PROTEINS: Structure, Function, and Genetics 22:119–131 (1995)

An Analysis of Side Chain Interactions and Pair Correlations Within Antiparallel β-Sheets: The Differences Between Backbone Hydrogen-Bonded and Non-Hydrogen-Bonded Residue Pairs

Merridee A. Wouters and Paul M. G. Curmi
Initiative in Biomolecular Structure, School of Physics, University of New South Wales, Sydney 2052, Australia

ABSTRACT **Cross-strand pair correlations are calculated for residue pairs in antiparallel β-sheet for two cases: pairs whose backbone atoms are hydrogen bonded together (H-bonded site) and pairs which are not (non-H-bonded site). The statistics show that this distinction is important. When glycine is located on the edge of a sheet, it shows a 3:1 preference for the H-bonded site. The strongest observed correlations are for pairs of disulfide-bonded cystines, many of which adopt a close-packed conformation with each cystine in a spiral conformation of opposite chirality to its partner. It is likely that these pairs are a signature for the family of small, cystine-rich proteins. Most other strong positive and negative correlations involve charged and polar residues. It appears that electrostatic compatibility is the strongest factor affecting pair correlation. Significant correlations are observed for β- and γ-branched residues in the non-H-bonded site. An examination of the structures shows a directionality in side chain packing. There is a correlation between (1) the directionality in the packing interactions of non-H-bonded β- and γ-branched residue pairs, (2) the handedness of the observed enantiomers of chiral β-branched side chains, and (3) the handedness of the twist of β-sheet. These findings have implications for the formation of β-sheets during protein folding and the mechanism by which the sheet becomes twisted.**

Key words: antiparallel β-sheet, twist, protein folding, side chain interactions, branched amino acids, cystine-rich proteins, side chain packing

INTRODUCTION

β-Sheets form an integral part of protein structures, often occurring in the interior core region. Of the regular secondary structure elements, only β-sheets involve long-range (in terms of sequence) interactions. In forming a sheet, stretches of amino acids align themselves against each other in a unique register. In this process, it is likely that interstrand side chain–side chain interactions determine the observed strand register. Thus, if we understood the factors that favored certain residue pairs as cross-strand neighbors, then we would make significant progress in the understanding of β-sheet formation. Furthermore, although protein design has resulted in the formation of molten globules with the correct topology, the final condensation to the native state has not yet been achieved.[1] It is highly likely that interactions between specific side chains must be correct before the protein will properly pack together. Thus, by examining the cross-strand residue pair correlations and preferred pair structures, we seek to identify the factors that determine pair interactions.

Another weakness in our understanding of β-sheets is our failure to accurately predict their occurrence from protein sequence data. Current algorithms that predict protein secondary structure tend to have a higher success rate in predicting α-helices than β-strands. A likely explanation for this is that helices form locally whereas interactions within β-sheets are nonlocal by nature. A reliable set of interstrand residue pair correlations should facilitate the development of better algorithms for the prediction of β-sheets. An algorithm based on interstrand correlations would allow the prediction of which residues form a strand as well as which strands interact with each other. Such information would ultimately lead to the prediction of the topology of the protein.[2]

Previous studies of interstrand residue pair correlations in β-sheets[3,4] found that the pair correlations differed from random. These studies were limited by the size of the structure database available at the time. In this paper, we reexamine the pair correlations in antiparallel β-sheets. The database has grown sufficiently to allow the use of a nonredundant dataset[5] without grouping residues.

Received November 4, 1994; revision accepted January 19, 1995.

Address reprint requests to Paul M. G. Curmi, Initiative in Biomolecular Structure, School of Physics, University of New South Wales, Sydney 2052, Australia.

