THE IDENTIFICATION OF RECURRENT TERTIARY MOTIFS BY INTERACTIONS OF PROTEIN SECONDARY STRUCTURE UNITS

A Senior Honors Thesis

by

HAMILTON COURTNEY HODGES

Submitted to the Office of Honors Programs
& Academic Scholarships
Texas A&M University
in partial fulfillment of the requirements of the

UNIVERSITY UNDERGRADUATE
RESEARCH FELLOWS

April 2003

Group: Life Sciences 2

THE IDENTIFICATION OF RECURRENT TERTIARY MOTIFS BY INTERACTIONS OF PROTEIN SECONDARY STRUCTURE UNITS

A Senior Honors Thesis

by

HAMILTON COURTNEY HODGES

Submitted to the Office of Honors Programs
& Academic Scholarships
Texas A&M University
in partial fulfillment of the designation of

UNIVERSITY UNDERGRADUATE
RESEARCH FELLOW

Approved as to style and content by:

Jerry W. Tsai

Edward A. Funkhouser

April 2003

Group: Life Sciences 2

ABSTRACT

The Identification of Recurrent Tertiary Motifs

by Interactions of Protein

Secondary Structure Units. (April 2003)

Hamilton Courtney Hodges
Department of Biochemistry & Biophysics
Texas A&M University

Fellows Advisor: Dr. Jerry W. Tsai
Department of Biochemistry & Biophysics

Proteins are the molecular machines that drive the processes of the cell; they carry out the functional and structural instructions outlined in an organism's genome. At their simplest, these biological catalysts are comprised of linear chains of amino acids that fold into unique three-dimensional structures. One of the goals of structural biology is to predict a protein's three-dimensional structure from its amino acid sequence. One important aspect of protein structure is the manner by which the non-covalent or weak interactions bring about a protein's fold. Often called tertiary interactions, these non-covalent interactions are often between amino acid residues that are distant in the linear sequence but close in three-dimensional space. Through an informatics analysis of recurrent tertiary contacts, we have derived a database of recurrent tertiary motifs. A group of 691 high-resolution, non-redundant protein structures was obtained. For each protein in this source data, we found all secondary structure units: alpha helices, beta strands, beta hairpins, and loops. We also identified three physical interactions between the secondary structure units: (1) hydrogen bonds were found by a continuous energy potential; (2) salt bridges were determined by a distance cutoff between oppositely charged atoms; and (3) hydrophobic contacts were derived from Voronoi polyhedra around carbon atoms. From the interactions between secondary structures, we identified the 21,100 protein substructures defined by tertiary interactions. These pieces of proteins were then clustered based on structural similarity into 4,039 groups. Each group represents a tertiary motif. Such a high number of recurrent contact pairs from a non-redundant sample source suggests that there is at least some level of redundancy for these non-covalent tertiary interactions. Applications for this tertiary motif database are currently being developed, with special interest in tertiary structure prediction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Protein structure is often thought of as a hierarchical system – one in which every level in the hierarchy is regulated by the chemistry and physics of the protein's amino acid sequence. Moving up these levels, it is relatively easy to see how local chemistry dictates local structures like α-helices and β-hairpins: hydrophobic collapse and main-chain hydrogen bonds bring about these secondary structures in cooperative thermodynamic steps. Characterizing the global fold or topology of a protein, however, is a much more complicated matter and is the subject of a great number of inquiries. The sheer complexity of predicting protein tertiary structure is due to the vast number of physical interactions that give rise to a given fold. Many aim to understand how proteins fold into their characteristic tertiary structures, and any insight into this complex biophysical problem would be of legitimate scientific value.

To better understand the predominant forces and principles in protein folding, a number of groups have adopted computational and informatics tools. By employing large data sets, researchers can analyze natural trends and evaluate fundamental hypotheses that would otherwise be difficult to examine. In the case of proteins, a few tools already exist to examine secondary structure and to a lesser extent, tertiary structure.

In 1983, Kabsch and Sander developed DSSP, which defines secondary structure for

---

This thesis follows the style and format of *Protein Science*.

each residue in a solved structure [1]. DSSP relies on definitions of secondary structure, which are based on hydrogen bonding patterns and torsional ϕ/ψ angles. These secondary structure assignments allow one to consistently define secondary structure across all classes of proteins. More recently, Gail Hutchinson and Janet Thornton incremented the usefulness of structure classification by including code for motif detection in their PROMOTIF utility [2]. Super secondary elements and other features like β-hairpins, β-bulges, α-β-α and Greek Key motifs are explicitly defined by PROMOTIF, and these can be obtained from a solved protein structure along with the same secondary structure definitions provided by DSSP.

Despite the amount of work completed thus far for the characterization of secondary structure, there exists a relative dearth of information about the organization of tertiary structure. This is not to say that tertiary structure is of no interest; on the contrary, the level of interest in tertiary structure is manifest by the biennial Critical Assessment of Structure Prediction (CASP, *http://predictioncenter.llnl.gov/*). In this assessment, experimenters are provided the sequences to proteins whose structures have yet to be released. These experimenters use tried as well as novel methods to predict the three-dimensional structures for these target sequences.

In the fourth CASP, David Baker and his group performed quite handily by employing a Monte Carlo-based fragment buildup routine called "Rosetta" [3]. Motivated perhaps by this method's success, a few groups have set about trying to obtain a minimum fragment

set necessary to describe backbone tertiary structure. For example, the Rosetta fragment set is a clustered set of 9mer fragments with an adjoining library of small 3mers for backbone refinements. Kolodny, et al, have also developed a similar library and have shown that these types of fragment sets are sufficiently diverse to describe the backbone topology of most proteins [4]. One of the problems associated with the fragment-based methods is the complexity cost associated with building up structures from shorter fragments [5]. Some have proposed to reduce this complexity by studying larger, super-secondary motifs instead of shorter fragments [6, 7, 8, 9]. But as the residue length of the fragments increases, the fragment set needed to describe known protein structures increases beyond what is useful. Others have therefore chosen to focus on the methods used to cluster these libraries to improve the selection of diverse fragments [5]. All of this overlooks a constant criticism of fragment-based tertiary structure prediction schemes: the fragments are only defined within a local, sequential scope. For this reason, it is virtually impossible for fragment-based methods to cope with explicit side-chain packing with residues more distant in sequence space. As evident from the recent CASP 5 novel fold and comparative modeling predictions [unpublished], this is the current bottleneck for the field.

The use of local fragments also frustrates many for a more philosophical reason: it fails to answer any biophysical questions. Recent work on biologically relevant fragments illustrates this understanding. Voigt, et al, in their work with hybrid β-lactamases, find that there are units of protein structure that are untouched by recombination events [10].

Unfortunately, these results are hardly useful for structure prediction because of the limited scope of their sample set. There is growing awareness that approaches that maintain a more physical view of protein structure will be better suited for computational studies. One suggested approach would take into account the interactions of secondary structure elements that give rise to topology [11]. Already there have been some attempts to categorize on a gross level all the possible β-strand pairing configurations [12, 13], and individual efforts to analyze helix:helix angle preferences [14], but an exhaustive catalogue of all possible secondary structure contact motifs has not yet been created.

Simple mathematical as well as all-atom models both suggest that a fundamental step in the folding process is the coupling of local and non-local interactions [15, 16]. Onuchic's work suggests that those local interactions that give rise to pockets of secondary structure early in the folding pathway are critical, but in order to bring about a stable tertiary fold, these must coincide with favorable non-local interactions that bring the secondary structural units together. Furthermore, it is assumed that divergent evolution would tend to stabilize the residues that form these non-covalent interactions. Any non-conservative mutation that destroyed a particular intramolecular contact would destabilize the fold by reducing the peptide's structural rigidity, thereby increasing its topological frustration.

For this reason, it is appropriate to inquire into the arrangement of protein structure at

these sites bridging secondary structure elements. In this study, I attempt to show that there are conserved sites of interactions between secondary structure units and that the side-chain packing arrangements at these points are critical for understanding the thermodynamic stability of natural proteins. Furthermore, a fragment library is created in this study, which will allow for an informatics analysis of these non-covalent "hinge" contacts.

MATERIALS AND METHODS

Source Data

In order to identify recurring tertiary motifs present in the PDB without overweighting a particular protein family or topology, we used Dunbrack's high-resolution subset of non-redundant proteins, culled-pdb (what is now called PISCES) [17]. Only crystallographic structures with resolutions better than 1.8 Å and sequence identity of less than 20% were chosen for this study. This set was chosen so as to limit the amount of redundancy in sequence space. Any peptides with chain breaks were rejected to simplify computation. In all, 691 protein structures were selected for analysis; a list of these structures is given in Table 1.

Computer Resources

The present study was run on a PC with dual 1 GHz Intel Pentium III CPUs running Red Hat Linux 7.0. The software that we developed employs the C library used by Gerstein in his earlier work [18].

