interpolated structures 0%, 33%, 67%, and 100% through the interpolation. Domain 2 can be seen to rotate down about 30 degrees. (E-F) are the same set of structures shown with a 90% rotation about the  $X$ -axis.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes the need for transparency and accountability in financial reporting.

2. The second part of the document outlines the various methods and techniques used to collect and analyze data. It includes a detailed description of the experimental procedures and the tools used for data collection.

3. The third part of the document presents the results of the study, including a comparison of the different methods and techniques used. It discusses the strengths and weaknesses of each method and provides a summary of the findings.

4. The fourth part of the document discusses the implications of the study and provides recommendations for future research. It highlights the need for further investigation into the effectiveness of the different methods and techniques used.

5. The fifth part of the document provides a conclusion and a summary of the key findings. It emphasizes the importance of maintaining accurate records and the need for transparency and accountability in financial reporting.

However, the most effective way to utilize RigiMOL output is to generate a trajectory of the domain motion. Four frames from such a trajectory are shown in Figure 6.10 for the  $V_L$  domains. RigiMOL can be used to generate animations of protein motion, which at 30 frames per second appear perfectly smooth. Several of these animations are included on the CD-ROM.

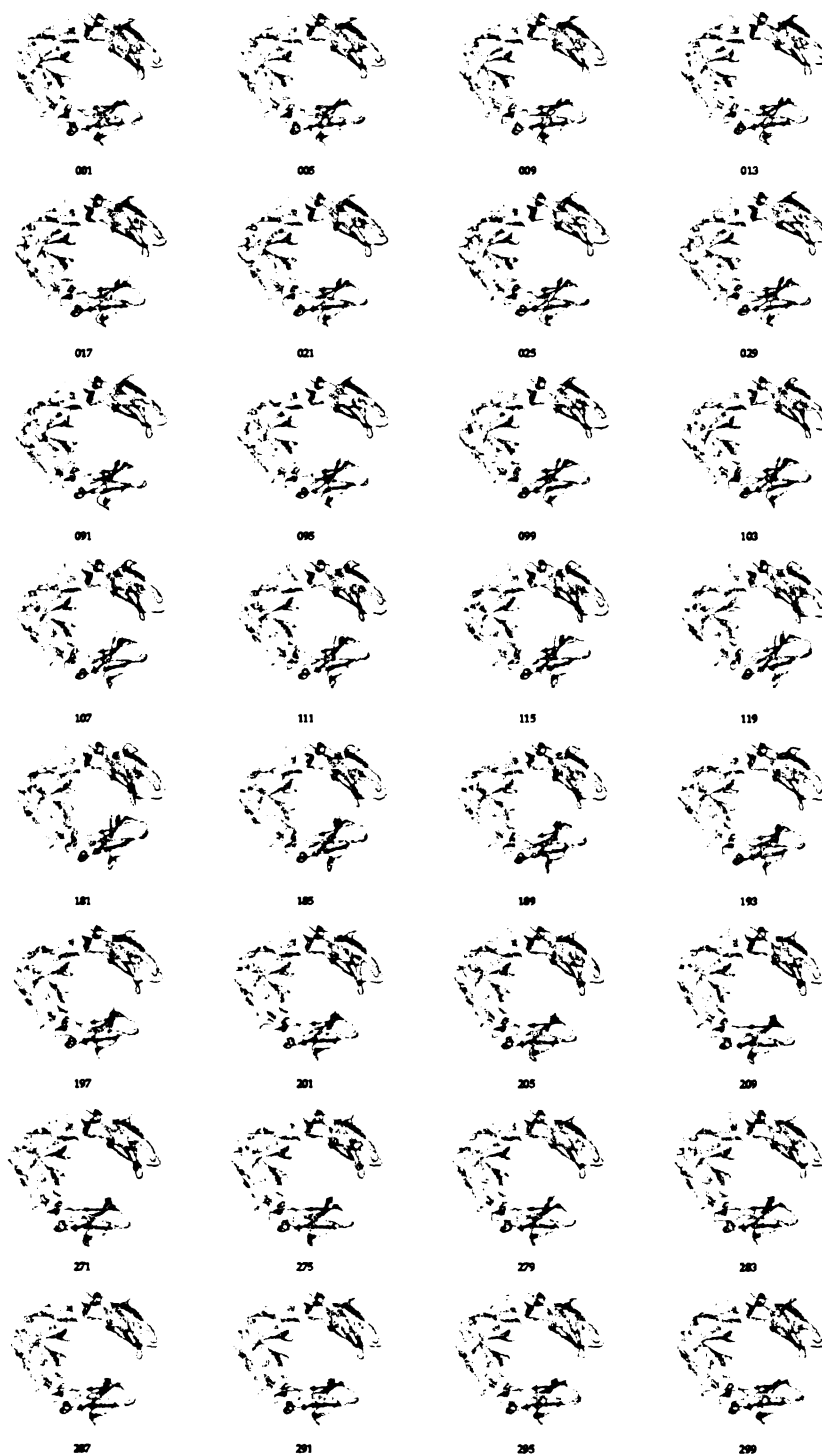
## 6.4 Applications

The potential applications of RigiMOL are many, but a few of the more promising ones are discussed here along with application to Fc.

### 6.4.1 Visualization of Global Conformational Changes in Fc

RigiMOL was applied to four of the Fc complex crystal structures in order to visualize how the global conformational changes would impact the consensus binding site. Selected frames from the movie are shown in Figure 6.11.

RigiMOL reveals that the dimeric unit of  $C_{H3}$  domains behaves almost as a single domain, with both subunits moving nearly in concert with one another. In contrast, the  $C_{H2}$  domains have no direct protein contacts with one another and move relatively independently. The  $C_{H2}$  domains show substantial movements and rotations relative to the  $C_{H3}$  domains, with the domain centers moving 2-3 Å and twisting as much as  $10^\circ$ . Naturally, these changes have a substantial effect on the  $C_{H2}$ - $C_{H3}$  domain cleft where the consensus binding site is located, since half of the binding site moves one way, while half moves another.



**Figure 6.11: Global Adaptability in the Fc Dimer.** Frames from a RigiMOL movie showing the adaptability of the Fc dimer bound to protein A (001), protein G (091), rheumatoid factor (181), and Fc receptor (271). The movie is included on the CD-ROM (Section C.2.1). In static form, the subtle differences are difficult to appreciate visually. For best results, directly compare the relative orientations of the three domains in frames 001, 091, 181, and 271.

24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200

201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300

### 6.4.2 Visualization of Adaptability in the Fc Binding Interface

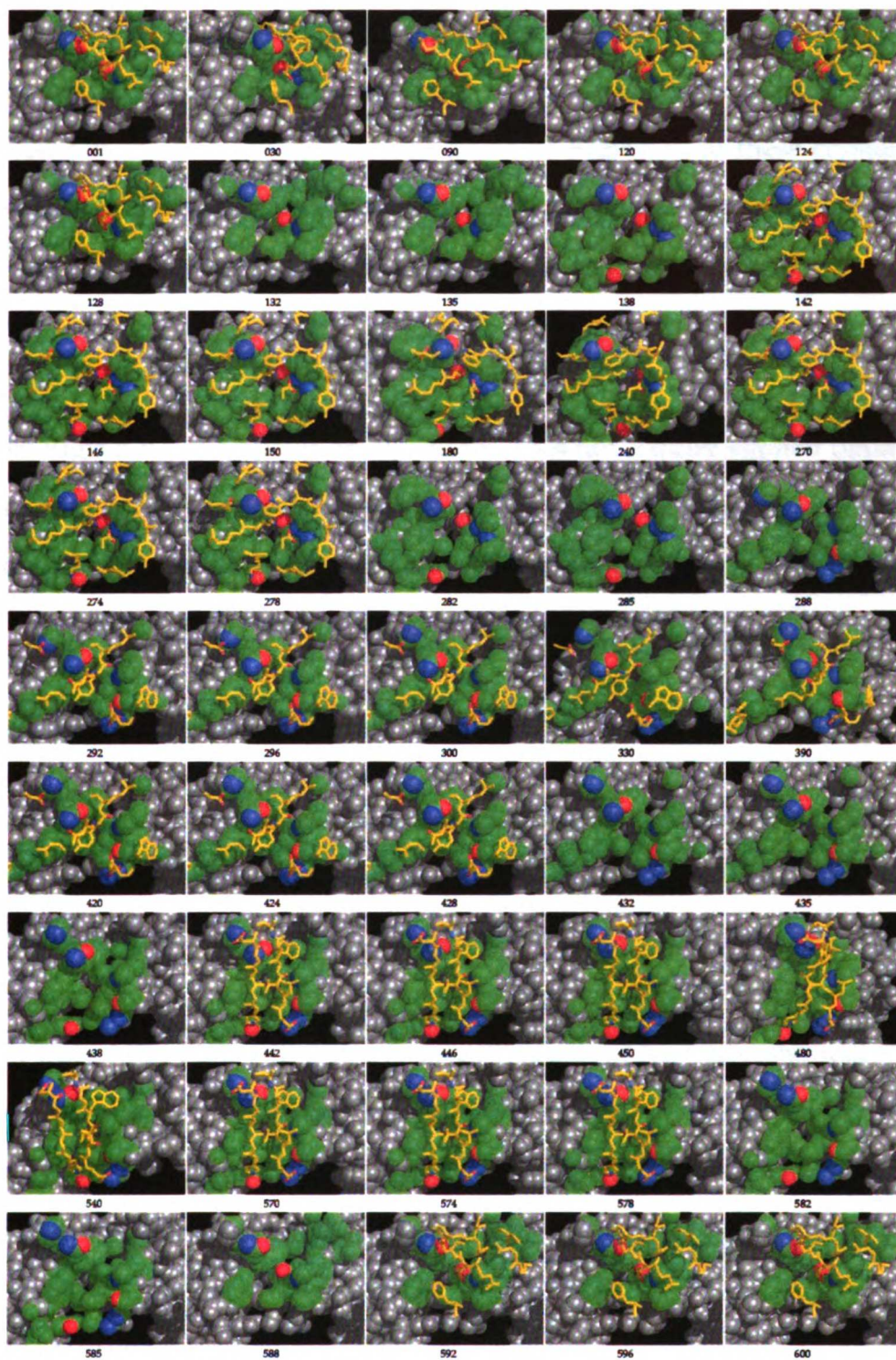
The adaptability of the consensus binding site on Fc was also visualized using RigiMOL (Figure 6.12). This movie demonstrates how the various side chains on Fc adapt to accommodate four different binding epitopes.

The relative motion the  $C_{H2}$  and  $C_{H3}$  domains mentioned above has considerable implications for the side chain residues which sit at the interface between these domains. Among these are Met252 and Met428, His 435, Glu380, Lys248, and Arg255. These predominantly long and flexible side chains accommodate the relative domain motions which lead to conformational changes at the bottom of the consensus binding site. The ability of Met252 and Met428 to form a flat surface for an aromatic ring such as Phe6 in protein A or Tyr98H in rheumatoid factor, or to form a pocket for Val10 in the peptide and Lys31 in protein G (Figure 5.7), appears to be tied in with this phenomenon.

The 250's loop on  $C_{H2}$  and the 430's loop on  $C_{H3}$ , which together line the sides of the consensus binding site, also show adaptability. Most of this is manifested in changes of side chain rotamer conformations of His433, Asn434, and Ile253, but the main chain scaffolding of these these loops also moves with the domains they are attached to. The adaptability of Asn434 in particular is important, since it adjusts to form buried hydrogen bonds with complementary groups in all four of the Fc binding interfaces.

### 6.4.3 Study of Enzyme Mechanisms

RigiMOL has already been applied to thymidylate synthase in order to visualize the link between the two active sites in the enzyme [Anderson, 1999], and it has also been used to visualize conformational changes resulting from inhibitor



**Figure 6.12: Adaptability of the Fc Consensus Binding Site.** Frames from a Rigi-MOL movies of the consensus binding region on Fc comparing related interactions of protein A (001), protein G (150), rheumatoid factor (300), and the peptide (450). The movie is difficult to appreciate here in static form, but can be viewed on the CD-ROM (Section C.2.1).



binding. The role of allostery in control of enzyme function has been known for decades, and many crystal structures have now been solved of allosteric enzymes in different states of regulation. Application of RigiMOL to these crystal structures could provide additional understanding about how the structural changes involved in allostery are propagated through the enzyme structure.

#### **6.4.4 Study of Protein/Nucleic Acid Complexes**

Because RigiMOL's interpolation algorithm does not require explicit parameterization, it can be applied to systems with non-protein groups such as nucleic acids, carbohydrates, or small molecule ligands. There are now many crystal structures available of such complexes which could be studied using RigiMOL.

#### **6.5 RigiMOL's Limitations**

Currently, the most significant limitations of RigiMOL are its requirement that the starting structures match with respect to residue and chain identifiers, and the  $N^2$  dependence on atom count. Incorporation of a protein sequence alignment algorithm could lift the first limitation and make it possible to compare structures straight out of the PDB without any modification. Fortunately, the latter limitation will become less important as computers have increasingly more memory. Currently, a 5000 atom molecule requires about 256MB of RAM and a half-hour to perform the analysis (virtual memory slows down the process 10 to 100-fold). A 10,000 atom structure would required a gigabyte of RAM and a couple hours to analyze. Fortunately, most scientific computing clusters already have this much RAM available.

## 6.6 Conclusion

The RigiMOL program is a powerful new tool for analyzing and visualizing structural changes in proteins. By breaking molecules down into rigid groupings, the process of both understanding and interpolating between structural changes can be greatly simplified. Applied to the Fc system, RigiMOL provides a clear picture of the conformational changes that occur in the Fc dimer as it binds to its various binding domains. We find that the global motions between protein domains are correlated with local changes in the protein structure at the consensus binding site. Since these structural changes appear necessary for the formation of specific interactions with the various Fc binding domains, the adaptability of Fc in this region probably plays an important role in making this binding site the preferred locus for binding on Fc.

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full. The list includes the names of the members of the committee, the names of the members of the sub-committee, and the names of the members of the advisory committee. The addresses are given in full, including the street name, the city, and the state.

2. The second part of the document is a list of the names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full. The list includes the names of the members of the committee, the names of the members of the sub-committee, and the names of the members of the advisory committee. The addresses are given in full, including the street name, the city, and the state.

3. The third part of the document is a list of the names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full. The list includes the names of the members of the committee, the names of the members of the sub-committee, and the names of the members of the advisory committee. The addresses are given in full, including the street name, the city, and the state.

## Chapter 7

### Conclusion

#### 7.1 Accomplishments

Using alanine scanning mutagenesis, we detected an energetic “hot spot” in the Protein A:Fc binding interface. A minimized locked-helix Protein A variant was designed to bind Fc by mimicking hot spot interactions. Although the peptide interacted with Fc as designed, much of the affinity was lost compared to the parent molecule. Apparently, simple mimicry of a hot spot is not sufficient to generate high affinity binding because the combined effects of many missing low-affinity interactions sap away much of the binding energy.

Next, a combinatorial library of disulfide constrained 20-mer peptides displayed on M13 bacteriophage was screened against Fc. From these libraries, peptides were identified which bound Fc with affinities 50-fold tighter than the designed locked-helix variants. Two subsequent rounds of optimization produced substantially improved peptides with affinities very close to that of Protein A, Z-domain ( $EC_{50}$ 's near 10 nM).

Analysis of the consensus sequence patterns and alanine scanning suggested that a short 13 residue peptide would also bind Fc with high affinity. This 1530 molecular weight peptide was prepared and found to exhibit an affinity for Fc of  $K_d=15$  nM. The crystal structure of this peptide was solved in complex with Fc,

revealing that the peptide sought out the same site for interaction on Fc as where the locked helix bound, and that it covered the same region where all four natural Fc binding proteins also bind.

Juxtaposition of the available Fc binding interface structures allowed identification of a set of conserved interactions, and showed that although the peptide was of a completely novel sequence, the interactions it used in binding Fc were highly analogous to those employed by the other natural Fc binding domains. The peptide essentially combines separate elements found across all of the Fc binders onto a compact and potent Fc binding molecule.

By comparing the overlapping binding interfaces of all the Fc binding proteins, a consensus binding site was detected. Two computer programs were created to analyze the surface properties and adaptability of the Fc dimer and of the consensus binding site in particular. The first program, SiteFinder, shows that the consensus binding site is one of only two large regions of highly accessible and non-polar surface on the Fc dimer. The second program, RigiMOL, provides a clear picture of how relative domain motions in the Fc fragment impact the shape of the consensus binding site, and how this adaptability enables the site to complement the diverse Fc binding proteins. Although created to analyze Fc, both programs should have broad applicability to other molecular systems as well.

## **7.2 What Was Learned?**

### **7.2.1 Molecular Engineering Requires Robust Technologies**

From our failure to generate potent minimized Fc binding peptides from protein A and from our subsequent success in generating such peptides using phage display, we learned the importance of applying the appropriate technology to a molecular

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full, including the street name, city, and state.

2. The second part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of chairman and vice-chairman.

3. The third part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of secretary and treasurer.

engineering problem. The simple reductionist approach of first finding the most critical elements of a protein interface, and then eliminating all other elements was found to be inadequate for creating molecules that bind Fc with high affinity. Instead, it was necessary to solve the problem in a much more robust fashion using combinatorial design and directed evolution.

Similarly, in trying to answer questions about the nature of the consensus binding site on Fc, we needed to resort to sophisticated computer programs capable of analyzing the motions of thousands of atoms and the properties of millions of surface patches. From these tools, we obtain clear conceptual suggestions regarding what features matter in binding to the Fc dimer. The hypotheses generated by this analysis can be tested and expanded upon in subsequent work.

### **7.2.2 Attractive Sites Exist on Protein Surfaces**

The remarkable convergence in binding both of natural proteins and selected peptides is a sure indicator that something distinguishes this particular site on the surface of Fc. Although this is not the first time that peptides have been shown to target protein binding sites, this is the first case where such a diverse panel of natural molecules has been shown to target such a site as well. Much work still needs to be done towards discovering what specific properties distinguish unusually attractive sites on protein surfaces. This consensus site on Fc appears to be distinguished by its solvent accessibility, non-polarity, and adaptability. Will these same properties be important in other attractive binding sites, or will other factors dominate? Answers to these questions should greatly improve our understanding of molecular recognition and enhance our ability to generate novel protein binding molecules.

### **7.2.3 Function Can Be Transferred Across Molecular Scaffolds**

The way in which the peptide has reproduced equivalent Fc binding interactions from the natural domains demonstrates convincingly that functional components in one system can be successfully utilized on the scaffold of another. This result validates the idea that protein interfaces can provide information useful in the development of smaller molecules which can bind proteins using protein-like interactions.

However, the large amount of surface area buried in the Fc:peptide complex suggests that large amounts of surface area may also need to be buried in the complexes of small molecules with protein interface sites. Compared to the protein A and protein G binding domains, the selected Fc binding peptides did not evolve a substantially smaller contact epitope on Fc. Instead they found a means of making more contacts with a very limited set of residues. Could a 500 molecular weight small molecule bury the same amount of surface area on Fc? It would seem difficult without exploiting some kind of concave crevice or cavity on the protein surface.

## **7.3 What Next?**

### **7.3.1 Practical Applications of the Fc Binding Peptides**

The engineered Fc binding peptides may have considerable practical use in purification of therapeutic antibodies. Currently, Protein A is the system of choice for this task, but it is fairly expensive. Perhaps these peptides would be a more cost effective tool for this procedure. In order to determine whether these peptides would be useful in antibody purification it will be necessary to find optimal immobilization substrates (such as resin) and the best conditions for both bind-



ing and elution. Given that the peptide shows a different pH dependence in Fc binding than protein A, the low pH shift used to elute IgG may not work for the peptides. From our knowledge of peptide binding, shifts to higher pH or to higher salt concentration would probably serve as alternative elution conditions.

### **7.3.2 Implications for Small Molecule Design**

The way in which the peptide re-used natural Fc binding interactions validates the idea of structure based design incorporating elements of protein recognition sites. In combinatorial design of a ligand targeted against a protein interface, this might involve including functional groups from the natural protein interface as substituents in the library.

A second theme coming out of this work is that, wherever possible, one should try to avoid strong adherence to a particular scaffold in targeting protein recognition sites. Instead, one should include a variety of scaffolds into the library (such as the various disulfide constrained loops we used), and then allow the screening process to pick out the best one. Once a good scaffold has been found, optimization should then be straightforward.

### **7.3.3 Identification of Potential Binding Sites**

The SiteFinder program may have considerable utility in identifying potent binding sites on protein surfaces. Such information could be used both in the process of target selection and in ligand design. In order to assess the potential value of SiteFinder for these applications, it will be necessary to perform a carefully controlled study with a variety of protein surfaces and binding sites to see how well it is able to pick out natural binding sites. In particular, it will be interesting to learn if SiteFinder can identify regions on protein surfaces which have also been

1. The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that this is crucial for ensuring the integrity and transparency of the financial system.

2. The second part of the document outlines the various methods used to collect and analyze data. It highlights the need for consistent and reliable data sources to support the analysis.

targeted by naive peptides, and perhaps monoclonal antibodies as well.

