PUZZLE PIECES DEFINED: LOCATING COMMON PACKING UNITS IN TERTIARY PROTEIN CONTACTS.

KM GERNERT, BD THOMAS, JC PLURAD, JS RICHARDSON, DC RICHARDSON
Duke University Medical Center, Biochemistry Dept., Durham, NC 27710
LD BERGMAN
IBM TJ Watson Research Ctr., PO Box 704, Yorktown Heights, NY 10598

ABSTRACT

Puzzle pieces are defined as small packing units which make up the unique tertiary interactions in proteins. Anti-parallel and perpendicular helix-helix contacts were broken down into basic puzzle-piece pairs in order to study the traits of such contacts: their limited geometry, preferred residue involvment, residue conformation and other common constraints. These traits can then be used for continued comparison of other protein structures, improving models of and designing proteins de novo and, in time, predicting 3D structure from primary sequence. Results from a small (100 proteins) database of anti-parallel helix-helix contacts and from preliminary work on a large database (600 proteins) of perpendicular helix-helix contacts are presented.

INTRODUCTION

Many experiments have shown that it is possible to design and synthesize a peptide or protein which will assume a particular secondary structure and even an approximately correct tertiary structure. However, designing a protein with a unique tertiary structure has proven difficult. Of the many approaches taken by several groups (α4 by DeGrado[1]; Betabellin[2] and Felix[3] by Richardsons; GCN4 redesigns by Sauer[4], Hodges[5] and Alber[6], to name a few (see[7] for a review)), only a few have claimed success (helix-turn-helix by Osterhout[8] and another by Ohkubo[9]). Since these successes consist of relatively short sequences (< 40 residues), the successful design of a protein with a native-like size and a unique, globular, tertiary structure remains to be accomplished.

To address the lack of uniqueness seen in the designed proteins, many groups have turned to the examination of tertiary contacts and have reconsidered the previous

studies of beta-beta[10], beta hairpin turns[11], beta-alpha[12], and alpha-alpha[13], and alpha and beta[14] contacts. These earlier studies have shown general properties of contacts such as interaction distances and angles and overall positioning of secondary structural elements. This study was designed to generate the more specific information necessary to design a helical primary amino acid sequence with the unique sidechain packing of the desired tertiary interactions (while excluding the alternatives).

There are two possible approaches to study the specifics of such complicated items as tertiary contacts in proteins. One is a multi-dimensional comparison of the contact as a whole, which requires an exponentially large number of occurrences of the contacts of interest for statistical significance. A more feasible method is to break the contacts down into smaller manageable pieces, which could then be compared. The smaller the piece, the fewer the attributes to be studied, the lower the number of occurrences for statistical results. Our study broke the specific tertiary interactions into such small, defined packing units, called "puzzle pieces".

We defined and located the puzzle pieces within known protein structures, using the program VIEW by Bergman[15], a molecular graphics program with an internal script language used to compile tools for viewing and studying a molecule of interest. Our approach was to look at the structures, derive the patterns of the contacts, and then analytically study these patterns for common characteristics.

The located puzzle pieces and interacting "puzzle-piece pairs" within similar protein structures were superimposed and analyzed statistically. A completed study of anti-parallel helix-helix contacts and a preliminary study on perpendicular helix-helix contacts reveal specific characteristics of interacting puzzle-piece pairs, including limited geometries, preferred residue involvement and common constrained contacts. We propose that common puzzle-piece pairs add uniqueness to protein structures due to their tight, constrained packing. One example is the overlapping puzzle-piece pairs described for the alacoil[16]. Proposed uses of these results include (1) continued comparison of protein structures and modeled structures, (2) revising the amino acid sequence of designed proteins, (3) design de novo of new protein structures, and in time, (4) predicting 3D structure from primary sequence.

MATERIALS AND METHODS

1. Protein files

Originally, protein files were taken from the Brookhaven Protein Data Bank [17], January 1992 version. Specific proteins used in the preliminary anti-parallel helix contact study were selected from a non-homologous list developed by Hobohm, et al.[18] (with a resolution higher than 2.0 Å), and additional, newer protein structures from the Brookhaven DataBank. The new structures were chosen by (1) known helical content, (2) non-redundancy to previously included structures, (3) resolution limit of 3.0 Å or better (one NMR structure was also included, 1bbn). The final list consisted of 96 protein chains.

More recently a larger protein database was generated from Hobohm and Sander[19], using the April 1995 version of the Brookhaven Protein Data Bank [17]. The additional restrictions applied included (1) a homology less than 45%, (2) a resolution greater than 3.0 Å, and (3) non-mutant structures; (4) for NMR structures the minimal average structure was choosen when available. The final list contained 597 protein chains.

2. Programs

The program Define_Structure (Define_S) by Richards et al.[20], with further assistance and associated programs from S.R. Presnell, was used to locate helices in each protein structure and to determine the intramolecular angle and distance between the helices. For this study, the preset limits of Define_S for α- and 3-10 helical structure were set as CHKHA = 0.70, and CHKH3 = 0.70. These limits gave helix ends that corresponded fairly closely to the ends as defined by the Richardsons[21].

