BIOMEDICAL APPLICATIONS OF BIOTECHNOLOGY

VOLUME 1

Biologically Active Peptides: Design, Synthesis, and Utilization

EDITED BY

William V. Williams
David B. Weiner

CRC Press
Taylor & Francis Group

Biologically Active Peptides: Design, Synthesis, and Utilization

— BIOMEDICAL APPLICATIONS OF BIOTECHNOLOGY —

VOLUME 1

Biologically Active Peptides: Design, Synthesis, and Utilization

EDITED BY

William V. Williams, M.D.
Director of Rheumatology Research
University of Pennsylvania

David B. Weiner, Ph.D.
Director of Biotechnology
The Wistar Institute

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

ISBN 13: 978-0-87762-935-1 (pbk)

Library of Congress Card Number 92-56712

Library of Congress Cataloging-in-Publication Data

Main entry under title:
Biologically Active Peptides: Design, Synthesis and Utilization

Full catalog record is available from the Library of Congress

DOI: 10.1201/9781003419853

Table of Contents

Introduction

THIS book marks the first in a series addressing the impact of the latest aspects of biotechnology on both clinical and experimental medicine. The biotechnology revolution currently underway has profoundly altered the approach of experimental biologists to their investigations. These changes are now reaching into clinical medicine with a barrage of novel diagnostic and therapeutic products. The pace of change has been both exhilarating and exhausting, as scientists and clinicians are challenged to keep abreast of developments which are more and more rapidly altering the study and practice of biology and medicine. The goal of this series is to acquaint the reader with some of the current approaches being taken in specific aspects of biotechnology as applied to experimental and clinical medicine.

The first book focuses on peptides, short strings of amino acids not large enough to warrant the term protein. This book is divided into three sections. The first section "Structural Approaches to Peptide Design" elaborates on theoretical aspects of peptide design for production of biologically active peptides and peptide mimetics. The second section "Chemical Synthetic Aspects of Peptide Production" explores the chemical synthetic strategies available for peptide production. The third section "Utilization of Bioactive Peptides" details several clinically relevant experimental systems where peptide use has led to either new therapeutic approaches or to important insights from the investigations. Thus, this book should lead the reader through the steps involved in development of biologically active peptides: design, synthesis, and evaluation.

Peptide design, as elaborated in the first section, has been greatly enhanced recently by the use of molecular structure analysis. These approaches are discussed in the first section. The chapter contributed by

Kieber-Emmons, "Strategies in Protein-Based Rational Drug Design," relates structural analysis of ligands and immunogens to rational peptide design. This chapter emphasizes structural homology for molecular modeling strategies related to peptide design. Recent advances in computational chemistry approaches with data base utilization are emphasized in a number of systems where bioactive peptides have been developed based on active site structures. While a peptide can exist in many conformations in solution, typically only a small set of conformations relate to the bioactive conformation of the peptide. Balaji and Ramnarayan explore their recent advances in biocomputational simulation of peptide energy minimization to determine the bioactive conformation of peptides and peptidomimetics. As discussed in "Computer Assisted Drug Design of Peptidomimetic Drugs," this approach has already led to the development of bioactive peptide analogs of relatively small size. Another aspect of peptide design of great interest is the development of small peptide analogs of larger molecules. This is aided by information relating to the active site of the molecule of interest. Von Feldt, Ugen, Kieber-Emmons, and Williams discuss the recently described approach of peptide design based on antibody structural analysis. This approach, as outlined in "Bioactive Peptide Design Based on Antibody Structure" capitalizes on the readily available information relating to antibody active sites. Critical information derived from peptide structural analysis can be applied to refinement of peptide structures. Shon and Opella, in "NMR Spectroscopy of Peptides and Proteins in Membrane Environments," discuss the principles of nuclear magnetic resonance spectroscopy as applied to the determination of peptide structure. They illustrate the use of two- and three-dimensional NMR to determine the structure of membrane associated peptides and proteins, including antibiotic magainin peptides and analogs as well as larger proteins. Thus, the first section elaborates on several approaches to molecular structural analysis as applied to bioactive peptide design.

The increasing utilization of peptides as both experimental tools and in drug design has been greatly facilitated by recent advances in synthetic strategies. The optimal synthetic approach for a particular peptide depends on several factors. This includes the amount desired, the length of the peptide, and the purpose for which it is intended. The synthetic strategies available include solid phase synthesis, liquid phase synthesis, and recombinant DNA approaches. These strategies are explored in depth in the chapter by Maria-Luisa Maccecchini, "Large Scale Peptide Production." The use of synthetic peptides as research tools has been greatly facilitated by solid-phase synthetic techniques. Pioneered by Merrienfeld, this approach has allowed short peptide synthesis to become a widely available experimental tool. Recent advances in this and related fields of peptide synthesis are discussed by Anwer and Kahn in "Advances in Peptide Syn-

thesis." It has been appreciated for some time that proteins and peptides in nature are typically modified. Phosphorylation, glycosylation, and many other modifications are commonly encountered. Otvos, and Hollosi, in their chapter "Development of Chemically Modified Peptides" discuss several recently described strategies for developing stable chemically modified peptides which employ naturally occurring modifications. Peptides, while of great utility for many applications, suffer from poor stability when given *in vivo*. This has led to a great deal of interest in the development of more stable peptide analogs or peptidomimetics. In his chapter "Design and Synthesis of Biologically Active Peptide Mimics," Williams discusses some of the more successful approaches taken in recent years in peptidomimetic design. This has led to the development of a variety of pharmaceutical agents currently in use or under evaluation. This section of the book provides an overview of a number of synthetic strategies for peptide and peptidomimetic development.

The use of synthetic peptides has allowed basic insights into a large number of clinically relevant experimental systems. Peptides have been increasingly used in immunology to determine the molecular nature of immune responsiveness. Peptides can serve as antigens for both cellular and humoral immune responses directed toward a number of pathogens. This has been increasingly applied to vaccine development. The basic immunology of pathogen-derived peptides is discussed by Levy and Weiner in "Synthetic Peptide-Based Vaccines and Antiviral Agents, Including HIV/AIDS as a Model System." A related chapter, discussing the rewards and pitfalls of using peptides as immunogens, is provided by Heber-Katz and Ertl. In "Peptides as Molecular Probes of Immune Responses," they discuss the general and specific aspects of peptide immunogens in several viral systems. The recent elucidation of the molecular basis for cellular immune responses has allowed novel approaches for modulating immune responses. Cellular immunity arises from the ability of major histocompatibility complex (MHC) molecules to bind peptide fragments of antigens, and this peptide-MHC complex to then trigger specific T cells via their T cell antigen receptors. One approach for modulating immune responses centers on developing high-affinity peptide ligands for MHC molecules. This approach is discussed by Sette, Lamont, and Grey in "The Design of MHC Binding Peptides." This highlights the ability of peptides to serve as ligands for MHC molecules, influencing the nature and extent of the interaction of the peptide-MHC complex with T cell receptors. One of the most renowned successes of peptides as molecular probes lies in the field of cellular adhesion molecules. The famous RGD (Arg-Gly-Asp) motif of many cellular adhesion molecules has been exploited by the development of synthetic peptide analogs. This, and many less well-known adhesion-related peptides, are discussed by Robey in "Biology and Chem-

istry of Extracellular Matrix Cell Attachment Peptides." Peptides can serve as substitute substrates for enzymes, or as enzyme inhibitors. This can affect a wide variety of physiologic and pathophysiologic functions. Grant, Meek, Metcalf, and Petteway discuss the development of such peptides in "Design of Peptide Analog Inhibitors of Proteolytic Processes."