Table 1. Source Protein Structures

1 19l	1bm8	1d0d	1e4m	1faz	1g8q	1hbn	1iat	1jf2	
16pk	1bn7	1d1q	1e58	1fcq	1g9o	1hd2	1id0	1jf8	
1a12	1bup	1d2s	1e5k	1fcy	1g9z	1hdh	1ido	1jfb	
1a3a	1bx4	1d2v	1e5m	1fe6	1ga6	1hdo	1ifc	1jfx	
1a4i	1bx7	1d4o	1e6u	1fg7	1gad	1heu	1igq	1jg1	
1a62	1bxa	1d4x	1e7l	1fgl	1gbg	1hfe	1iho	1jg8	
1a6m	1bxo	1d5n	1e85	1fgy	1gbs	1hg7	1ihr	1jh6	
1a73	1byi	1d5t	1eb6	1fi2	1gci	1hlr	1ii5	1jhd	
1a8d	1byq	1d7p	1edm	1fiu	1gcq	1hp1	1iib	1jhf	
1a8e	1c0p	1d8w	1eex	1fjj	1gcu	1hq1	1ij2	1jhg	
1a8o	1c1k	1dbf	1eg9	1fk5	1gd0	1hqk	1ijq	1jhj	
1a9x	1c11	1dbo	1egw	1flm	1gj7	1hqs	1ijv	1jid	
1aba	1c24	1dc1	1ej8	1flt	1gk8	1hsr	1ijy	1jiw	
1afw	1c3p	1dci	1ejg	1fm0	1gk9	1htr	1ikh	1jjt	
1agj	1c3w	1dcs	1elk	1fn8	1gkl	1hty	1ikp	1jjy	
1ah7	1c4q	1df4	1elu	1fn9	1gkm	1hvb	1ikt	1jk3	
1aho	1c52	1dfm	1elw	1fna	1gmi	1hw1	1im5	1jke	
1aie	1c5e	1dg6	1en2	1fo8	1gmu	1hx0	1in4	1jks	
1ajj	1c75	1dgf	1eon	1fp2	1gmx	1hx6	1inl	1jkx	
1ajs	1c7k	1dgw	1ep0	1fpo	1gnl	1hxi	1ioo	1jl0	
1aoh	1c8c	1din	1epx	1fqt	1gnu	1hxn	1iq5	1jl1	
1aop	1c9o	1dj0	1eqj	1fs1	1go3	1hxr	1iqz	1jm0	
1aqu	1cc8	1dk0	1eqo	1fs5	1gp0	1hyo	1irq	1jmk	
1aqz	1ccw	1dl2	1erz	1fs7	1gp6	1hyp	1isu	1jni	
1arb	1ccz	1dlf	1es9	1fsg	1gpe	1hz4	1itx	1jp3	
1atg	1cex	1dlw	1et1	1ft5	1gpi	1hzt	1iu8	1jp4	
1atl	1cg5	1dmg	1eu1	1fvg	1gtv	1i0d	1iua	1jqc	
1avw	1chd	1dnl	1euj	1fvk	1gut	1i0h	1ixh	1jr8	
1axn	1cip	1dos	1euv	1fw9	1gvp	1i0r	1j77	1jsr	
1ayl	1cjc	1dow	1evl	1fx2	1gx1	1i0v	1j79	1jtg	
1ayx	1cmc	1doz	1evy	1fxm	1gx5	1i12	1j7x	1juh	
1azo	1cnv	1dp7	1ew4	1fye	1h2r	1i19	1j83	1jw9	
1b0u	1cq4	1dpj	1ewf	1fzk	1h4g	1i1j	1j8r	1jx6	
1b2p	1cqm	1dps	1eyh	1g2b	1h4r	1i27	1j8u	1jy1	
1b3a	1cru	1dqe	1ezm	1g2r	1h4x	1i2h	1j96	1jy2	
1b6a	1cs1	1dqz	1ezw	1g2y	1h5q	1i2t	1j98	1jya	
1b8z	1cse	1ds1	1f0j	1g3p	1h5u	1i40	1j9b	1jye	
1b9w	1csh	1dsz	1f1e	1g4i	1h61	1i4f	1j9e	1jyh	
1bb1	1ctj	1dtd	1f2t	1g4y	1h6f	1i4u	1jak	1jyk	
1bbz	1ctq	1dvj	1f3u	1g55	1h6h	1i52	1jat	1jz8	
1bd0	1cv8	1dwk	1f46	1g5a	1h6u	1i5g	1jay	1jzg	
1bdo	1cxq	1dxg	1f5n	1g5t	1h70	1i60	1jb3	1k0i	
1bcb	1cy5	1dy5	1f5w	1g60	1h72	1i6w	1jb9	1k0m	
1beh	1cy9	1dyp	1f60	1g61	1h75	1i71	1jbe	1k20	
1bfg	1cyo	1dzk	1f7d	1g66	1h7n	1i88	1jcl	1k2y	
1bgc	1czf	1dzo	1f7l	1g6s	1h80	1i8f	1jd0	1k3i	
1bgf	1czp	1e29	1f86	1g6u	1h8d	1i8o	1je0	1k4g	
1bkb	1d02	1e2k	1f8e	1g6x	1h8u	1i9s	1jek	1k4i	
1bkf	1d06	1e30	1f94	1g7a	1h97	1i9z	1jer	1k4v	
1bkr	1d0c	1e4c	1f9z	1g8e	1h99	1iab	1jet	1k51	

Table 1 Continued.

1k55	1lmb	1qcx	1slu	2erl	8abp
1k6f	1ln4	1qcz	1sml	2fcb	
1k6w	1lo7	1qd1	1svf	2fdn	
1k6x	1lpl	1qdd	1swu	2hft	
1k75	1lri	1qe3	1t1d	2igd	
1k7c	1m6p	1qfm	1tca	2ilk	
1k92	1mfa	1qft	1tfe	2lis	
1k94	1mfm	1qge	1thf	2mcm	
1ka1	1mgt	1qgi	1thv	2mhr	
1kaf	1mla	1qgv	1thx	2nac	
1kbq	1mml	1qgw	1tif	2nlr	
1kcq	1mof	1qh4	1tml	2pth	
1kgd	1mol	1qh5	1toa	2pvb	
1khc	1mpg	1qh8	1tvx	2rmc	
1khx	1mrj	1qhv	1tx4	2sga	
1kic	1mrp	1qj4	1tyv	2sic	
1kid	1msk	1qj5	1ubi	2sns	
1kk1	1mty	1qjc	1ugi	2spc	
1kko	1mug	1qjp	1unk	2tgi	
1koe	1mun	1qkr	1uro	2tps	
1kp6	1mwp	1ql0	1ute	2vhb	
1kpf	1nbc	1qlw	1utg	3bam	
1kpt	1nfp	1qmv	1vcc	3cao	
1kq3	1nkd	1qna	1vfy	3chb	
1kqf	1nkr	1qnf	1vhh	3cla	
1kqr	1nls	1qnr	1vie	3cyr	
1kr7	1nox	1qop	1vns	3eip	
1ks9	1npk	1qq5	1vsr	3ezm	
1ksh	1nps	1qq9	1wap	3grs	
1ktg	1nul	1qqf	1wer	3hts	
1ktp	1nxb	1qqq	1wfb	3lzt	
1ku3	1oaa	1qre	1whi	3nul	
1kv5	1opd	1qs1	1yge	3pnp	
1kv7	1or3	1qst	1zin	3pro	
1kve	1orc	1qtn	256b	3pvi	
1kwf	1pa2	1qto	2a0b	3pyp	
1kyp	1pcf	1qts	2acy	3seb	
1kzk	1pda	1qtw	2ahj	3sil	
1l1l	1pgs	1qu9	2arc	3std	
1l3k	1pgt	1qus	2bbk	3vub	
1l6x	1pin	1ra9	2bdp	4eug	
1l7m	1pmi	1rb9	2bop	4gcr	
1l7u	1ppn	1rge	2btc	4uag	
1lam	1ppt	1rhs	2cpg	4ubp	
1lbu	1psr	1rie	2ctc	4xis	
1lbv	1pym	1sbp	2cua	6rlx	
1lj5	1qau	1sbw	2dpm	7a3h	
1lkk	1qb7	1sgp	2eng	7odc	

Partitioning of Secondary Structure

For each peptide chain, secondary structure was defined by PROMOTIF [2]. PROMOTIF identifies the secondary structure for each residue, and was used rather than DSSP [1] so that β-hairpins would also be identified. We chose a four-state secondary structure definition, consisting of (1) hairpins, (2) helices, (3) β-strands, and (4) loops. Each structure was then cut into smaller fragments according to secondary structure assignments, such that each break was located at the interface between two secondary structure segments. Only secondary structure units containing 4 or more residues were considered. The β-hairpins were defined to be two consecutive anti-parallel β-strands with less than 9 intervening residues between them. Because we also desired to capture the loop regions, contiguous turn and coil residues were merged and identified as single loops. These filters limited the noise of the secondary definitions and ensured that the loop regions were not broken.

Contact Determination

The contacts (hydrophobic interactions, hydrogen bonds, and salt bridges) between each cut segment were then determined. This was accomplished by a program that was written in-house, which we call "ssContacts," for secondary structure contacts. For ssContacts, a pseudo-potential was developed that contains a hydrogen bond term, a salt bridge term, and a van der Waals interaction term. Each of these is described in the sections below.

Hydrogen Bonds

To identify and measure the interaction of hydrogen bonds, we used the potential developed by Fabiola, et al [19]. We used this implicit hydrogen bond potential rather than an explicit one, because crystallographic structures do not resolve protons, due to their lack of electron density. This potential is based upon the distance between a donor atom (e.g. a nitrogen) and an acceptor atom (e.g. an oxygen), as well as the C-D..A bond angle, where C, D and A denote the carbon attached to the donor atom, the donor atom, and the acceptor atom, respectively. The computation of the potential is given below.

$$E_{hb} = \varepsilon\left[\left(\frac{\sigma}{R_{DA}}\right)^6 - \left(\frac{\sigma}{R_{DA}}\right)^4\right]\cos^4(\theta - \theta_0)$$

In the above expression, ε and σ are weighting factors, set at 13.5 kcal mol^{-1} and $\sqrt{2/3}R_0$, respectively, with R_0 being the optimal distance between the donor and acceptor (2.9 Å). Also, R_{DA} is the distance (in Å) between the donor and acceptor atoms. θ is the C-D..A bond angle, while θ_0 is chosen to be 115° or 155°, whichever is closest to the measured θ. These values were chosen so that the ideal H-bond had a value of 2.0 kcal mol^{-1}. This potential is double-welled, centered about 115° and 155°, with an ideal distance of 2.9 Å.