#### **7.3.4 Analysis of Structural Movement in Proteins**

The RigiMOL program has obvious applications throughout the field of structural biology. At present, the highest priority in development is on improving the quality of the program's user interface so that researchers around the world can easily apply the program to their unique problems. Currently, RigiMOL works in a stand alone fashion, and the output must be manually sent to a molecular graphics program for visualization and rendering of results. In the future, it should be possible to directly interface RigiMOL with a graphics program so that minimal work will be required to apply it.

#### **7.4 In Closing**

In this work, we have demonstrated the value of powerful technologies in tackling difficult biological programs. We have shown how combinatorial design can succeed where manual design fails. Then, by augmenting our experimental results with computational analysis, we have obtained a more complete understanding than we could have otherwise achieved. We are now faced with exciting new questions to pursue. Indeed this is very much in keeping with patterns that have emerged recently in the biological sciences: powerful experimental technologies generate tons of data, sophisticated analysis reduce the data down to meaningful concepts, and then humans apply the results and generate new hypotheses. This is true for a range of fields such as genomics, proteomics, crystallography, or drug design. The cycle of research continues...

## **Appendix A**

### **A Theoretical Guide to the Practice of Binding Measurements**

#### **A.1 Introduction**

The binding assay is one of the central tools employed in studying macromolecular interactions. Binding assays in combination with other methods can tell us what molecules interact, how tightly these molecules interact, what specific parts of the molecules matter most to binding, and what other molecules are capable of disrupting the binding.

Although the fundamental mathematics governing binding equilibria have been worked out many years ago, a practical gap still exists between the theory described in the average biophysics textbook and the actual assays performed in a biochemistry lab. This is to be expected since textbooks have to handle the general form of a problem and can not be burdened by specific details that exist in any given situation.

This appendix is intended to bridge the gap between theory and practice in one specific area by taking an in-depth look at several common types of binding assays employed in the typical biochemistry lab. Emphasis is on development of rigorous theoretical models that exactly describe the real binding assays, and on what those exact models tell us about how the assays should best be practiced, how the results should be interpreted, and what problems often arise. This

appendix should be useful to students seeking to develop a complete understanding of binding assays as well as current researchers in molecular recognition that would like to achieve a complete sense of the problems that can arise.

Although this appendix contains a number of mathematical formulas and derivations, an attempt has been made to isolate the complex math from the resulting conclusions. This should help someone seeking to grasp the practical implications of this work since he or she can skip over the derivations and focus on the discussions and conclusions. However, those seeking a complete understanding of how these results were obtained will appreciate the explicit derivations. Furthermore, many useful formulas and derivations have been included in order to encourage and assist others who might want to extend this work by developing models specific to their own unique binding assays.

**Assumptions:**

1. Binding is one-to-one.
2. Chemical activities are 1.0.
3. Non-specific binding is ignored.
4. The standard state is 1 molar concentration.

## A.2 Binding Equilibria

### A.2.1 Biomolecular Systems

#### Introduction

Bimolecular systems are easy to describe, and their behavior is governed by relatively simple relationships. A thorough understanding of how bimolecular systems behave will help greatly in trying to understand more complicated systems.

#### Definitions



The equilibrium dissociation constant is defined as:

$$K = \frac{[A][B]}{[AB]} \quad (\text{A.2})$$

Conservation requires:

$$[A]_T = [A] + [AB] \quad (\text{A.3})$$

$$[B]_T = [B] + [AB] \quad (\text{A.4})$$

#### Derivations

Solving for [AB]:

$$K = \frac{([A]_T - [AB])([B]_T - [AB])}{[AB]} \quad (\text{A.5})$$
$$[AB]K = ([A]_T - [AB])([B]_T - [AB])$$
$$[AB]K = [A]_T[B]_T - [AB][A]_T - [AB][B]_T + [AB]^2$$

$$0 = [AB]^2 + (-K - [A]_T - [B]_T)[AB] + [A]_T[B]_T$$

$$[AB] = \frac{K + [A]_T + [B]_T \pm \sqrt{(K + [A]_T + [B]_T)^2 - 4[A]_T[B]_T}}{2}$$

Since  $[AB]$  must be less than  $[A]_T$  and  $[B]_T$ , the relevant root can be identified:

$$[AB] = \frac{K + [A]_T + [B]_T - \sqrt{(K + [A]_T + [B]_T)^2 - 4[A]_T[B]_T}}{2} \quad (\text{A.6})$$

Equation (A.6) can be used to obtain  $[A]$  for known  $[A]_T$ ,  $[B]_T$ , and  $K$ . This enables direct computation of  $[A]$  and  $[B]$ .

$$[A] = [A]_T - [AB] \quad (\text{A.7})$$

$$[B] = [B]_T - [AB] \quad (\text{A.8})$$

There are some other relationships which prove handy from time to time in computing the fourth parameter of the system when given the other three:

$$K = \frac{[A]([B]_T - [AB])}{[AB]}$$

$$K[AB] = [A]([B]_T - [AB])$$

$$[A] = \frac{K[AB]}{[B]_T - [AB]}$$

$$[A] = \frac{K}{\frac{[B]_T}{[AB]} - 1} \quad (\text{A.9})$$

$$[B] = \frac{K}{\frac{[A]_T}{[AB]} - 1} \quad (\text{by symmetry}) \quad (\text{A.10})$$

$$K[AB] = [A][B]_T - [A][AB]$$

$$[AB](K + [A]) = [A][B]_T$$

$$[AB] = \frac{[A][B]_T}{K + [A]}$$

$$[AB] = \frac{[B]_T}{\frac{K}{[A]} + 1} \quad (\text{A.11})$$

$$[AB] = \frac{[A]_T}{\frac{K}{[B]} + 1} \quad (\text{by symmetry}) \quad (\text{A.12})$$

The fraction bound of either component can be obtained from equations (A.11) and (A.12) by dividing through by  $[B]_T$  and  $[A]_T$  respectively.

$$f_B = \frac{1}{\frac{K}{[A]} + 1} \quad (\text{A.13})$$

$$f_A = \frac{1}{\frac{K}{[B]} + 1} \quad (\text{A.14})$$

These equations can be rearranged to give  $K$  from the fraction bound of one component and the free concentration of the other.

$$\frac{K}{[A]} + 1 = \frac{1}{f_B}$$

$$\frac{K}{[A]} = \frac{1}{f_B} - 1$$

$$K = \frac{[A](1 - f_B)}{f_B} \quad (\text{A.15})$$

$$K = \frac{[B](1 - f_A)}{f_A} \quad (\text{by symmetry}) \quad (\text{A.16})$$

Alternatively, the free concentration of one component can be obtained from  $K$  and the fraction bound of the second component.

$$\frac{[A]}{K} = \frac{f_B}{1 - f_B}$$

$$[A] = \frac{K f_B}{1 - f_B}$$



$$[A] = \frac{K}{\frac{1}{f_B} - 1} \quad (\text{A.17})$$

$$[B] = \frac{K}{\frac{1}{f_A} - 1} \quad (\text{by symmetry}) \quad (\text{A.18})$$

## Discussion

Here are some basic data points to keep in mind:

1. At very low concentrations of one component, binding is completely governed by the concentration of the second component. For example, with very low  $[A]$ , 50% of A will be bound when  $[B] = K$ , 90% of A will be bound when  $[B] = 10K$  and 99% of A will be bound when  $[B] = 100K$
2. When  $[A] = [B] = K$ , only 38.2% of each component will be bound. At ten-fold higher concentrations, 73.0% will be bound. Even at 100-fold higher concentrations, still only 90% of each component will be bound.
3. A 10-fold reduction in the free concentration of one component coupled with a ten-fold increase in the other will give the same amount of bound complex.

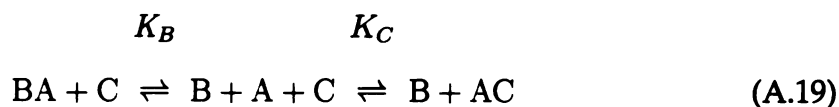
## A.2.2 Trimolecular Systems

### Introduction

Analysis of bimolecular systems is useful for binding assays where there is some kind of internal readout of the bound species (for example, the change in tryptophan fluorescence upon binding). However, many binding assays commonly used employ binding of a third ligand as a surrogate for readout of the bound concentrations of the first two components. In order to describe these assays, it is necessary to study the equilibria of trimolecular systems.

### Definitions

B and C are in competition for binding to A.



The equilibrium dissociation constants are defined as:

$$K_B = \frac{[A][B]}{[AB]} \quad (\text{A.20})$$

$$K_C = \frac{[A][C]}{[AC]} \quad (\text{A.21})$$

Conservation requires:

$$[A]_T = [A] + [AB] + [AC] \quad (\text{A.22})$$

$$[B]_T = [B] + [AB] \quad (\text{A.23})$$

$$[C]_T = [C] + [AC] \quad (\text{A.24})$$

Combining (A.20) and (A.23) and solving for [AB]:

$$\begin{aligned}
 K_B &= \frac{[A]([B]_T - [AB])}{[AB]} \\
 [AB]K_B &= [A][B]_T - [AB][A] \\
 [AB](K_B + [A]) &= [A][B]_T \\
 [AB] &= \frac{[A][B]_T}{K_B + [A]} \\
 [AB] &= \frac{[B]_T}{\frac{K_B}{[A]} + 1} \quad (\text{A.25}) \\
 [AC] &= \frac{[C]_T}{\frac{K_C}{[A]} + 1} \quad (\text{by symmetry}) \quad (\text{A.26})
 \end{aligned}$$

Combining (A.22), (A.25), and (A.26):

$$[A]_T = [A] + \frac{[B]_T}{\frac{K_B}{[A]} + 1} + \frac{[C]_T}{\frac{K_C}{[A]} + 1} \quad (\text{A.27})$$

The relevant root to this cubic equation (A.27) is (courtesy of Mathematica):

$$[A] = \frac{-\alpha}{3} - \frac{2^{\frac{1}{3}}(3\beta - \alpha^2)}{3 \left[ 27\gamma + 9\alpha\beta - 2\alpha^3 + \sqrt{4(3\beta - \alpha^2)^3 + (27\gamma + 9\alpha\beta - 2\alpha^3)^2} \right]^{\frac{1}{3}} + \left[ 27\gamma + 9\alpha\beta - 2\alpha^3 + \sqrt{4(3\beta - \alpha^2)^3 + (27\gamma + 9\alpha\beta - 2\alpha^3)^2} \right]^{\frac{1}{3}}} \quad (\text{A.28})$$

$$\begin{aligned}
 \alpha &= K_B + K_C - [A]_T + [B]_T + [C]_T \\
 \beta &= K_B K_C - [A]_T(K_C + K_B) + [B]_T K_C + [C]_T K_B \\
 \gamma &= [A]_T K_B K_C
 \end{aligned}$$

Equation (A.28) can be used to obtain [A] for known [A]<sub>T</sub>, [B]<sub>T</sub>, [C]<sub>T</sub>, K<sub>B</sub>, and

$K_C$ . All of the other parameters for the system can be computed:  $[AB]$  and  $[AC]$  from equations (A.25) and (A.26) and  $[B]$ ,  $[C]$ ,  $f_B$  and  $f_C$ , from the relationships below.

$$[B] = [B]_T - \frac{[B]_T}{\frac{K_B}{[A]} + 1} \quad (\text{A.29})$$

$$[C] = [C]_T - \frac{[C]_T}{\frac{K_C}{[A]} + 1} \quad (\text{A.30})$$

$$f_B = \frac{1}{\frac{K_B}{[A]} + 1} \quad (\text{A.31})$$

$$f_C = \frac{1}{\frac{K_C}{[A]} + 1} \quad (\text{A.32})$$

Provided that  $[A]$  and one of the dissociation constants is known, the second dissociation constant can be obtained using the following relationships derived from equation (A.27):

$$K_B = \frac{[A] \left[ [A]^2 + (K_C - [A]_T + [B]_T + [C]_T)[A] + ([C]_T - [A]_T)K_C \right]}{[A]_T K_C + ([A]_T - [C]_T - K_C)[A] - [A]^2} \quad (\text{A.33})$$

$$K_C = \frac{[A] \left[ [A]^2 + (K_B - [A]_T + [B]_T + [C]_T)[A] + ([B]_T - [A]_T)K_B \right]}{[A]_T K_B + ([A]_T - [B]_T - K_C)[A] - [A]^2} \quad (\text{A.34})$$

## Discussion

Unfortunately, solving trimolecular systems in the general case can be difficult because equation (A.28) involves complex intermediates that can not be handled by a pocket calculator or spreadsheet. Analysis programs such as Mathematica are required for this task. Many of the graphs shown in the following sections rely heavily on being able to evaluate this equation. However, in any particular situation, there are usually aspects which make it possible to solve the system

### A.2.3 Trimolecular Systems (Special Case $K = K_B = K_C$ )

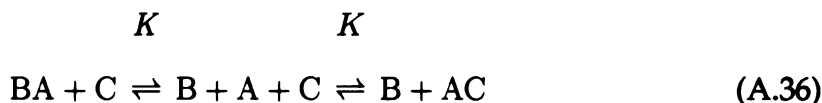
#### Introduction

In certain kinds of trimolecular assays, the only difference between one ligand and another will be the presence of radioactive or spectroscopic label. Often these labels have little or no effect on the affinity, so we can solve the system under the assumption that the two dissociation constants are equal. This assumption makes the problem much easier.

#### Definitions

$$K = K_B = K_C \quad (\text{A.35})$$

B and C are in competition for binding to A.



#### Derivations

(A.27) reduces to:

$$[\text{A}]_T = [\text{A}] + \frac{[\text{B}]_T + [\text{C}]_T}{\frac{K}{[\text{A}]} + 1} \quad (\text{A.37})$$

which can be expanded into a quadratic:

$$\begin{aligned} [\text{A}]_T &= [\text{A}] + \frac{[\text{A}]( [\text{B}]_T + [\text{C}]_T )}{K + [\text{A}]} \\ K[\text{A}]_T + [\text{A}]_T[\text{A}] &= [\text{A}]^2 + [\text{A}]K + [\text{A}]( [\text{B}]_T + [\text{C}]_T ) \\ 0 &= [\text{A}]^2 + (K - [\text{A}]_T + [\text{B}]_T + [\text{C}]_T)[\text{A}] - [\text{A}]_T K \end{aligned}$$

with the relevant solution:

$$[A] = \frac{[A]_T - [B]_T - [C]_T - K + \sqrt{(K - [A]_T + [B]_T + [C]_T)^2 + 4[A]_T K}}{2} \quad (\text{A.38})$$

which can be easily computed from known  $[A]_T$ ,  $[B]_T$ ,  $[C]_T$ , and  $K$ . The remaining parameters of the system can be obtained using equations (A.25-A.26) and (A.29-A.32). Also note that equation (A.37) can be solved to give  $K$  as a function of  $[A]$  and the other parameters:

$$\begin{aligned} K([A]_T - [A]) &= [A]^2 + (-[A]_T + [B]_T + [C]_T)[A] \\ K &= \frac{[A]^2 + ([B]_T + [C]_T - [A]_T)[A]}{[A]_T - [A]} \end{aligned} \quad (\text{A.39})$$

### A.3 Direct Binding Assays: Two Component Systems

#### A.3.1 Dose-Response



**Summary:** With a fixed target concentration ( $[T]_T$ ), the ligand concentration ( $[L]_T$ ) is varied, and the response due to the formation of complex ( $[TL]$ ) is recorded. The ligand concentration  $[L]_{T(50)}$  that gives a half-maximal response is used to estimate  $K$ .

**Background:** This is the simplest way to determine a binding affinity, but it is only reliable under conditions where the target  $T$  is present at a concentration substantially below the dissociation constant ( $[T]_T \ll K$ ). It also relies on the assumption that the response being recorded is directly proportional to the concentration of complex over all concentrations.

**Examples:**

1. Measuring cell growth as function of cytokine concentration.
2. Measuring a change in target fluorescence as a function of ligand binding.

**Discussion:** The response at a given ligand concentration is usually recorded as a fraction of the maximal response obtained and is assumed to correspond to the following:

$$f = \frac{[TL]}{[T]_T} \quad (\text{A.41})$$



At half maximal response, when  $[L]_T$  has been titrated to the  $[L]_{T(50)}$ , the concentration of complex should equal half of the total target concentration.

$$f = \frac{1}{2} = \frac{[TL]}{[T]_T} \quad (\text{A.42})$$

$$[TL] = \frac{[T]_T}{2} \quad (\text{A.43})$$

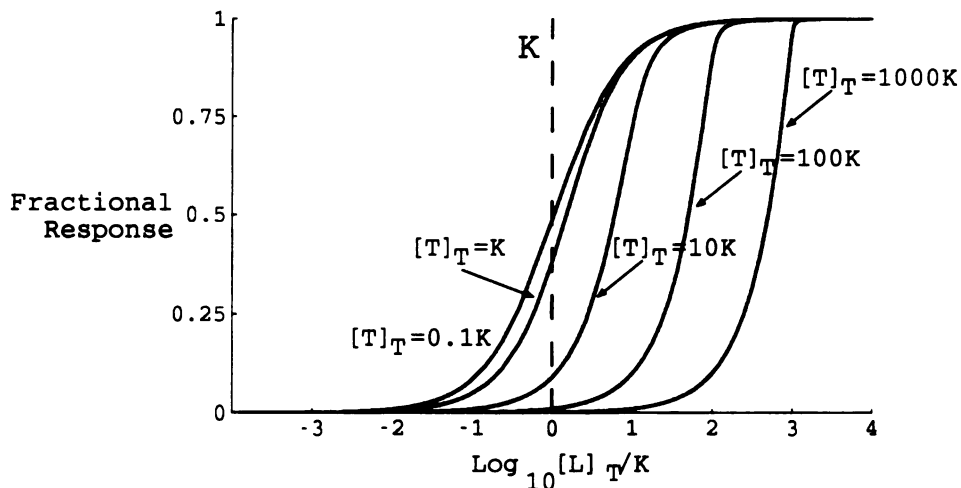
combining (A.5) with (A.43) gives:

$$\begin{aligned} K &= \frac{([T]_T - [TL])([L]_T - [TL])}{[TL]} \\ K &= \frac{([T]_T - \frac{[T]_T}{2})([L]_{T(50)} - \frac{[T]_T}{2})}{\frac{[T]_T}{2}} \\ K &= [L]_{T(50)} - \frac{[T]_T}{2} \\ [L]_{T(50)} &= K + \frac{[T]_T}{2} \end{aligned} \quad (\text{A.44})$$

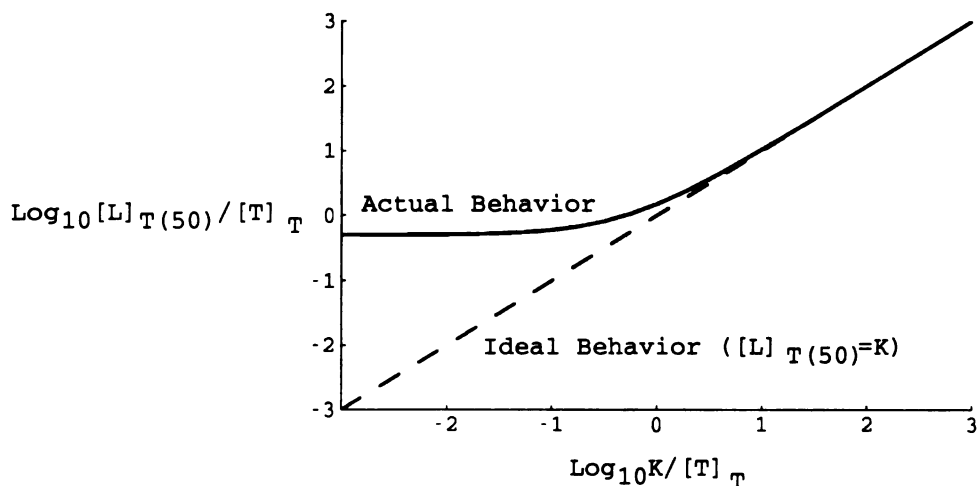
The  $[L]_{T(50)}$  clearly depends on both the affinity of the interaction and the total concentration of target in the experiment.

A series of dose-response curves is shown in Figure A.1 for different target concentrations ( $[T]_T$ ). Notice how the shape of the curve changes as  $[T]_T$  ranges above  $K$ , becoming steeper and asymmetric. In cases when the data is particularly good, this behavior might be useful as a diagnostic. A good rule of thumb is that a titration curve should span two orders of magnitude in ligand concentration from the 5% to 95% signal range.