While many subjects related to biologically active peptides are not addressed in this book, the purpose was not to exhaustively cover the field. Instead, this book highlights some of the basic and advanced aspects of peptide design, synthesis, and utilization. The reader should come away with an appreciation for peptide development strategies that have been successfully employed by a number of investigators. While all of these may not apply to a particular experimental system, it is likely that aspects of the strategies discussed herein should have relevance to any given experimental system employing biologically active peptides.

VOLUME 1

Biologically Active Peptides: Design, Synthesis, and Utilization

SECTION I

STRUCTURAL APPROACHES TO PEPTIDE DESIGN

DOI: 10.1201/9781003419853-1

CHAPTER 1

Strategies in Protein-Based Rational Drug Design

THOMAS KIEBER-EMMONS[1,2]

INTRODUCTION

Of the strategies being considered for disease management, the use of native proteins as templates to develop therapeutic agents continues to be touted as one of the most powerful. This notion has grown from identification of many peptides with potent stimulatory and inhibitory effects on cell proliferation in the past ten years. Growth factors represented by interleukins are a good example. These polypeptides were first defined as signalling molecules controlling activities of cells in the immune system. However, it has now been shown that many of these peptides are multifunctional in that any one peptide in this set may exhibit proliferative effects, anti-proliferative effects, and effects unrelated to proliferation [1]. An appreciation of the ligand-receptor interactions underlying these functionalities can lead to a variety of synthetic chemical agents for medical applications.

More recently, protein-based drug design that uses the shape of pharmacologically active antibodies and other proteins to create new active drugs is increasing in popularity (see Chapter 3). Recent advances in genetic engineering coupled with immunoglobulin (Ig) sequence analysis and x-ray crystallography has made it possible to contemplate designing or building functionally better antibodies [2] for example. The design and synthesis of single-chain antigen binding proteins or polypeptides are expected to have advantages in some clinical applications because of their relatively small

[1]The Wistar Institute of Anatomy and Biology, 3601 Spruce Street, Philadelphia, PA 19104, U.S.A.
[2]Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

DOI: 10.1201/9781003419853-2

size and fast serum clearance [3]. These molecules might have primary uses that parallel those applications for which monoclonal antibodies are in current use such as imaging, cancer therapy, and vaccine development.

Biological activities residing in the Fc region of Igs has contributed to the notion of using peptides derived from Igs to modulate immune responses [4]. The structural analysis of antibodies which are mimics of receptors and ligands, provides a further avenue for the development of novel drugs [5–7]. Conformational analysis of bioactive peptides can lead to suggestions for synthetic structures that lock in particular conformations that may enhance peptide efficacy [8–13]. Such ideas are consistent with medicinal chemistry practices spanning the last 20 years, but appear to be just discovered by immunologists and biologists alike [7]. From a structural standpoint, protein-based drug design may be viewed as a continuum in geometry, with the objective of mimicking the spatial positioning of reactive functional groups in large molecules by those of small molecules.

A multidisciplinary approach including knowledge-based systems involving crystallographic and sequence data bases, experimentally determined constraints and molecular modeling, provides a viable strategy for studying structure/function relationships necessary for protein-based drug design. In this chapter various strategies commonly used in molecular modeling and conformational analysis are briefly reviewed as an introduction to those uninitiated in the art. Practical aspects of the theory and practice of conformational energy analysis is first explored, followed by an overview of energy minimization strategies and design applications and approaches. Emphasis is placed upon structural design considerations and approaches for targeting effector molecules and cells, drawing upon available evidence for the value of antibodies and synthetic peptides in specific diseases.

THEORY AND METHODOLOGY

POTENTIAL ENERGY FUNCTIONS AND MINIMIZATION

Molecular simulation calculations have become a standard approach to describe the conformational properties of macromolecules and to test structural hypotheses of designed molecules. This approach coupled with Nuclear Magnetic Resonance (NMR) has become the *sine qua non* for studying peptide/protein and nucleic acid structure. From a biophysical perspective these combined approaches try to define the rules that determine macromolecular conformational properties. For example can we simulate how a linear amino acid sequence folds into a protein.

At the heart of such approaches is the mathematical description of the classical energy analog of quantum chemical treatments of a complex molecular system as a function of its coordinates [14,15]. In order to describe the classical energy, potential energy functions or force fields were formulated based upon physico-chemical descriptions of molecular interactions. The early molecular energy descriptions included classical energy functions for the electrostatic interactions, van der Waals interactions defined in terms of dispersion and repulsion components, and torsional terms about rotatable dihedral angles. Latter terms accounted for the Hookian behavior of a molecule's internal coordinates defined as bond lengths, bond angles, and dihedral angles. The potential energy of the system, as described by the various components is expressed as an analytical function and the internal coordinates of the molecule and the distances between atoms. The analytical function is also called a target function.

The potential energy of the system as a function of its coordinates describes the multidimensional energy hyperspace of the system [16]. By calculating the energy of the system at a particular set of coordinate values, one explores the multidimensional energy surface. An important method for exploring the energy surface is to find configurations for which the energy is a minimum. This method finds a point in configurational space where all the forces on an atom are balanced. By minimizing the energy of a molecule, stable molecular conformations can be identified. The analysis of minimized structures provides detailed structural information, while the calculated energy can be partitioned into contributions from specific interactions.

In the last thirty years force field descriptions have grown in their complexity. In general, the accuracy required in a force field depends on the properties of the system that are of interest. A force field may reproduce structural information compared to crystal structures very well, although the calculated energies may not be accurate enough for quantitative energy descriptions in terms of enthalpies. Several force fields are routinely used in molecular energy calculations. These include ECEPP [17], AMBER [18], CHARMM [19], CVFF [20] and MM2/3 [21]. ECEPP was one of the earliest force fields described for analysis of peptides and proteins [17]. The ECEPP force field did not include terms for evaluating strain energies resulting from bond stretching and angle deformations. The function was well parameterized based upon the physico-chemical information. A vast literature is available on the parameterization process of the force field by Scheraga and co-workers [17]. Like ECEPP, the MM2/3 force field is also well parameterized. This force field has been mainly used in conformational energy calculations of hydrocarbons. AMBER, CHARMM and CVFF force fields have become widely used due to the commercialization of software packages that contain these and derivatized versions of these

force fields. CHARMM and CVFF have been principally used for peptide/protein studies, with AMBER also used in nucleic acid studies.