Salt Bridges

In our structures, electrostatic interactions were computed in a much simpler fashion. Our method was adapted from Kumar and Nussinov [20], in which both positively and

negatively charged atoms are first identified. For simplicity, a neutral pH is assumed, so that the N-terminal nitrogen, the ζ-nitrogen atoms of lysine, as well as the ε-, the η^1-, and the η^2-nitrogen atoms of arginine are considered positively charged nitrogens. Negatively charged oxygens are defined to be the last oxygen of the C-terminus, the ε^1- and ε^2-oxygen atoms of glutamate, as well as the δ^1- and δ^2-oxygen atoms of aspartate. The potential for salt bridges is given below.

$$E_{sb} = 1 \text{ kcal mol}^{-1} \times \frac{2.6 \text{ Å}}{R_{SB}}$$

This results in an ideal energy of 1 kcal mol^{-1}, centered at a distance of 2.6 Å between the oppositely charged atoms. This potential diminishes with $1/R_{SB}$. In the case of salt bridges that also have hydrogen bonds, both the electrostatic salt bridge contributions and the hydrogen bond values are considered.

Van der Waals Contacts

For greater accuracy in the identification of hydrophobic contacts, Voronoi polyhedra [21, 22] were employed to pinpoint the exact neighbor and hydrophobic surface area of each hydrophobic interaction. These polyhedra are used to divide the three-dimensional space around each atom into atomic volumes. These are used because the atomic volumes are not consistent across all atoms, and because packing in proteins is asymmetric [22]. By using Voronoi polyhedra, each hydrophobic contact could be weighted according to the amount of shared surface area between the polyhedra that surround two carbon atoms. The use of Voronoi polyhedra has been shown to be more

precise and accurate than traditional radial cutoffs in this context [22]. The contribution of van der Waals interactions to the contact potential is given below.

$$E_{vdw} = 0.045 \text{ kcal mol}^{-1} \text{ Å}^{-2} \times \phi_{C-C}$$

In the above expression, ϕ_{C-C} denotes the shared face-surface area (in Å^2) between the polyhedra surrounding two carbon atoms. The scaling constant 0.045 kcal mol^{-1} was found in to be consistent with experimentally determined values for hydrophobic interactions [23].

Defining a Motif with Energetic Potentials

To ensure that only those interactions that give rise to tertiary structure were considered, only the contacts between atoms greater than 10 residues apart are considered; thus, the local $i \rightarrow i+4$ contacts that appear in α-helices are not considered. Furthermore, especially with van der Waals contacts, there were more than a few interactions whose energies were quite small ($\ll$ 0.1 kcal mol^{-1}). The aim of this project is to look at only those interactions that significantly contribute to the native topology of the protein, so we considered two secondary structure segments to be in contact if their energies of interaction sum to at least 2.0 kcal mol^{-1}. This ensures that those segments with negligible interactions are overlooked, in favor of those segments that are held more strongly together.

Clustering

The resulting motifs were first separated according to each segment's secondary structure type: helix:hairpin contact pairs were partitioned from helix:helix contact pairs, and so on. In addition to this first partitioning, we tried two different more refined clustering methods. For the first method, we clustered the structures based on the number of residues prior to clustering by RMSD. In contrast, the second method relied on a difference-in-length term coupled with an α-carbon RMSD term. In all, our hierarchical clustering algorithm was similar to methods used in previous studies [24, 25]. Both of these methods are described further in the sections below.

Method 1: Clustering by Length First

After the initial separation based on secondary structure assignments, the contact pairs were partitioned by the number of residues each segment contained. The bins were defined with bins at residue length cutoffs of $4n$ (n is an integer ≥ 1) for each segment; for example, a contact pair of residue lengths n_1=6 and n_2=18 would be separated from another contact pair of lengths n_1=6 and n_2=20. Thus the residue lengths of each segment were considered. The clustering based on residue length is needed before the final clustering, which is based off of α-carbon RMSD. These contact pairs were clustered using a greedy multi-centered clustering algorithm developed in-house. With this scheme, each motif is compared against each of the cluster centers. If no valid match is found, that motif becomes the first member of a new cluster. With the addition

of any new motif to a cluster, the cluster center is recomputed. This center is defined to be the "most average" motif – that is, it has the lowest RMSD score when compared to all members of its own cluster.

The score used in the clustering algorithm is composed of a structural term, defined by the α-carbon root mean squared deviation (RMSD). Since clusters contained diverse sequences, we could not perform an all-atom RMSD calculation. Initially, the alignment utility DALI [26] was presumed to be best for this purpose; however, the number of residues of each segment was often below the threshold required for the DALI algorithm to function. Therefore, the least-squares RMSD between two contact pairs (each of which contain two segments) was chosen to measure structural similarity. The RMSD is defined by the following expression:

$$\text{RMSD} = \sqrt{\frac{\sum_{i,j=1}^{N}(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}{N}}$$

In the above equation, the sum of the squared distances between the α-carbons in three-dimensional Cartesian space is divided by the total number of comparisons; the root of this value yields the RMSD.

However, determining the RMSD was made somewhat more complicated by the fact that each motif contains two segments of varying lengths. To overcome this, the RMSD was calculated only for the maximum common number of residues for each segment. For the larger of the two segments, the middle residues were chosen to best represent

that segment. The threshold RMSD value to use when clustering was determined empirically by finding a value that resulted in 20% singletons. In other words, 20% of the structures would not cluster into groups at the RMSD chosen for each secondary-structure and residue-length bin. This was chosen in order to account for the fact that some contact types are much more restricted in space (eg, strand-strand interactions) than others (like helix-loop interactions).

Method 2: Difference-in-Length Term

Although clustering only by RMSD resulted in clustered motifs of similar orientations, it did not discriminate between structures of vastly different lengths. For this reason, we tried adding a difference-in-length term to the overall clustering score. The overall score for a given motif when compared to a cluster center was defined by:

$$\text{score} = RMSD + \frac{1}{2}(\Delta l_1 + \Delta l_2),$$

where RMSD is in units of Å, and $\Delta\lambda_1$ and $\Delta\lambda_2$ represent the integer differences in residue lengths between the center and the member for the first and second segments, respectively. This added term results in an extra penalty when comparing two motifs that contain segments of differing lengths. For this method, the cutoff score used was 8.0; this was found to best cluster similar motifs together while ensuring small structural anomalies (e.g. β-bulges) were ignored. Each of these motifs was ultimately clustered by secondary structure type, three dimensional similarity (RMSD), and length of each segment.

Contact Maps

Contact maps were made for each cluster. These maps show the placement and type of contact for each motif. In order to combine the contacts for all the members of a cluster into a single contact map, each member was superimposed onto the cluster center.

For each structural contact, both segments were superimposed independently onto their respective cluster center segment, and equivalent residues were defined by the smallest α-carbon to α-carbon distance, if that distance was less than 2 Å. If the smallest distance is greater than 2 Å, then that residue was considered to have no equivalent on the cluster center. After the superposition and defining of equivalent residues relative to the cluster center, the contact map was created based on that cluster center. For this, each contact on a member was evaluated as if it occurred on the equivalent residues of the cluster center. Each contact map is simply the sum of the contacts for a given motif. Since each tertiary motif contains structures of nearly identical configuration, the contact points are expected to overlap considerably.

RESULTS

By using ssContacts, we obtained a total of 21,100 tertiary contact pairs. Each of these contact pairs was then clustered by using the first algorithm described in the Methods section. Table 2 summarizes the data for these clusters in addition to their energetic parameters. The RMSD cutoffs used in this method were variable; their histogram is given in Figure 1. It is clear from an inspection of the energy data that our potentials yield reasonable results – the helix:helix motifs display a high level of hydrophobic packing (through van der Waals interactions), while strand:strand motifs yield far more hydrogen bonding. This is consistent with generally understood packing arrangements for each of these types of secondary structures.

The distribution of sequence separations between each contact pair is given in Figure 2 with the top curve representing all contact pairs, and the bottom curve representing only the cluster singletons. In this histogram, we see that the bulk of the contact pairs have intervening segments of less than 50 residues. The distribution also shows that the probability of two segments being in contact decreases with their sequence separation.