Bottoming out of the assay occurs when  $K$  is larger than  $[T]_T$ . The solid line in Figure A.2 shows the actual measured  $[L]_{T(50)}$  as a function of the ratio between  $K$  and the target concentration ( $[T]_T$ ). The unfortunate implication of this behavior is that improvements in ligand affinity that push  $K$  below  $[T]_T$  will be partially or



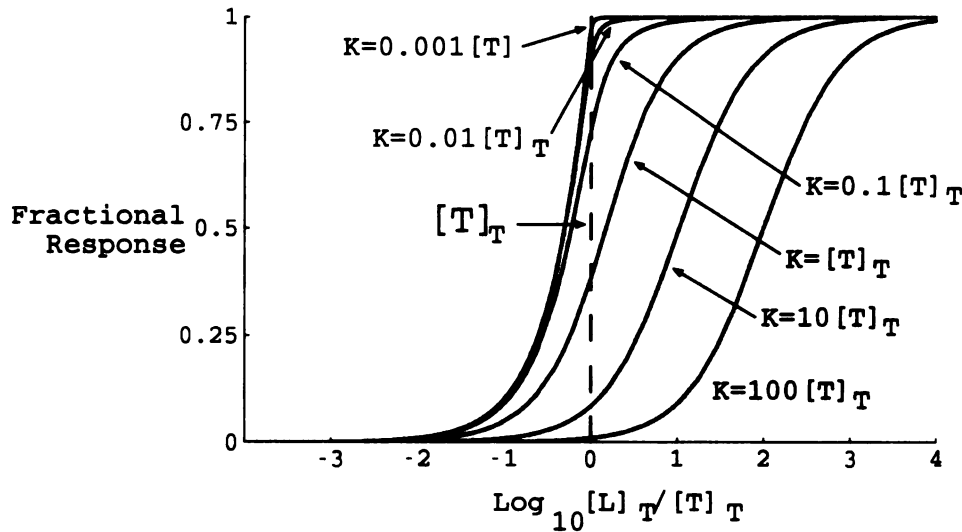
**Figure A.1: Dose Response Measurements.** Fractional response curves are shown as a function of ligand concentration ( $[L]_T$ ) at various target concentrations ( $[T]_T$ ), with all concentrations relative to the the dissociation constant ( $K$ ).



**Figure A.2: Limitations of Dose-Response Measurements.** The ligand concentration ( $[L]_{T(50)}$ ) that gives a half-maximal response is shown as a function of the dissociation constant ( $K$ ), with both axes relative to the total target concentration ( $[T]_T$ ). Actual behavior (solid line) is shown to deviate from ideal linear behavior (dashed line).

completely masked in this kind of assay.





**Figure A.3: Dose-Response as a Function of Ligand Affinity.** Fractional response curves for ligands with various dissociation constants ( $K$ ) are shown as a function of total ligand concentration ( $[L]_T$ ), with both values relative to the total target concentration ( $[T]_T$ ).

Figure A.3 contains a series of curves that might be obtained in an experiment where ligands of various affinities are compared. Notice the close similarity between the curve obtained for a ligand with an affinity 10-fold below  $[T]_T$  compared to one 1000-fold below  $[T]_T$ . In a real experiment with noise in the data, these curves would be indistinguishable. One must therefore be highly suspect of binding affinities obtained by this method if the curves show any symptoms of bottoming out, such as becoming steep and narrow. In such a case, one should view the measured affinities merely as upper bounds on the true affinities.

### A.3.2 Scatchard Analysis



**Summary:** Measure  $[TL]$  and  $[L]$  over a range of conditions that achieve good sampling of  $[TL]$  over the interval 0 to  $[T]_T$ . Graph  $\frac{[TL]}{[L]}$  vs.  $[TL]$ . If binding is one-to-one and non-cooperative, the data should fit to a straight line with a slope of  $-\frac{1}{K}$  and an  $x$ -intercept of  $[T]_T$ .

**Background:** This is the classic method for determining binding affinities and it is reliable even when the target T is present at a concentration at or above the dissociation constant ( $[T]_T \approx K$ ). However, if  $[T]_T$  is much greater than  $K$ , or if  $K$  itself is rather large (weak binding) then it will be difficult to obtain a well-distributed sampling of  $[TL]$  since saturating conditions may not be achievable.

In systems that do not involve simple one-to-one binding, the linearity of the curve and the  $x$ -intercept can also provide information about cooperativity and stoichiometry of the interaction.

**Discussion:** The derivation of the Scatchard Equation follows. Combining Equations (A.2) and (A.4):

$$K = \frac{[L]([T]_T - [TL])}{[TL]}$$
$$[TL] = \frac{[L]([T]_T - [TL])}{K}$$
$$\frac{[TL]}{[L]} = \frac{[T]_T - [TL]}{K}$$
$$\frac{[TL]}{[L]} = \frac{[T]_T}{K} - \frac{[TL]}{K} \quad (A.46)$$

Equation (A.46) is the classic form of the Scatchard Equation. Dividing by  $[T]_T$  gives an alternate form:

$$\begin{aligned}\frac{[TL]/[T]_T}{[L]} &= \frac{1}{K} - \frac{[TL]/[T]_T}{K} \\ \frac{f_T}{[L]} &= \frac{1}{K} - \frac{f_T}{K}\end{aligned}\tag{A.47}$$

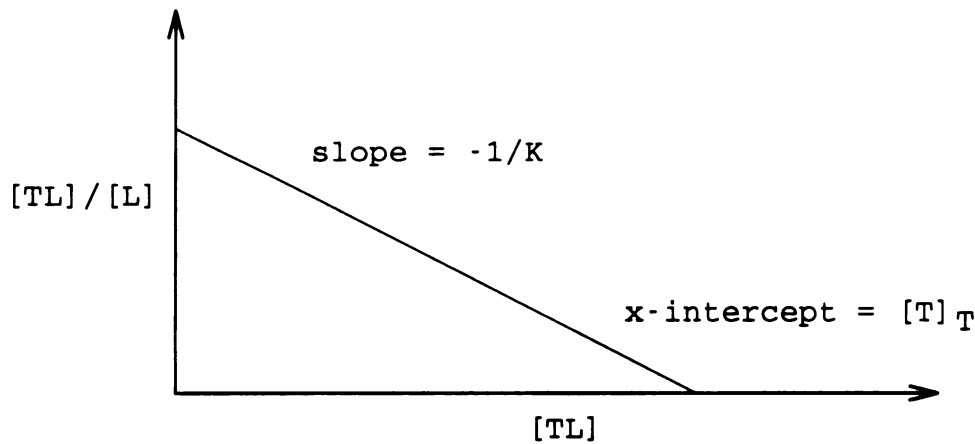
where  $f_T$  is the fraction of  $[T]_T$  bound.

$$f_T = \frac{[TL]}{[T]_T}\tag{A.48}$$

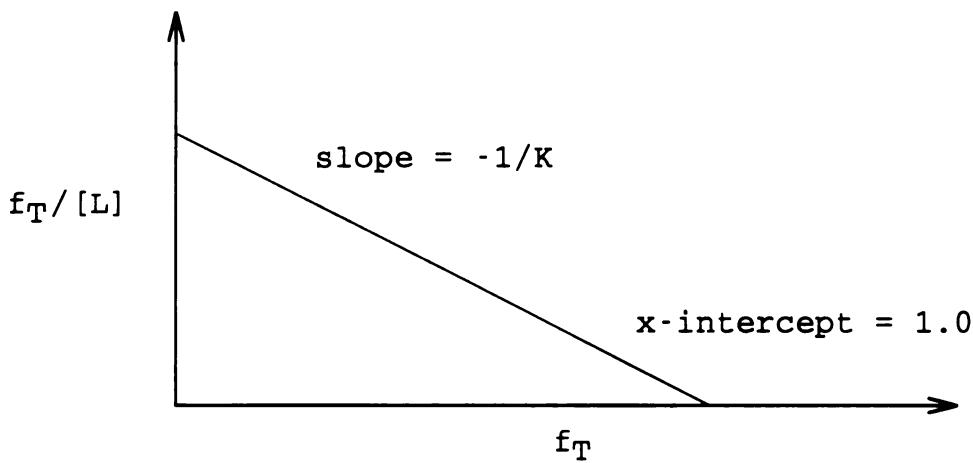
Graphs of the two modes of Scatchard analyses corresponding to equations (A.46) and (A.47) are shown in Figures A.5 and A.5 respectively. The first mode requires knowledge of  $[TL]$  and  $[L]$  in order to calculate  $K$ , which usually means that some separation of the bound target from the unbound target must be accomplished as part of the experiment. The difficulty of this step varies significantly from system to system.

The second mode only requires knowledge of  $f_T$  and  $[L]$ .  $f_T$  may be easier to obtain in various biophysical experiments (i.e. fluorescence quenching) so long as  $f_T$  can be reliably measured over the full range. However, one must be able to observe saturation behavior in the system so that an unambiguous signal level for 100% fractional binding can be established and used to calculate  $f_T$  at lower signal levels.

Unfortunately, deviations from linearity can sometimes occur when trying to achieve very high concentrations of one component due to aggregation and precipitation phenomena. Since these deviations occur near the extreme points of the linear fit, they can seriously affect the measured affinities. For this reason, the



**Figure A.4: The Scatchard Plot.** The concentration ratio of bound target over free ligand ( $[TL]/[L]$ ) is graphed as a function of the bound target ( $[TL]$ ) concentration. The  $x$  intercept gives the total target concentration ( $[T]_T$ ) and the slope gives the negative inverse of the dissociation constant ( $K$ ).



**Figure A.5: An Alternative Scatchard Plot.** The fractional saturation of target divided by the free ligand concentration ( $f_T/[L]$ ) is graphed as a function of fractional target saturation ( $f_T$ ). The  $x$  intercept should be 1.0 and the slope gives the negative inverse of the dissociation constant ( $K$ ).

Scatchard linearization is no longer considered the optimum method for measuring affinities.

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full, including the street name, city, and state.

2. The second part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of chairman and vice chairman.

3. The third part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of secretary and treasurer.



### A.3.3 Direct Fitting of Titration Curves



**Summary:** Measure  $[TL]$  and  $[L]$  over a range of conditions that achieve good sampling of  $[TL]$  over the interval 0 to  $[T]_T$ . Fit the data to equation (A.50) below using curve fitting software in order to obtain the dissociation constant  $K$ .

**Background:** Now that virtually every scientist has access to computerized fitting programs, it is no longer necessary to employ linearizations that over-emphasize certain data points by applying linearizations such as those required for Scatchard Analysis. Instead, one can simply make measurements and then fit them to the appropriate model.

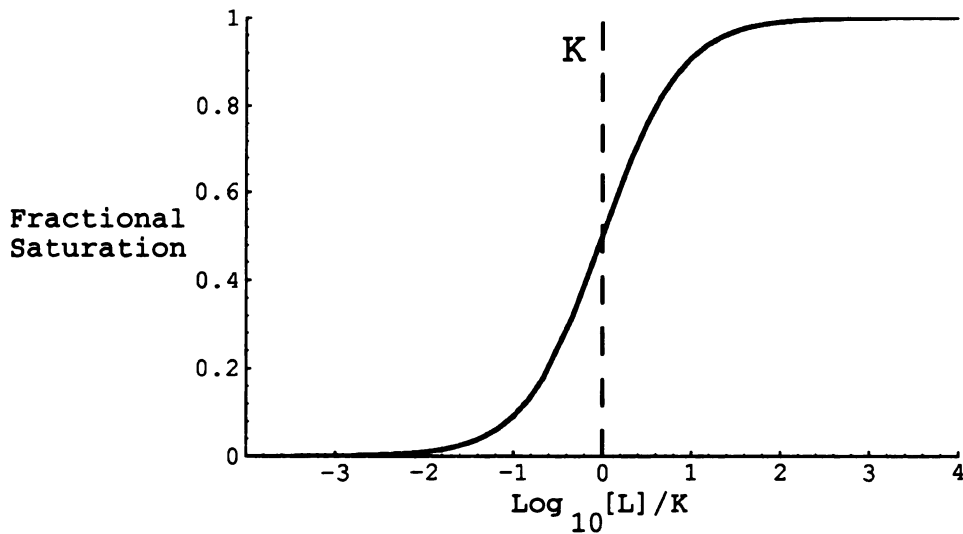
**Discussion:** From Equations (A.11) and (A.13), we know:

$$[TL] = \frac{[T]_T}{\frac{K}{[L]} + 1} \quad (A.50)$$

To make an affinity determination, all that is required is that we know  $[T]_T$  and that we have sufficient sampling of  $[TL]$  as a function of known  $[L]$ . As with Scatchard analysis, we can also just use the fraction of bound target as a function of known  $[L]$ .

$$f_T = \frac{1}{\frac{K}{[L]} + 1} \quad (A.51)$$

In order to accurately determine dissociation constants, it is important to carefully measure the *free* ligand concentration,  $[L]$  (not total ligand concentration  $[L]_T$ ), and to get data points that cover intermediate target saturation levels (10-90% satura-



**Figure A.6: Direct Fitting of Titration Curves.** At 50% saturation, the free concentration of ligand ( $[L]$ ) will equal the dissociation constant ( $K$ ).

tion). A true one-to-one titration curve should cover two orders of magnitude in  $[L]$  from the 9% to 91% saturation levels, and it should have a slope of 0.25 at  $[L] = K$  (or of 0.576 when graphed versus  $\log [C]_T$ ).

Note that under conditions where  $[T]_T \ll K$ , one can make the approximation that  $[L] = [L]_T$ , and  $K$  can be determined by dose-response.

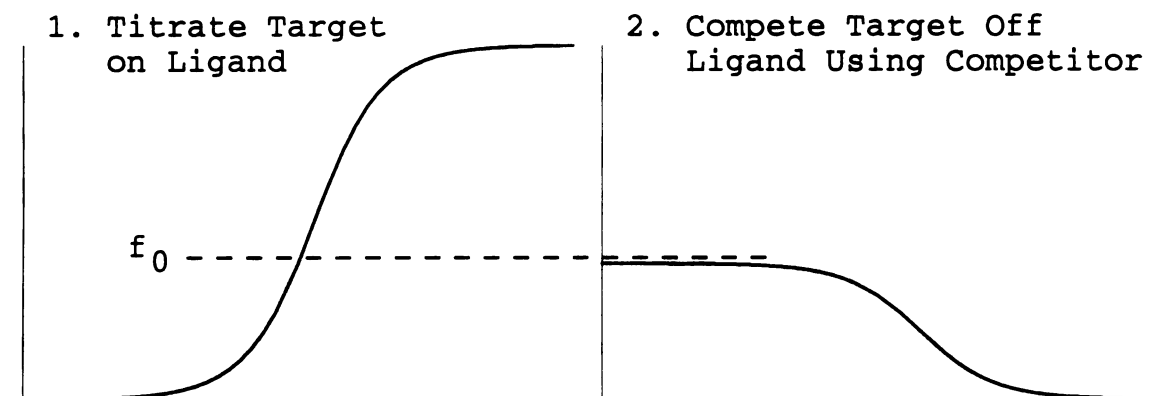
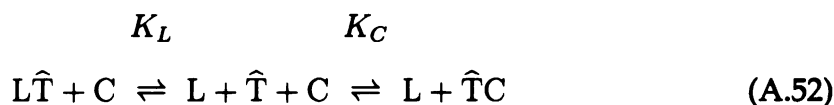


Figure A.7: **Assays with Labeled Target ( $EC_{50}$ 's).** In (1), labeled target ( $\hat{T}$ ) is titrated onto the ligand (L) to a saturation level  $f_0$ . In (2), competitor (C) is added to compete the target off of the ligand. The mid-point of this curve is the  $EC_{50}$ .

#### A.4 Competition Binding Assays: Three Component Systems

##### A.4.1 Assays with Labeled Target ( $EC_{50}$ 's)



**Summary:** The goal of this assay is to obtain an estimate of the affinity of the competitor C for the target  $\hat{T}$ . The target is labeled, as indicated by the ( $\hat{\cdot}$ ). A ligand L for  $\hat{T}$  is usually immobilized or captured in some manner so as to enable direct measurement of  $[\hat{T}L]$ .

A summary of the assay is shown in Figure A.7. First, in the absence of competitor C, the total ligand concentration  $[L]_T$  is fixed and the total target concentration  $[\hat{T}]_T$  is varied to achieve a low fractional saturation  $f_L^0$  of L with bound  $\hat{T}$ . Next, the total competitor concentration  $[C]_T$  is titrated to obtain a competition curve of the competitor C blocking binding of ligand L to the target  $\hat{T}$ . The total competitor concentration  $[C]_{T(50)}$  that results in 50% competition is referred to as the  $EC_{50}$ . It is related to  $K_C$  and can be used to estimate  $K_C$  under certain condi-

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full. The list includes the names of the members of the committee, the names of the members of the sub-committee, and the names of the members of the advisory committee. The addresses are given in full, including the street name, the city, the state, and the zip code.

2. The second part of the document is a list of the names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full. The list includes the names of the members of the committee, the names of the members of the sub-committee, and the names of the members of the advisory committee. The addresses are given in full, including the street name, the city, the state, and the zip code.

tions.

**Discussion:** The general relationship between the  $EC_{50}$  and the dissociation constant of the competitor  $K_C$  is derived here. Starting with the following two relationships:

$$[L]_T = [L] + [\hat{T}L] \quad (A.53)$$

$$K_L = \frac{[\hat{T}][L]}{[\hat{T}L]} \quad (A.54)$$

we can obtain the concentration of unbound  $\hat{T}$  as a function of the fractional saturation of L:

$$f_L^0 = \frac{[\hat{T}L]}{[L]_T} \quad (A.55)$$

and the dissociation constant  $K_L$ :

$$[\hat{T}] = \frac{K_L[\hat{T}L]}{[L]} \quad (A.56)$$

$$[\hat{T}] = \frac{K_L[\hat{T}L]}{[L]_T - [\hat{T}L]} \quad (A.57)$$

$$[\hat{T}] = \frac{K_L}{\frac{[L]_T}{[\hat{T}L]} - 1} \quad (A.58)$$

$$[\hat{T}] = \frac{K_L}{\frac{1}{f_L^0} - 1} \quad (A.59)$$

Following titration of  $\hat{T}$  binding to L at a saturation level of  $f_L^0$ , the concentration

1. The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that this is crucial for ensuring the integrity of the financial data and for facilitating the audit process.

2. The second part of the document outlines the specific procedures that should be followed when recording transactions. It details the steps from identifying the transaction to the final entry in the accounting system.

of free  $\hat{T}$  will be  $[\hat{T}]_{L(0)}$ :

$$[\hat{T}]_{(0)} = \frac{K_L}{\frac{1}{f_L^0} - 1} \quad (\text{A.60})$$

and the concentration of bound species can be computed directly from the fractional saturation (assuming that the total ligand concentration is known):

$$[\hat{TL}]_{(0)} = f_L^0 [L]_T \quad (\text{A.61})$$

These two values sum to the total concentration of target  $[\hat{T}]_T$  in the experiment:

$$[\hat{T}]_T = [\hat{T}]_{(0)} + [\hat{TL}]_{(0)} \quad (\text{A.62})$$

$$[\hat{T}]_T = \frac{K_L}{\frac{1}{f_L^0} - 1} + f_L^0 [L]_T \quad (\text{A.63})$$

The next step is to add competitor  $[C]_T$  until the amount of target bound to the ligand has been reduced by 50%. This is the  $EC_{50}$  point where the fractional saturation of L will have been halved to  $f_L^{50} = f_L^0/2$ :

$$[\hat{TL}]_{50} = f_L^{50} [L]_T = \frac{f_L^0 [L]_T}{2} \quad (\text{A.64})$$

and the concentration of free target  $[T]_{50}$  will be:

$$[\hat{T}]_{50} = \frac{K_L}{\frac{1}{f_L^{50}} - 1} = \frac{K_L}{\frac{1}{f_L^0/2} - 1} = \frac{K_L}{\frac{2}{f_L^0} - 1} \quad (\text{A.65})$$

At the  $EC_{50}$  point, the concentration of target-competitor complex can be calcu-





lated from the conservation equation as follows:

$$[\hat{T}]_T = [\hat{T}]_{50} + [\hat{T}C]_{50} + [\hat{T}L]_{50} \quad (\text{A.66})$$

$$[\hat{T}C]_{50} = [\hat{T}]_T - [\hat{T}]_{50} - [\hat{T}L]_{50} \quad (\text{A.67})$$

$$[\hat{T}C]_{50} = \frac{K_L}{\frac{1}{f_L^0} - 1} + f_L^0[L]_T - \frac{K_L}{\frac{2}{f_L^0} - 1} - \frac{f_L^0[L]_T}{2} \quad (\text{A.68})$$

$$[\hat{T}C]_{50} = \frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} + \frac{f_L^0[L]_T}{2} \quad (\text{A.69})$$

The only remaining parameter to be defined is the free competitor concentration at the EC<sub>50</sub> point [C]<sub>50</sub>, which can also be obtained using a conservation equation:

$$[C]_{T(50)} = [C]_{50} + [\hat{T}C]_{50} \quad (\text{A.70})$$

$$[C]_{50} = [C]_{T(50)} - [\hat{T}C]_{50} \quad (\text{A.71})$$

$$[C]_{50} = [C]_{T(50)} - \frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} - \frac{f_L^0[L]_T}{2} \quad (\text{A.72})$$

Finally, we are ready to relate all of these terms together using the equilibrium constant for the competitor  $K_C$ :

$$K_C = \frac{[\hat{T}]_{50}[C]_{50}}{[\hat{T}C]_{50}} \quad (\text{A.73})$$

$$K_C = \frac{\left[ \frac{K_L}{\frac{2}{f_L^0} - 1} \right] \left[ [C]_{T(50)} - \frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} - \frac{f_L^0[L]_T}{2} \right]}{\frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} + \frac{f_L^0[L]_T}{2}} \quad (\text{A.74})$$

$$K_C = \frac{f_L^0 K_L \left[ [C]_{T(50)} - \frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} - \frac{f_L^0[L]_T}{2} \right]}{(2 - f_L^0) \left[ \frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} + \frac{f_L^0[L]_T}{2} \right]} \quad (\text{A.75})$$

$$K_C = \frac{[C]_{T(50)} - \frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} - \frac{f_L^0[L]_T}{2}}{\frac{1}{1 - f_L^0} + \frac{(2 - f_L^0)[L]_T}{2K_L}} \quad (\text{A.76})$$

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The list includes the names of the members of the committee, the names of the members of the sub-committee, and the names of the members of the advisory committee. The addresses are listed in the same order as the names.