The development of potential energy functions is a discipline in its own right emphasizing the theoretical basis of force fields, their parameters, and the parameterization process. There are many limitations in the use of energy calculations and the work is incomplete. However, strategies involving conformational energy calculations can provide valuable information in probing conformations with the aim of developing novel molecules. There is a wealth of software and methods available for computer-assisted drug design (recently reviewed by Cohen et al. [22]) that is either the same or complementary to protein-based drug design.

DESIGN STRATEGIES USING MINIMIZATION AND DYNAMICS

At the heart of protein-based drug design is the description of the conformational properties of a protein or bioactive peptide. The essential difficulty in using potential energy approaches to study conformational properties is the multiple-minimum problem [16]. It is conceptually well noted that if a minimization is started from a conformation chosen at random, the probability of reaching its global minimum is essentially zero. Subsequently many computational chemists would agree that free space minimization, the process of starting with a random geometry and minimizing its free energy, would not necessarily lead to structures that would otherwise be observed by crystallography. However, this view is not universally shared (see Chapter 2). There are many applications in which the biasing of a target function is appropriate to test whether a particular conformation can be populated or to impose constraints on the molecular system so as to influence the energy pathway during the computing of conformational states. A target function is a term used to describe the function for which a minimum value is sought. Basically it is a customization of the function to address a specific modeling question. Such customization takes the form of constraints and restraints.

A constraint is the fixing of atoms in space. In examining the loop conformation of turn regions in antibodies for example, one popular constraint is to fix atoms at the base of the loop [23]. This fixing of atoms can be used to simulate the affect on the tertiary environment in preserving the beginning and the end of the turn region. In contrast, restraints can take many forms. Torsional restraints bias the harmonic torque about a bond to force it to adopt a new value. A good example of using such an approach is to study energy pathways across a barrier [24]. Another example of a restraint is the use of distances between atoms derived from NMR experiments or coupling constants used as torsional restraints [25].

One of the pioneering efforts in modifying potential functions is the strategy of template forcing [24]. In this approach one molecule is forced

to adopt the conformation of another. This approach is useful in evaluating whether a particular peptide can adopt a binding mode conformation of a target peptide. This information is particularly useful in drug design applications. Atom pairs between the molecule under study are considered as flexible and the target molecule is assigned a Hookian type potential function, pulling the atoms in the flexible molecule toward the target molecule. The energy required to force the study molecule into the required conformation can be used to evaluate how easily an analog can adopt the conformation of a given template.

While strategies using energy minimization can lead to meaningful results, molecular dynamics approaches provide a means to simulate the motion of atoms. This approach allows for the study of the fluctuation in the relative positions of atoms in a macromolecule as a function of time. Molecular dynamics has been used to elucidate structures from NMR experiments, to refine x-ray crystallographic structures or molecular models from poor starting structures, and to calculate the free energy change resulting from mutations in proteins [26]. Molecular dynamics provides information on the accessibility of conformational states available to a molecule.

A combination of molecular dynamics simulations and minimization is useful to search for structural features that may be significant for the binding of peptides to receptors. Both low temperature (room temperature 300 K) and high temperature (900 K) dynamics is typically employed. High temperature dynamics greatly increases the efficiency of producing conformational transitions, however there are structural caveats that must be considered in such calculations. For example the omega or peptide bond typically deviates from its normal values under such conditions. Therefore caution is advised to the novice user in performing such calculations to be sure that there is geometrical consistency. In peptide design applications a good prototype example is provided by the identification of constrained analogs of gonotropin releasing hormone (GnRH) as antagonists using a combined strategy incorporating energy minimization, template forcing, molecular dynamics, and NMR information [10,27]. Again, numerous treatments have been presented on molecular dynamic simulations in biology [27–29].

While the solution conformation(s) of peptides is accessible for examination by physical techniques, interpretation of such observations is often difficult. Most peptides of interest like peptide hormones and neurotransmitters are small linear peptides and thereby have a considerable conformational flexibility in solution. Furthermore, the suggested conformer(s) can be dramatically affected by the environment [30]. Subsequently structures observed by physical techniques such as NMR or Circular Dichromism (CD) are relevant only to the condition under study—the physical properties of the solvent.

In contrast, theoretical calculations are typically performed *in vacuo*, where the influence of the solvent on the conformation(s) is largely ig-

nored. The aim of these studies is nonetheless, to gain insight into the structure of a given peptide in a very particular milieu – that of the receptor where solvent is typically excluded. Subsequently such *in vacuo* type calculations are deemed as being representative of possible ligand binding modes for receptor binding. Nevertheless, the question still arises whether the conformation(s) determined in solutions or *in vacuo* have any relevance to the conformation of the peptide at the receptor in the bound state or even in the close vicinity of the receptor. However, in spite of these conceptual difficulties and the inherent limitations of available physical and theoretical methods, they could, particularly in conjunction with chemical modifications of the peptides, help to gain insight into the complex phenomena of ligand-receptor interaction. This information can then be further translated into testable synthetic compounds for evaluation that may lead to therapeutics.

COMPARATIVE MODELING TO DEFINE PROTEIN AND PEPTIDE STRUCTURES

The prediction of protein and peptide structure has a rich history. From the early work of Anfinsen showing that the inherent information for protein folding lies in the amino acid sequence of a protein [31], a large effort has been directed toward developing methods to predict structure [32]. The ability to predict structure has evolved to a state of heavy reliance on the use of data bases. The prediction of tertiary structure of proteins is cast in a knowledge-based approach depending on the identification of analogies in secondary structures, motifs, domains, or ligand interactions between a protein being modeled with structures already available [32,33]. The concept has evolved from early work on the prediction of protein structure based upon distance geometry relationships established from crystallographic data bases [34–38]. This approach is a topic called comparative model building or modeling by homology and discussed in various forms over the years [39–42].

The first step in this approach is to search structural data bases looking for sequence similarities between the protein or fragment being modeled with data base members. The prediction of an unknown protein structure by comparative modeling requires the correct correspondence between residues. Sequence alignments of proteins have become a standard approach to establishing both the family identity of a protein and as a method to define structural elements to be used in comparative model building ([43] and references therein). The approach however is not without caveats that can mislead the structural elucidation effort. Procedures such as those of Needleman and Wunsch [44] or Smith and Waterman [45], with scoring parameters expressed as mutation data matrix, correctly place proteins

into phylogenetic families when the proteins are closely related. It has been long recognized that for distantly related proteins these procedures are not as accurate. One of the reasons attributed to this problem is that sequence alignment methods typically apply a gap penalty uniformly throughout a sequence [46]. When sequence alignments are compared with alignments based upon three-dimensional structures, insertion and deletions identified by sequence alignments alone have been observed to occur in the middle of structural elements that are important for protein structural organization. Sequence alignments with uniform gap penalties can change the pattern of residue relationships that affect packing organization, destroying the complementarity of occluded surfaces.