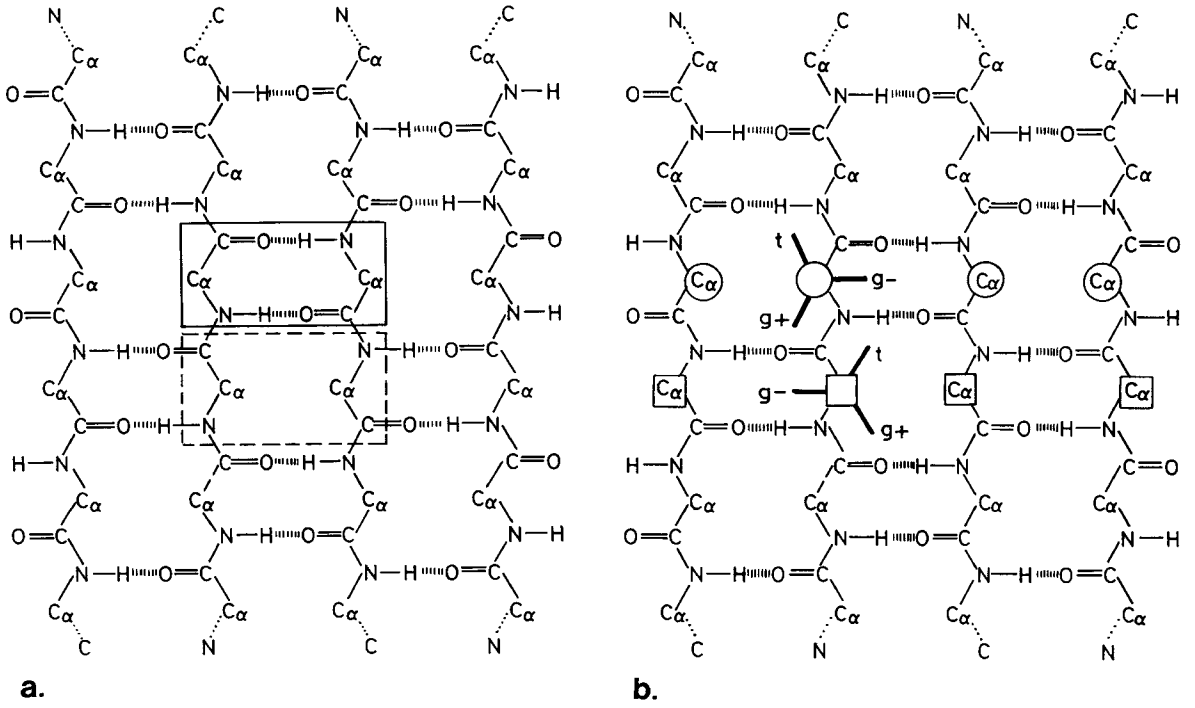

Fig. 1. (**a**) An antiparallel β-sheet showing the H-bonded (unbroken box) and non-H-bonded (dashed box) sites. Each residue within a site is equivalent due to the 2-fold symmetry. Side chain rotamers for pairs of residues in the two sites are different. (**b**) Rotamers of non-β-branched residues are shown for one member of the pair in the H-bonded (○) and non-H-bonded (□) sites. The rotamers are indicated by g−, gauche−; g+, gauche+; and t, trans. Side chains of circled C_αs point out of the page while those of boxed C_αs point into the page.

We have investigated whether the backbone site influences pair correlations. In antiparallel β-sheets, there are two distinct residue pair types (Fig. 1a): the pair where the backbone atoms of the residues are hydrogen bonded to each other (H-bonded site) and the pair where the backbone atoms of the residues are not hydrogen bonded to each other (non-H-bonded site). This distinction has been noted by other authors: the "large" and "small" hydrogen-bonded rings of Salemme and Weatherford[6] and the "narrow pair" and "wide pair" of hydrogen bonds of Richardson and colleagues.[1] The backbone of each of the residues within a pair is equivalent, being related by a 2-fold axis perpendicular to the sheet. Physical differences between these pairs include (1) C_αs of members of H-bonded pairs are separated by about 5.5 Å whereas C_αs for non-H-bonded pairs are separated by about 4.5 Å; and (2) the preferred rotamer configurations for the two pair types will be significantly different (Fig. 1b). For example, in non-β-branched residues, the gauche-conformation points the C_β–C_γ bond towards its partner residue in the H-bonded site but away from its partner in the non-H-bonded site (Fig. 1b). Thus, in this study, we further classified antiparallel β-sheet residue pairs into H-bonded and non-H-bonded sites.