2. The second part of the document is a list of the names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The list includes the names of the members of the committee, the names of the members of the sub-committee, and the names of the members of the advisory committee. The addresses are listed in the same order as the names.

Equation (A.76) is the general form for the dissociation constant as a function of the EC<sub>50</sub> value [C]<sub>T(50)</sub>, the total ligand concentration [L]<sub>T</sub>, the affinity of the ligand K<sub>L</sub>, and the fractional saturation f<sub>L</sub><sup>0</sup> arrived at during the titration step.

This equation can be rearranged to give the EC<sub>50</sub> as a function of the dissociation constant K as follows:

$$K_C \left[ \frac{1}{1 - f_L^0} + \frac{(2 - f_L^0)[L]_T}{2K_L} \right] = [C]_{T(50)} - \frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} - \frac{f_L^0 [L]_T}{2} \quad (\text{A.77})$$

$$[C]_{T(50)} = K_C \left[ \frac{1}{1 - f_L^0} + \frac{(2 - f_L^0)[L]_T}{2K_L} \right] + \frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} + \frac{f_L^0 [L]_T}{2} \quad (\text{A.78})$$

Let us consider a few limiting cases in order to get a better feeling of how this equation behaves. First, if [L]<sub>T</sub> vanishes:

$$\lim_{[L]_T \rightarrow 0} [C]_{T(50)} = K_C \left[ \frac{1}{1 - f_L^0} \right] + \frac{f_L^0 K_L}{(2 - f_L^0)(1 - f_L^0)} \quad (\text{A.79})$$

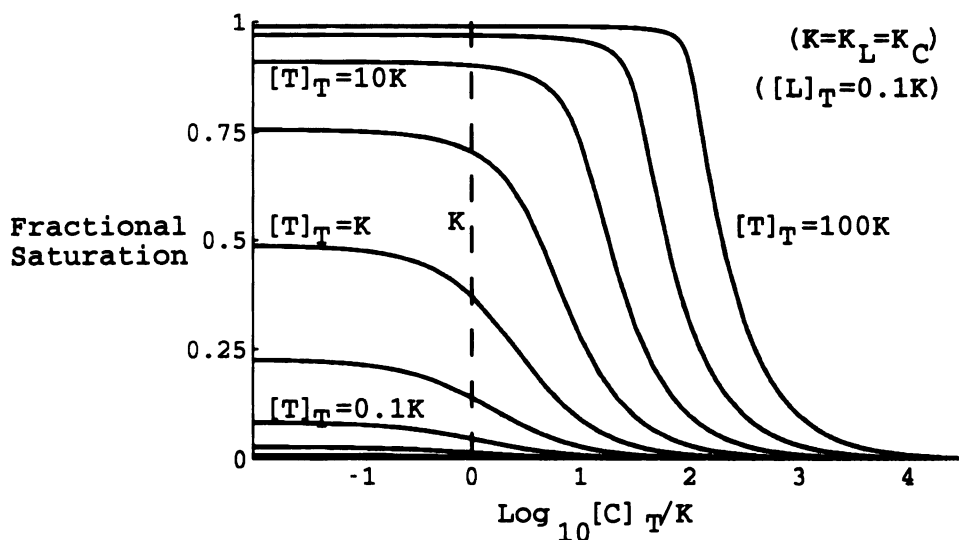
then the EC<sub>50</sub> will still have some dependence on the affinity of the ligand for the target, but this dependence will be minimized by low fractional saturation of the target.

If f<sub>L</sub><sup>0</sup> vanishes:

$$\lim_{f_L^0 \rightarrow 0} [C]_{T(50)} = K_C \left[ 1 + \frac{[L]_T}{K_L} \right] \quad (\text{A.80})$$

then the EC<sub>50</sub> will track linearly with K<sub>C</sub>, and the multiplier will be closer to one if the amount of immobilized ligand is below the affinity of the ligand for the target.

Also of particular interest is the special case where the ligand and the competi-



**Figure A.8: Effects of Target Concentration and Ligand Saturation.** If there is too much target ( $[\hat{T}]_T$ ), the initial fractional saturation ( $f_0$ ) will be too high, resulting in an  $EC_{50}$  ( $[C]_{T(50)}$ ) substantially above  $K_C$ .

tor are the same molecule, so  $K_C = K_L = K$ :

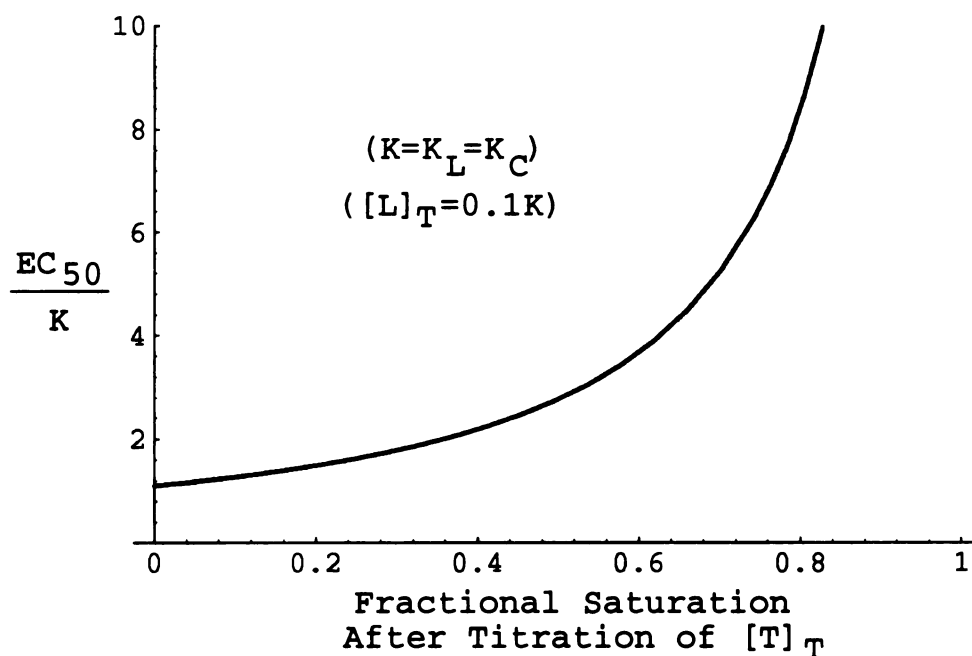
$$[C]_{T(50)} = \frac{K}{1 - f_L^0} + \frac{(2 - f_L^0)[L]_T}{2} + \frac{f_L^0 K}{(2 - f_L^0)(1 - f_L^0)} + \frac{f_L^0 [L]_T}{2} \quad (\text{A.81})$$

$$[C]_{T(50)} = \frac{2K}{(2 - f_L^0)(1 - f_L^0)} + [L]_T \quad (\text{A.82})$$

In this case, the  $EC_{50}$  will depend only on the fractional saturation and the total ligand concentration. Ratios of  $EC_{50}$ 's for a panel of variants obtained using this format over a uniform fractional saturation and a ligand concentration below  $K$  should be very close to the ratios of the corresponding dissociation constants.

One of the key parameters to control in an  $EC_{50}$  measurement is the degree of saturation ( $f_L^0$ ) of the target by the ligand during the initial titration. In general, the lower the  $f_L^0$ , the closer the  $EC_{50}$  will be to the dissociation constant  $K_C$ . Figure A.8 shows how the competition curves are shifted towards higher  $EC_{50}$  values when  $f_L^0$  is high. Also note how the competition curves get much steeper at high





**Figure A.9: Dependence of the  $EC_{50}$  on Ligand Saturation.** The  $EC_{50}$  increases asymptotically as the fractional saturation ( $f_0$ ) approaches unity.

$f_L^0$ . Figure A.9 shows how the ratio of the  $EC_{50}$  to  $K$  increases asymptotically as  $f_L^0$  approaches 1.

In cases where the ligand is a different molecule than the competitor, or if the ligand has reduced affinity for the target, the  $EC_{50}/K$  dependence on  $f_L^0$  can increase dramatically. This behavior is shown in Figure A.10. In general, only ligands with affinities greater than or equal to that of the competitor should be used in an  $EC_{50}$  assay.

Another parameter that can significantly alter the  $EC_{50}/K$  relationship is the amount of ligand immobilized. This amount must be kept below the dissociation constant of both the ligand and the competitor. Figure A.11 shows how the total ligand concentration ( $[L]_T$ ) can have a dramatic effect on the measured  $EC_{50}$  regardless of the saturation level. Ideally, the total ligand concentration should be 10-fold less than the smallest of the two dissociation constants.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that proper record-keeping is essential for the integrity of the financial system and for the ability to detect and prevent fraud.

2. The second part of the document outlines the specific procedures for recording transactions. It details the steps involved in the accounting cycle, from identifying the transaction to posting it to the appropriate ledger account.

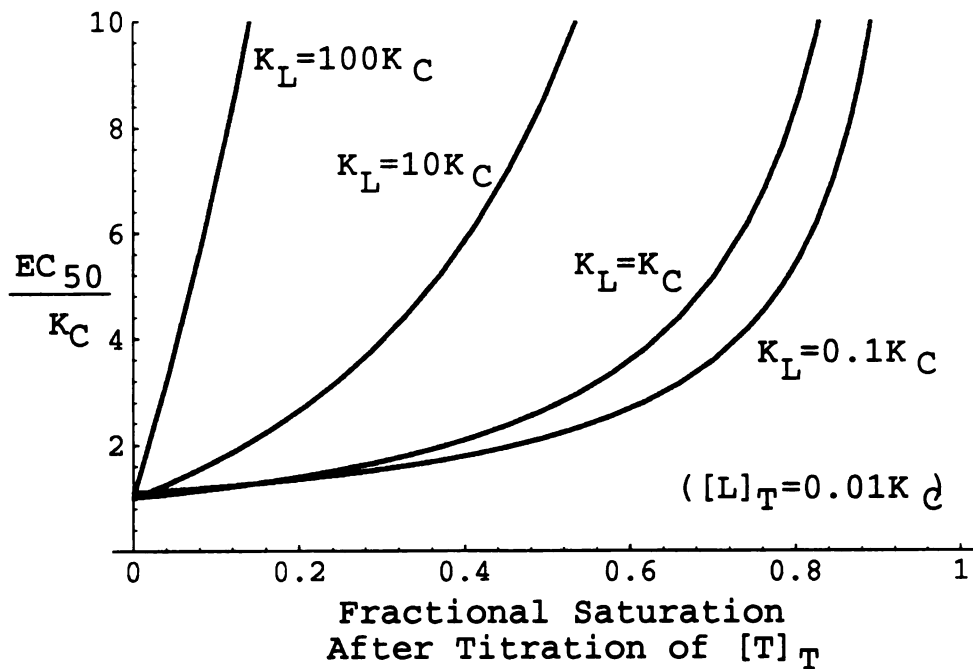


Figure A.10: Dependence of the  $EC_{50}$  on Ligand Affinity. The  $EC_{50}$  will show a stronger dependence on the fractional saturation ( $f_0$ ) if the ligand affinity ( $K_L$ ) is much weaker than that of the competitor ( $K_C$ ).

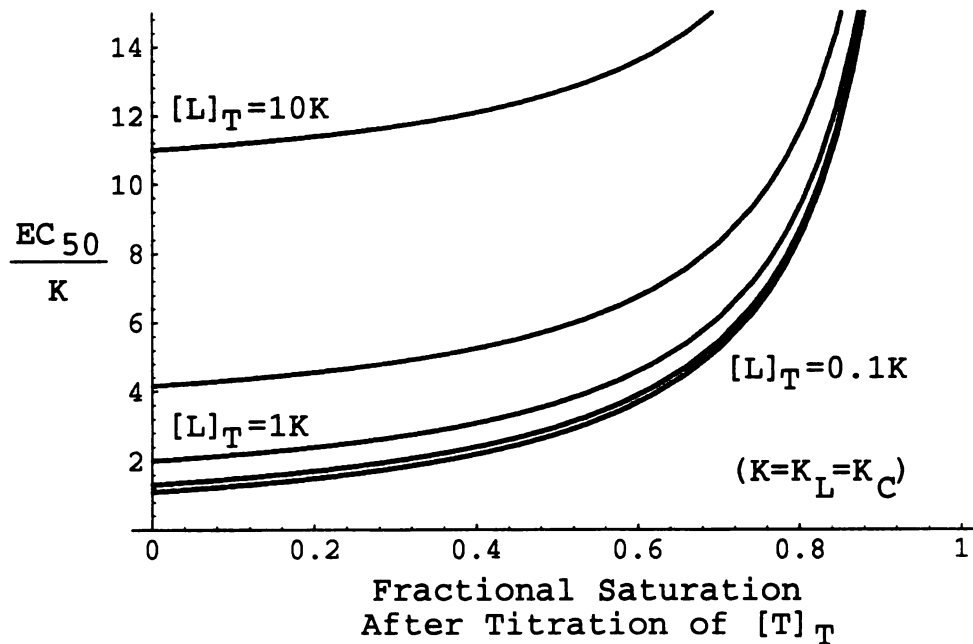


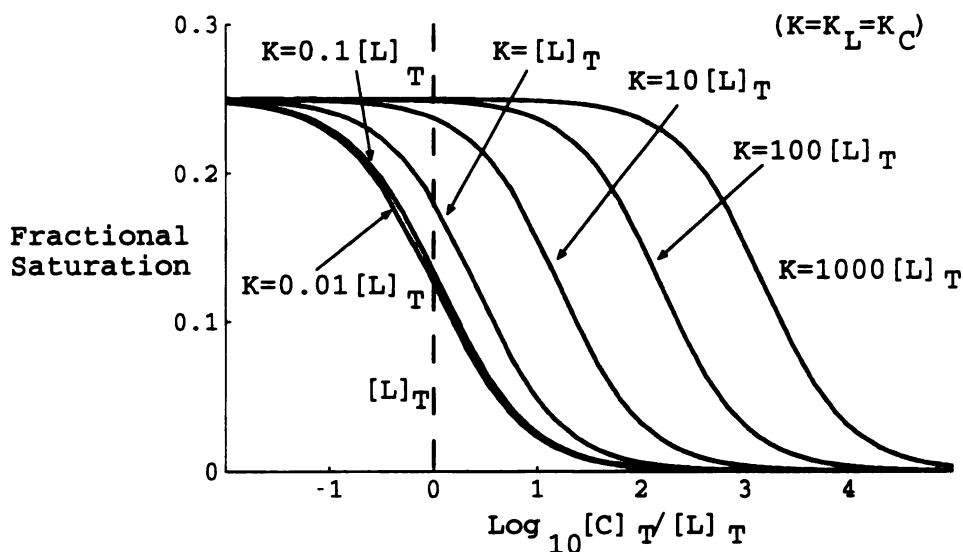
Figure A.11: Dependence of the  $EC_{50}$  on Ligand Concentration. The  $EC_{50}$  will be shifted upwards if the ligand ( $[L]_T$ ) is present at a concentration above  $K$ .



1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full, including the street name, number, and city.

2. The second part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of chairman.

3. The third part of the document is a list of the names and addresses of the members of the committee who have been elected to the office of secretary.



**Figure A.12: Bottoming Out of an  $EC_{50}$  Competition Assay.** No changes in the curve shape will occur if an  $EC_{50}$  competition assay bottoms out due to excess ligand concentration ( $[L]_T$ ) relative to  $K$ .

Unfortunately,  $EC_{50}$  assays can bottom out without showing any obvious signs of having done so. Figure A.12 illustrates the behavior by which curves remain essentially undistorted even though the true affinity is several orders of magnitude lower than the  $EC_{50}$ .

The lower limit on measurable affinities in  $EC_{50}$  assays is the concentration of ligand (Figure A.13).

Prerequisites for successful  $EC_{50}$  measurements are:

1.  $f_L^0$  should be as low as possible, but with good signal (25% is good).
2.  $f_L^0$  should be uniform across all  $EC_{50}$ 's to be compared (important!).
3.  $[L]_T < K_L$  ( $EC_{50}$ 's essentially scale as  $K_C$  multiplied by a factor of  $1 + [L]_T/K_L$ )
4.  $K_C \geq K_L$  (If  $K_C < K_L$  then the  $EC_{50}$  dependence on  $f_L^0$  gets very steep.)

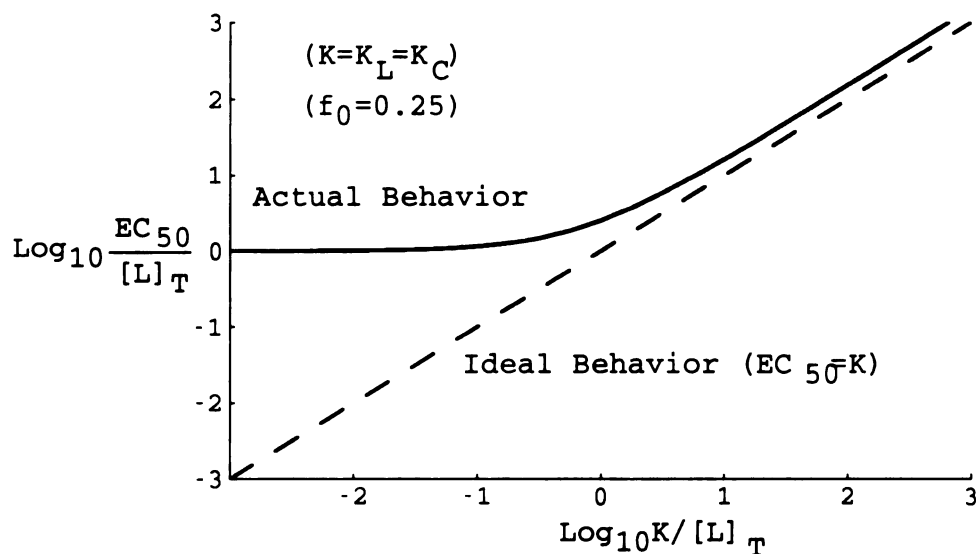
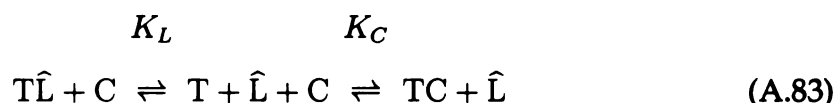


Figure A.13: **Limitations of EC<sub>50</sub> Competition Assays.** The total ligand concentration ( $[L]_T$ ) places an effective lower limit on the affinity which can be measured in an EC<sub>50</sub> competition assay.

#### A.4.2 Assays with Labeled Ligand (IC<sub>50</sub>'s)



**Summary:** The goal of this assay is to measure the ability of a competitor C to bind target T. The target T is usually immobilized or captured, and the ligand  $\widehat{L}$  is labeled so as to enable direct measurement of  $[\widehat{T}\widehat{L}]$ .

A summary of the process is shown in Figure A.14. First, in the absence of competitor C, the total target concentration  $[T]_T$  is fixed and the total ligand concentration  $[\widehat{L}]_T$  is varied to achieve a low fractional saturation  $f_L^0$  of  $\widehat{L}$  binding to T. Next, the total competitor concentration  $[C]_T$  is titrated to obtain a competition curve of the competitor C blocking binding of ligand  $\widehat{L}$  to the target T. The total competitor concentration  $[C]_{T(50)}$  that results in 50% competition is referred to as the IC<sub>50</sub>. It is related to  $K_C$  and can be used to estimate  $K_C$  under certain condi-

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The names are: Mr. J. H. Smith, Mr. J. B. Jones, Mr. W. C. Brown, Mr. T. A. White, Mr. R. M. Green, Mr. L. K. Black, Mr. D. E. Gray, Mr. F. G. White, Mr. H. I. Black, Mr. J. L. Gray, Mr. M. N. White, Mr. O. P. Black, Mr. Q. R. Gray, Mr. S. T. White, Mr. U. V. Black, Mr. W. X. Gray, Mr. Y. Z. White.

2. The second part of the document is a list of the names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The names are: Mr. J. H. Smith, Mr. J. B. Jones, Mr. W. C. Brown, Mr. T. A. White, Mr. R. M. Green, Mr. L. K. Black, Mr. D. E. Gray, Mr. F. G. White, Mr. H. I. Black, Mr. J. L. Gray, Mr. M. N. White, Mr. O. P. Black, Mr. Q. R. Gray, Mr. S. T. White, Mr. U. V. Black, Mr. W. X. Gray, Mr. Y. Z. White.