The use of variable gap penalties to improve the alignment of amino acid sequences of distantly related proteins has been proposed [46,47]. Such an approach to analyze relationships of distantly related proteins depends on the availability of three-dimensional structural information of family members. Three-dimensional information can be used to redistribute gap penalties, increasing penalties for insertions and deletions within definable secondary structural elements or motifs that are highly conserved among a protein family [48]. Family members can be of diverse structural nature. The Ig super-family is such a set of proteins with members that differ appreciably in the number of α helices and β strands. In addition, regions of secondary structure in related proteins do not always retain the same boundaries.

With the ever burgeoning number of available sequences, attention has been turned to the problem of multiple sequence alignment. The most common approach is based upon iterative pair wise alignment [49]. This method rests on linking a representation of a set of multiple aligned sequences as a profile, where the alignment column comprises the frequency of the amino acids at a given position with scoring based on this distribution. This approach is often relied upon to detect distantly related sequences. In general this approach has numerous caveats also associated with it [43,47–49].

Recently much effort has been directed to identifying structural relationships that may be utilized in relational data base schemes. Many approaches rely on pattern analysis of one form or another, such as neural network models [50], dot matrix methods [43], distance geometry [51], etc. Most of these approaches rely on searching for sequences that are similar to the sequence of a protein whose structure is known. This strategy works well for closely related structures but structural similarities can go undetected as the level of sequence identity drops. Alternative approaches have been presented that relate rules of 1D sequence information with 3D structure, by considering that nature preserves structure and not sequences [52]. Proteins that fold into similar structures can have large differences in their sizes and shapes of residues at equivalent residue positions.

The alignment of sequences based upon structural elements that are conserved within a family can be extended to include structural elements that are conserved across protein families. One such approach has been referred to as profile analysis [53]. Profile analysis can be used as a method for detecting secondary structure tendencies of protein regions. Such an approach could follow the general method outlined by Gribskov et al. [53] but reflects a knowledge-based strategy to identify the local structural order of a protein. This use of the profile method infers information on three-dimensional structure from primary sequence. The structural information can be contained in a database of structural motifs developed by considering differential geometry relationships among proteins containing the motifs [54,55]. The sequences of the geometrically similar motifs can then be aligned in order to establish a sequence profile that relates the motifs. Creation of a set of profiles for a variety of super-secondary structures offers a library of structural motifs that can be searched by typing in a sequence. Sequences that define a motif are related by assessing an evolutionary scoring matrix that considers the conservation of amino acids critical to maintaining the folding pattern of the motif.

Comparison of protein sequences with the library can yield information on structural motifs within a given protein. This approach is different from considering a sequence alignment of one member of a protein family with a given protein. Multiple sequences intrinsic to a particular fold define the alignment. Extensions of the multiple sequence alignment problem has been discussed [47,48]. Since there are usually insertions and deletions between sequences, it is difficult to match the backbone conformation of the unknown protein exactly with those of the known proteins and considerable care is required in order to correctly model the conformations of inserted loops as well as side chain conformations [56–61].

ANTIBODIES AS BIOSYNTHETIC TEMPLATES

A milieu of receptors have now been cloned that are potential targets for the development of pharmaceutics. Among the pharmaceutics are antibodies and compounds based upon the structural properties of antibodies. Antibodies serve as a well-characterized example of comparative modeling approaches with application to pharmaceutic and peptide design.

The therapeutic applications of Igs is continually increasing. However, a major limitation in the clinical use of rodent hybridomas in human therapy is the anti-globulin response during therapy. In addition to the anti-mouse response, another potential problem with these antibodies is the anti-idiotype (Id) response, blocking the binding of the antibodies to the target effector molecule. A partial solution to this problem is the con-

struction of chimeric antibodies. Historically, such chimerics evolved by replacing the rodent constant domain with a human derived constant region [62]. This ability allowed for isotype tailoring to participate in antibody-dependent cellular cytotoxicity (ADCC) and complement-dependent cytotoxicity. However such antibodies can still induce anti-globulin affects that are directed toward the variable domain of antibodies [63].

The next level of designing chimerics involved transplanting complementary determining regions (CDRs) of rodent antibodies onto an antibody scaffold of human origin [64]. An example of this approach is most noted by the designing of CAMPATH-1 [65]. More recently this approach has been shown effective in the "humanizing" of a murine antibody to the interleukin-2 (IL-2) receptor as a potential immunosuppressive [66], a murine monoclonal directed toward the V3 domain of HIV-1 as a possible therapeutic [67], and a murine anti-CD4 antibody to be used as an immunosuppressive agent [68]. In all these studies three basic concepts were employed: (1) the human Frameworks (FRs) selected as the basis for designing the reshaped human antibodies should be as homologous as possible to the FRs of the rodent antibody; (2) the amino acid residues of the rodent antibody that are members of the canonical structures for CDR loops should be conserved; and (3) additional amino acid residues in the FRs that show nonconservative changes between the rodent and human sequences, particularly those in structurally relevant positions need to be examined for influence of antibody packing or combining site stabilization.

Ig TEMPLATE PROPERTIES

The structure of Igs is a topic much studied. Salient themes derived from these studies include the identification of residue side chains that are conserved in various regions of the V_L and the V_H primary structures as a means to preserve the geometry of the V_L-V_H interface [69], structural variability as it relates to sequence diversity [70,71], and the effect of packing of various side chains has on the conformational properties of CDR regions [72] and how in turn this affects antigen specificity. The variation in CDR topography which is dictated by the lengths of turn regions as well as the amino acid diversity of Igs [73,74], is fundamental to the antigen specificity of Igs. Proteins that express the Ig folding behavior may have different lengths associated with turn regions and exhibit a broader spectrum of amino acid variation within its sequence compared with other family members. Comparative modeling approaches are based upon defining structural elements that are transferable from members of a protein family to the unknown protein.

In comparative modeling approaches we first define structural elements that are conserved between members of the family. These structurally con-

served regions define the basic template from which to build a structural perspective of the unknown protein. In Igs, the simplest structurally conserved regions are β strands that make up the scaffold. However, proteins can also be dissected into consensus micro-structures consisting of only several residues that are structurally discernable or dissected into consensus macro-structures which may span several secondary structural units. In a broader context consensus alignments can be utilized to engineer human antibodies from murine antibodies. For example, a generic template for antibodies can be generated by considering the average spatial position of Cα backbone atoms under energy minimization constraints of superimposed structures. For β structures the spatial deviations are not large. We have recently extended this approach to develop consensus structures for turn regions including CDR domains.

In Figure 1.1 an illustrative example of sequence alignments involving a micro-definition of β type structures found in both heavy and light chains of Igs is presented. These regions are grouped based upon a least squares fitting procedure showing a minimum in the root-mean-square (RMS) deviations among the antibody crystal structures between heavy and light chains. Least squares fitting (superpositioning) is used as a measure to determine how close two structures are. Cα crystal coordinates of respective structures are spatially positioned with respect to each other to determine whether the structures are superimposable. The lower the RMS the more alike are the two structures. The grouping in Figure 1.1 separates intervening secondary structural units such as CDR regions and Id determining regions (IDR) [70] or turn regions in the FR portion of the variable domains. The definition of consensus β structures define the beginnings and ends of loop or turn regions that may be of different lengths within antibodies and in molecules that belong to the Ig super-family.