Finally, for the pairs that show significant correlations, we have examined the high stereochemical quality structures in our dataset in order to determine the physical basis of their interactions.

METHODS

Dataset

A set of 253 nonredundant protein structures[5] from the Brookhaven protein data bank[7] (PDB) was used, based on the December 1993 list from EMBL.[8] No pair of chains in the set exceeds 25% sequence identity over subsequence lengths of 80 residues or greater after optimal sequence alignment. On initially using this dataset, we found that certain structural motifs produced anomalously high pair correlations. These occurred in homologous structures such as the immunoglobulin folds where the three residues: lysine, cystine, and tryptophan, were found on adjacent strands in most examples even when the rest of the sequence had less than 25% identity. To remove this bias, no domain of any protein was included in our dataset if it had more than one corresponding pair in antiparallel β-sheet in common with any other homologous domain. The proteins or parts of proteins which were deleted are (1) the immunoglobulin folds: 1cd8, 1fc1, 1mam, 3cd4; (2) the blue copper proteins: 1aoz (domain 1) and 2aza; and (3) 1atx, 1bbt (chain 2), 1cau (chain B), 1cdt, 1nxb, 1rbp, 1smr, 2pia (domains 1 and 2),

2por, 2tbv (domain 1), and 9wga (domains 2, 3, and 4).

Secondary structure was determined using the program DSSP.[9] A program was written to extract all residues classified as being in extended sheet and to count pairs of residues in both sites in antiparallel sheet.

Pair Classification

We distinguished between residue pairs in the H-bonded and non-H-bonded sites by determining whether the C=O bond for a particular residue, i, was parallel or antiparallel to the vector joining the C_α of i to the C_α of its partner, j, by calculating the scalar product

$$\overrightarrow{C_iO_i} \cdot \overrightarrow{C_{\alpha i}C_{\alpha j}} \tag{1}$$

For H-bonded pairs, the result should be positive. For non-H-bonded pairs, it should be negative.

Pair Correlations

Pair correlations were calculated for all pairs of residues in the same fashion as Lifson and Sander.[4] For two residues i and j, the pair correlation is given by

$$g_{ij} = \frac{n_{ij}}{E_{ij}} \tag{2}$$

where n_{ij} is the number of times residue types i and j occur as a pair in the dataset and E_{ij} is the expected number of pairs of i and j if such a pair occurred randomly. Note that, for the purposes of correlation in antiparallel β-sheet, the pair ij is indistinguishable from the pair ji.

A $g_{ij} = 1$ is indistinguishable from random; $g_{ij} < 1$ indicates the pair occurs less frequently than expected randomly; $g_{ij} > 1$ indicates the pair occurs more frequently than expected.

E_{ij} is calculated using

$$E_{ij} = \frac{n_i n_j}{n} \tag{3}$$

where

$$n_i = \sum_{j=1}^{20} n_{ij} \tag{4}$$

is the number of pairs containing the residue type i with the summation carried out over the 20 amino acids.

The total number of counts is

$$n = \sum_{i=1}^{20} \sum_{j=1}^{20} n_{ij} \tag{5}$$

which sums over all residues. It is equal to twice the number of pairs.

Statistical Significance

We used the χ^2 test to determine statistical significance. Pair counts were compared with their expectation values by calculating

$$\chi^2 = \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \tag{6}$$

The χ^2 value was then used to determine the confidence level at which a pair count deviates from random.