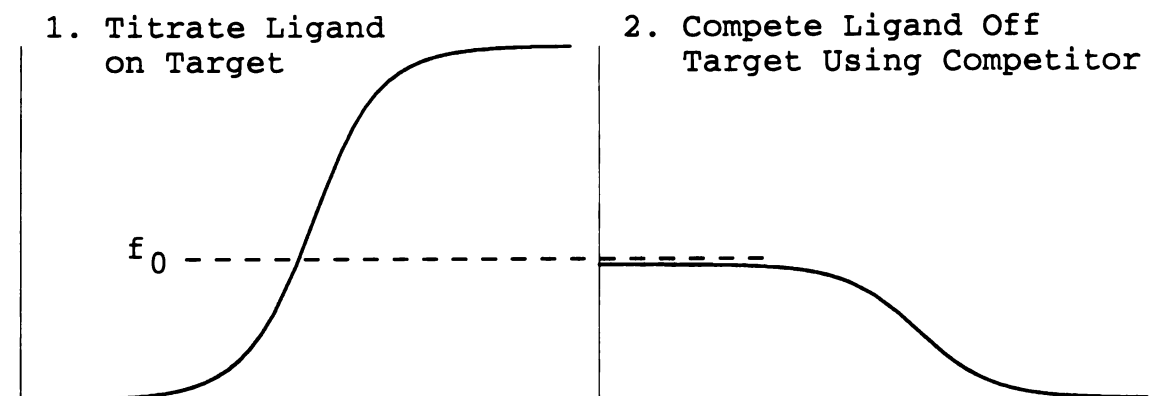


Figure A.14: Assays with Labeled Ligand ( $IC_{50}$ 's). In (1), labeled ligand ( $\hat{L}$ ) is titrated onto the target to a saturation level  $f_0$ . In (2), competitor ( $C$ ) is added to compete the ligand off of the target. The mid-point of this curve is the  $IC_{50}$ .

tions.

**Discussion:** The general relationship between the  $IC_{50}$  and the dissociation constant  $K_C$  is derived here. Starting with the following two relationships:

$$[\hat{L}]_T = [\hat{L}] + [T\hat{L}] \quad (\text{A.84})$$

$$K_L = \frac{[T][\hat{L}]}{[T\hat{L}]} \quad (\text{A.85})$$

we can obtain the concentration of unbound  $\hat{L}$  as a function of the fractional saturation of  $\hat{T}$

$$f_T^0 = \frac{[T\hat{L}]}{[T]_T} \quad (\text{A.86})$$

and the dissociation constant  $K_L$ :

$$[\hat{L}] = \frac{K_L[T\hat{L}]}{[T]} \quad (\text{A.87})$$

$$[\widehat{L}] = \frac{K_L[\widehat{TL}]}{[T]_T - [\widehat{TL}]} \quad (\text{A.88})$$

$$[\widehat{L}] = \frac{K_L}{\frac{[T]_T}{[\widehat{TL}]} - 1} \quad (\text{A.89})$$

$$[\widehat{L}] = \frac{K_L}{\frac{1}{f_T^0} - 1} \quad (\text{A.90})$$

Following titration of  $\widehat{L}$  binding to T at a saturation level of  $f_T^0$ , the concentration of free  $\widehat{L}$  will be  $[\widehat{L}]_0$ :

$$[\widehat{L}]_0 = \frac{K_L}{\frac{1}{f_T^0} - 1} \quad (\text{A.91})$$

and the concentration of bound species can be computed directly from the fractional saturation (assuming that the total target concentration is known):

$$[\widehat{TL}]_0 = f_T^0 [T]_T \quad (\text{A.92})$$

These two values sum to the total concentration of ligand  $[\widehat{L}]_T$  in the experiment:

$$[\widehat{L}]_T = [\widehat{L}]_0 + [\widehat{TL}]_0 \quad (\text{A.93})$$

$$[\widehat{L}]_T = \frac{K_L}{\frac{1}{f_T^0} - 1} + f_T^0 [T]_T \quad (\text{A.94})$$

The next step is to add competitor  $[C]_T$  until the amount of ligand bound to the target has been reduced by 50%. This is the  $IC_{50}$  point where the fractional saturation of T will have been halved to  $f_{50} = f_T^0/2$ :

$$[\widehat{TL}]_{50} = f_{50} [T]_T = \frac{f_T^0 [T]_T}{2} \quad (\text{A.95})$$

The concentration of free target  $[T]_{50}$  can be obtained using the  $[\widehat{L}]_T$  conservation

equation and the ligand dissociation constant:

$$[\widehat{L}]_{50} = [\widehat{L}]_T - [T\widehat{L}]_{50} \quad (\text{A.96})$$

$$[T]_{50} = \frac{[T\widehat{L}]_{50} K_L}{[\widehat{L}]_{50}} = \frac{[T\widehat{L}]_{50} K_L}{[\widehat{L}]_T - [T\widehat{L}]_{50}} \quad (\text{A.97})$$

$$[T]_{50} = \frac{\frac{f_T^0 [T]_T K_L}{2}}{\frac{K_L}{f_T^0 - 1} + f_T^0 [T]_T - \frac{f_T^0 [T]_T}{2}} \quad (\text{A.98})$$

$$[T]_{50} = \frac{f_T^0 [T]_T K_L}{2 \left[ \frac{f_T^0 K_L}{1 - f_T^0} + \frac{f_T^0 [T]_T}{2} \right]} \quad (\text{A.99})$$

$$[T]_{50} = \frac{[T]_T K_L}{\frac{2K_L}{1 - f_T^0} + [T]_T} \quad (\text{A.100})$$

$$[T]_{50} = \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1} \quad (\text{A.101})$$

At the IC<sub>50</sub> point, the concentration of target-competitor complex can be calculated from the [T]<sub>T</sub> conservation equation as follows:

$$[T]_T = [T]_{50} + [TC]_{50} + [T\widehat{L}]_{50} \quad (\text{A.102})$$

$$[TC]_{50} = [T]_T - [T]_{50} - [T\widehat{L}]_{50} \quad (\text{A.103})$$

$$[TC]_{50} = [T]_T - \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1} - \frac{f_T^0 [T]_T}{2} \quad (\text{A.104})$$

$$[TC]_{50} = \frac{(2 - f_0)[T]_T}{2} - \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1} \quad (\text{A.105})$$

The only remaining parameter to be defined is the free competitor concentration at the IC<sub>50</sub> point [C]<sub>50</sub>, which can also be obtained using a conservation equation:

$$[C]_{T(50)} = [C]_{50} + [TC]_{50} \quad (\text{A.106})$$

$$[C]_{50} = [C]_{T(50)} - [TC]_{50} \quad (\text{A.107})$$

$$[C]_{50} = [C]_{T(50)} - \frac{(2 - f_T^0)[T]_T}{2} + \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1} \quad (\text{A.108})$$

Finally, we are ready to relate all of these terms together using the equilibrium constant for the competitor  $K_C$ :

$$K_C = \frac{[T]_{50}[C]_{50}}{[TC]_{50}} \quad (\text{A.109})$$

$$K_C = \frac{\left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1} \left[ [C]_{T(50)} - \frac{(2 - f_T^0)[T]_T}{2} + \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1} \right]}{\frac{(2 - f_T^0)[T]_T}{2} - \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1}} \quad (\text{A.110})$$

$$K_C = \frac{[C]_{T(50)} - \frac{(2 - f_T^0)[T]_T}{2} + \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1}}{\left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right) \left( \frac{(2 - f_T^0)[T]_T}{2} \right) - 1} \quad (\text{A.111})$$

$$K_C = \frac{[C]_{T(50)} - \frac{(2 - f_T^0)[T]_T}{2} + \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1}}{\frac{2 - f_T^0}{1 - f_T^0} + \frac{(2 - f_T^0)[T]_T}{2K_L} - 1} \quad (\text{A.112})$$

$$K_C = \frac{[C]_{T(50)} - \frac{(2 - f_T^0)[T]_T}{2} + \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1}}{\frac{1}{1 - f_T^0} + \frac{(2 - f_T^0)[T]_T}{2K_L}} \quad (\text{A.113})$$

Equation (A.113) is the general form for the dissociation constant as a function of the  $IC_{50}$  value  $[C]_{T(50)}$ , the total target concentration  $[T]_T$ , the affinity of the ligand  $K_L$ , and the fractional saturation  $f_T^0$  arrived at during the titration step.

This equation can be rearranged to give the  $IC_{50}$  ( $[C]_{T(50)}$ ) as a function of the other parameters as follows:

$$K_C \left[ \frac{1}{1 - f_T^0} + \frac{(2 - f_T^0)[T]_T}{2K_L} \right] = [C]_{T(50)} - \frac{(2 - f_T^0)[T]_T}{2} + \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1} \quad (\text{A.114})$$



$$[C]_{T(50)} = K_C \left[ \frac{1}{1 - f_T^0} + \frac{(2 - f_T^0)[T]_T}{2K_L} \right] + \frac{(2 - f_T^0)[T]_T}{2} - \left( \frac{2}{(1 - f_T^0)[T]_T} + \frac{1}{K_L} \right)^{-1} \quad (\text{A.115})$$

Let us consider a few limiting cases in order to get a better feeling of how this equation behaves. First, if  $[T]_T$  vanishes:

$$\lim_{[T]_T \rightarrow 0} [C]_{T(50)} = K_C \left[ \frac{1}{1 - f_T^0} \right] \quad (\text{A.116})$$

then the  $IC_{50}$  will depend only on the affinity of the competitor and the fractional saturation of the target by the ligand.

If  $f_T^0$  vanishes:

$$\lim_{f_T^0 \rightarrow 0} [C]_{T(50)} = K_C \left[ 1 + \frac{[T]_T}{K_L} \right] + [T]_T - \frac{K_L [T]_T}{2K_L + [T]_T} \quad (\text{A.117})$$

$$\lim_{f_T^0 \rightarrow 0} [C]_{T(50)} = K_C \left[ 1 + \frac{[T]_T}{K_L} \right] + \frac{[T]_T^2 + K_L [T]_T}{2K_L + [T]_T} \quad (\text{A.118})$$

then the  $IC_{50}$  will equal a constant term plus a multiple of  $K_C$ .

As with  $EC_{50}$ 's,  $IC_{50}$ 's exhibit strong dependence on the initial level of titration  $f_T^0$ , in this case of the ligand binding to the target. However, unlike with  $EC_{50}$ 's,  $IC_{50}$  curves obtained at high  $f_T^0$  look fairly normal, as shown in Figure A.15.

The actual dependence of the  $IC_{50}$  to  $K_C$  ratio on  $f_T^0$  is virtually identical that of  $EC_{50}$ 's (Figure A.16).

However,  $IC_{50}/K_C$  does not have any strong dependence on the affinity of the ligand  $K_L$  in the same way that  $EC_{50}/K_C$  is dramatically altered by a ligand of lower affinity (Figure A.17).

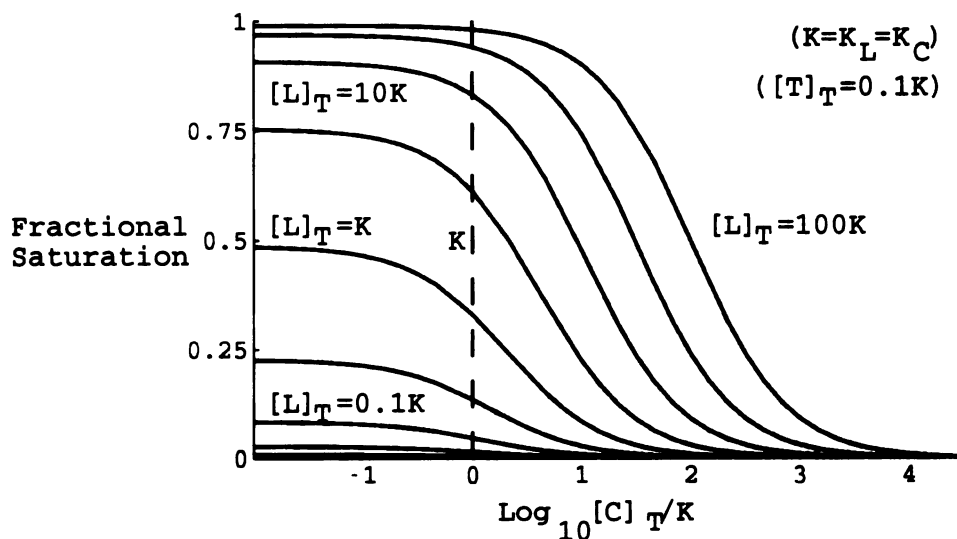


Figure A.15: **Effects of Ligand Concentration and Target Saturation.** If there is too much ligand ( $[\hat{L}]_T$ ), the initial fractional saturation ( $f_T^0$ ) will be too high, resulting in an  $IC_{50}$  ( $[C]_{T(50)}$ ) substantially above  $K$ .

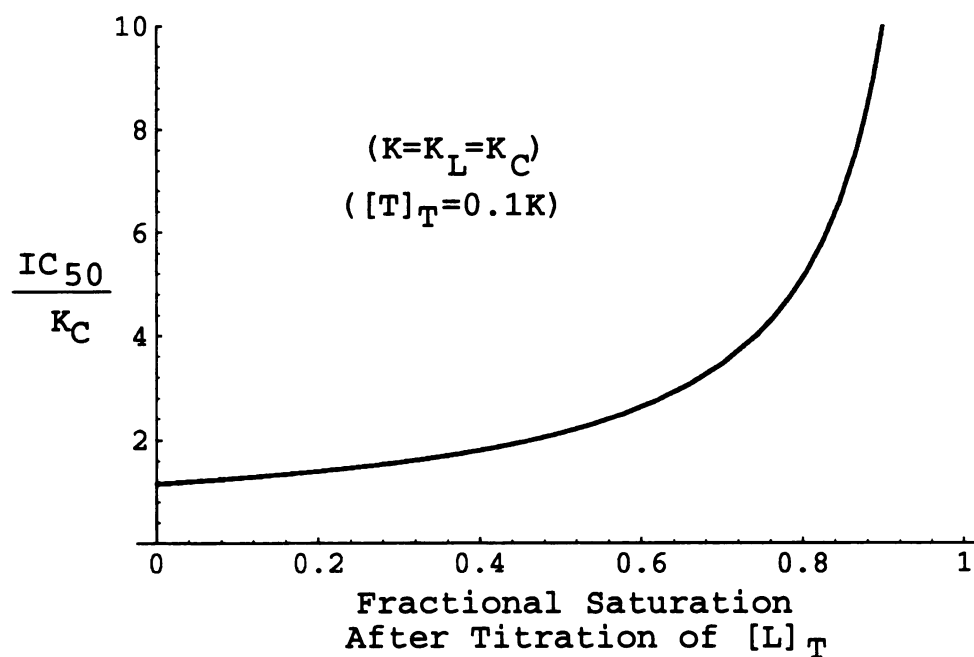


Figure A.16: **Dependence of the  $IC_{50}$  on the Fractional Target Saturation.** The  $IC_{50}$  increases asymptotically as the fractional saturation ( $f_0$ ) approaches unity.

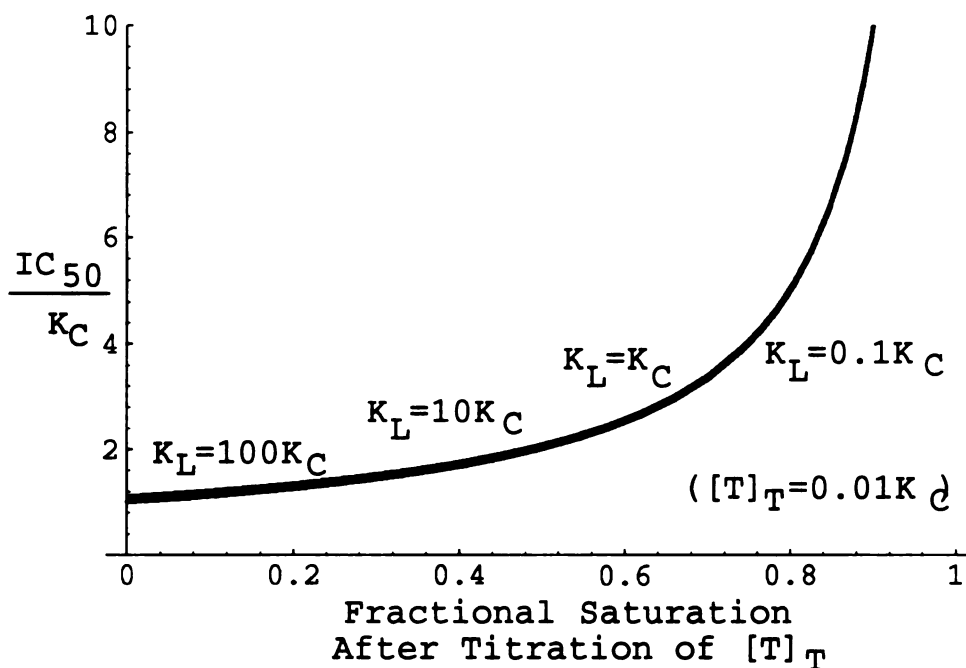


Figure A.17: **Dependence of the  $IC_{50}$  on Ligand Affinity.** Unlike  $EC_{50}$  binding assays,  $IC_{50}$ 's are not affected by the relative affinities of the ligand ( $K_L$ ) and competitors ( $K_C$ ) for the target.

In contrast,  $IC_{50}/K$  is modulated by increased target concentration in a manner resembling how  $EC_{50}/K$  is affected by an increase in total ligand concentration. In order to get reliable  $IC_{50}$  values, the target concentration should be 10-fold lower than the smaller of the two dissociation constants (Figure A.18).

Unlike with  $EC_{50}$  measurements, curves obtained from  $IC_{50}$  measurements will show symptomatic distortion when the bottom of the assay's usable range is reached. As shown in the next figure, the curves will become steeper and asymmetric (Figure A.19).

Ultimately, the total target concentration sets the lower limit of the assay's usable range (Figure A.20).

Prerequisites for successful  $IC_{50}$  measurements are:

1.  $f_T^0$  should be as low as possible, but with good signal (25% is good).

1. The first part of the document is a list of names and addresses of the members of the committee.

2. The second part of the document is a list of the names and addresses of the members of the committee.

3. The third part of the document is a list of the names and addresses of the members of the committee.

4. The fourth part of the document is a list of the names and addresses of the members of the committee.

5. The fifth part of the document is a list of the names and addresses of the members of the committee.

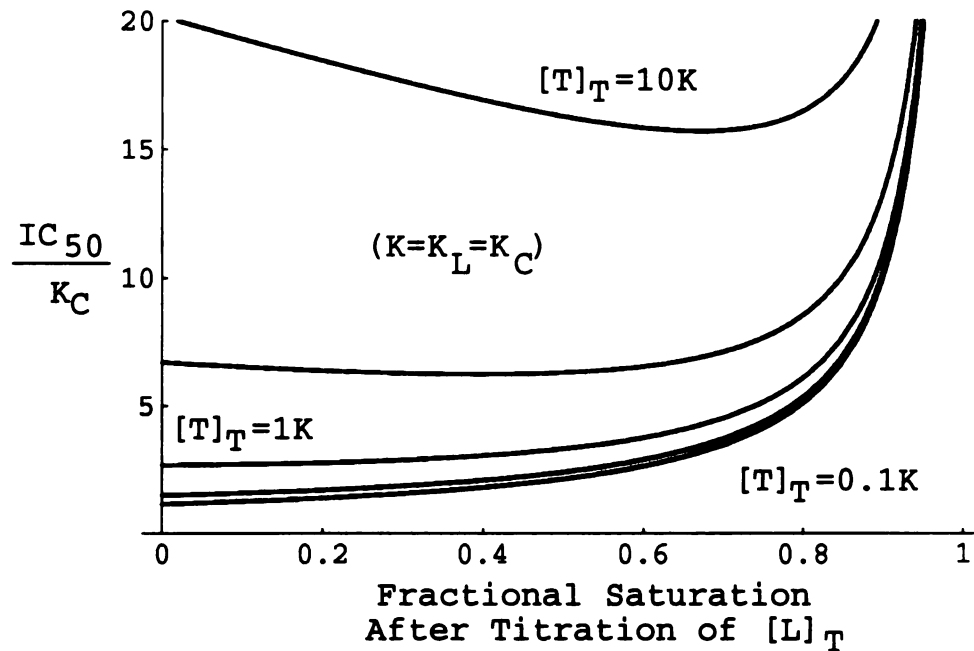


Figure A.18: **Dependence of the  $IC_{50}$  on the Total Target Concentration.** The  $IC_{50}$  will be shifted upwards if the target ( $[T]_T$ ) is present at a concentration above  $K$ .