The program Access by Richards, et al.[22], was used to determine the solvent-accessible surface area of the proteins or pieces of the proteins.

Other programs written by K.P. Murphy and modified for this project were used to summarize the residues buried, and the residues buried in a specific contact. The percentage buried, for a given sidechain, is in comparison with the static accessibility for individual amino acid residues, X, in model extended tripeptides Ala - X - Ala[23].

Dihdrl.f, a Fortran program written by E. Abola and supplied on the Brookhaven Protein Data Bank tapes, was used to calculate the chi values of the amino acids.

VIEW [15], a molecular graphics program written by L.D. Bergman (University of North Carolina, Chapel Hill), was also used in study of the helix-helix contacts.

Programs and scripts, summarized in the Results section, implemented under the VIEW system were used to evaluate details of the predefined low-angle and perpendicular helix-helix contacts. VIEW is available for public use via anonymous ftp (ftp.cs.unc.edu or 152.2.128.159).

Most of the computing was done on an SGI4D/440 or ESV workstation at the Macromolecular Graphics Shared Resource, Duke Comprehensive Cancer Center.

RESULTS AND DISCUSSION

For the most conclusive comparisons one would prefer to superimpose all possible examples of a contact and study the images as a whole (for instance superimpose all anti-parallel helix-helix contacts). This however requires a large net of comparisons since each contact covers a large surface area and many different amino acids, and an unbelievably large database is required before statistical significance on any one characteristic could be calculated. In order to keep the comparisons within reach, the contacts were broken down into smaller pieces. Thus, a single piece can be compared to others, and statistic preferences of a small number of different characteristics could be calculated for the set.

Low angle helix-helix contacts from a variety of proteins with different overall structures were broken down into small pieces, "puzzle pieces", as described below. Each piece consisted of a small number of residues from one turn of a helix, and two pieces interacted across the contact to form a "puzzle pair". One anti-parallel helix-helix contact consisted of several layers of interacting puzzle pairs. Such pairs could be isolated, superimposed and compared. Perpendicular helix-helix contacts consisted of one puzzle pair which covered the center of the contact and a second type of puzzle pair which covered the edges of the contact. In this case, only one instant of the central puzzle pair was located and superimposed per helix-helix pair.

The definition and location of the puzzle pieces of the two different types of contacts studied (anti-parallel and perpendicular) follow the same concepts, but very different procedures. Both procedures will be described along with the development of the individual puzzle piece definitions.

1. Contacts: Locating Helix-helix Contacts

A. Anti-parallel Helix-helix Contacts.

First, the Define_S program was used to locate helix pairs which were at a low-angle to one another, anti-parallel (an omega angle of -180° to -140°). Helices were determined to be in contact if there was a decrease in the solvent accessible surface area for the two helices as calculated together versus individually. Helix pairs with contact surface area below 100 Å^2 were excluded from this study. Such contacts were either between helices that stacked end-to-end or between helices which had only the tips of a few sidechains in contact and thus were positioned as a result of other parts of the protein. In the 96 proteins, a total of 92 anti-parallel helix-helix contacts were found, 79 of which were usable in this study.

B. Perpendicular Helix-helix Contacts.

Define_S and Access were also used to locate perpendicular contacts. Helices with an angle of association between 85 and 125 and between -85 and -125 and a contact surface area between 325 and 1000 Å^2 were considered. (The limits of the contact surface area cutoff is discussed in Section 3B.) In the 597 proteins studied, 319 had perpendicular helices, for a total of 299 perpendicular helix-helix contacts.

2. Definitions: Puzzle Pieces and Puzzle Pairs

A. Anti-parallel Helix-helix Contacts.

The overall geometries of the anti-parallel helix-helix contacts were broken down into individual turns of a helix (one turn is four consecutive residues A, B, C, D), plus the turn of the neighboring helix which it contacted.

The first script included four specific steps.

i. A search located all of the C(α) of helix 1 which were less than 7.5 Å away from an C(α) of helix 2. This located the contacts between the helices in classic low-angle contacts, but did not include atoms from the backs of the helices, or from helix ends that bent away from one another.

ii. Each C(α) of one turn from helix 1 has the potential to come within 7.5 Å of two different turns from helix 2. The list is sorted such that parallel contacts are grouped together. The group with the shortest average contact distance is labeled as "primary" contacts (approximately horizontal), while the other group is called the "secondary" contacts (vertical).

iii. For each contacting C(α), the four C(α) (2 from each helix) which have the

shortest total separations across the contact, are called the "central residues" of that turn. The two central residues are labeled B and C in amino-acid sequence order.

iv. Of the two helices in a turn pair, that one is identified as "1" which has its B residue most central to the contact (as defined by the smallest angle B1; the angle between C(α) of B1, the projection of B1 onto its helix axis, and the projection of this point onto the other helix axis, Fig. 1). That residue, B1, is often the one with the most interactions.