Structural Comparisons

To determine physical interactions, a subset of high stereochemical quality structures was selected from the dataset based on the output of PROCHECK and the criteria given by Morris and co-workers,[10] viz the resolution should by 2.0 Å or better; the R-factor should be ≤ 0.20; 75% or more of the (ϕ, ψ) angles should be within allowed regions of the Ramachandran plot; and the pooled χ_1 angle deviations no greater than 16°. To compare structural features for a family of pairs, the high quality structures were superposed using the program SUPPOS[11] which implements the algorithm of Kabsch.[12] For the superposition, the structures were aligned using the backbone atoms of each member of the pair (i.e., N, C_α, C, and O). Superposed pairs were viewed on a graphics system using InsightII.[13] Figures were produced with the program MOLSCRIPT.[14]

RESULTS

Differences Between Edge and Interior Strands

Table I shows the distribution of residues in antiparallel β-sheet. The distribution is similar to that of previous studies.[15] It shows that disulfide-bonded cystines, aromatic and β-branched residues are more abundant in antiparallel β-sheets. The data in the left-hand side of Table II show how often residues are found in the interior of a β-sheet compared to the edge strand or in bulges. (An "edge" is defined as any position where the residue has a partner only on one side.) Hydrophobic and aromatic residues have a higher tendency to reside in the interior of the sheet (> 30%) than do charged or polar residues (< 25%). In this regard, glycine and cystine (but not cysteine) partition with the polar amino acids. Residues with the highest propensity for bulges are cysteine, aspartate, asparagine, glycine, and proline. Proline is never found in the interior because the proline ring incorporates the backbone. The proline ring lies in the plane of the sheet, sterically disrupting strand pairing as well as not being able to donate an N–H for hydrogen bonding.

All residues in the interior of the β-sheet are involved in both an H-bonded and a non-H-bonded interaction. In contrast, edge residues can occupy only one of these sites. The data on the right-hand side of

TABLE I. Residue Abundances in Antiparallel β-Sheet Compared to the Global Distribution*

	Global (%)	Antiparallel (%)	β_A
Ala	8.44 ± 0.12	5.97 ± 0.28	0.71
Arg	4.53 ± 0.09	4.51 ± 0.24	1.00
Asn	4.63 ± 0.09	2.78 ± 0.19	0.60
Asp	5.93 ± 0.10	2.92 ± 0.20	0.49
Cys	1.05 ± 0.04	1.11 ± 0.12	1.06
SS-cys	0.78 ± 0.04	1.49 ± 0.14	1.92
Phe	4.06 ± 0.09	6.17 ± 0.28	1.52
Gln	3.58 ± 0.08	3.15 ± 0.21	0.88
Glu	6.15 ± 0.10	4.41 ± 0.24	0.72
Gly	7.98 ± 0.12	5.13 ± 0.26	0.64
His	2.26 ± 0.06	2.19 ± 0.17	0.97
Ile	5.38 ± 0.10	7.83 ± 0.32	1.45
Leu	8.21 ± 0.12	9.11 ± 0.34	1.11
Lys	6.12 ± 0.10	5.31 ± 0.26	0.87
Met	2.13 ± 0.06	2.32 ± 0.18	1.09
Pro	4.70 ± 0.09	2.27 ± 0.18	0.48
Ser	6.10 ± 0.10	5.82 ± 0.28	0.95
Thr	6.03 ± 0.10	8.17 ± 0.32	1.36
Trp	1.49 ± 0.05	2.12 ± 0.17	1.42
Tyr	3.72 ± 0.08	5.91 ± 0.28	1.59
Val	6.72 ± 0.11	11.31 ± 0.37	1.68
Σ	52357	7231	

*A comparison of the global abundance of the amino acids with that found in antiparallel sheet. The last column is the propensity for antiparallel β-sheet. It is calculated by dividing the antiparallel frequency for a residue *i* by the global frequency. Residues which are more abundant in antiparallel-sheet than globally are the β-branched and aromatic residues as well as disulfide-bonded cystines. The elevated level of the latter is principally due to their abundance in small proteins with minimal hydrophobic cores. Residues which do not favor antiparallel β-sheet include proline and aspartate, followed by asparagine, glycine, alanine, and glutamate.

Table II show the results of a χ^2 test to determine whether the partitioning of edge residues between the H-bonded and non-H-bonded sites is significantly different from random. Four residues show a nonrandom partitioning at a confidence level greater than 0.95. Proline is restricted to the non-H-bonded site (as discussed above). Glycine strongly favors the H-bonded site with a ratio of approximately 3:1. The basis of this preference is not clear. Due to the flexibility of glycine, it may require further constraint if it is to reside on the edge of a sheet. Thus, it may favor the H-bonded site. Lysine and arginine also favor the H-bonded site.