We also found a very strong linear correlation between the number of residues in a protein and the number of tertiary contact pairs (Figure 3). Initially this seemed trivial, but after noting that these contact pairs have wildly variable lengths (compare the 150-residue coiled-coil motifs with a simple 8-residue strand:strand motif), this finding is

Table 2. Cluster Data and Energetic Analysis

Secondary Structure Type	Number of Contact Pairs	Number of Clusters	Number of Singles	Avg Hbond Energy	Avg SB Energy	Avg VDW Energy
Helix:Helix	3195.	1467.	735.	0.97 (1.60)	0.22 (0.66)	4.36 (2.98)
Helix:Hairpin	1401.	615.	311.	1.08 (1.77)	0.13 (0.50)	3.37 (1.93)
Helix:Strand	2178.	838.	446.	1.24 (1.85)	0.07 (0.35)	2.26 (1.09)
Helix:Loop	5978.	2301.	1231.	3.21 (2.84)	0.11 (0.46)	1.88 (1.41)
Hairpin:Hairpin	526.	231.	121.	2.89 (3.55)	0.12 (0.53)	3.96 (1.99)
Hairpin:Strand	920.	358.	193.	4.12 (3.06)	0.07 (0.30)	3.41 (1.74)
Hairpin:Loop	1512.	637.	328.	2.21 (1.93)	0.12 (0.46)	2.03 (1.29)
Strand:Strand	1794.	599.	366.	4.23 (2.17)	0.05 (0.30)	2.70 (1.28)
Strand:Loop	2028.	797.	412.	2.47 (1.54)	0.05 (0.33)	1.21 (0.91)
Loop:Loop	1568.	679.	340.	2.46 (1.90)	0.08 (0.38)	1.59 (1.09)

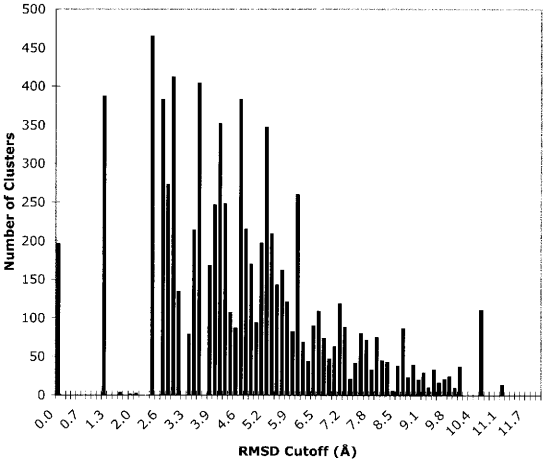

Figure 1. RMSD Cutoff Distribution for Clusters

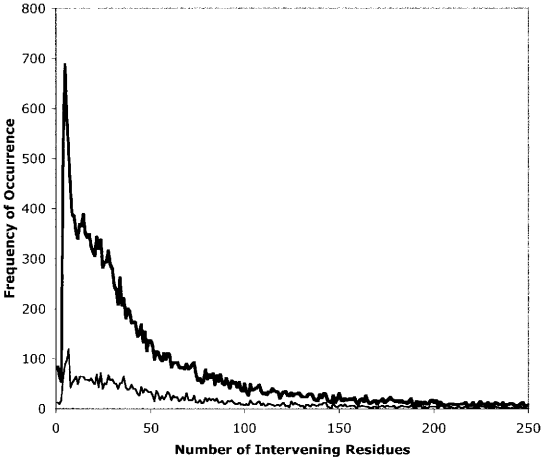

Figure 2. Distribution of Sequence Separations

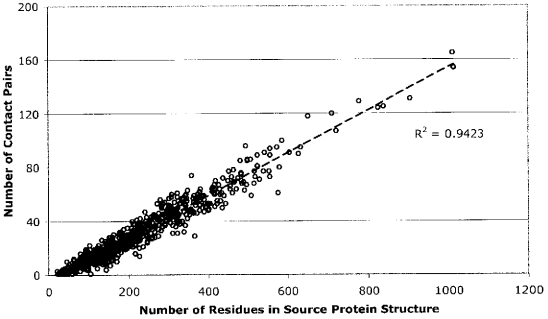

Figure 3. Relationship to Number of Original Residues

seen to be a bit more interesting. Such a tight correlation suggests that these tertiary contact motifs be considered as modular units of tertiary structure.

The results from the second method of clustering were discarded because the similarity score used turned out not to be a very good metric. For instance, in comparing two contact pairs, if they contained the same number of residues, the RMSD cutoff effectively became 8 Å. Conversely, if the two pairs were structurally identical, but one was longer than the other, the score might not fall beneath the threshold required for similarity. The difference-in-length term only served to perturb the RMSD cutoff in a way that was not always desired. Mathematically, we were reducing RMSD and residue length into one unidimensional score; since these two terms are not orthogonal or linear in our metric space, this was not a good choice. For this reason, we decided to stay with the original method of first clustering by length, then by α-carbon RMSD.

The top ten most populous clusters for each secondary structure class were then analyzed. Descriptions of each class of motif are given in the following sections.

Helix:Helix

The helix:helix motifs were the second most common contact pairs in all of the library. There were 3,195 contact pairs, and 1,467 recurrent motif structures were obtained from this data set. The most common helix:helix motif (00.012012.0071) is shown in Figure 4, with its corresponding contact map in Figure 5. As seen on the contact map, and as

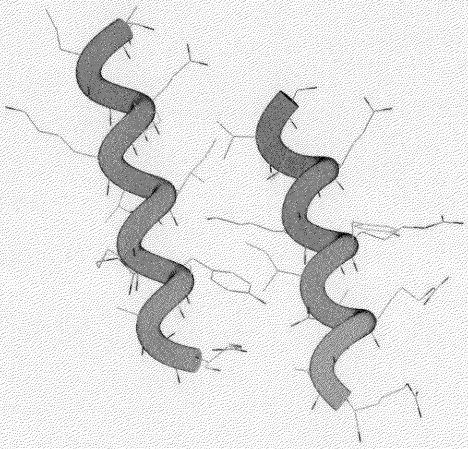

Figure 4. Helix:Helix 00.012012.0071 Cluster Center

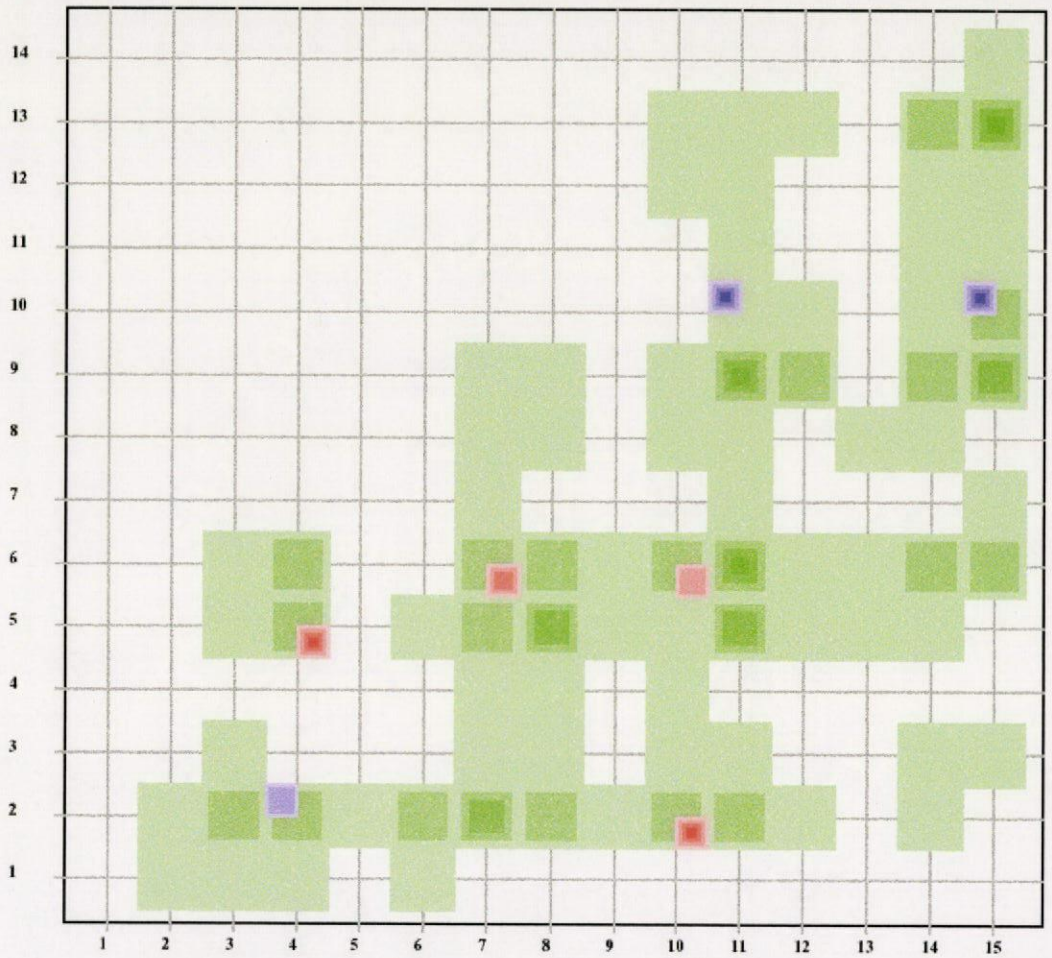

Figure 5. Helix:Helix 00.012012.0071 Contact Map

The above contact map and those that follow represent the atomic interactions between two secondary structure units. The numbers along the *x*-axis represent the residues in the first segment, and those along the *y*-axis represent the residues on the second segment. Green defines a van der Waals contact, red denotes a salt bridge, and blue defines a hydrogen bond. Intensity of color defines the strength of the interaction.

shown on Table 2, these helix:helix motifs displayed a much larger tendency for hydrophobic van der Waals contacts than polar contacts. Previous work by Bowie [14] illustrated that helix packing favored orthogonal helix:helix structure packing arrangements, where the two axes of the helices are perpendicular (90°) to each other. We therefore expected that our clustering bin with the highest number of members would be filled with a helix:helix motif with orthogonal packing.