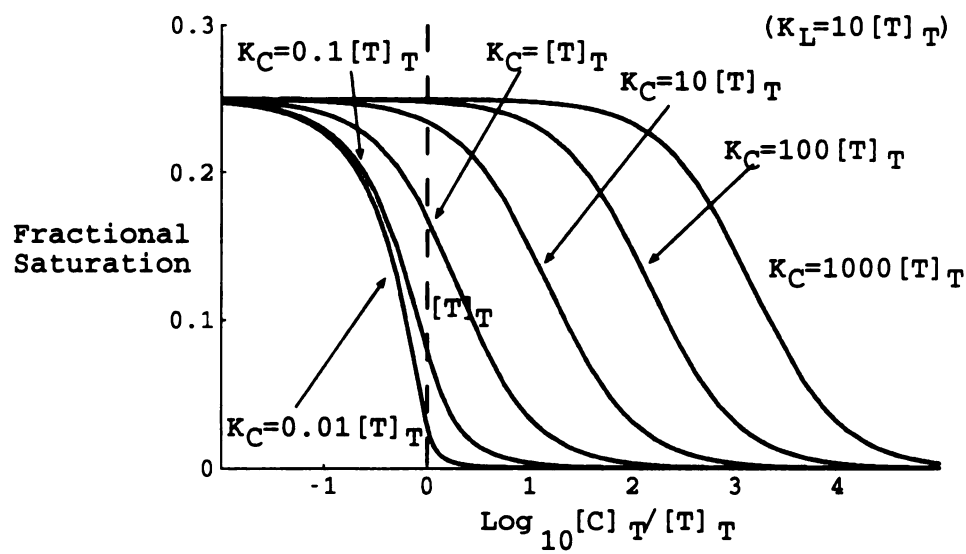
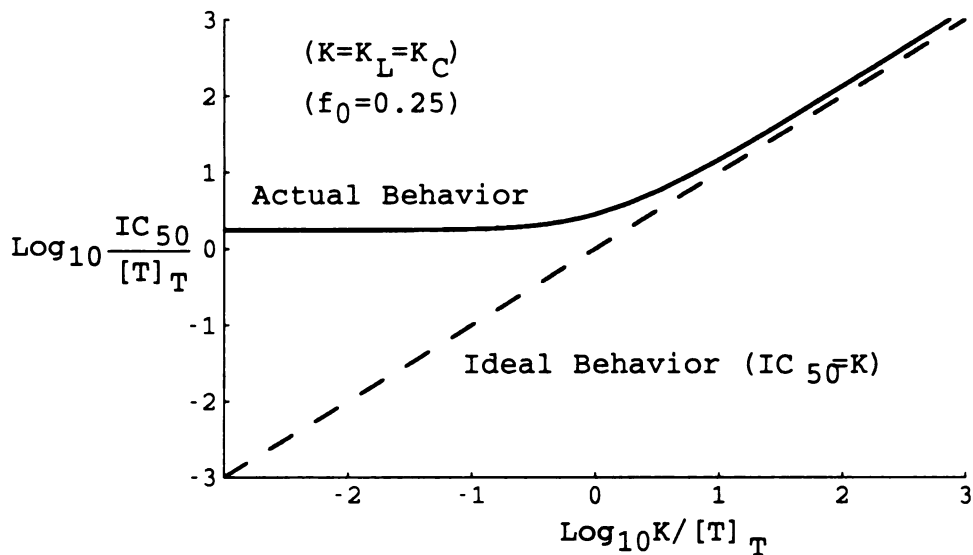


Figure A.19: **Bottoming Out of an  $IC_{50}$  Competition Assay.** When the useful lower range of an  $IC_{50}$  binding assay is reached because the target concentration ( $[T]_T$ ) exceeds  $K$ , the curves will become excessively steep and asymmetric.



**Figure A.20: Limitations of  $IC_{50}$  Competition Assays.** The total target concentration places and effective lower limit on the inhibition which can be measured in an  $IC_{50}$  competition assay.

2.  $f_T^0$  should be uniform across all  $IC_{50}$ 's to be compared (important!).
3.  $[T]_T < K_L$  ( $IC_{50}$ 's essentially scale as  $K_C$  multiplied by a factor of  $1 + [T]_T / K_L$ )

### **$IC_{50}$ Competition Assays Performed Under Near Saturating Conditions**

Sometimes it is necessary or convenient to perform competitive binding assays under highly saturating conditions. In this case, the ratio of the  $IC_{50}$  to the total ligand concentration can be used to estimate the ratio of the dissociation constants  $K_C / K_L$ .

From Figure A.15, we know that the  $IC_{50}$  binding assay will exhibit a normal inhibition curve even when competing against very high concentrations of ligand (note that this is *not* true for the  $EC_{50}$  format binding assays). Reproducible  $IC_{50}$  values can therefore be extracted from those curves, even though the  $IC_{50}$ 's will be far from the true  $K_C$ 's.

The question to be answered is: under what conditions does the ratio of the  $IC_{50}$  to the ligand concentration approach the ratio of the dissociation constants of the competitor over the ligand?

$$\text{when does } \frac{IC_{50}}{[\widehat{L}]_T} \approx \frac{K_C}{K_L} ? \quad (\text{A.119})$$

This ratio can be derived by combining equations (A.94) and (A.115). For the sake of brevity, the derivation is omitted. The resulting expression is:

$$\frac{IC_{50}}{[\widehat{L}]_T} = \frac{(2 - f_T^0)(K_C + K_L)}{2f_T^0 K_L} + \frac{2(1 - f_T^0)K_L}{f_T^0(f_T^0[T]_T - 2K_L - [T]_T)} + \frac{K_L - K_C}{2(f_T^0[T]_T - K_L - [T]_T)} \quad (\text{A.120})$$

Again, we can use limiting cases to understand the behavior of this rather complex equation. As the total target concentration approaches zero:

$$\lim_{[T]_T \rightarrow 0} \frac{IC_{50}}{[\widehat{L}]_T} = \frac{(2 - f_T^0)(K_C + K_L)}{2f_T^0 K_L} - \frac{2(1 - f_T^0)K_L}{2f_T^0 K_L} - \frac{K_L - K_C}{2K_L} \quad (\text{A.121})$$

$$\lim_{[T]_T \rightarrow 0} \frac{IC_{50}}{[\widehat{L}]_T} = \frac{2K_C + 2K_L - f_T^0 K_C - f_T^0 K_L - 2K_L + 2f_T^0 K_L - f_T^0 K_L + f_T^0 K_C}{2f_T^0 K_L} \quad (\text{A.122})$$

$$\lim_{[T]_T \rightarrow 0} \frac{IC_{50}}{[\widehat{L}]_T} = \frac{K_C}{f_T^0 K_L} \quad (\text{A.123})$$

then the  $IC_{50}/[\widehat{L}]_T$  ratio will equal the ratio of the dissociation constants divided by the fractional saturation of the target.

When the initial fractional saturation approaches unity:

$$\lim_{f_T^0 \rightarrow 1} \frac{IC_{50}}{[\widehat{L}]_T} = \frac{K_C + K_L}{2K_L} + 0 - \frac{K_L - K_C}{2K_L} \quad (\text{A.124})$$

$$\lim_{f_T^0 \rightarrow 1} \frac{IC_{50}}{[\widehat{L}]_T} = \frac{K_C}{K_L} \quad (\text{A.125})$$

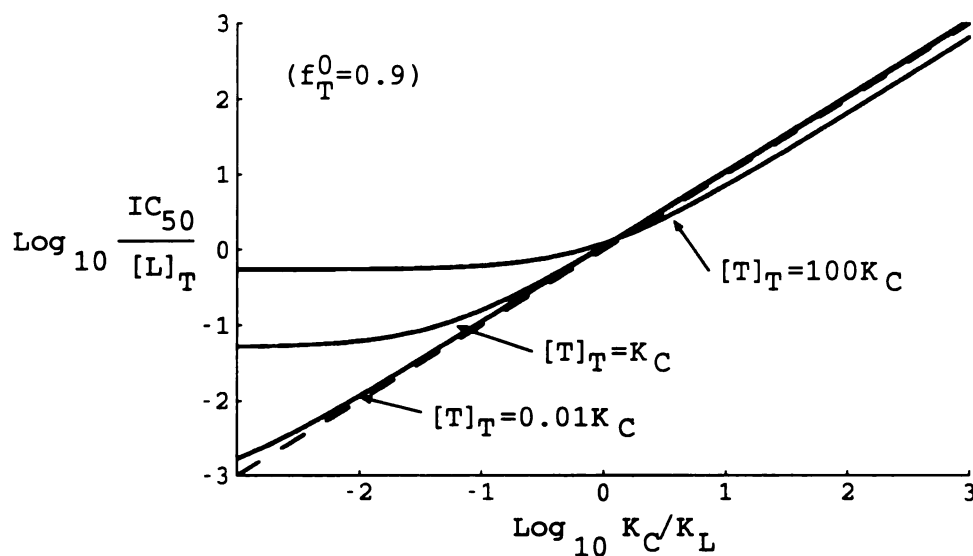


Figure A.21: **Measurements Under Saturating Conditions: Target Concentration.** High target concentrations have a detrimental effect on the observed  $IC_{50}/[\hat{L}]_T$  ratio.

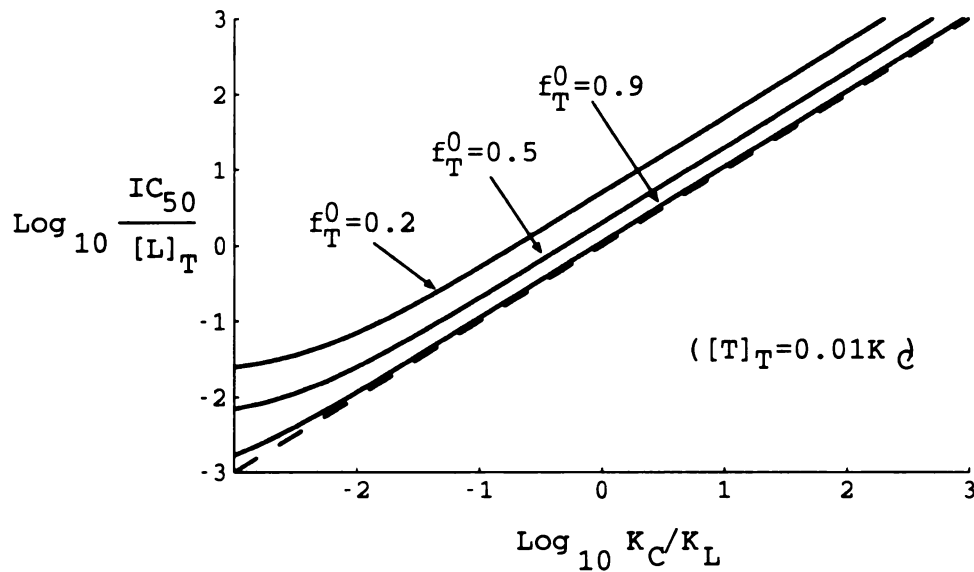
then the  $IC_{50}/[\hat{L}]_T$  ratio will equal the ratio of the dissociation constants. Of course, unrealistically high concentrations of both ligand and competitor would be required to achieve this state. One must not forget that the asymptotic curve in Figure A.16 still applies, and that  $IC_{50}$  and  $[L]_T$  will both approach infinity as the initial fractional saturation approaches one. The major limitation of this kind of assay is the upper limit on the achievable concentrations of C and L in solution.

The behaviors of the  $IC_{50}/[\hat{L}]_T$  ratio, compared to  $K_C/K_L$  can also be described graphically. Figure A.21 illustrates how the observed ratio is affected by high target concentrations, and Figure A.22 shows how the observed ratio is affected by different saturation levels.

Tips for performing assays under near-saturating conditions:

1. The total target concentration ( $[T]_T$ ) should be below the dissociation constant of the labeled ligand ( $K_L$ ) by at least 10-100 fold, especially if the competitor binds tighter to the target than does the ligand.





**Figure A.22: Measurements Under Saturating Conditions: Target Saturation.** In this assay format, high fractional saturations will give  $IC_{50}/[\hat{L}]_T$  ratios closer to the true ratio of the affinities.

2.  $f_T^0$  should be high, but not so high that the expected  $IC_{50}$  will exceed what can be achieved in the system. Lower saturation levels will give a lower absolute  $IC_{50}$  but the  $IC_{50}/[\hat{L}]_T$  ratio will suggest that the competitor binds with a weaker affinity than it actually does.

### A.4.3 Fitting Curves from Competition Assays

Because  $EC_{50}$ 's and  $IC_{50}$ 's depend on several additional parameters beyond just the inhibition constant of the competitor ( $K_C$ ), there is no simple way to rigorously formulate the fractional inhibition as a function of the  $EC_{50}$  or  $IC_{50}$ . However, when competitions are performed under appropriate conditions (described in sections A.4.1 and A.4.2), they will give symmetric curves that can be fit to a standard sigmoidal function:

$$y = \frac{1}{\left(\frac{a}{x}\right)^b + 1} \quad (\text{A.126})$$

where  $x$  is the independent variable,  $a$  is the mid point of the curve, and  $b$  is the slope constant. Note that a standard one-to-one binding curve (equation A.13) shares this form, but in this case  $b$  is unity:

$$f_B = \frac{1}{\frac{K}{[A]} + 1} \quad (\text{A.127})$$

Unlike binding curves,  $EC_{50}$  and  $IC_{50}$  competitions will sometimes have midpoint slopes that may vary depending on the conditions of the assay. It is therefore appropriate to include  $b$  as an adjustable parameter in the fitting process.

$$\begin{aligned} f_L &= \frac{1}{\left(\frac{EC_{50}}{[C]_T}\right)^b + 1} \\ f_T &= \frac{1}{\left(\frac{IC_{50}}{[C]_T}\right)^b + 1} \end{aligned} \quad (\text{A.128})$$

Some curve fitting programs work better with these equations reformatted using logs:

$$f_L = \frac{1}{10^{b(\log EC_{50} - \log [C]_T)} + 1} \quad (\text{A.129})$$

$$f_T = \frac{1}{10^{b(\log IC_{50} - \log [C]_T)} + 1}$$

or even better:

$$f_L = \frac{1}{10^{b(a-x)} + 1} \quad (\text{A.130})$$

where  $x$  is the log of the total competitor concentrations ( $\log [C]_T$ ), and  $a$  is a fitting parameter used to compute the  $EC_{50}$  or  $IC_{50}$  ( $= 10^a$ ) after fitting.

For analyzing real data, the two-parameter fits can be expanded into four-parameter fits which take into account minimum and maximum inhibition levels.

General form:

$$y = \frac{c - d}{\left(\frac{a}{x}\right)^b + 1} + d \quad (\text{A.131})$$

in practice:

$$\begin{aligned} \text{inhibition of } [\hat{T}L] &= \frac{\max - \min}{\left(\frac{EC_{50}}{[C]_T}\right)^b + 1} + \min & (\text{A.132}) \\ \text{inhibition of } [T\hat{L}] &= \frac{\max - \min}{\left(\frac{IC_{50}}{[C]_T}\right)^b + 1} + \min \end{aligned}$$

or using logs:

$$\begin{aligned} \text{inhibition of } [\hat{T}L] &= \frac{\max - \min}{10^{b(\log EC_{50} - \log [C]_T)} + 1} + \min & (\text{A.133}) \\ \text{inhibition of } [T\hat{L}] &= \frac{\max - \min}{10^{b(\log IC_{50} - \log [C]_T)} + 1} + \min \end{aligned}$$

or parameterized as in (A.130) (most stable):

$$\text{inhibition of } [\hat{T}L] = \frac{\max - \min}{10^{b(a-x)} + 1} + \min \quad (\text{A.134})$$

If all of your data is background subtracted, you can carry out a three-parameter fit by setting *min* to zero in the above equations.

In general, sigmoidal fits will give a good approximation of the mid-point value even in cases where the curve exhibits some asymmetry. However, as discussed in sections A.4.1 and A.4.2, asymmetric or steep curves will sometimes (but not always) indicate problems with an assay. Excessively steep competition curves will have slope parameters (*b*) greater than 1.0.

## Curve Fitting Using KalidaGraph

When using KaliedaGraph (Synergy Software), I recommend that you work with concentrations in logarithms of molarities (see next page). That way the fits will converge more often and the plots will be easier to read. However, if you prefer to work in molarities, the equations are (from A.132 and A.133):

$$y = ((m_4 - m_3)/((m_1/m_0)^{m_2 + 1}) + m_3 \quad (\text{A.135})$$

$$y = ((m_4 - m_3)/(10^{(m_2 * (\log(m_1) - \log(m_0)))}) + 1)) + m_3$$

$y$  = Assay Signal (Dependent Variable)

$m_0$  = Total Competitor Concentrations (Independent Variable)

$m_1$  = IC<sub>50</sub> or EC<sub>50</sub>

$m_2$  = Slope Parameter (should fit to 1.0)

$m_3$  = Signal at Zero Competitor

$m_4$  = Signal at Infinite Competitor

For a typical competition assay with a signal range of 0.1 to 0.9 and a micromolar IC<sub>50</sub> or EC<sub>50</sub>, you would plug the following equation and initial values into KaliedaGraph's fitting routine:

$$( (m_4 - m_3) / ( (m_1/m_0)^{m_2 + 1} ) + m_3 ; m_1 = 0.000001 ; m_2 = 1.0 ; m_3 = 0.9 ; m_4 = 0.1$$

Inverse forms of (A.135) include:

$$m_0 = m_1 \left[ \frac{m_4 - m_3}{y - m_3} - 1 \right]^{-1/m_2}$$
$$m_1 = m_0 \left[ \frac{m_4 - m_3}{y - m_3} - 1 \right]^{1/m_2}$$

If you use logarithms of molarities, Kaliedagraph's curve fitting routine will exhibit robust convergence and give many fewer "Singular Coefficient Matrix" errors. You will need to convert your concentrations into logarithms before graphing. From (A.134):

$$y = ((m_4 - m_3)/(10^{(m_2 * (m_1 - m_0)) + 1})) + m_3 \quad (\text{A.136})$$

$y$  = Assay Signal (Dependent Variable)

$m_0$  = log Total Competitor Concentrations (Independent Variable)

$m_1$  = log IC<sub>50</sub> or EC<sub>50</sub>

$m_2$  = Slope Parameter (should fit to 1.0)

$m_3$  = Signal at Zero Competitor

$m_4$  = Signal at Infinite Competitor

For a typical competition assay with a signal range of 0.1 to 0.9 and a micromolar IC<sub>50</sub> or EC<sub>50</sub>, you would plug the following equation and initial values into KaliedaGraph's fitting routine:

$$((m_4 - m_3) / (10^{(m_2 * (m_1 - m_0)) + 1})) + m_3 ; m_1 = -6 ; m_2 = 1.0 ; m_3 = 0.9 ; m_4 = 0.1$$

After fitting, you would compute the IC<sub>50</sub> or EC<sub>50</sub> as:

$$\text{IC}_{50} \text{ or } \text{EC}_{50} = 10^{m_1} \quad (\text{A.137})$$

Inverse forms of (A.136) include:

$$m_0 = m_1 + \log \left[ \frac{m_4 - m_3}{y - m_3} - 1 \right]^{-1/m_2}$$

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are given in full, including the street name, city, and state.

2. The second part of the document is a list of the names and addresses of the members of the committee who have been appointed to the various subcommittees. The names are listed in alphabetical order, and the addresses are given in full, including the street name, city, and state.

#### A.4.4 Estimating EC<sub>50</sub>'s and IC<sub>50</sub>'s from Single Point Inhibition Measurements

NOTE: This is usually a bad idea and should only be done when real competition curves can not be obtained. Understand that the values obtained will be *very* approximate. One typical application of this would be in comparing single point screening data from two screens performed at different concentrations.

Solving for  $a$  in (A.126) gives:

$$\begin{aligned}a &= x \left( \frac{1}{y} - 1 \right)^{1/b} \\ \text{EC}_{50} &= [\text{C}]_{\text{T}} \left( \frac{1}{f_L} - 1 \right)^{1/b} \\ \text{IC}_{50} &= [\text{C}]_{\text{T}} \left( \frac{1}{f_T} - 1 \right)^{1/b}\end{aligned}$$

where  $f$ 's are fractional inhibition, and  $[\text{C}]_{\text{T}}$  is the competitor concentration.

If  $b$  is unknown, it can be assumed to be unity as in the ideal case, which gives:

$$\begin{aligned}a &= x \left( \frac{1}{y} - 1 \right) \\ \text{EC}_{50} &= [\text{C}]_{\text{T}} \left( \frac{1}{f_L} - 1 \right) \\ \text{IC}_{50} &= [\text{C}]_{\text{T}} \left( \frac{1}{f_T} - 1 \right)\end{aligned}$$

These relationships will really only give useful results in the range ( $0.1 \leq f \leq 0.9$ ). EC<sub>50</sub>'s and IC<sub>50</sub> obtained from  $f$ 's outside this range really just represent huge magnifications of experimental error. You have been warned!