An example of discordance between sequence alignment and structural alignment is observed in the consensus structure labeled A in Figure 1.1. In the comparison of 1REI and 1F19 structures which are of the same length, an insertion is observed in the TT tract of F19 V_L compared with the substance P (SP) tract found in REI V_L. Calculation of the RMS value for the homologous SSLSAS residue tract found in these two proteins is 1.8 A, compared with .74 A for the alignment shown in Figure 1.1. It is evident that the length of the turn region connecting structures A and B varies, being compensated by the consensus structures.

There is generally a structural distinction made between antibody heavy and light chains. From a comparative modeling perspective, these chains are structural analogs of each other. Analysis of crystallographically determined light chain dimers has shown that the light chain can adopt structural conformations that are similar to heavy chain conformations [75]. We have included structural consensus alignments in Figure 1.1 for heavy

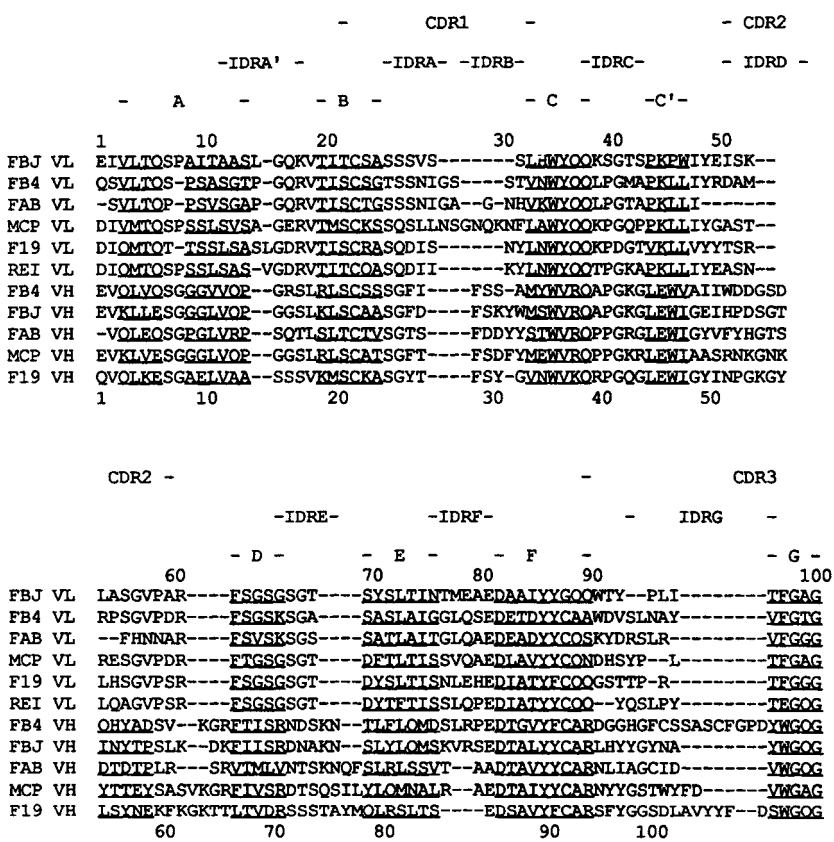

Figure 1.1 Sequence alignments implied by profile analysis and superpositioning of crystallographically known light and heavy chains. The following V_L and V_H sequences were extracted from the respective Brookhaven Protein Databank; KOL (1FB4), NEWM (3FAB), Bence-Jones protein REI (1REI), MCPC603 (2MCP), J539 (1FBJ), RI9.9 (1FI9). Underlined sequence fragments signify conserved β strands among the structures. Top numbering corresponds to established light chain scheme with bottom numbering corresponding to heavy chain scheme. Strands are conventionally labelled A–G and CDR regions illustrated. IDRs are labeled with the prefix IDR. Both CDR and IDR beginnings and ends are not to scale with respect to both heavy and light chains. CDRs and IDRs are variable in length between the chains.

chains with light chains. Such analysis contributes to our understanding of the molecular mechanisms involved in the generation of antibody diversity, extending the rules governing sequence-structure correlations of loop structures, and improving the accuracy of predicted Ig structures.

One of the more important features of examining residue substitutions in these structures is their effect on the conformational properties associated with CDR regions and turn packing [72]. Turns in general exhibit struc-

tural properties that can be both sequence-dependent and -independent [73]. It has been shown that relatively few residues through their packing, hydrogen bonding, or ability to assume unusual backbone angle conformations, are primarily responsible for the main-chain conformations of hypervariable regions [76]. These residues are found to occur within the hypervariable region and conserved β sheet FR [76]. The identification of conserved residues in multiple sequences of Igs suggest that these residues play an important role in preserving structure. It has been suggested that the repertoire of conformations appear to be limited to a relatively small number of discrete structural classes [76]. Chothia and Lesk [76] have referred to these commonly occurring main-chain conformations of the hypervariable region as "canonical structures."

Structural analysis of Igs indicates that FR residues play an important role in determining the conformational properties of hypervariable regions and are directly related to the class of chain [72]. For example, CDR2 loops in heavy chains are closely associated with the size of the residue at position 71. This association is also dependent upon the size of the CDR2 domain (loop length). Extending this original analysis [72], we find that there are 47 Ig sequences [77] that have CDR2 regions with three-residue lengths. Of these sequences there is either a Gly or Asp residue at position 55. This finding can be correlated with the subsequent identification of having an Arg or Lys residue at position 71 in 43 of these sequences and a Val or Leu residue at this position in the remaining four sequences. Based upon what we know about the role of position 71 in affecting the CDR2 conformation [72] the expected canonical conformation [76] for the majority of these three residue CDR2 regions would be that observed for the antibody NEWM and HY-HEL-10.

Continuing this type of analysis we find that there are 194 Ig sequences with CDR2 regions that are four residues in length. Of these there are 35 sequences with Arg or Lys at position 71, Gly, Asn, or Asp at position 54. The canonical conformation is expected to be like that found in KOL. Ninety-nine sequences have Pro at position 52a [77], Gly and in a few cases Asn or Asp at position 55. These sequences have Val or Leu at position 71. The expected conformation for these sequences is that observed for HY-HEL-5. There are 61 sequences with six residue CDR2 regions. All have Tyr at position 55, Arg at 71 and all but two have Gly, Asn, or Asp at position 54. The expected conformation for these CDRs would be that of MCPC603.

These aspects of CDR conformations and the role of packing residues may be transferable to light chains. One of the more notable effects would be the substitution of the highly conserved Gly residue at position 66 observed in human and murine light chain sequences [77]. Lys and Arg residues can be found in this position [77]. However, Gly substitution is

likely to affect the hairpin turn at positions 68-69 and the conformational properties of CDR1 and CDR2 in light chains.