We were obviously surprised to see that two of the three most recurrent motifs yielded an angle of nearly 0° – seemingly contrary to previous work. This result can be rationalized though, by remembering that our clustering considers the whole of structural similarity, with such details as relative locations and three-dimensional orientation. That said, what our helix:helix motif data suggests is that those helix packing arrangements centered near 0° are more similar to *each other* structurally, than the motifs at the more preferred angles are similar to each other.

It could be that, while the Ω angle developed by Bowie displays certain preferences, it may not take into account three-dimensional similarity: two motifs, both at 90°, could have other significant differences, for example the relative location of the contacts between the two helices. In other words, the single-dimensional index Ω may not provide a full picture of how helices pack against each other. More exhaustive studies focusing exclusively on the helix:helix motifs in our library would be necessary to provide an in-depth analysis of the discrepancies between this and previous work.

Helix:Hairpin

The helix:hairpin motifs are also of interest, since the packing arrangements between α and β secondary structure elements had been well-characterized in the 1980s [27, 28]. In the work of Scheraga, et al, they found four classes of energetically favorable helix:hairpin arrangements, each characterized by the orientation of the axis of the helix relative to the direction of the β-strands. The most favorable of their four interactions was an axis of the helix roughly parallel with the direction of the strands. Also low in energy was the arrangement roughly perpendicular to the strands, as well as a diagonal packing arrangement. According to their work, each of these was a low-energy configuration because of the attractive non-covalent side-chain-side-chain interactions present between the two secondary structure elements.

In our motif library, the most common helix:hairpin motifs were of the parallel variety, as seen in Figure 6 (01.016020.0002, also 01.016016.0012). Also, the diagonal arrangement also appears to be quite common (motif 01.008012.0003). As Alan Fersht noted in his study with barnase [27], the interdigitated (or 'knobs in grooves') residue packing results in a high amount of van der Waals interactions between the α-helix and the anti-parallel β-sheet. Our data suggests that these complementary hydrophobic interactions seem to be the most significant in securing these α/β intramolecular contacts, as illustrated by the contact map shown in Figure 7, and summarized also in Table 2.

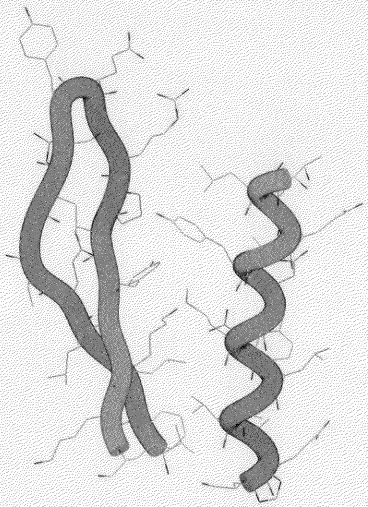

Figure 6. Helix:Hairpin 01.016020.0002 Cluster Center

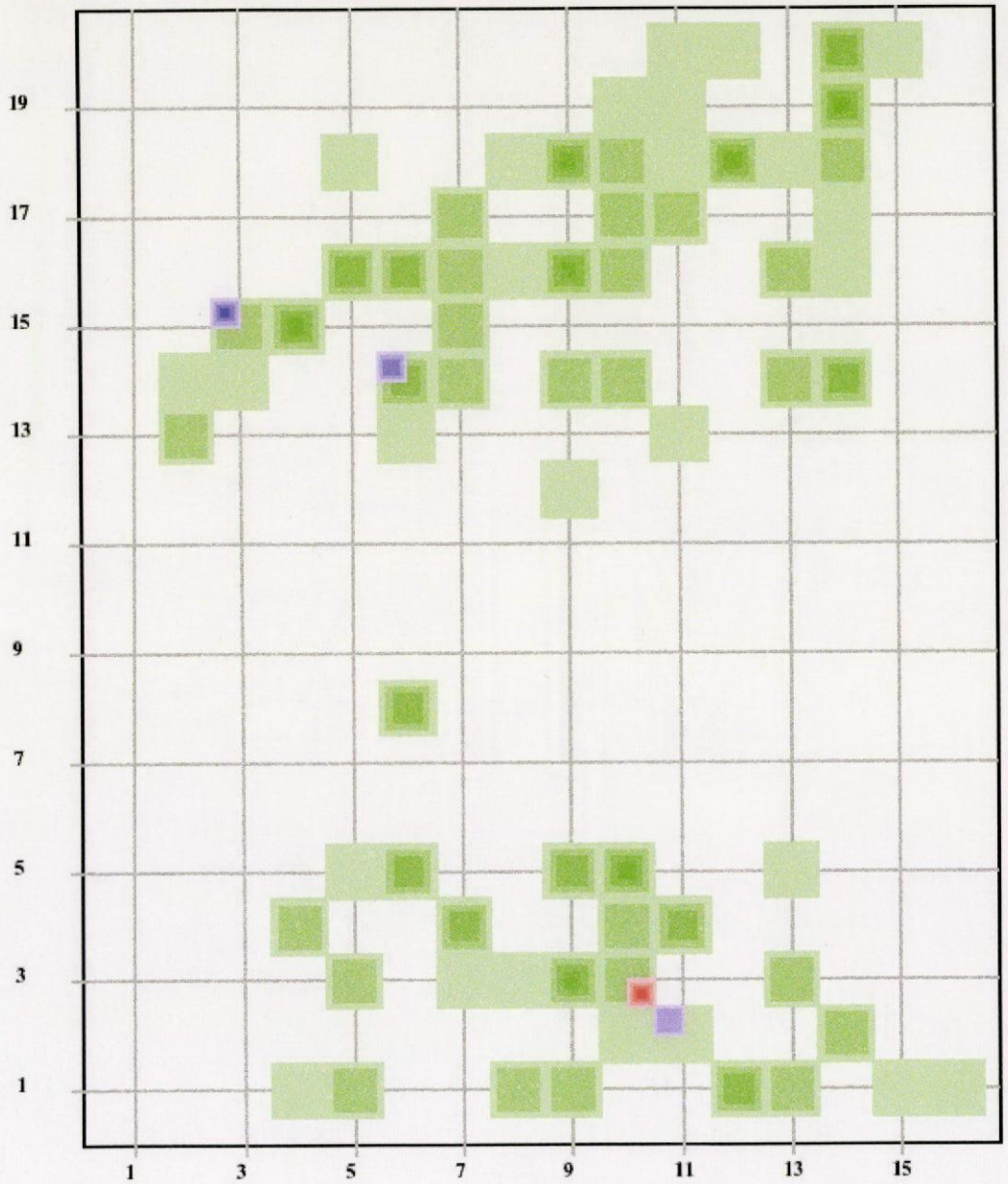

Figure 7. Helix:Hairpin 01.016020.0002 Contact Map

Helix:Strand

As would be expected, the helix:strand motifs maintain much of the same type of packing arrangements as the helix:hairpin motifs. The predominant form each of the motifs takes is an α-helix whose axis is almost parallel to a β-strand, as seen in Figure 8 (motif 02.012004.0152, also 02.016004.0122 and 02.020004.0025). As noted above, Scheraga, et al, calculated this to be the lowest energy configuration, due to favorable side-chain-side-chain interactions between the two secondary structure segments [28]. We do note the presence of some diagonally oriented motifs (for example, 02.020004.0040), but interestingly these seem to appear with less regularity than in the helix:hairpin motifs. This may be due to the fact that we considered hairpins independently of "plain" β-strands – thus, the β-segments in the helix:strand bins may over-represent the parallel β-sheets, simply because the consideration of hairpins as separate would remove those anti-parallel strands from consideration.

Owing to the somewhat constant nature of the orientations of these helix:strand motifs, the largest partitioning seems to be occurring at the clustering by residue length stage. It is interesting to note that while any strand could be considered in contact with a helix by simply one good hydrogen bond, the bulk of the motifs in our library show the entire length of the extended strand to run along the helix (see Figure 9). Obviously, there seems to validate Scheraga, et al, in that these motifs do appear to be common, which does imply some sort of structural stability.

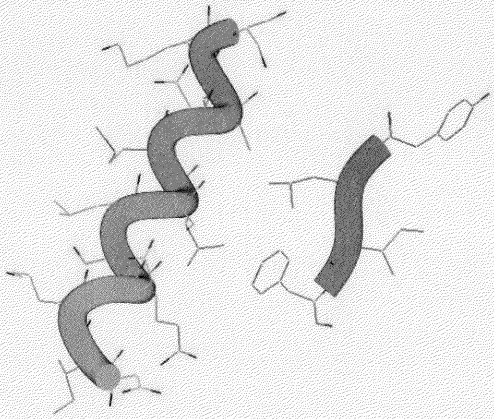

Figure 8. Helix:Strand 02.012004.0152 Cluster Center

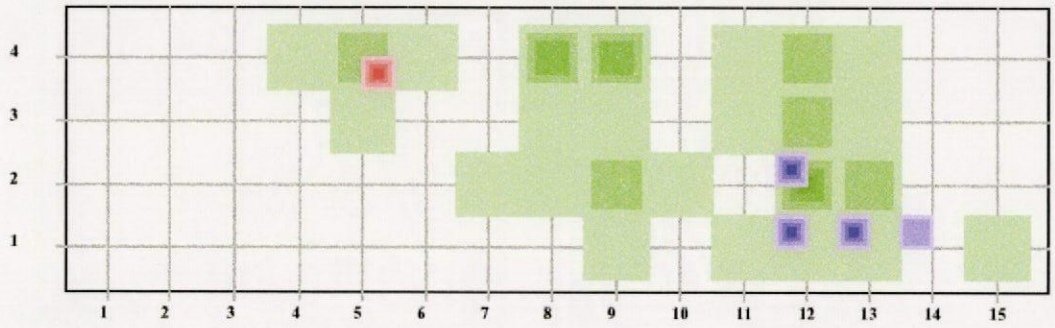

Figure 9. Helix:Strand 02.012004.0152 Contact Map

Helix:Loop

If we consider only the highly recurrent motifs in our sample, the helix:loop structures were typically of very little sequence separation. Most of these highly populated clusters motifs involved loops that trailed or preceded α-helices, which is striking because only 28% of our starting data had sequence gaps of less than 3 residues. This suggests that those helix:loop motifs with greater sequence gaps between the two segments are much more variable than those with no sequence gaps. A typical helix:loop motif of the more recurrent variety displays one or more hydrogen bond between the loop and the helix, and the two secondary structure elements have zero residues separating them. Such a motif is illustrated by Figures 10 and 11 (motif 03.012004.0204). More motifs maintaining this type of configuration are 03.012004.0054, 03.012004.0057, 03.008004.0140, 03.012004.0225, and 03.016004.0002.