#### A.4.5 Predicting Saturation Levels at Alternate Concentrations



Given a saturation of  $f_T^0$  at some free ligand concentration  $[L]$ , the expected saturation  $f_T^c$  at a concentration ratio  $c[L]$  can be derived. From (A.17):

$$[L] = \frac{K}{\frac{1}{f_T^0} - 1} \quad (\text{A.139})$$

And from (A.13):

$$f_T^c = \frac{1}{\frac{K}{c[L]} + 1} \quad (\text{A.140})$$

Combining (A.139) and (A.140):

$$f_T^c = \frac{1}{K \left( \frac{cK}{\frac{1}{f_T^0} - 1} \right)^{-1} + 1}$$

$$f_T^c = \frac{1}{\frac{\frac{1}{f_T^0} - 1}{c} + 1}$$

$$f_T^c = \frac{c}{\frac{1}{f_T^0} - 1 + c} \quad (\text{A.141})$$

When  $[T]_T$  is much less than  $K$ ,  $[L]$  can be approximated using  $[L]_T$ , and  $c$  can then be set to the ratio of the total ligand concentrations.

## **Appendix B**

### **Laboratory Protocols**

#### **B.1 Cleavage and Purification of Fc' from IgG<sub>1</sub> fusion protein**

##### **Protocol**

1. Combine:  
75 mgs CD4-Fc fusion protein (Genentech)  
20 mM MES pH 6.0  
230 mM Mannitol  
23 mM Glycine  
2  $\mu$ g Papain (AmSO<sub>4</sub> precipitate, Worthington)  
Incubate 2 hr at 37 °C with gentle shaking.
2. Quench with a pinch of leupeptin to inactivate papain and incubate 15' more.
3. Purify by column chromatography over a 10 mL Protein A sepharose column (Pharmacia) using Pierce's Gentle Ag/ Ab binding and elution buffers. This will remove the CD4.
4. Purify by gel filtration over a 1 m Superdex 75 column (Pharmacia) at 0.5 mL/min in 20 mM MES pH 6.0. This will enable separation of cut Fc from uncut CD4-Fc fusion. Analyze fractions by SDS-PAGE and pool.
5. Concentrate to at least 1-2 mg/mL using a CentriPrep 10.
6. Store at 4°C.

## B.2 Preparation of M13 Bacteriophage

### Solutions

PBS Phosphate Buffered Saline (1 L)

8 g NaCl

0.2 g KCl

1.43 g Na<sub>2</sub>HPO<sub>4</sub>

0.2 g KH<sub>2</sub>PO<sub>4</sub>

H<sub>2</sub>O to 1 L, pH will be 7.2

PPS Phage Precipitation Solution

20% polyethylene glycol (8k)

2.5 M NaCl

### Protocol

1. Streak XL1 blue (Stratagene) *E. coli* on LB/tetracycline agar and incubate overnight.
2. Inoculate 15 ml 2YT with 10 µg/mL tetracycline using a single colony from the plate. Shake in 125 mL flask, 37°C, 200 RPM.
3. Grow to OD<sub>600</sub> = 1.0 which is about 5x10<sup>8</sup> cells/mL.
4. Infect 2 mL of the culture with a maximum of 5x10<sup>9</sup> bacteriophage particles in a 7 mL miniprep tube. Shake for 60-90 minutes.
5. Add 5x10<sup>9</sup> M13 VCS "helper" phage particles (Stratagene). Shake for 30 minutes.
6. Prepare 250 mL flask with 25 mL 2YT, 5 µg/mL tetracycline, and an appropriate amount of antibiotic resistance marker (usually 5-10 µg/mL chloramphenicol, or 50 µg/mL carbenicillin). Inoculate with the entire 2 mL culture and shake 16 to 24 hours until cell growth is complete but before the cells being to lyse.
7. Transfer culture to 50 mL centrifuge tubes and spin at 8k RPM (SS-34 rotor) for 10' to pellet cells.
8. Transfer supernatant to fresh 50 mL tube with 5 mL PPS. Mix thoroughly. Centrifuge at 10k RPM for 10' to pellet phage. Aspirate supernatant, respin for 3', and reaspirate.
9. Resuspend phage in 1 mL PBS. Transfer to a 1.5 mL microfuge tube. Spin for 3' at 14k to pellet particulate matter. Transfer supernatant to a fresh microfuge tube with 155 µL PPS and mix thoroughly.
10. Spin for 5' at 14k to pellet phage. Aspirate supernatant, respin for 1', and aspirate remaining supernatant.
11. Resuspend in buffer appropriate for the desired usage. For storage purposes, resuspend in 250 µL PBS and store at 4°C.

### **B.3 Titration of M13 Bacteriophage**

#### **Protocol**

- 1 . Streak XL1 blue (Stratagene) *E. coli* on LB/tetracycline agar and incubate overnight.
- 2 . Inoculate 25 ml 2YT with 10  $\mu\text{g}/\text{mL}$  tetracycline using a single colony from the plate. Shake in 250 mL flask, 37°C, 200 RPM.
- 3 . Grow to  $\text{OD}_{600} = 1.0$  which is about  $5 \times 10^8$  cells/mL.
- 4 . Prepare 10-fold serial dilutions of your phage stocks in PBS on a standard (non-stick) 96-well microtiter plate (200  $\mu\text{L}$  per well). Be sure to dilute the stocks down to a level where you expect that there will be <50 phage per well.
- 5 . Dilute the growing culture two-fold with 2YT and then fill 96-well plates with 180  $\mu\text{L}$  culture per well. Use a multichannel pipetteman to transfer 20  $\mu\text{L}$  of diluted phage onto these wells.
- 6 . Cover the plate and incubate for 60' on a rotary shaker.
- 7 . Prepare LB agar plates with the appropriate antibiotic (usually carbenicillin or chloramphenicol) for your phage construct, and warm to 37°C. You will need at least one plate for every three stocks to be titrated in order to perform duplicate measurements.
- 8 . Use a multichannel pipettor to transfer 10  $\mu\text{L}$  of infected culture onto the agar plates in 6x6 matrixes. Incubate overnight.
- 9 . In the morning, count the colonies and back-compute the dilutions to determine the original stock concentrations (in colony forming units per mL).

## **B.4 Preparation of Single Strand Template DNA (for Kunkel Mutagenesis)**

### **Solutions**

**PBS** Phosphate Buffered Saline (1 L)

8 g NaCl

0.2 g KCl

1.43 g Na<sub>2</sub>HPO<sub>4</sub>

0.2 g KH<sub>2</sub>PO<sub>4</sub>

H<sub>2</sub>O to 1 L, pH will be 7.2

**PPS** Phage Precipitation Solution

20% polyethylene glycol (8k)

2.5 M NaCl

### **Protocol**

1. Inoculate 1 mL 2YT with 10 µg/mL chloramphenicol and 50 µg/mL carbenicillin with a single colony of CJ236 from a fresh streak. Grow 8 hours. Add 5x10<sup>9</sup> VCS helper phage particles (Stratagene) and grow 30'. Use this culture to inoculate 400 mL 2YT (with antibiotics) and grow overnight.
2. Spin down the culture in two GSA centrifuge tubes at 10k RPM for 10'. Transfer the supernatant to a fresh tubes with 40 mL PPS. Mix thoroughly. Spin 15k RPM for 15'. Carefully decant supernatant and respin at 15k RPM for 5'. Aspirate the remaining supernatant, and resuspend the pellet in 25 mL PBS. Transfer to a 50 mL centrifuge tube and spin 10k RPM for 10' in an SS-34 rotor to pellet insoluble material. Transfer supernatant to a fresh tube with 4 mL PPS and mix thoroughly. Spin 15k RPM for 15'. Decant supernatant, respin for 1', and aspirate remaining supernatant. Resuspend pellet in 10 mL PBS.
3. Add 10 mL phenol/CHCl<sub>3</sub>. Mix thoroughly then spin at 15k RPM for 20'. Remove top 8 mL of the aqueous phase and transfer to a fresh tube. Add 800 µL 3 M NaOAc (pH 5.2) and 20 mL ethanol at 4°C. Mix and chill at -20°C for 15'. Spin 15k RPM for 20' and decant supernatant. Respin 1' and aspirate remaining supernatant.
4. Resuspend pellet in 8 mL QX1 buffer from the QIAquick Gel Extraction Kit (Qiagen). Purify the DNA using spin columns and the standard protocols and then elute with H<sub>2</sub>O. Final DNA yield should be 900 µL DNA at 0.1-0.4 µg/mL.

## B.5 Preparation of M13 Bacteriophage Display Libraries

### Solutions

TM Tris/Magnesium Buffer  
0.5 M Tris pH 7.5  
0.1 M MgCl<sub>2</sub>

### Protocol

1. **Kinase:**  
2  $\mu$ L oligo (10 OD/mL stock)  
2  $\mu$ L TM buffer  
2  $\mu$ L 10 mM ATP  
2  $\mu$ L 100 mM DTT  
12.5  $\mu$ L H<sub>2</sub>O  
0.5  $\mu$ L Kinase  
Incubate for 40 minutes at 37°C.
2. **Anneal:**  
20  $\mu$ g kunkel template (see Protocol B.4)  
20  $\mu$ L kinased oligo ( $\approx$  0.6  $\mu$ g)  
25  $\mu$ L TM buffer  
H<sub>2</sub>O to 250  $\mu$ L  
anneal from 95°C to 40°C over 60' in PCR machine.
3. **Fill-in:** to the above add  
1.5  $\mu$ L 100 mM ATP  
10  $\mu$ L 25 mM dNTPs  
3.0  $\mu$ L TM  
15  $\mu$ L 100 mM DTT  
chill on ice, then add  
6  $\mu$ L T4 ligase (NEB)  
3  $\mu$ L T7 polymerase (NEB)  
incubate at 15°C overnight  
run 1  $\mu$ L on TAE agarose gel to verify fill-in
4. **Purify** DNA by extracting with phenol/CHCl<sub>3</sub> and then Tris-EDTA saturated CHCl<sub>3</sub>. Then add 1 mL QX1 buffer (Qiagen) and purify over two QI-Aquick columns, eluting with 30  $\mu$ L H<sub>2</sub>O.
5. **Electroporate** into 350  $\mu$ L competent cells, recover 90' in 35 mL SOC in a 250 mL flask shaking at 200 RPM in a 37°C incubator. At this point, remove an aliquot of cells and titer a small aliquot by serial dilution onto an agar plate with appropriate antibiotics.
6. **Grow** library by adding 1 mL VCS helper phage (Stratagene) and shaking overnight. Harvest phage as usual (Protocol B.2 step 7 on).

## B.6 Affinity Panning of M13 Bacteriophage Libraries

### Solutions

**CB Coating Buffer**

100 mM sodium carbonate, pH 9.6

**CBB Casein Blocking Buffer (1 L)**

3.55 g Na<sub>2</sub>HPO<sub>4</sub>

10 g Hammersten grade casein

11.6 g NaCl

0.02 % thimerisol (optional preservative)

H<sub>2</sub>O to 1 L, pH will be 7.1

**DEB Dithiothreitol Elution Buffer**

50 mM DTT (diluted fresh in water from frozen 1 M stock)

**GEB Glycine Elution Buffer**

100 mM glycine, pH 2.0

**PT Phosphate Buffered Saline with Tween (1 L)**

8 g NaCl

0.2 g KCl

1.43 g Na<sub>2</sub>HPO<sub>4</sub>

0.2 g KH<sub>2</sub>PO<sub>4</sub>

500 μL Tween-20

H<sub>2</sub>O to 1 L, pH will be 7.2

### Protocol

1. Immobilize target protein on Maxisorp plates (Nunc). To avoid contamination, use one plate per library. Prepare protein solutions in CB at 5-10 μM. Transfer 100 μL to each well (1 well per library). For each coated well, prepare a control well with 100 μL CB alone. Incubate overnight at 4°C.
2. Wash plates 10X with PT and then incubate wells with 200 μL CBB for 30' to prevent non-specific binding.
3. Prepare fresh phage as described in B.2 and resuspend in 250 μL CBB.
4. For each library, transfer 100 μL to each of the experimental and control wells. Incubate at R.T. on an rotary shaker (gentle) for at least one hour.
5. Wash plates by hand 20X with PT. To avoid contamination, do not use a automatic plate washer.
6. Elute phage using either GEB or DEB. Add 100 μL to each experimental and control well. Shake vigorously for 30 sec. on rotary shaker. Transfer solution to a microfuge tube. If using GEB, neutralize with 15 μL Tris base.
7. Titer the starting stock and outputs from the control and experimental wells using protocol B.3. Sequencing can be performed off of the titration plates.

## B.7 Phage-ELISA Binding Assay ( $EC_{50}$ 's)

Use of a 96-well plate washer in the following assay is highly recommended to insure uniform results. **NOTE:** In this assay, the phage is the target, the ligand is the immobilized protein, and the competitor is the protein which binds to the phage in solution.

### Solutions

#### CB Coating Buffer

100 mM sodium carbonate, pH 9.6

#### CBB Casein Blocking Buffer (1 L)

3.55 g  $Na_2HPO_4$

10 g Hammersten grade casein

11.6 g NaCl

0.02 % thimerisol (optional preservative)

H<sub>2</sub>O to 1 L, pH will be 7.1

#### PT Phosphate Buffered Saline with Tween (1 L)

8 g NaCl

0.2 g KCl

1.43 g  $Na_2HPO_4$

0.2 g  $KH_2PO_4$

500  $\mu$ L Tween-20

H<sub>2</sub>O to 1 L, pH will be 7.2

### Protocol

1. Immobilize ligand protein on Maxisorp plates (Nunc). Allocate one row or column of an ELISA plate for each phage variant to be measured, and prepare identical plates for the titration and competition stages. Prepare ligand protein solutions in CB at 5-10  $\mu$ M. Transfer 100  $\mu$ L to each well (1 well per library). Incubate for 60' at RT or overnight at 4°C.
2. Wash plates 10X with PT and then incubate wells with 200  $\mu$ L CBB for 30' to prevent non-specific binding.
3. Prepare fresh phage as described in B.2 and resuspend in 250  $\mu$ L CBB. If comparing a family of mutants, be sure to prepare enough phage to allow for a control measurement on each plate.
4. **Titration:** Prepare serial dilutions of phage in CBB on 96-well non-stick plates (one row or column per phage variant). 3-fold dilutions over 12 wells is usually optimal (160  $\mu$ L final volume per well).



5. Wash half of the ligand coated plates 5X with PT, then transfer 100  $\mu$ L of the phage dilutions onto these plates. Incubate 60' at RT. Wash 20X with PT then develop using  $\alpha$ -M13 HRP conjugate and a colorimetric substrate such as OPD.
6. For each variant, plot the signal as a function of phage dilution. Use graphing software to fit each curve to a sigmoidal function and determine the phage concentration for each variant that will give a uniform signal at 25% of maximum (this is the fractional saturation of the immobilized ligand). Prepare 2X concentrations of the phage variants in CBB at the concentration appropriate for each variant.
7. **Competition:** Prepare serial dilutions of competing ligand protein in CBB on a non-stick 96-well plate. 3- or 4-fold dilutions spread over 12 wells is usually optimal starting from a concentration expected to inhibit 95% of binding.
8. Wash the remaining ligand coated plates 5X with PT, then transfer 50  $\mu$ L of competing ligand protein dilutions and 50  $\mu$ L of phage dilutions onto each plate. Incubate 60' at RT, then wash 20X with PT and develop using  $\alpha$ -M13 HRP conjugate and a colorimetric substrate such as OPD (Sigma).
9. For each variant, plot the signal as a function of competing ligand protein dilution. Using curve-fitting software, find the mid-point concentration of the curves. This will be the  $EC_{50}$  value for the variant.

## B.8 ELISA Inhibition Binding Assay (IC<sub>50</sub>'s)

Use of a 96-well plate washer in the following assay is highly recommended to insure uniform results. **NOTE:** In this assay, the immobilized protein is the target, the ligand is the labeled, and the competitor is the protein which binds to the immobilized target.

### Solutions

#### CB Coating Buffer

100 mM sodium carbonate, pH 9.6

#### CBB Casein Blocking Buffer (1 L)

3.55 g Na<sub>2</sub>HPO<sub>4</sub>

10 g Hammersten grade casein

11.6 g NaCl

0.02 % thimerisol (optional preservative)

H<sub>2</sub>O to 1 L, pH will be 7.1

#### PT Phosphate Buffered Saline with Tween (1 L)

8 g NaCl

0.2 g KCl

1.43 g Na<sub>2</sub>HPO<sub>4</sub>

0.2 g KH<sub>2</sub>PO<sub>4</sub>

500 μL Tween-20

H<sub>2</sub>O to 1 L, pH will be 7.2

### Protocol

1. Immobilize target protein on Maxisorp plates (Nunc). Allocate one row or column of an ELISA plate for each competition experiment. Also allocate one plate for titration of labeled ligand protein. Prepare target protein solutions in CB at 5-10 μM. Transfer 100 μL to each well (1 well per library). Incubate for 60' at RT or overnight at 4°C.
2. Wash plates 10X with PT and then incubate wells with 200 μL CBB for 30' to prevent non-specific binding.
3. **Titration:** Prepare serial dilutions of the labeled ligand in CBB on 96-well non-stick plates (prepare several replicate rows). 3-fold dilutions over 12 wells is usually optimal (160 μL final volume per well).
4. Wash the target coated plates for titration 5X with PT, then transfer 100 μL of the diluted labeled ligand onto these plates. Incubate 60' at RT. Wash 20X with PT then develop as appropriate for the label.

- 5 . Plot the signal as a function of label dilution. Use graphing software to fit the curve to a sigmoidal function and determine the labeled ligand concentration that give a signal around 25% of maximum (this is the fractional saturation of target). Prepare 2X stocks of the label in CBB at this concentration.
- 6 . **Competition:** Prepare serial dilutions of the competitors in CBB on a non-stick 96-well plate. 3- or 4-fold dilutions spread over 12 wells is usually optimal starting at a concentration expected to inhibit at least 95% of binding.
- 7 . Wash the remaining target coated plates 5X with PT, then transfer 50  $\mu\text{L}$  of label dilutions and 50  $\mu\text{L}$  of competitor dilutions onto each plate. Incubate 60' at RT, wash 20X with PT and develop as appropriate for the label.
- 8 . For each variant, plot the signal as a function of competitor dilution. Using curve-fitting software, find the mid-point concentration of the curves. These will be the  $\text{IC}_{50}$  values for the competitors.

## **B.9 Protein Expression from a Phagemid Vector**

### **Solutions**

TE 100 mM Tris-HCl pH 8.0  
1 mM EDTA

### **Protocol**

1. Transform DNA in to in PEG/DMSO competent cells (34B8 or 27C7) using standard protocols, plate on agar containing antibiotic appropriate for the phagemid marker. The phagemid must contain an amber (TAG) stop codon at the end of the protein and before M13 gene III.
2. Use a toothpick to inoculate 2 mL LB (plus antibiotic) in a miniprep tube with a colony from the plate. Grow for 5-10 hours.
3. Gently spin down cells at 3k rpm for 5' and remove supernatant. Replace with 1 mL AP5 media and resuspend cells. Then inoculate 50 mL AP5 (plus antibiotic) in a 500 mL flask with the 1 mL culture. Grow 16-24 hours at 37°C 250 RPM. Be careful not to over-grow. If a larger culture is required, after about 8 hours, this 50 mL culture can be used to inoculate a 1 L culture in a 4 L flask, which should be grown another 16 hours. Final OD600s should be around 1.2
4. Spin down the cells at 7k RPM for 10'. Decant supernatant. Respin briefly then aspirate.
5. Freeze the cells overnight at -20°C .
6. Thaw and resuspend cells in 1/500th of the original culture volume in ice cold TE, with 1 mM PMSF and 1 mM benzamidine.
7. Mix for 2 hours at 4°C on an orbital shaker.
8. Spin at 10k for 15' to pellet fractured cells. Initiate protein-specific purification procedure on the supernatant (i.e. ammonium sulfate precipitation, column chromatography, etc.).

## **Appendix C**

### **Contents of the CD-ROM**

#### **C.1 Introduction**

As we advance forward into the information age, it seems likely that soon most graduate theses will be available via the internet and will be accompanied by all kinds of additional information in the form of interactive graphics, animations, and supplemental information.

This thesis is supplemented by a CD-ROM disk which contains electronic versions of the thesis and a collection of movies generated using RigiMOL. If you do not have a copy of the disk, you can request one by sending an email to:

`warren@delanoscientific.com`

#### **C.2 Contents**

To access information on the CD-ROM, insert the disk and open the **index.html** file using a web browser such as Netscape or Internet Explorer. This file will open automatically on some computers.

### **C.2.1 RigiMOL Movies**

A collection of movies generated using RigiMOL is included on the CD-ROM. The following systems are visualized:

- IgG-Fc
- Human Growth Hormone Hormone/Receptor
- Thymidylate Synthase
- HIV Protease

### **C.2.2 Thesis in PDF Format**

The CD-ROM contains a copy of this entire document in PDF format. You will need Adobe Acrobat version 4.0 or higher to view or print this document.

### **C.2.3 Thesis in Postscript Format**

This file is intended for printing only. It may require a printer with at least 64 MB of RAM, and it will reproduce best on a color postscript laser printer.