CDR LOOPS AS STRUCTURAL ANALOGUES

Competitive binding assays between analogous peptide ligands can be viewed in the context of identifying similarities in molecular structure between numbers of a congeneric series of drug compounds. Various atom types or functional group sites can be identified at particular relative geometrical positions to establish steric or electronic relationships between the molecular structures.

The molecular mimicry of one peptide by another can hinge on the ability of two analogues to adopt similar conformations, tantamount to illustrations of designing antagonists to peptide hormones [10,24]. It is also possible for two active analogues to have conformations that are relatively dissimilar with their activity the result of a shared configuration of functional reactive moieties. In other words, functional groups on amino acid side chains can be recognized independently of the backbone conformation. This result would be similar to that for active analogues of enkephalin where it was found that for several active conformers the backbone conformations were quite different, yet the spatial orientations of the side chains were similar [78].

Previous studies of the conformational properties of CDR loops indicate that loop conformations can be both sequence-dependent and -independent [73]. Those studies that have emphasized the sequence-independent nature of particular CDR loops have concluded that CDR length and base geometry are primary factors in determining CDR conformation [23,73,79–81]. Analysis of crystal structures of antibodies indicates that regions that are conserved in sequence are conserved in structure. Therefore, an analysis of CDR loops that are of similar length and share amino acid homology reduces to a study of the conformational attributes of structural analogues. However, analysis of CDR loops that are of the same length but lack sequence homology indicates that side-chain orientations are to a large extent preserved.

The superpositioning of CDR loops of equivalent length and varying sequence homology emphasizes the orientational importance of side chains in spite of differences in backbone conformation. This obviously indicates that packing considerations heavily influence the conformation of the backbone, with the conformations limited due to the constraints imposed by the antibody tertiary structure. If there is conservational pressure to preserve packing, and thus side-chain orientations, then the diversity in sequence within a hypervariable segment alone will not generate the topographical diversity required for anti-protein specificity. The increase in

specificity will require CDRs of differing length as first suggested by Rees and co-workers [73], as well as some ability for local mobility within CDRs leading to a notion of inducible complementarity in antibody binding.

The importance of these observations has ramifications in understanding the structural and ultimately the evolutionary arrangement of CDRs within antibodies. An early assertion had been that the beta-barrel FR of antibodies constitutes a beta sheet scaffold onto which binding sites may be built, implying that the structure of CDRs is relatively independent of the FR context. However, this simple notion is not generally the case as cited in developing humanized antibodies. This observation reemphasizes that the packing arrangement of the FR with respect to the CDRs is an important determinator in directing the folding behavior of CDR turn regions.

BIOACTIVE PEPTIDE DESIGN AND ACTIVE SITE STRUCTURE

The identification of peptide forms having profound effects on the immune system has naturally led to strategies using peptides to modulate the immune response. Two early examples are peptides derived from Ig and complement. Peptides derived from Igs may provide a novel source of peptide-based drugs for the treatment of immunologically mediated diseases. The tetrapeptide tuftsin derived from the CH_2 domain of IgG (residues Thr-Lys-Pro-Arg) has a host of associated biological activities [82]. Structure/function and conformational studies of the tuftsin molecule predicted that another peptide, Gly-Gln-Pro-Arg, would mimic tuftsin activity. This peptide, called rigin, was synthesized and found to extend phagocytosis-stimulating activity towards heterologous erythrocytes and bacteria [83]. The immunomodulatory activity of these peptides is unrelated to their native parental form. Monomeric IgG does not stimulate phagocytosis, suggesting that peptides such as tuftsin are enzymatically released before expressing activity [84].

The agonist and antagonist activity of such immunoregulatory hormones may result from binding to a plethora of immunoregulatory receptors. The therapeutic activity of such peptides has been demonstrated *in vivo*. One Ig-derived peptide, human IgE pentapeptide (Asp-Ser-Asp-Pro-Arg) has been shown to have some therapeutic activity in humans [85]. The potential agonist/antagonist relation exhibited by these short peptides may be identified by considering the sequence similarities between peptides. The IgG-derived tuftsin resembles the N-terminus of Substance P (SP) (Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met-NH_2) and has shown the ability to bind to SP receptors [84]. When injected intracerebroventricularly, tuftsin induces analgesia (SP antagonism). In turn, SP has similarly been demonstrated to have tuftsin-like phagocytosis-stimulatory activity.

SP is a neuropeptide that is a well-characterized member of the tachykinin family [86]. Sensory neuropeptide release by a sensory neuron may signal tissue damage in the spinal cord and participate in regulating the inflammation, immune response, and, ultimately, wound healing in the affected peripheral tissue [86]. Conformational analysis of tachykinins was first performed on the SP fragment [87]. Constrained agonists and antagonists of tachykinins have demonstrated the potential utility of cyclic peptides in modeling the interactions between a peptide and its binding site [88–90] as with GnRH [10,27]. Antibodies that mimic SP have been described [91]. The identification of sequence relationships between SP and these antibodies, coupled with comparative modeling and the conformational and analogue studies of SP as previously discussed, can lead to further design of SP and other tachykinin analogues [92,93].

Examples of similarity between Ig sequences and bioactive molecules have been demonstrated. Information derived from anti-receptor antibody sequences or related biologically significant proteins can lead to the development of peptides that bind the active sites of receptors, and possess biological activity. Such relationships suggest that Igs may serve as design templates for more classes of compounds for immunoregulation, thereby serving as potential therapeutic drugs. It is possible to produce or design fully synthetic peptides using essential sequence information obtained from anti-Id hybridomas. The conserved nature of antibody structures demonstrated by x-ray crystallography and amino acid sequence analysis is well suited to the process of comparative molecular model building. Since the structures of the light and heavy chains are known, FR region conformations can be used as constraints in performing energy minimization modeling of hypervariable regions. Therefore, comparative modeling of antibodies provide geometrical information that can be translated into designing therapeutic peptides by considering shape, charge distribution, and chemical functionality. Thus, in an antibody-directed approach to peptide and pharmacophore design, the amino acid sequences of the hypervariable region of antibodies that mimic agonist or antagonist properties can be used as starting points to develop peptide and nonpeptide agents.

ANTIBODY DERIVED PEPTIDES AS PROBES FOR RECEPTORS

In prior work, we have successfully developed receptor binding peptides based on analysis of antibody structure and protein loop structures (see Chapter 3). In one set of studies we utilized an experimental system involving a murine antibody mimic of a reovirus protein antigen that binds to specific receptors on murine neurons and lymphocytes [94–96]. Structural relationships were first established between the antibody and the reovirus antigen [94]. Peptides were developed to encompass reverse turn regions comprising the respective hypervariable loops that showed biolog-

ical activity [95,96]. Implications of these studies include the ability to analyze antibody binding at a molecular level, the observation that antibodies and other ligands may utilize similar binding strategies for a common receptor site, and the ability to develop peptides with defined specificity directly from information on the molecular structure of the antibodies.