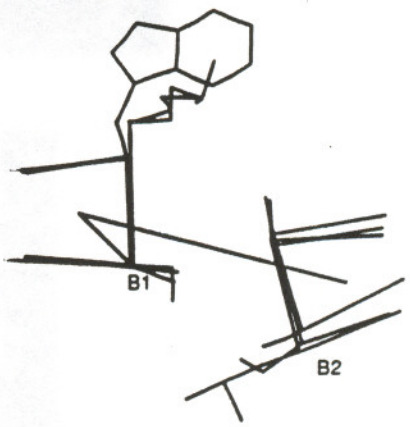

Figure 1. Overlay of 2 square puzzle pairs: 2ccy 100&40, 2ccy 51&89, and 1prc 99&37. The 2 central sidechains of each turn, the B residue labels, and the C(α) backbone are shown. The lines of the B1 angle are drawn for helix 1.

A set of angles and distances which fully define each contact pair was then calculated. A list of the measurements includes: (1) distance across the contact at B1, (2) angle B1 (residue B from helix 1) and angle B2 (residue B from helix 2), and (3) fractional offset (defined below). Many of the measurements require the projection of points onto the helix axes. A short axis was generated centered around the central B residue of each turn. This assured an axis that was meaningful at the point of interest, the turn.

A fractional-offset was defined as the vertical offset of the turn on helix 1 relative to the primary and secondary turns on helix 2. The average position of B1 and C1 was projected onto the line midway between the axis of helix 1 and the axis of helix 2; and the average position of B2 and C2 for both the primary and secondary turn were each also projected onto this mid-line.

All of these measurements along with the residue type of each position were recorded in a large database which could then be analyzed statistically by other programs or which could be searched by other VIEW scripts to display related turn pairs. The turn pairs (puzzle pairs) were then overlaid onto an idealized turn 1 and were then analyzed according to: (1) overall contact geometry, (2) residue content, (3) sidechain conformation of each residue involved, (4) resulting geometry of contact, etc.

B. Perpendicular Helix-helix Contacts.

The overall geometries of the perpendicular helix-helix contacts were broken down into two types of puzzle pieces. The first type consists of the central portion of the contact, a triangular surface on each helix. Each triangle highlighted three C(α)'s in either a (i, i+1, i+4) or (i, i+3, i+4) sequence order. The second type consists of the ends of the contact, are different from the central pieces, and will be studied separately.

i. In defining the primary, central puzzle pieces, the initial search located all of the C(α) of helix 1 less than 7.5 Å away from an C(α) of helix 2, and sorted the contacts by distance. The three shortest contacts between unique C(α) determined the triangles for the individual helices (Fig. 2).

Figure 2. Perpendicular helices from 2ccy, 10-20&95-111. The C(α) backbone is solid; the triangles, stippled, represent the 3 C(α) closest to the contact. The triangles overlap; the front triangle packs one of its vertices (top) into the center of the second (back) triangle.

ii. The distance between the helices in contact was measured across the line which was the common perpendicular of the two helix axes.

iiii. This common perpendicular and the helix axis on helix 1 was used to superimpose the contacts for visualization and further examination.

A similar set of geometrical measurements (as recorded for the anti-parallel contacts) are needed to describe the perpendicular puzzle pairs. These measurements include: (1) distance across the contact, (2) angle of the residues to the contact center, and (3) translation of the residues from the contact center.

3. Process: Derivation of the Patterns and Definitions.

A. Anti-parallel Helix-helix Contacts.

i. The definition of puzzle pieces was based on the relative residue positions. Such definitions based on the relative angle, separation and translation of the C(α)

positions were both simple and intuitive.

ii. The interaction between puzzle pieces was studied as pairs of single opposing turns, rather than sets of threes or fours, as an initial simple approach and for an initial significant number of occurrences of puzzle pairs. The turn pairs which were the closest had the most surface area in contact and thus were assumed to be most influential on one another. The recorded secondary contacts can be used for further definition of contact pairs, which with a reasonable number of occurrences can be studied along with the primary contacts as triplets, etc. to build back up the contact.

iii. In defining the fractional offset, as well as other geometrical properties, one must keep the measurement in perspective with the contact. Initially the residue positions were projected onto the individual helix axes. This gave confusing results, since in some cases the helices were quite distant and at varying roll angles to one another. Using the midline common axis averaged out these differences and related the translation of the helices with respect to the contact rather than to the helices themselves.

B. Perpendicular Helix-helix Contacts.

i. In definition of the perpendicular contacts, a finer filter was required for the contact surface area. Our study is of the perpendicular helices which cross each other and not the helices which meet at their ends or are connected by only a few residues. To remove this group from the list, a minimal surface area in contact of 300 $Å^2$ was required. An upper limit of 1000 $Å^2$ was placed to remove the perpendicular helices which bent to form a lower contact angle and a larger contact area at one end. Final acceptance of the perpendicular helix-helix contacts was made through visual inspection.

ii. Definition of the puzzle pieces was more complicated for the perpendicular helix-helix contacts due to the fact that such contacts are not repetitive internally. The center of the contact was separated as one (the "central") puzzle pair and the edges of the contacts were treated as second (the "edge") puzzle pieces and interacting pairs. The central and edge puzzle pairs are quite different; however, the edge pieces along the N-termini of helix 1 can be compared with those along the N-termini of helix 2.