This lack of diversity in the conserved motifs could simply be due to the fact that a loop is classified as such precisely because it is not a fixed, rigid structure. If it were the case that a loop had enough contact along the face of an α-helix, the extended segment might instead have been classified as a β-strand. In other words, the presence of interactions that fix the segment in place may be a critical factor in fixing the extended β configuration instead of the more unordered loop structure.

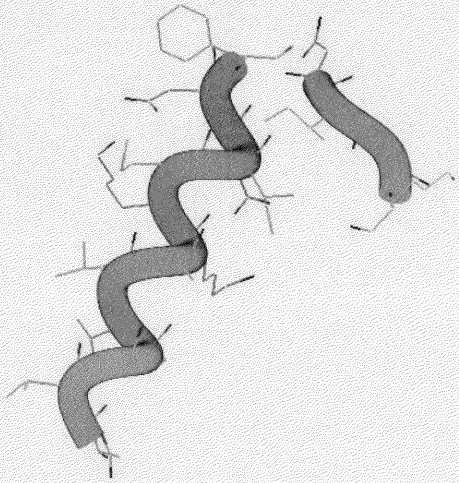

Figure 10. Helix:Loop 03.012004.0204 Cluster Center

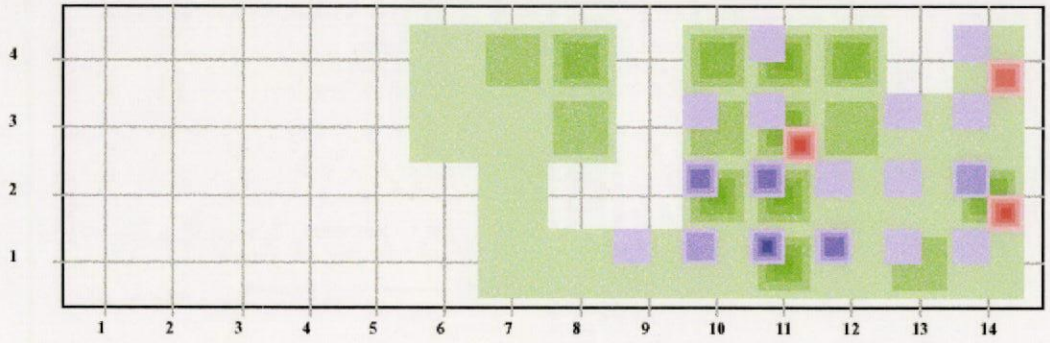

Figure 11. Helix:Loop 03.012004.0204 Contact Map

Hairpin:Hairpin

These motifs were perhaps the most irregular group on our sample set. The top two most recurrent motifs (11.016016.0022 and 11.012012.0018) formed stacked β motifs, in a structure that could be described as the stacking of two small sheets, one on top of another. It is interesting to note that in both of these structures, the β-hairpins are parallel to each other; that is, their turns are both pointing in the same direction (see Figures 12 and 13). It is also worth noting that this type of orientation allows for relatively simple packing – the top hairpin must simply be shifted by one residue's length to be in register with the bottom hairpin for complementary grooves-in-ridges side-chain packing. The third most populated cluster, motif 11.012012.0005, represented a structural motif of the more expected variety. This configuration is a single long sheet, brought about by the interaction of two hairpins. In this case, the hairpins are in an anti-parallel orientation with main-chain hydrogen bonding, and the sheet has the familiar propeller-twist architecture seen in other large β-sheets.

Altogether, the hairpin:hairpin motifs were decidedly the least common type of interactions between secondary structure elements. Only 526 out of 21,100 contacts (or under 3%) of the total intramolecular contacts were of this type. This sort of data might be useful in scoring novel folds in tertiary structure prediction – the lack of consistency for this type of motif might imply energetic instability, but it could just as easily represent an evolutionary happenstance.

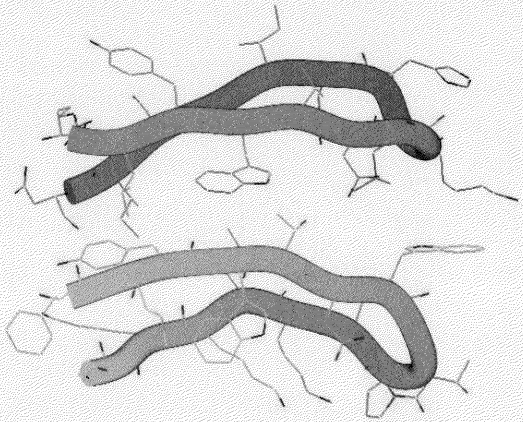

Figure 12. Hairpin:Hairpin 11.016016.0022 Cluster Center

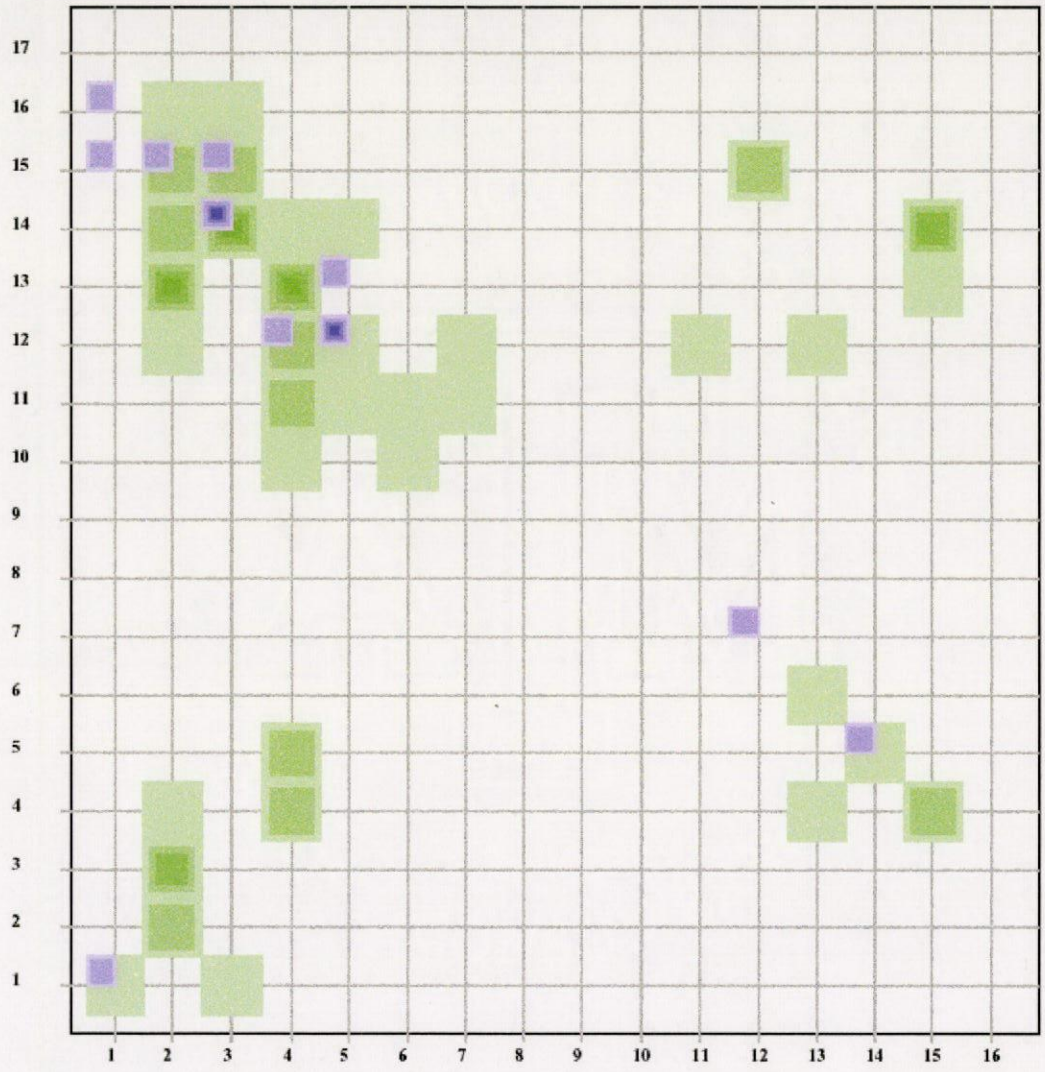

Figure 13. Hairpin:Hairpin 11.016016.0022 Contact Map

Hairpin:Strand

The most populated clusters in our database for hairpin:strand motifs are filled with anti-parallel sheet-like structures. In this case, the β-strand is oriented perpendicular with respect to the β-hairpin, forming nice anti-parallel propeller-twist β-sheet motifs (12.016004.0041, 12.012004.0049, 12.020004.0009, 12.016008.0004, and 12.020008.0011). This is illustrated in Figures 14 and 15 by motif 12.016004.0041. As summarized on Table 2, these structures have a slightly higher preference for hydrogen bonding and slightly lower preference for van der Waals interactions than the related hairpin:hairpin motifs. This may be because of the increased propensity to form anti-parallel β-sheets with more main-chain hydrogen bonding, rather than to stack on top of each other, as noted above for some of the hairpin:hairpin clusters.