More recently, the design of peptide ligands based on antibody structure has progressed. These studies include the ability to apply analysis of the molecular structure of antibodies and structurally related proteins to the development of specific receptor binding peptides, even when these receptors are represented by chemically discrete, small molecules such as sialic acid. Variant peptides lacking specific side-chain hydroxyl groups postulated to be critical for binding to the reovirus receptor were deleted in these peptides, and their ability to interact with the receptor assessed [97]. Varying affinity for the receptor was displayed by these peptides, allowing us to assign a hierarchy for potential intermolecular interactions.

When coupled with prior studies implicating sialic acid as a potential reovirus receptor, it was possible to develop molecular models of the antibody-receptor interaction [97]. The models were useful in the subsequent development of peptides with higher affinity for receptor binding. Cyclic peptide analogs of the peptides were developed based on the molecular models [74]. This allowed determination of the optimal conformation for receptor interaction by the peptides. A cyclic peptide with an optimized conformation was demonstrated to possess $> 40\times$ higher affinity for the receptor than the linear analog. The conformational properties of the cyclic peptide has been further defined by a constrained synthetic mimetic [7].

We have further utilized the notion of using antibody and protein structures as templates to develop antibodies and peptides with specific binding properties and biological activity in studies of the interaction of HIV-1 with specific cellular structures. The significance of these studies is that biologically active peptides were developed in the absence of direct crystallographic information but made use of structural templates to design the active peptides. These templates were chosen based upon sequence relationships with protein substructure data bases. Much like the reovirus system we have shown that regions of the envelope protein of HIV-1 and HIV-2 exhibit sequence homologies and folding properties with members of the Ig gene family [98]. Engineered peptides derived from these regions of the envelope protein were shown to modulate CD4 dependent cellular functions (block virus infectivity and are immuno-suppressive) and the respective anti-peptide antibodies recognize the native envelope protein [99–102]. Conformational calculations of the bioactive peptides were utilized in these studies to correlate peptide conformational properties with peptide biological activities and immunogenicity [99–102].

ANTIBODY DERIVED PEPTIDES AS PROBES OF FUNCTIONAL GROUPS

Monoclonal antibodies have been useful in the identification of reactive functional groups. In the studies of Taub et al. [103], analysis of antibody hypervariable loop structures of an anti-fibrinogen receptor monoclonal antibody referred to as PAC-1, resulted in identification of a sequence (RYDT) similar to the RGDS binding sequence of many adhesion molecules, and a peptide derived from that hypervariable loop sequence inhibited platelet aggregation in response to fibrinogen [103]. Adhesion receptors, principally integrins, play a prominent role in cellular and developmental biology. As a family, integrins promote cell attachment to fibronectin, vitronectin, laminin, collagen, fibrinogen, and von Willebrand factor [104]. The structural domains of integrins have been correlated with ligand binding by cross-linking to peptides containing the sequence Arg-Gly-Asp (RGD), a ligand recognition motif for several but not all integrins [104]. Presumably, recognition is influenced by the dipolar character of this motif.

Recent studies (unpublished) indicate that platelet activation can be achieved with certain monoclonal antibodies reactive with $\beta 3$ component of integrins. These antibodies termed OPG2 and CP3 activated the solubilized receptor, thus precluding a critical role for the membrane microenvironment, while the antibody PAC-1 does not. We have been modeling the antibodies OPG2, CP3, and PAC-1 to determine why they have different functionality in binding to platelets. All three antibodies contain an RYD tract in the CDR3 region of their heavy chain. In our effort to deduce possible conformations of the H3 domain shared by the respective antibodies, we have utilized a comparative modeling procedure previously used to model regions of antibodies and related molecules [74].

To develop possible models for the localized structural folds of the H3 domain we first examined the sequence and folding pattern of H3 regions of crystallographically known antibodies. These comparisons focused on identifying geometric constraints as defined by the spatial positions of the flanking invariant portions of the H3 region. The H3 domain of several antibodies of known crystal structure were least squares superimposed to define these residue positions (Table 1.1). The invariant positions define the amino-terminal beginning and the carboxy-terminal end that are shared among the putative H3 domains of varying lengths. Least squares fitting (superpositioning) of the Cα coordinates (positions) is used in this manner to determine the degree of similarity between the H3 structures. The systematic superpositioning of the H3 domain over short sequences defines a consensus region where the structure is conserved among the antibody templates. This consensus region defines the FR onto which a model can

TABLE 1.1. Comparison of Ig Templates Used to Define Starting Geometries.

X-Ray	N-Terminal	Putative Turn	C-Terminal
1F19	CARSF	YGGSDLAV	YYF-DSWG
2FB4	CARDG	GHGFCSSASCF	GPD-YWG
2MCP	CARNY	YGSTW	YFD-VWG
1FBJ	CARLH	YYGY	N-A-YWG
2FAB	CARNL	IAGC	I-D-VWG
Antibody			
OPG2	CTRHP	FYRYDGGN	YYAMDHWG
CP3	CARGR	NRNRYDGD	YYAMDYWG
PAC-1	CARRS	PSYYRYDGAGP	YYAMDYWG

VH sequences for the crystallographically known Igs were extracted from their respective Brookhaven Protein Databank entries; R19.9 (1F19), MCPC603 (2MCP), J539 (1FBJ), NEWM (3FAB), KOL (2FB4). Dashes in the sequence represent insertions identified by least squares fitting the N- and C-terminal structurally conserved regions. The fitting procedure allows for choosing an appropriate putative turn length for the target antibodies.

be built for the various loops. The procedure then is to search the crystallographic data base for loops of the same size as the putative loop to be modeled.

In this way we have defined the alignments in Table 1.1 to provide information on loop length of crystallographically known Igs compared with those of the antibodies under study. In this context the loop length of PAC-1 is that of the antibody FB4, while OPG2 and CP3 have loop lengths reflective of the F19 structure. The F19 template structure indicates that the alignment for CP3 and OPG2 are not in register, a result that has been previously discussed in comparative modeling procedures. In the search procedure, the spatially conserved Cα positions at the N- and C-terminal regions were held fixed. A Cα distance matrix was constructed for combinations of these positions and compared to precalculated Cα distance matrices made from high resolution protein structures. The 20 best matches were examined and an appropriate choice was made based upon similarities in chiralities of side-chains and at the junctures of the loops, and displayed sequence similarities between alternative loops and the H3 loop of the respective antibodies.

The strategy that we have employed was to test whether conformations for the H3 domain of OPG2 and CP3 could be shared between these antibodies. Our structural analysis indicates that the orientation of the RYD tract in PAC-1 is different from that of OPG2 and CP3 suggesting that the conformational properties of OPG2 and CP3 for the CDR3 region is more similar between these antibodies than that for PAC-1. This would suggest that the similar biological activity shared between OPG2 and CP3 might be directly related to the conformational properties of the CDR3 region of

these two antibodies. It is also possible that the differences in activity between the antibodies is a function of the depth of the RGD binding pocket on GPIIb-IIIa accessible to the antigen binding surface of the antibodies. Recent studies of immobilized RGD peptides of varying lengths indicate that platelet interactions are sensitive to the length of the RGD presenting peptides.