A triangle of residues (3 C(α)) were used rather than a pair of C(α) or a

parallelogram (4 C(α)), because in sorting the closest contacts between the two helices the first unique ones always described a small, compact triangle of the sequence order (i, i+1, i+4) or (i, i+3, i+4). The next unique contact fell in any number of places which did not form a consistent geometric figure that could be compared among many contacts. This central triangle, unlike a pair of C(α), covered the major portion of the contact, and thus could be defined as the simplest central puzzle piece. This suggested that the three closest residues were sufficient in forming a unique contact, and the others were involved in stabilization.

iii. Superposition of the triangular puzzle pairs of the perpendicular helices was based on the helix axis and common perpendicular of the contact, rather than the C(α) positions. Superposition due to the C(α) resulted in an overlay of the triangles, without the overlay of the helices, thus, showing a distorted picture of the contacts. The superposition of one helix axis and the common perpendicular resulted in the overlay of the contact, with the relative position of the triangles as the variable. Thus, the triangles were studied to relate their different properties (ie. residue type, closest C(α) - C(α), and smallest C(α) angle) and their respective positions to the overall contact.

4. Results: Puzzle Piece Characteristics

A. Anti-parallel Helix-helix Contacts.

i. Clusters. In the preliminary study of anti-parallel helix-helix contacts, 79 helix-helix pairs were analyzed, and 218 contacting puzzle pairs were identified. 16% of the puzzle pairs were listed with only primary and no secondary turn pairs, showing that the primary turn pair is at the end of a helix. For primary puzzle pairs, the spatial distribution of turn 2 relative to turn 1 showed a low, but significant clustering. Fig. 3 shows the plot of B1 angle versus B2 angle. The most prominent grouping of the B1 angle fell within the range of 20° to 40° degrees, and the most prominent peak of the B2 angle fell within the range of 70° to 95°.

Clusters were expected, since constrained systems such as a peptide chain and sidechain conformation could not assume all possible conformations. Steric hinderance was expected to be the main reason for the exclusion of some positions of turn 2. However, it was also expected that the most populated positions of turn 2 would be more stable, having more positive packing influences and interactions.

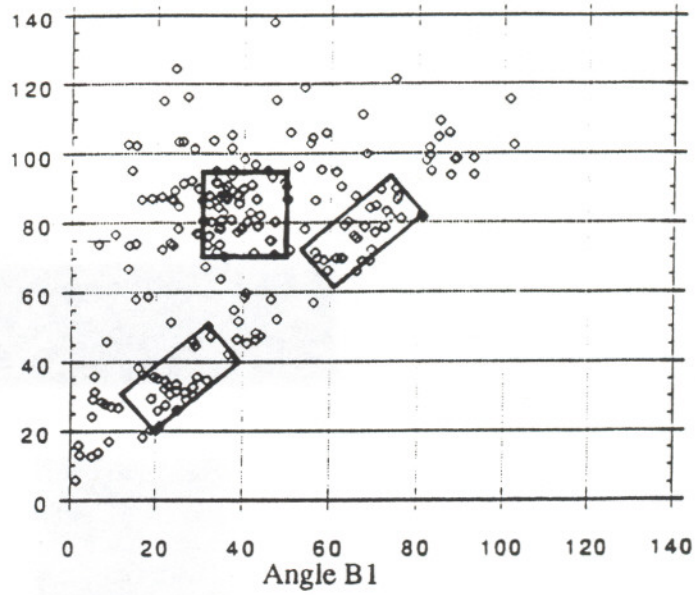

Figure 3. A plot of the B1 angle versus the B2 angle of defined anti-parallel puzzle pairs. The clusters of puzzle pairs are highlighted: square puzzle pairs B1 angle 30° - 50°, B2 angle 70° - 95°; B diagonal puzzle pairs B1 angle 10° - 30°, B2 angle 30° - 40°; C diagonal contacts B1 angle 60° - 80°, B2 angle 65° - 90°

ii. Square Contacts. The most prominent cluster seen in Fig. 3 is classified as square contacts (Fig. 1). Isolating this cluster of puzzle pairs revealed a group of turn pairs with distance separation between the helices of 7.5 - 9.5 Å and with a characteristic fractional separation of 0.0 - 0.05. These contacts are called square contacts since the face of helix 1 (peptide bond between the B1 and C1 residue) is parallel to the face of helix 2, like the sides of a square. These square puzzle pairs were seen to group by the residue type or more importantly by the sidechain conformation present in the central B1 position, which suggests, in this case, that the B1 residue is involved in the specific interatomic contacts that place turn 2. The preferred residues in the B1 position for these clustered puzzle pieces were leucine, alanine, methionine, phenylalanine and serine. Additional requirements for the neighboring sidechains (C1, B2 and C2) and their conformations were seen. C1 and B2 are often long sidechains which bend and pack around C2 and B1 respectively. C2 is often a small hydrophobic residue which tightly packs along the helix backbone between B1 and C1, or a residue with a long sidechain to pack along C1. (Statistics of such secondary interactions will be studied on the larger database, now being completed.)