Hairpin:Loop

Much like the helix:loop motifs, the hairpin:loop clusters tend to be well-populated by contiguous segments, that is, segments with zero intervening residues between them. Of the highly populated clusters, roughly half of them have loops that precede the hairpins (e.g. motifs 13.020004.0007 and 13.012008.0011); the other half contains loops that immediately follow hairpins (e.g. motifs 13.016004.0054 and 13.016004.0077). In both cases, there seems to be a moderate amount of hydrogen bonding and van der Waals interactions (cf. Table 2). The hairpin:loop motif 13.020004.0007 is shown in Figure 16, as well as its contact map in Figure 17.

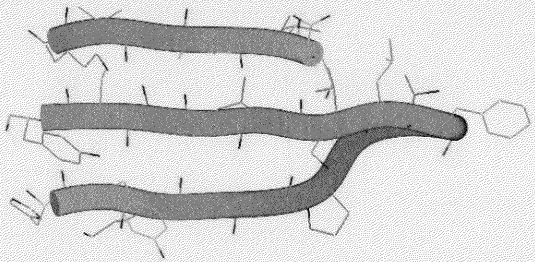

Figure 14. Hairpin:Strand 12.016004.0041 Cluster Center

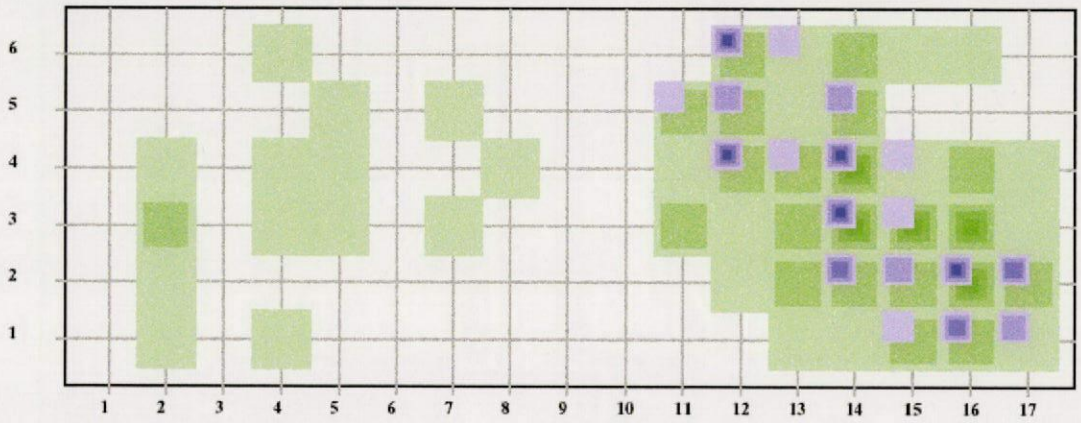

Figure 15. Hairpin:Strand 12.016004.0041 Contact Map

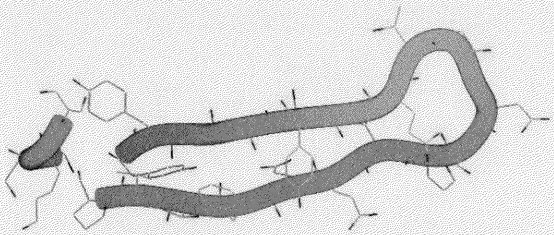

Figure 16. Hairpin:Loop 13.020004.0007 Cluster Center

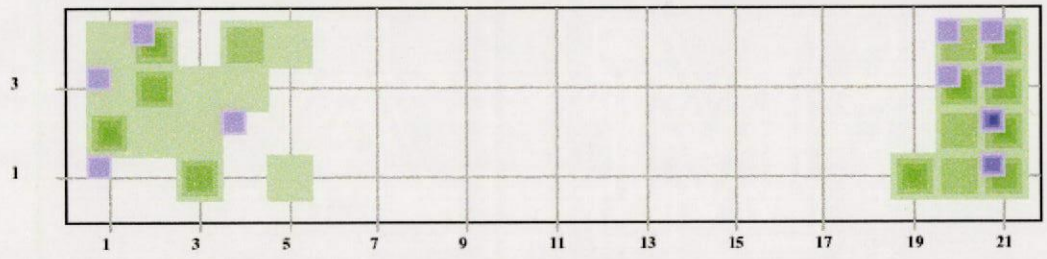

Figure 17. Hairpin:Loop 13.020004.0007 Contact Map

The only structurally consistent arrangement these motifs share is that the loop tends to attach one strand of the hairpin to the other strand, opposite the side of the hairpin's turn. This closure of the β-hairpin may serve to further stabilize the hairpin structure. Although the interactions seen in this group are not typically electrostatic in nature, this "anti-unzipping" type of hairpin stabilization would be consistent with previous electrostatic studies on the β_1 region in the IgG-binding domain of protein G [29]. These intramolecular contacts may be important for certain hairpins that lack stabilizing interactions at their ends to prevent unzipping.

Strand:Strand

The strand:strand motifs consistently had the highest populated clusters of any other group. This is probably due to the quite fixed, consistent conformations that β-sheets adopt. Only one of the ten most highly populated clusters formed an anti-parallel β-sheet (motif 22.004004.0121), the others were all parallel (eg, motifs 22.004004.0003, 22.004004.0198, and 22.004004.0181). An example of the parallel sheet is given in Figure 18, with its contact map in Figure 19. The top ten highly recurrent clusters all contained contact pairs from the n=1 bin (i.e. the residue lengths were between 4 and 7 for both segments), and each maintained the typical main-chain hydrogen bonding pattern seen in their respective types of β-sheets. The anti-parallel motifs maintained straight-on main-chain hydrogen bonds, while the parallel motifs displayed bifurcated main-chain hydrogen bonds across the strands.

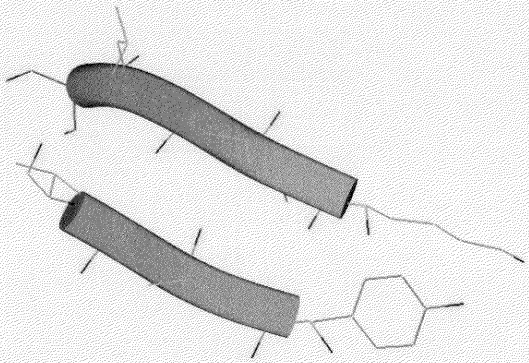

Figure 18. Strand:Strand 22.004004.0003 Cluster Center

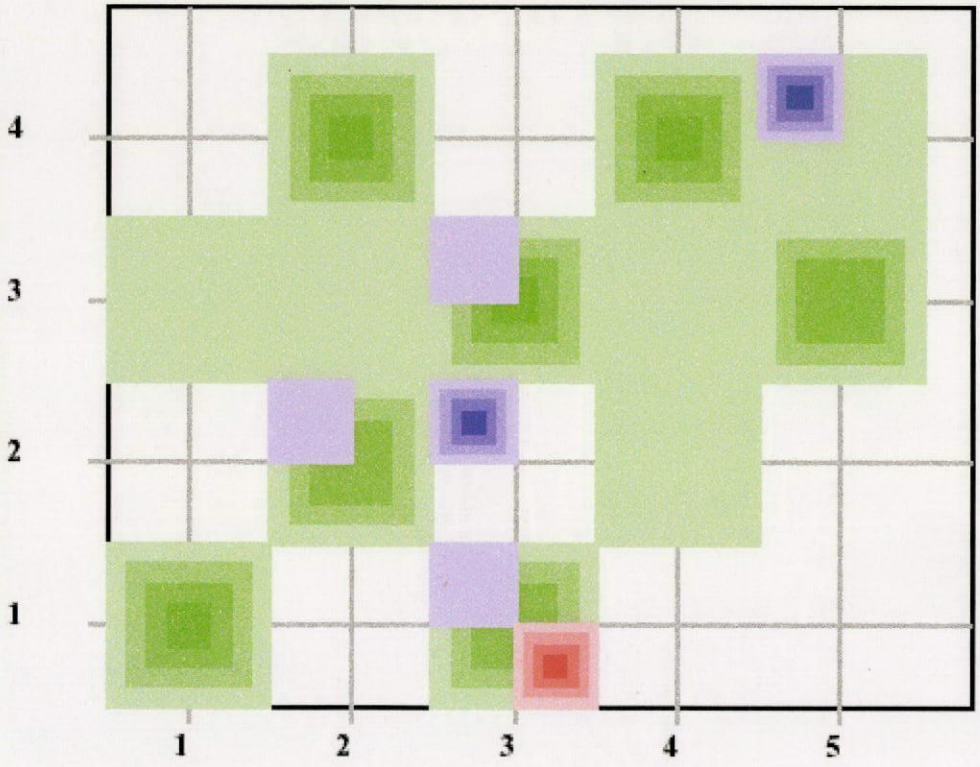

Figure 19. Strand:Strand 22.004004.0003 Contact Map

The high selectivity for residue length may be due to under-representation of longer strands in our sample set, or may be an artifact of using RMSD as a structural metric. As the length of an extended structure grows, a "lever-arm effect" can take place, where local deviations far from the ends make it difficult to globally superimpose the two structures well. This gives rise to higher RMSD scores, and may in part explain the preference for short strands in this group.