## References

- [Anderson, 1999] Anderson, A. C., O'Neil, R. H., DeLano, W. L., Stroud, R. M. *The structural mechanism for half-the-sites reactivity in an enzyme, thymidylate synthase, involves a relay of changes between subunits.* *Biochemistry* **38**, 13829-13836 (1999).
- [Argos, 1988] Argos, P. *An investigation of protein subunit and domain interfaces.* *Protein Eng.* **2**, 101-113 (1988).
- [Arza, 1998] Arza, B., F  lez, J. *The emerging impact of phage display technology in thrombosis and haemostasis.* *Thromb. Haemost.* **80**, 354-365 (1998).
- [Bartsch, 1974] Bartsch, H.-J. *Handbook of mathematical formulas.* Academic Press, Inc.: New York (1974).
- [Bellamacina, 1996] Bellamacina, C. R. *The nucleotide binding motif: a comparison of nucleotide binding proteins.* *FASEB J.* **10**, 1257-1269 (1996).
- [Bernstein, 1977] Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M. *The protein data bank: a computer-based archival file for macromolecular structures.* *J. Mol. Biol.* **112**, 535-542 (1977).
- [Bogan, 1998] Bogan A. A., Thorn K. S. *Anatomy of hot spots in protein interfaces.* *J. Mol. Biol.* **280**, 1-9 (1998).
- [Bottomley, 1994] Bottomley, S. P., Popplewell, A. G., Scawen, M., Wan, T., Sutton, B. J., Gore, M. G. *The stability and unfolding of an IgG binding protein based upon the B domain of protein A from Staphylococcus aureus probed by tryptophan substitution and fluorescence spectroscopy.* *Protein Eng.* **7**, 1463-1470 (1994).
- [Bracken, 1994] Bracken, C., Gulyas, J., Taylor, J. W., Baum, J. *Synthesis and nuclear magnetic resonance structure of an  $\alpha$ -helical, bicyclic, lactam-bridged hexapeptide.* *J. Am. Chem. Soc.* **116**, 6431-6432 (1994).
- [Braisted, 1996] Braisted, A. C., Wells, J. A. *Minimizing a binding domain from protein A.* *Proc. Natl. Acad. Sci.* **93**, 5688-5692 (1996).
- [Braisted, 1996] Braisted, A. C. *Unpublished results.* Genentech, Inc., South San Francisco, California.

- [Browner, 1992] Browner, M. F., Fletterick, R. J. *Phosphorylase: a biological transducer*. Trends. Biochem. 17, 66-71 (1992).
- [Brünger, 1992] Brünger, A. T. *X-PLOR Version 3.1. A system for X-ray crystallography and NMR* Yale Univeristy Press: New Haven (1992).
- [Brünger, 1998] Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. *Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination*. Acta Cryst. D. 51, 740-748 (1998).
- [Burmeister, 1994a] Burmeister, W. P., Huber, A. H., Bjorkman, P. J. *Crystal structure of the complex of rat neonatal Fc receptor with Fc*. Nature 372, 379-383 (1994).
- [Burmeister, 1994b] Burmeister, W. P., Gastinel, L. N., Blum, M. L., Bjorkman, P. J. *Crystal structure at 2.2 Å resolution of the MHC-related neonatal Fc receptor*. Nature 372, 336-343 (1994).
- [Capon, 1989] Capon, D. J., Chamow, S. M., Mordenti, J., Marsters, S. A., Gregory, T., Mitsuya, H., Byrn, R. A., Lucas, C., Wurm, F., Groopman, J. E., Broder, S., Smith, D. H. *Designing CD4 immunoadhesins for AIDS therapy*. Nature 337, 525-531 (1989).
- [Carr, 1993] Carr, C. M., Kim, P. S. *A Spring-loaded mechanism for the conformational change of influenza hemagglutinin*. Cell 73, 823-832 (1993).
- [Carter, 1996] Carter, P. *Personal communication*. Genentech, Inc., South San Francisco, California.
- [Chothia, 1975] Chothia, C., Janin, J. *Principles of protein-protein recognition*. Nature 256, 705-708 (1975).
- [Clackson, 1995] Clackson, T., Wells, J. A. *A hot spot of binding energy in a hormone-receptor interface*. Science 267, 383-386 (1995).
- [Conte, 1999] Conte, L. L., Chothia, C., Janin, J. *The atomic structure of protein-protein recognition sites*. J. Mol. Biol. 285, 2177-2198 (1999).
- [Corper, 1997] Corper, A. L., Sohi, M. K., Bonagura, V. R., Steinitz, M., Jefferis, R., Feinstein, A., Beale, D., Taussig, M. J., Sutton, B. J. *Structure of human IgM rheumatoid factor Fab bound to its autoantigen IgG-Fc reveals a novel topology of antibody-antigen interaction*. Nature Struct. Biol. 4, 374-381 (1997).
- [Creighton, 1993] Creighton, T. E. *Proteins* W. H. Freeman and Company: New York (1993).
- [Crippen, 1978] Crippen, G. M. *The tree structural organization of proteins*. J. Mol. Biol. 126, 315-332 (1978).



- [Cunningham, 1990] Cunningham, B. C., Bass, S. J., Fuh, G., Wells, J. A. *Zinc mediation of the binding of human growth hormone to the human prolactin receptor.* *Science* **250**, 1709-1712 (1990). J. Mol. Biol.234554563
- [Cunningham, 1993] Cunningham, B. C., Wells, J. A. *Comparison of a structural and a functional epitope.* *J. Mol. Biol.* **234**, 554-563 (1993).
- [Cunningham, 1997] Cunningham, B. *Unpublished results.* Genentech, Inc., South San Francisco, California.
- [Dalgarno, 1998] Dalgarno, D. C., Botfield, M. C., Rickles, R. J. *SH3 Domains and drug design: ligands, structure, and biological function.* *Biopolymers* **43**, 383-400 (1998).
- [Deisenhofer, 1981] Deisenhofer, J. *Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9-Å and 2.8-Å resolution.* *Biochemistry* **20**, 2361-2370 (1981).
- [de Vos, 1998] de Vos, A. M. *Unpublished results: crystal structure of Protein A B-Domain in complex with IgG-Fc at 2.6 Å resolution.* Genentech, Inc., South San Francisco, California.
- [Dodson, 1998] Dodson, G., Wlodawer, A. *Catalytic triads and their relatives.* *Trends. Biochem.* **23**, 347-352 (1998).
- [Fahnestock, 1990] Fahnestock, S. R., Alexander, P., Filpula, D., Nagle, J. *Structure and evolution of the streptococcal genes encoding protein G.* *Bacterial Immunoglobulin-Binding Proteins* **1**, 133-148 (1990).
- [Ferrin, 1988] Ferrin, T. E., Huang, C. C., Jarvis, L. E., Langridge, R. *The MIDAS display system.* *J. Mol. Graphics* **6**, 13-27 (1988).
- [Frick, 1992] Frick, I. M., Wikström, M., Forsén, S., Drakenberg, T., Gomi, H., Sjöbring, U., Björk, L. *Convergent evolution among immunoglobulin G-binding bacterial proteins.* *Proc. Natl. Acad. Sci.* **89**, 8532-8536 (1992).
- [Fuh, 1992] Fuh, G., Cunningham, B. C., Fukunaga, R., Nagata, S., Goeddel, D. V., Wells, J. A. *Rational design of potent antagonists to the human growth hormone receptor.* *Science* **256**, 1677-1680 (1992).
- [Gerstein, 1998] Gerstein, M., Krebs, W.G. *A database of macromolecular motions.* *Nucleic Acids Res.* **26**, 4280-4290 (1998).
- [Gerstein, 1999] Gerstein, M., Krebs, W. G. *The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework.* Submitted (1999).
- [Ghetie, 1997] Ghetie, V., Ward, E. S. *FcRn: the MHC class I-related receptor that is more than an IgG transporter.* *Immuno. Today* **18**, 592-598 (1997).

- [Gouda, 1992] Gouda, H., Torigoe, H., Saito, A., Arata, Y., Shimada, I. *Three dimensional solution structure of the B domain of Staphylococcal protein A: comparisons of the solution and crystal structures.* *Biochemistry* **31**, 9665-9672 (1992).
- [Guss, 1990] Guss, B., Lindberg, M., Uhlén, M. *The gene for staphylococcal protein A.* *Bacterial Immunoglobulin-Binding Proteins: Academic Press, Inc.*, 29-39 (1990).
- [Heizmann, 1991] Heizmann, C. W., Hunziker, W. *Intracellular calcium-binding proteins: more sites than insights.* *Trends. Biochem.* **16**, 98-103 (1991).
- [Langone, 1982] Langone, J. L. *Protein A of Staphylococcus aureus and related immunoglobulin receptors produced by Streptococci and Pneumococci.* *Adv. Immuno.* **32**, 157-252 (1982).
- [Jackson, 1991] Jackson, D. Y., King, D. S., Chmielewski, J., Singh, S. *General approach to the synthesis of short  $\alpha$ -helical peptides.* *J. Am. Chem. Soc* **113**, 9391-9392 (1991).
- [Jin, 1994] Jin, L., Wells, J. A. *Dissecting the energetics of an antibody-antigen interface by alanine shaving and molecular grafting.* *Protein Science* **3**, 2351-2357 (1994).
- [Janin, 1990] Janin, J., Chothia, C. *The structure of protein-protein recognition sites.* *J. Biol. Chem.* **265**, 16027-16030 (1990).
- [Janin, 1996] Janin, J. *Protein-protein recognition.* *Prog. Biophys. Mol. Biol.* **64**, 145-165 (1996).
- [Jenderberg, 1995] Jenderberg, L., Persson, B., Andersson, R., Karlsson, R., Uhlén, M., Nilsson, B. *Kinetic analysis of the interaction between protein A domain variants and human Fc using plasmon resonance detection.* *J. Mol. Recog.* **8**, 270-278 (1995).
- [Jenderberg, 1996] Jenderberg, L., Tashiro, M., Tejero, R., Lyons B. A., Uhlén, M., Montelione G, T., Nilsson, B. *The mechanism of binding staphylococcal protein A to immunoglobulin G does not involve helix unwinding.* *Biochemistry* **35**, 22-31 (1996).
- [Jones, 1996] Jones, S., Thornton, J. M. *Principles of protein-protein interactions.* *Proc. Natl. Acad. Sci.* **93**, 13-20 (1996).
- [Karlsson, 1995] Karlsson, R., Jendeberg, L., Nilsson, B., Nilsson, J., Nygren, P. *Direct and competitive kinetic analysis of the interaction between human IgG1 and a one domain analogue of protein A* *J. Immuno. Methods* **183**, 43-49 (1995).
- [Kleywegt, 1996] Kleywegt, G. J., Brünger, A. T. *Checking your imagination: applications of the free R value.* *Structure* **4**, 897-904 (1996).
- [Kunkel, 1987] Kunkel, T. A., Roberts, J. D., Zakour, R. A. *Rapid and efficient site-specific mutagenesis without phenotypic selection.* *Methods Enzymol.* **154**, 367-382 (1987)..

- [Lee, 1983] Lee, C. H., Moseley, S. L., Moon, H. W., Whipp, S. C., Gyles, C. L., So, M. *Characterization of the gene encoding heat-stable toxin II and preliminary molecular epidemiological studies of enterotoxigenic Escherichia coli heat-stable toxin II producers.* *Infect. Immun.* **42**, 264-268 (1983).
- [Levitt, 1985] Levitt, M., Sander, C., Stern P. S. *Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease, and lysozyme.* *J. Mol. Biol.* **181**, 423-447 (1985).
- [Livnah, 1996] Livnah, O., et al. *Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å.* *Science* **273**, 464-471 (1996).
- [Lowman, 1991] Lowman, H. B., Bass, S. H., Simpson N., Wells, J. A. *Selecting high-affinity binding proteins by monovalent phage display.* *Biochemistry* **30**, 10832-10838 (1991).
- [Lowman, 1993] Lowman, H. B., Wells, J. A. *Affinity maturation of human growth hormone by monovalent phage display.* *J. Mol. Biol.* **234**, 564-578 (1993).
- [Lowman, 1998] Lowman, H. B., Chen, Y. M., Skelton, N. J., Mortensen, D. L., Tomlinson, E. E., Sadick, M. D., Robinson, I. C. A. F., Clark, R. G. *Molecular mimics of insulin-like growth factor 1 (IGF-1) for inhibiting IGF-1: IGF-binding protein interactions.* *Biochemistry* **37**, 8870-8878 (1998).
- [Lupas, 1997] Lupas, A. *Predicting coiled-coil regions in proteins.* *Curr. Opin. Struct. Biol.* **7**, 388-393 (1997).
- [Kay, 1998] Kay, B. K., Kurakin, A. V., Hyde-DeRuyscher, R. *From peptides to drugs via phage display.* *Drug Disc. Today* **3**, 370-378 (1998).
- [Koivunen, 1999] Koivunen, E., Arap, W., Rajotte, D., Lahdenranta, J, Pasqualini, R. *Identification of receptor ligands with phage display peptide libraries.* *J. Nucl. Med.* **40**, 863-888 (1999).
- [McDowell, 1998a] McDowell, R. S., Gadek, T. R., Barker, P. L., Burdick, D. J., Chan, K. S., Quan, C. L., Skelton, N., Struble, M., Thorsett, E. D., Tischler, M., Tom, J. Y. K., Webb, T. R., Burnier, J. P. *From peptide to non-peptide. 1. The elucidation of a bioactive conformation of the arginine-glycine-aspartic acid recognition sequence.* *J. Am. Chem. Soc.* **116**, 5069-5076 (1998).
- [McDowell, 1998b] McDowell, R. S., Blackburn, B. K., Gadek, T. R., McGee, L. R., Rawson, T., Reynolds, M. E., Robarge, K. D., Somers, T. C., Thorsett, E. D., Tischler, M., Webb, R. R., Venuti, M. C. *From peptide to non-peptide. 2. The de novo design of potent, non-peptidal inhibitors of platelet aggregation based on a benzodiazepinedione scaffold.* *J. Am. Chem. Soc.* **116**, 5077-5083 (1998).
- [Mire-Sluis, 1999] Mire-Sluis, A. R. *Cytokines: from technology to therapeutics.* *TIBTECH* **17**, 319-325 (1999).

- [Muller, 1998] Muller, Y. A., Chen, Y., Christinger, H. W., Li, B., Cunningham, B. C., Lowman, H. B., de Vos, A. M. *VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 Å resolution and mutational analysis of the interface*. *Structure* **6**, 1153-1167 (1998).
- [Navaza, 1994] Navaza, J. *Amore - an automated package for molecular replacement*. *Acta Cryst.* **A50**, 157-163 (1994).
- [Nichols, 1995] Nichols, W. L., Rose, G. D., Eyck, L. F. T., Zimm, B. H. *Rigid domains in proteins: an algorithmic approach to their identification*. *Proteins* **23**, 38-48 (1995).
- [Nilson, 1995] Nilson, B. H. K., Frick, I., Åkesson, P., Försten, S., Björk, L., Åkerström, B., Wikström, M. *Structure and stability of Protein H and the M1 protein from Streptococcus pyogenes. Implications for other surface proteins of gram-positive bacterial immunoglobulin G-binding bacterial proteins*. *Proc. Natl. Acad. Sci.* **89**, 8532-8536 (1995).
- [Nilsson, 1987] Nilsson, B., Moks, T., Jansson, B., Abrahamsén, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A., Uhlén, M. *A synthetic IgG-binding domain based on Staphylococcal protein A*. *Protein Eng.* **1**, 107-113 (1987).
- [Nord, 1995] Nord, K., Nilsson, J., Nilsson, B., Uhlén, M., Nygren, P. *A combinatorial library of an  $\alpha$ -helical bacterial receptor domain*. *Protein Eng.* **8**, 601-608 (1995).
- [Pavone, 1992] Pavone, V., Di Blasio, B., Santiti, A., Benedetti, E., Pedone, C., Crisma, M. *The longest, regular polypeptide 3(10) helix at atomic resolution*. *J. Mol. Biol.* **214**, 633-635 (1992).
- [Pearce, 1996] Pearce, K. H., Ultsch, M. H., Kelley, R. F., de Vos, A. M., Wells, J. A. *Structural and mutational analysis of affinity-inert contact residues at the growth hormone receptor interface*. *Biochemistry* **35**, 10300-10307 (1996).
- [Phelan, 1997] Phelan, J. C., Skelton, N. J., Braisted, A. C., McDowell, R. S. *A general method for constraining short peptides to an alpha-helical conformation*. *J. Am. Chem. Soc.* **119**, 455-460 (1997).
- [Read, 1986] Read, R. J. *Improved Fourier coefficients for maps using phases from partial structures with errors*. *Acta Cryst.* **A42**, 140-149 (1986).
- [Rossman, 1974] Rossman, M. G., Liljas, A. *Recognition of structural domains in globular proteins*. *J. Mol. Biol.* **85**, 177-181 (1974).
- [Russel, 1999] Russel, C. S., Clarke, L. A. *Recombinant proteins for genetic disease*. *Clinical Genetics* **55**, 389-394 (1999).
- [Sandberg, 1993] Sandberg, W. S., Terwilliger, T. C. *Engineering multiple properties of a protein by combinatorial mutagenesis*. *Proc. Natl. Acad. Sci.* **90**, 8367- (1993).



- [Sanger, 1977] Sanger, F., Nicklen, S., Coulson, A. R. *DNA sequencing with chain terminating inhibitors*. Proc. Natl. Acad. Sci 74, 5463-5467 (1977).
- [Sauer, 1996] Sauer, R. T., Milla, M. E., Waldburger, C. D., Brown, B. M., Schildbach, J. F. *Sequence determinants of folding and stability for the P22 Arc repressor dimer*. FASEB J. 10, 42-48 (1996).
- [Sharp, 1990] Sharp, K. A., Honig, B. *Electrostatic interactions in macromolecules*. Annu. Rev. Biophys. Chem. 19, 301-332 (1990).
- [Sauer-Eriksson, 1995] Sauer-Eriksson, A. E., Kelywegt, G. J., Uhlen, M., Jones, T. A. *Crystal structure of the C2 fragment of Streptococcal protein G in complex with the Fc domain of human IgG*. Structure 3, 265-278 (1995).
- [Skinner, 1996] Skinner, M. M., Terwilliger, T. C. *Potential use of additivity of mutational effects in simplifying protein engineering*. Proc. Natl. Acad. Sci. 93, 10753-10757 (1996).
- [Stone, 1987] Stone, G. C., Sjöbring, U., Björck, L., Sjöquist, J., Barber, C. V., Nardella, F. A. *The Fc binding site for streptococcal protein G is in the C $\gamma$ 2-C $\gamma$ 3 interface region of IgG and is related to the sites that bind staphylococcal protein A and human rheumatoid factors*. J. Immunol. 143, 565-570 (1987).
- [Towbin, 1987] Towbin, H., Staehelin, T., Gordon, J. *Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications*. Proc. Natl. Acad. Sci. 76, 4350-4354 (1987).
- [Tsai, 1997] Tsai C., Lin, S. L., Wolfson, H. J., Nussinov, R. *Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect*. Protein Science 6, 53-64 (1997).
- [Uhlén, 1984] Uhlén, M., Guss, B., Nilsson, B., Gatenbock, S., Philipson, L., Lindberg, M. *Complete sequence of the staphylococcal gene encoding protein A*. J. Biol. Chem. 3, 1695-1702 (1984).
- [Ultsch, 1999] Ultsch, M. H., de Vos, A. M. *Manuscript in preparation*. Genentech, Inc., South San Francisco, California.
- [Vaughan, 1993] Vaughan, J. H. *Pathogenic concepts and origins of rheumatoid factor in rheumatoid arthritis*. Arthritis Rheum. 36, 1-6 (1993).
- [Vaughn, 1997] Vaughn, D. E., Bjorkman, P. J. *High-affinity binding of the neonatal Fc receptor to its ligand requires receptor immobilization*. Biochemistry 36, 9374-9380 (1997).
- [Weismann, 1998] Weismann, C., Christinger, H. W., Cochran, A. G., Cunningham, B. C., Fairbrother, W. J., Keenan, C. J., Meng, G., de Vos, A. M. *Crystal structure of the complex between VEGF and a receptor-blocking peptide*. Biochemistry 37, 17765-17772 (1998).

- [Wells, 1996] Wells, J. A., de Vos, A. M. *Hematopoietic receptor complexes*. *Annu. Rev. Biochem.* **65**, 609-634 (1996).
- [Wriggers, 1997] Wriggers, W., Schulten, K. *Protein domain movements: detection of rigid domains and visualization of hinges in comparisons of atomic coordinates*. *Proteins* **29**, 1-14 (1997).
- [Wrighton, 1996] Wrighton, N. C., et al. *Small peptides as potent mimetics of the protein hormone erythropoietin*. *Science* **273**, 458-463 (1996).
- [Young, 1997] Young, A. C. M., Valadon, P., Casadevall, A., Scharff, M. D., Sacchettini, J. C. *The three-dimensional structures of a polysaccharide binding antibody to *Cryptococcus neoformans* and its complex with a peptide from a phage display library: implications for the identification of peptide mimotopes*. *J. Mol. Biol.* **1997**, 274-622 (1997).630





# For reference

Not to be taken  
from the room.

7063772



3 1378 00706 3772