Many different sidechains can occur in the B1 position of a square contact; however, as the size increases or the conformation of the sidechain shifts, increasingly larger changes are made in the contact.

Alanine in B1 results in the most ideal square contact, as shown in Fig. 2. Most of the puzzle pieces with the smallest allowed distance separation (7.5 - 8.0 Å)

belong to this group. To complete the contact, C1 and B2 are often long sidechains which bend and pack around C2 and B1 respectively. C2 is often a small hydrophobic residue which tightly packs along the helix backbone between B1 and C1.

Next in importance is the occurrence of leucine in the B1 position of a square contact. It often assumes a chi 1 = -60, and either packs along the B2 sidechain, or with a slightly smaller B1 angle it packs along the backbone between the B2 and C2 residue. Phenylalanine also is common in the B1 position with a chi 1 = -60. The C(β) of phenylalanine often packs against the backbone between B2 and C2, while its ring is packed along the B2 residue, which presents a flat parallel surface to the aromatic ring if it is a chi 1 = -60 conformer. Although much less common, phenylalanine and leucine with a chi 1 = 180 in B1 can also make another kind of square contact, which extends the distance separation out beyond that typical for square contacts to 8.5 to 11.0 Å. This requires long extended sidechains in several of the other contact positions to form a strong contact; these are often leucines.

iii. Specific Cases (ROP-type and FHA-type). [16] The very tightest helix contacts (<8.5 Å) were examined specifically, because of their potential role in encouraging unique packing. Looking at only the tight contacts of helices, the ideal square puzzle pair with alanine or other small residue in B1 (previously described) and another tight square puzzle pair with alanine in C2 were found (discussed in more detail in Ref. [16]). Close ROP-type puzzle pairs (ideal square with alanine in B1) have a distance separation approximately 8 to 10 Å, and a small fractional separation of 0 to 0.2. This fractional separation fits B1 into the triangle formed by B2 and C2 of turn 2 and the B1 of the following turn. Two consecutive puzzle pairs can stack to form a tight contact, or as found in 2rop, 6 such puzzle pairs may stack, alternating the B1 residue from helix 1 to helix 2. In the case of the FHA-type puzzle pair (tight square with alanine or serine and seldomly threonine or cysteine in C2) the distance between helix axes is 7.5 to 8.5 Å with a fraction separation of 0.3 to 0.5. With this fraction separation the C(α) of the B residues are fairly distant, yet the C(β)s point towards one another in a distinct diagonal layer, and C2 packs into the triangle between B1, C1 and B1 of the preceding turn. It is common to have two adjacent C2 alanine type puzzle pairs at the close contact of two straight, anti-parallel helices; 2fha has 5. Such stacks of these puzzle pairs, alternating from helix 1 and

helix 2, again form long close contacts between the helices, resembling an anti-parallel coiled-coil with the leucines replaced by alanines.

iv. Diagonal Contacts. Highlighted during graphical display and suggested by the B1 - B2 angle plots of Fig. 3, two other clusters were seen at (1) angle B1 of 10° to 30° and angle B2 of 30° to 40°, and (2) angle B1 of 60° to 80° and angle B2 of 65° to 90°. Both of these clusters resulted in a diagonal contact with either the B1 and B2 residues forming the major contact, as in the case of cluster (1) (Fig. 4), or the C1 and C2 residues contacting, cluster (2). These contacts resulted in longer distance separations between the helices than for the square contacts: 8.5 - 10.0 Å and 9.0 - 11.0 Å, respectively. The primary fractional separation is typically higher, approximately 0.25 - 0.35 and 0.40 - 0.55, respectively. Isoleucine, leucine, phenylalanine and tyrosine were particularly common as the residues interacting across diagonal contacts.

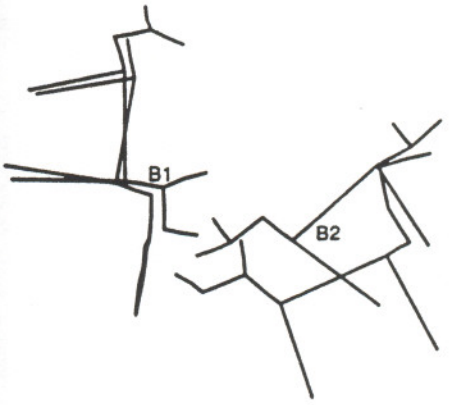

Figure 4. Overlay of 2 diagonal puzzle pairs B's touching: 256b 72&30 and 1ann 251&272. The 2 central sidechains of each turn, the B residue labels, and the C(α) backbone are shown.