Strand:Loop

The following last two groups were very diverse; in evaluating the highly conserved motifs, we realized that there is much diversity in the strand:loop clusters. Very few trends stick out, except to say that, on average, more of the interactions were from hydrogen bonding instead of van der Waals packing.

Loop:Loop

Likewise, the loop:loop clusters were very diverse, with little consensus in their configurations. These clusters also tended to be constrained mainly by hydrogen bonds more than van der Waals interactions. The lack of a clear trend in either of these last motifs may be indicative of the sheer number of proteins we are sampling. Evolution may not tend to conserve loop regions in particular; on the contrary, it is commonly thought that loop regions rather than scaffold regions tend to confer specificity to a given protein.

DISCUSSION

The results from our clustering suggest that there is some level of redundancy for tertiary contact motifs. And this should probably be expected: since the evolution of protein structure will tend to maintain stability of a given fold, it is not at all surprising that certain motifs would thus appear regularly. As we have defined them, these tertiary contact motifs are contacts between residues distant in sequence space but are nevertheless near in three-dimensional space – thus they tend to be the points that confer topology to a protein structure. Divergent evolution, it is assumed, would tend to conserve the types of interactions at these contact points, so that the topology would not change significantly as the protein evolves. Therefore, these points along the protein backbone must play a critical role in securing the protein's structure.

Now that the packing of residues at these conserved points can be systematically analyzed, this will undoubtedly help to predict how specific mutations might alter the physical stability of a protein. Furthermore, the problem of predicting idealized packing arrangements can now be probed from an informatics perspective with our fragment library, since our motif library contains a great deal of information about how specific residues pack in three-dimensional space against other residues further down in sequence.

In addition, the methods developed in this study will also be used in study of protein interaction sites. The connectivity of secondary structure that presents an oligomerization domain can be catalogued, and this connectivity can be probed in other proteins to scan for potential interactions. Since proteins are thought to have co-evolved [30], divergent evolution can be assumed; this would be expected to ease the prediction of interactions for a given protein system. This type of approach could also obviate many of the problems associated with induced-fit interactions by analyzing the protein structure only at these conserved, non-covalent hinge contact sites.

REFERENCES

[1] Kabsch W and Sander C. Dictionary of protein structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22 (12): 2577-2637, 1983.

[2] Hutchinson EG and Thornton JM. PROMOTIF – A program to identify and analyze structural motifs in proteins. *Protein Science* 5: 212-220, 1996.

[3] Bonneau R; Tsai J; Ruczinski I; Chivian D; Rohl C; Strauss CE; and Baker D. Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins* Suppl 5: 119-126, 2001.

[4] Kolodny R; Koehl P; Guibas L; and Levitt M. Small libraries of protein fragments model native protein structures accurately. *Journal of Molecular Biology* 323 (2): 297-307, 2002.

[5] Hunter CG and Subramaniam S. Protein fragment clustering and canonical local shapes. *Proteins: Structure, Function, and Genetics* 50: 580-588, 2003.

[6] Richards FM and Kundrot CE. Identification of structural motifs from protein coordinate data: secondary structure and first-level super-secondary structure. *Proteins* 3: 71-84, 1988.

[7] de Brevern AG; Etchebest C; and Hazout S. Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. *Proteins: Structure, Function, and Genetics* 41 (3): 271-287, 2000.

[8] Salem GM; Hutchinson EG; Orengo CA; and Thornton JM. Correlation of observed fold frequency with the occurrence of local structural motifs. *Journal of Molecular Biology* 287: 969-981, 1999.

[9] de la Cruz X; Hutchinson EG; Shephard A; and Thornton JM. Toward predicting protein topology: an approach to identifying beta hairpins. *Proceedings of the National Academy of Sciences USA* 99 (17): 11157-11162, 2002.

[10] Voigt CA; Martinez C; Wang Z-G; Mayo S; and Arnold F. Protein building blocks preserved by recombination. *Nature* 9 (7): 553-558, 2002.

[11] Garratt RC; Thornton JM; and Taylor WR. An extension of secondary structure prediction towards the prediction of tertiary structure. *FEBS Letters* 280 (1): 141-146, 1991.

[12] Ruczinski I; Kooperberg C; Bonneau R; and Baker D. Distributions of beta sheets in proteins with application to structure prediction. *Proteins* 48 (1): 85-97, 2002.

[13] Steward RE and Thornton JM. Prediction of strand-pairing in antiparallel and parallel β-sheets using information theory. *Proteins: Structure, Function, and Genetics* 48: 178-191, 2002.

[14] Bowie JU. Helix packing angle preferences. *Nature Structural Biology* 4 (11): 915-917, 1997.

[15] Nelson ED and Onuchic JN. Proposed mechanism for stability of proteins to evolutionary mutations. *Proceedings of the National Academy of Sciences USA* 95: 10682-10686, 1998.

[16] Clementi C; García AE; and Onuchic JN. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: All-atom representation study of Protein L. *Journal of Molecular Biology* 326: 933-954, 2003.

[17] Wang G and Dunbrack RL Jr. PISCES: a protein sequence culling server. *Bioinformatics* submitted 2002.

[18] Gerstein M; Tsai J; and Levitt M. The volume of atoms on the protein surface: Calculated from simulation, using Voronoi polyhedra. *Journal of Molecular Biology* 249: 955-966, 1995.

[19] Fabiola F; Bertram R; Korostelev A; and Chapman MS. An improved hydrogen bond potential: impact on medium resolution protein structures. *Protein Science* 11 (6): 415-423, 2002.

[20] Kumar S and Nussinov R. Salt bridge stability in monomeric proteins. *Journal of Molecular Biology* 293 (5): 1241-1255, 1999.

[21] Voronoi G. Nouveles applications des parametres continus a la theorie des formes quadratique. Journal für Reine und Angewandte Mathematik 134: 198-287, 1908.

[22] Tsai J; Taylor R; Chothia C; and Gerstein MB. The packing density in proteins: Standard radii and volumes. *Journal of Molecular Biology* 290: 253-266, 1999.

[23] Raschke TM; Tsai J; and Levitt M. Quantification of the hydrophobic interaction by simulations of the aggregation of small hydrophobic solues in water. *Proceedings of the National Academy of Sciences USA* 98 (11): 5965-5969, 2001.

[24] Shortle D; Simons KT; and Baker D. Clustering of low-energy conformations near the native structures of small proteins. *Proceedings from the National Academy of Sciences USA* 95 (19): 11158-11162, 1998.

[25] Karpen ME; Tobias DJ; and Brooks CL 3rd. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. *Biochemistry* 32 (2): 412-420, 1993.

[26] Holm L and Sander C. DALI: A network tool for protein structure comparison. *Trends in Biochemical Science* 20 (11): 478-480, 1995.

[27] Kellis JT Jr; Nyberg K; Sali D; and Fersht AR. Contribution of hydrophobic interactions to protein stability. *Nature* 333 (6175): 784-786, 1988.

[28] Chou KC; Nemethy G; Rumsey S; Tuttle RW; and Scheraga HA. Interactions between an alpha-helix and a beta-sheet. Energetics of alpha/beta packing in proteins. *Journal of Molecular Biology* 186 (3): 591-609, 1985.

[29] Tsai J and Levitt M. Evidence of turn and salt bridge contributions to beta-hairpin stability: MD simulations of C-terminal fragment from the B1 domain of Protein G. *Biophysical Chemistry* 101-102 (1): 187-201, 2002.

[30] Bennett MJ; Schlunegger MP; and Eisenberg D. 3D domain swapping: a mechanism for oligomer assembly. *Protein Science* 4: 2455-2468, 1995.

VITA

Hamilton Courtney Hodges
313 Lincoln Ave, Apt 136
College Station, TX 77840

Education

1995 – 1999 Dallas Christian High School
Graduated with Highest Honors

1999 – 2003 Texas A&M University
Major: Biochemistry
Minor: Mathematics

Awards and Honors

1999	Association of Former Students Scholarship
2000	Barnes & Noble Academic Excellence Scholarship
2000	Member, National Society of Collegiate Scholars
2001	Charles J. Koerth, Sr., Memorial Scholarship
2002	College of Agriculture and Life Sciences Scholarship
2002	Honors Undergraduate Research Fellow / Biochemistry Research Scholar
2003	Honorable Mention for NSF Graduate Research Fellowship

Research Experience

Summer 2001 *Neuroimmunology*
Jane Welsh / Department of Veterinary Immunology
Mary Meagher / Department of Psychology

2001 – 2003 *Computational Proteomics & Bioinformatics*
Jerry Tsai / Department of Biochemistry & Biophysics

Scientific Posters and Conferences

Apr 2002	ASBMB Annual Meeting
Aug 2002	Protein Society Annual Symposium
Dec 2002	The Fifth Critical Assessment of Structure Prediction (CASP)
Apr 2003	Texas A&M University Student Research Week

Memberships

1999	Phi Eta Sigma Honor Society, Member
2001	American Society for Biochemistry and Molecular Biology, Student Member
2002	The Protein Society, Student Member
2003	Sigma Xi Scientific Research Society, Associate Member