Previous studies on the mimicking of haptens by antibodies suggested that RGD analogs such as KGD and KGGD might mimic the dipolar character of phosphorylcholine (PC) [105,106]. In fact recent analysis of the viper venom Barbourin shows that this integrin specific antagonist contains a KGD tract substituted for the RGD tract found in previous viper venoms [107]. The hapten, PC, is bound with high affinity and specificity by members of the Ig S107 germ line family. The crystal structure of one member of this family, MCPC603, has been well studied with its three-dimensional complex with PC known. Anti-Id antibodies have been developed that mimic the immunogenicity of PC containing formulations [106], that compete with PC for binding to the TEPC-15 myeloma protein. These anti-Id monoclonal antibodies were originally characterized as near-site specific (F6-3) and PC-binding site specific (4C11).

Sequence analysis of F6-3 and 4C11 [106] revealed several amino acid residues which might mimic the three-dimensional structure of PC in binding to S107 family members. The principal dipolar tract involved the sequence EKFKD [106]. Utilizing the information on the three-dimensional structure of the MCPC603-PC complex, *de novo* designed model synthetic peptides were developed of the form KGD and KGGD using molecular graphics and conformational energy analysis [105]. The tetramer KGGD was shown to inhibit 4C11 binding to TEPC-15 over the range of inhibitor used [105]. The tetramer was not however as effective as the hapten based upon the inhibitor molarity. Nevertheless, these studies were the first to indicate the feasibility of developing peptides from antibodies which mimic the structural and binding features of small chemical groups.

These studies were extended to identify peptides derived from anti-PC antibodies that would compete with PC for PC specific binding sites [108]. In these studies a 26 residue peptide derived from TEPC-15 was shown to be an effective inhibitor of PC. This peptide spanned the CDR2 region of TEPC-15 and part of the FR 3 region. This region shares discontinuous homology with human C-reactive protein (CRP) (Table 1.2). The PC-binding region of CRP appears to share an epitope with the mouse T-15 Id as shown by recognition of both the TEPC-15 myeloma protein and CRP by monoclonal antibody to the T-15 Id [109,110]. The TEPC-15 anti-Ids F6-3 and 4C11 have been shown to bind to CRP. The CRP1 derived peptide and the TEPC-15 derived peptide in Table 1.2 both inhibit anti-Id binding

TABLE 1.2. Binding Domain Sequence Similarities.

TEPC15	50	ASRNKANDYTTEYSASVKGRFIVSR
CRP1	56	ATKRQDNEIL
CRP2	83	SASGIVEFWV
β5	303	NEAN----EYTAS
β3	302	GSDN----HYSAS
β1	538	CRKRDNTNEIY
β5	238	RVSRNDRDA
β3	237	SVSRNRDA

Regions of local residue similarity (bold face) between the T-15 Id self-binding region, CRP, and beta subunits. Residue numbering is at the left.

to TEPC-15 [109]. The CRP1 peptide is suggested to bind PC directly [109] and has been shown to bind both fibronectin and laminin presumably through phosphorylated amino acids. F6-3 binds to unconjugated CRP peptide directly while 4C11 apparently binds to conjugated (KLH) CRP peptide [109]. In this context the anti-Ids F6-3 and 4C11 act as surrogates for both fibronectin and laminin. Sequence homology is observed between these antibodies and integrin regions which needs to be examined further in a structural context.

We have shown that the TEPC-15 derived sequence is responsible for TEPC-15 self-binding and participates in defining the T-15 Id [108]. However our work also suggests that the binding properties of the TEPC-15 derived peptide defines a complementary hydropathy relationship required for protein recognition. The CRP1 peptide tract may contain an RGD analog defined as RQD. We have redesigned the T-15 derived 26 mer peptide to perturb the conformational properties of the CDR2 turn region by eliminating the KAN residues (Table 1.2, TEPC-15). This peptide is more effective than the 26 mer in inhibiting T15-PC binding (unpublished). In this redesigned peptide, the primary N-terminal sequence tract is AASRNDYTT. The RNDY tract might in turn be an analog of the RGDY peptide shown to be even more effective than RGD in inhibiting platelet aggregation. While our peptide was designed as a mimic for PC in binding to a PC specific binding site, it might be more representative of an alternative adhesion motif.

In Table 1.2 integrin sequences from $\beta 1$, $\beta 3$, and $\beta 5$ subdomains are listed that exhibit sequence and hydropathic relationships similar to the TEPC-15 and CRP sequence. These sites have not been described in the literature as possible sites for association either to ligands or between respective integrin α and β chains. Initial studies of the conformational properties of the integrin sequences in Table 1.2 suggest that the $\beta 3$ structures are more flexible. From a functional point of view, $\beta 3$ structures are more promiscuous in their association with integrin subdomains.

The molecular basis for the self-binding characteristics of the T-15 Id has been shown to be associated with the charge spacing of a putative turn region involving the CDR2 region of T-15 (Table 1.2). Reverse turn regions of antibodies are typically associated with the immunoreactivity of monoclonal antibodies. This self-binding region has been implicated as a site of cross-reactivity between CRP and T-15 antibodies [110]. Subsets of charge spacings are also observed in suggested DNA binding antibodies [111]. This spacing is also observed in the β chain of T cell receptors (TCRs) associated with IL-2 receptor positive synovial T cells in rheumatoid arthritis patients [111]. This site has been implicated in binding to superantigen, suggesting a role for superantigen mediation in rheumatoid arthritis. In addition the relationships also suggest that the charge spacing may be inherent in the TCR lineage just as they are in antibody structures.

Adhesion motifs may be targets for complementary interactions involving autoantibodies and receptors. Adhesion motifs like that found in laminin may also be mimicked by antibodies. It is possible that the adhesion motif RGD found on a variety of integrins may be mimicked by autoantibodies. The analysis of such antibodies can provide information at several levels including information on how to target such autoantibodies to suppress their occurrence as well as learning the geometrical features required for binding such receptors. The geometrical information can be used to develop agents that block the fibrinogen receptor for example which could be useful in treating hypercoagulatable states and preventing myocardial infarctions. The study of antibody complexes with endothelium components may also lead to understanding how to develop small molecules that might interfere with immune complexes associated with some autoimmune diseases.

CONFORMATIONALLY CONSTRAINED PEPTIDES AND MIMETICS

Searching the protein sequence data base indicates a sampling of sequence homologies of varying degrees between viral, native host proteins and reverse turn regions of antibodies [111]. This type of analysis points out that viral sequences as well as native host proteins are readily encoded in the Ig repertoire. The homologies imply that local regions of sequence homology can be found principally due to the fact that there is a limited number of amino acid residues that form such turns. Turns as a recognition unit, are conserved evolutionarily among proteins of different types. The specificity of the recognition features are modulated by selecting amino acids that can change the conformational features of these turns, as well as selecting the appropriate contact residues.

In general terms, the topography of a binding site of a ligand can be viewed on the imprint of the general three-dimensional structure of the