The secondary contact of a diagonal turn was typically a distant square contact. The sidechains of the secondary turn reach up or down to the neighboring sidechains of the diagonal contact. It is such neighboring contacts that add stability to the diagonal contact, as well as restrain the possibilities of the central diagonal residue. In hindsight, the diagonal turn is the result of a square contact which has an increased fractional separation, such that the square contact becomes the secondary contact and the diagonal contact becomes the primary. If the offset is up relative to helix 1 (in the N to C direction), a C1 - C2 diagonal pair will result; if the offset is down, a B1 - B2 diagonal pair will result. The larger database now being studied will allow a statistical look at the secondary contacts (which are often square contacts) surrounding the diagonal pairs.

v. Residue Types. Table 1 lists the preferences for residues involved in anti-parallel contacts, and Table 2 lists the number of each residue type found in each of

Table 1
Preferred Residues in Anti-parallel and Perpendicular Contacts

AA	Helical Residues	Ave. %	Anti-parallel count	Anti-parallel pref.	Perpendicular count	Perpendicular pref.
ALA	467	11.5	112	0.801	310	*0.788*
ARG	174	4.28	65	1.25	169	1.15
ASN	181	4.45	33	0.61	119	0.781
ASP	250	6.15	43	*0.575*	151	*0.717*
CYS	60	1.48	19	1.06	44	0.869
GLN	169	4.16	40	0.792	102	*0.717*
GLU	321	7.89	56	*0.583*	172	*0.637*
GLY	214	5.26	45	0.703	124	*0.689*
HIS	86	2.11	33	1.28	72	0.997
ILE	212	5.21	86	1.36	264	**1.48**
LEU	380	9.34	220	**1.94**	533	**1.67**
LYS	339	8.34	55	*0.543*	160	*0.561*
MET	92	2.26	58	**2.11**	103	1.33
PHE	159	3.91	72	**1.51**	240	**1.79**
PRO	91	2.24	12	0.441	60	0.783
SER	246	6.05	48	0.653	147	*0.71*
THR	199	4.89	54	0.908	147	0.878
TRP	48	1.18	31	**2.16**	83	**2.06**
TYR	114	2.8	44	1.29	172	**1.8**
VAL	265	6.52	90	1.14	250	1.12
	4067		1216		3422	

* Bold and italics are >3 std. dev. from 1.

Table 2
Preferred Residues in the Central Residue Positions in Defined Anti-parallel Puzzle Pieces

AA	B1 pref.	C1 pref.	B2 pref.	C2 pref.
ALA	1.13	0.91	1.13	1.36
ARG	1.15	0.75	1.15	1.25
ASN	0.60	0.31	1.40	0.30
ASP	0.36	0.82	0.51	0.44
CYS	0.60	0.93	1.21	1.51
GLN	1.18	1.10	1.29	0.43
GLU	0.62	0.64	0.51	0.40
GLY	1.02	1.04	0.51	1.02
HIS	0.64	0.65	1.48	0.64
ILE	1.54	0.88	1.20	1.37
LEU	1.63	**1.76**	**1.96**	**1.72**
LYS	0.43	0.77	0.54	0.48
MET	2.17	1.21	1.38	1.78
PHE	1.94	1.05	0.91	0.80
PRO	0.40	0.41	1.40	0.60
SER	0.96	1.43	0.59	0.44
THR	0.82	1.49	0.91	1.00
TRP	1.51	2.71	1.13	2.65
TYR	0.64	1.14	0.96	**2.39**
VAL	0.96	0.70	0.75	1.23

* Bold is >3 std. dev. from 1.

the central residue positions in puzzle pieces (B1, C1, B2, C2) and highlights those preferred relative to their percentage in helical structure overall (AH residues). The preferred residues in anti-parallel helix contacts are: leucine, methionine, tryptophan, phenylalanine and isoleucine. The highest occurrences of residues are leucine (220), alanine (112) and valine (90) out of 1216 residues. Arginine also has a slight preference for involvment; it is the only hydrophilic residue preferred. The highly preferred methionines are often the last residue of an i+4 row at the edge of the contact. More specific correlations of preferred residue types and preferred residue conformations will obviously be found once the dataset is broken down into the clusters of similar geometry groups or clusters with a specific residue in the B1 position, for instance. This will be performed on the larger dataset now being studied.

B. Perpendicular Helix-helix Contacts.

Table 2 lists the preferences for residues involved in perpendicular contacts. Like anti-parallel contacts, leucine, tryptophan, phenylalanine and isoleucine are preferred in perpendicular contacts. Tyrosine is also preferred in perpendicular contacts. Perpendicular contacts have a decreased occurrence of many of the charged residues along with alanine, glycine and serine.

A tabulation of the statistics on the defined puzzle pairs for perpendicular contacts is in progress.

CONCLUSION

General characteristics of anti-parallel helix contacts include: (1) the importance of the B1 residue in establishing the B1 angle (rotation of helix 1 about its axis in the contact), (2) the importance of the C2 residue in establishing the B2 angle (rotation of helix 2 about its axis in the contact), and (3) the involvement of both in setting the fractional separation between the helices. With this in mind, a list of residues preferred in each central position of turn pairs of known geometry (square vs. diagonal B1 - B2 vs. diagonal C1 - C2) is given in Table 2. A small residue in B1 sets a tight, square contact; a small residue in C2 often does likewise. A step-wise progression of increasing size of B1 in square contacts results in an decrease in the B1 angle, an increase in the distance separation and fractional separation. The

requirement of different sidechains in B2 also changes with this progression. A larger residue in B1 with a chi of 180 sets a crooked, distant or diagonal contact. A C1 residue with a chi 1 of -60 can pack around C2, unless C2 is long and has a chi 1 of 180; such a C2 will form a long parallel packing contact with C1 (chi 1 = 180) instead.

These clusters of puzzle pairs show that proteins do use common packing motifs and geometries to position their secondary structural elements uniquely. These packing motifs are assumed to have characteristics which give them a stabilizing factor, be it a preferred unique arrangement or simply a reduction in other possible geometries and conformations.

This initial study included only a small number of helix-helix contacts. A significantly larger database (now being processed) will result in additional statistically significant characteristics of the contacts. With a larger number of puzzle pairs it is expected that the clusters will become more defined, that the residues will show strong primary occurrences (i.e. alanine in the B1 position prefers a tight square contact) and also strong secondary occurrences (i.e. if a turn with a leucine in B1 contacts a turn with an isoleucine in B1, it will prefer to form a diagonal contact). Additional studies are now underway, using a database of 597 proteins.

The accumulation of such general characteristics of perpendicular helix contacts is still underway. It is expected to show not only particular characteristics for perpendicular contacts, but also highlight the differences of the perpendicular from the anti-parallel contacts.

Understanding the specifics of these tertiary interactions, will allow their application to (1) continued comparison of protein structures and modeled structures, (2) revising the amino acid sequence of designed proteins for specific, unique tertiary contacts, (3) design de novo of additional protein structures, and in time, (4) predicting 3D structure from primary sequence.

A. Analysis.

The database of sidechains, sidechain position and sidechain conformation can be used during the model building of structure determination. In studying modeled sequences, such as the original Felix models, the data from native proteins can highlight which characteristics of the models are far from the ideal native structures.

For instance, in the hydrophobic core of Felix, changes were suggested on the overall surface of helix-helix contacts: an increase in the area of contact and an increase in the interdigitation of sidechains. In the helix-helix contacts of Felix, changes were suggested on actual amino acid composition of puzzle pieces: an increase of the use of preferred puzzle pair residues and the common geometries.

B. Design.

The statistics gathered on preferred puzzle pairs could be used to design a protein sequence based on a new structural model. It is proposed that using preferred puzzle pieces would first constrain the tertiary interaction and that completing the contact with other common puzzle pieces would stabilize it. Choosing puzzle pieces with the preferred residues and residue conformations for such helix-helix contacts would increase the likelihood of the choosen structure, and in turn reduce the alternate conformations of the residues in the actual protein structure. Building on a defined three-dimensional geometry with the most preferred residues in each position for that three-dimensional (and not just two-dimensional) structure will greatly increase the chance of acquiring that structure in a stable form. (An example of the detailed look at the redesign of Felix DEL 3 is given in Ref. 3.)

With the completion of the study of perpendicular helix-helix contacts, one is expected to distinguish anti-parallel from perpendicular tendencies. One may then be able to use specific sidechain conformations preferred in that helix-helix contact in building protein structures, or even design in preferred residue types and residue neighbors for that specific helix-helix contacts, or predict tertiary interactions from primary sequence.

It is also expected that other protein motifs, other α-α, α-β, β-β, etc., will have defineable puzzle pieces and puzzle piece interactions. A statistical study of these other puzzle pieces will create another database of the preferred residues and interactions of those tertiary contacts, creating additional information for the distinguishing of the contacts and of the design of such contacts.

ACKNOWLEDGMENTS

This work was done with the support of NIH research grant GM-15000; NASA

contract NRA-91-OSSA-18; the GRIP project, NIH research grant RR02170, at University of North Carolina, Chapel Hill for LDB; and the use of the Macromolecular Graphics Shared Resource of the Duke Comprehensive Cancer Center.

1. Handel, TM, SA Williams, WF DeGrado. (1993) Metal Ion-Dependent Modulation of the Dynamics of a Designed Protein. Science. 261 879-885.

Regan, L, A Rockwell, Z Wasserman, W DeGrado. (1994) Disulfide Crosslinks to Probe the Structure and Flexibility of a Designed Four-helix Bundle Protein. Protein Science. 3 2419-2427.

Handel, TM, SA Williams, D Menyhard, WF DeGrado. (1993) Introduction of a Trp Residue into alpha4 as a Probe of Dynamics. J.Am.Chem.Soc. 115 4457-4460.

2. J. S. Richardson et al. (1992) Looking at Proteins: Representations, Folding, Packing and Design. Biophysics Journal. 63 1186-1191.

3. Gernert, KM. (1994) Structural Characterization and Redesign of Felix, a Designed Four-helix Bundle Protein. Chapter 1 and 2, Dissertation Thesis, Duke University, Department of Biochemistry.

4. Hu, JC, NE Newell, B Tidor, RT Sauer. (1993) Probing the Roles of Residues at the "e" and "g" Positions of the GCN4 Leucine Zipper by Combinatorial Mutagenesis. Protein Science 2 1072-1084.

5. Monera, OD, NE Zhou, CM Kay, RS Hodges. (1993) Comparison of Anti-parallel and Parallel Two-Stranded Alpha-helical Coiled-coils. J.Biol.Chem. 268 19218-19227.

Zhu, B-Y, NE Zhou, CM Kay, RS Hodges. (1993) Packing and Hydrophobicity Effects on Protein Folding and Stability: Effects of Beta-branched Amino Acids, Val and Ile, on the Formation and Stability of Two-stranded Alpha-helical Coiled-coils/ Leucine Zippers. Protein Science. 2 383-394.

6. Harbury, PB, T Zhang, PS Kim, T Alber. (1993) A Switch Between Two-, Three- and Four- Stranded Coiled-coils in GCN4 Leucine Zipper Mutants. Science. 262 1401-1407.

7. Richardson, JS, and DC Richardson. (1989) The de novo Design of Protein Structures. TIBS 14 304-309.

8. Fezoui, Y, DL Weaver, JJ Osterhout. (1994) de Novo Design and Structural Characterization of an Alpha-helical Hairpin Peptide- A Model System for the Study of Protein Folding Intermediates. PNAS. 91 3675-3679.

9. Kuroda, Y, T Nakai, T Ohkubo. (1994) Solution Structure of a de Novo Helical Protein by 2D-NMR Spectroscopy. J.Mol.Biol. 236 862-868.

10. Lifson, S, C Sander. (1979) Anti-parallel and Parallel Beta-strands Differ in Amino Acid Residue Preferences. Nature. 282 109-111.

Salemme, FR. (1983) Structural Properties of Protein Beta-sheets. Prog. Biophys. Molec. Biol. 42 95-133.

Chothia, C, J Janin. (1981) Relative Orientation of Close Packed Beta-Pleated Sheets in Proteins. PNAS. 78(7) 4146-4150.

11. Sibanda, BL, JM Thornton. (1985) Beta Hairpin Families in Globular Proteins. Nature. 316 170-174.

12. Edwards, MS, MJE Sternberg, JM Thornton. (1987) Structural and Sequence Patterns in the Loops of Beta-alpha-beta Units. Prot. Eng. 1:(3) 173-181.

Janin, J, C Chothia. (1980) Packing of alpha-helices onto Beta Pleated Sheets and the anatomy of alpha/beta Proteins. JMB. 143 95-128.

13. Chothia, C, M Levitt, D Richardson. (1981) Helix to Helix Packing in Proteins. JMB. 145, 215-250.

14. Chothia, C, M Levitt, D Richardson. (1977) Structure of Proteins: Packing of α-helices and Pleated Sheets. PNAS. 74(10) 4130-4134.

15. Bergman, LD. (1993) VIEW: A System for Prototyping Scientific Visualizations. Dissertation Thesis, University of North Carolina, Chapel Hill, Department of Computer Science.

16. Gernert, KM, TH LaBean, MC Surles, JS Richardson, and DC Richardson. (1995) The Alacoil: A Very Tight, Anti-parallel Coiled-coil of α-helices. Protein Science (in press).

17. Bernstein, et al. Protein Data Bank, Chemistry Dept., Bldg. 555, Brookhaven National Lab., P.O. Box 5000, Upton, NY 11973-5000.
e-mail: pdb@bnl.bitnet, pdb@bnl.gov

18. Hobohm, U, M Scharf, R Schneider, C Sander. (1992) Selection of Representative Protein Data Sets. Protein Science. 1 409-417.

19. Hobohm, U., C. Sander. (1994) Enlarged Representative Set of Protein Structures. Prot. Sci. 3 522-524. Revised list of June 1995.

20. Richards, FM, CE Kundrot. (1988) Identification of Structural Motifs from Protein coordinate Data: Secondary Structure and First - Level Supersecondary Structure. Proteins. 3 71-84.

21. Richardson, JS, D.C. Richardson. (1988) Amino Acid Preferences for Specific Locations at the Ends of alpha Helices. Science. 240 1648-1652.

22. Lee, B, FM Richards. (1971) The Interpretation of Protein Structures: Estimation of Static Accessibility. J. Mol. Biol. 55 379-400.

23. Ponder, JW, FM Richards. (1987) Tertiary Templates for Proteins: Use of Packing Criteria in the Enumeration of Allowed Sequences for Different Structural Classes. J. Mol. Biol. 193 775-791.