Pair Preferences in Antiparallel β-Sheet

Table III shows the correlations for pairs of residues in the H-bonded and non-H-bonded sites. For each pair, the table entries give the pair correlation, g_{ij}, the confidence level given by the χ^2 test and the number of counts. Correlations favored or disfavored at the 99% confidence level are highlighted. The subsequent analysis will mainly concentrate on these pairs.

Strong correlations: electrostatic compatibility

Table IV lists the pairs that have a strong positive correlation, $g_{ij} > 2.0$, which are in the 99% confidence interval. What is striking about this list is that almost all residues are charged or polar and they are coupled in pairs that are compatible in terms of both electrostatics and physical dimensions (i.e., charged residues couple with either opposite charges or polar groups; long residues couple with long or medium length residues etc.). An examination of the negative pair correlations (99% confidence interval) in Table III shows that in all cases, the pairs involve one charged/polar residue coupled with a nonpolar residue. Thus, it appears that electrostatic compatibility is the strongest determinant of residue pairing.

The long polar residues Glu, Gln, Lys, and Arg appear to form a family in terms of their pair correlations (Table IV). Both Glu and Gln show favorable correlations with Lys and Arg in the H-bonded site, with weaker correlations in the non-H-bonded site (Table III). This weak backbone site discrimination may be due to the length and flexibility of these side chains which allows them to accommodate both types of backbone structure.

An examination of the high stereochemical quality structures shows that, in general, these residues do not form specific intrapair interactions. Glu–Lys form ion pairs in three out of eight and four out of 12 examples in the H-bonded and non-H-bonded sites, respectively, while Glu–Arg in the H-bonded site forms seven ion pairs in eight examples. High quality pairs between Gln–Lys and Gln–Arg show no particular pattern of interaction.

Similar observations are made for other high correlation polar residue pairs. In the H-bonded site, two out of eight Thr–Asn pairs are hydrogen bonded although six of the eight threonines are in the unusual gauche− rotamer which points the hydroxyl group towards its partner. In three of these cases where there is no direct hydrogen bond, the side chains are linked via a water molecule. None of the side chains in the four high quality Lys–Asp pairs in the H-bonded sites actually interact with each other. Similarly, no serine in the four high quality Ser–Ser pairs in the H-bonded site interacts with its partner. In contrast, all four Ser–Asn pairs in the non-H-bonded site form side chain hydrogen bonds, with serine in the trans rotamer and asparagine adopting gauche+.

In general, polar and charged residues are on the surface of the protein and are interacting with the solvent and other polar surface residues. The examples of strongly correlated pairs do not tend to show specific intrapair interaction patterns. Thus, the origin of these high correlations resides in more gen-

TABLE II. The Location of Specific Residues in Antiparallel β-Sheet*

	Residue locations			Edge residue distribution			
	Interior (%)	Bulge (%)	Edge (%)	H-bonded (%)	Non-H-bonded (%)	Total counts	χ^2
Ala	31.2 ± 2.1	9.1 ± 1.3	59.8 ± 2.2	55.3 ± 3.0	44.7 ± 3.0	284	3.17
Arg	22.1 ± 2.2	6.6 ± 1.3	71.3 ± 2.4	56.2 ± 3.1	43.8 ± 3.1	249	3.86
Asn	17.2 ± 2.4	21.5 ± 2.6	61.3 ± 3.0	51.6 ± 4.0	48.4 ± 4.0	157	0.16
Asp	14.7 ± 2.2	20.7 ± 2.5	64.7 ± 2.9	48.8 ± 3.8	51.2 ± 3.8	172	0.09
Cys	32.7 ± 4.7	18.4 ± 3.9	49.0 ± 5.0	41.7 ± 7.1	58.3 ± 7.1	48	1.33
SS-cys	22.1 ± 3.8	11.5 ± 2.9	66.4 ± 4.3	39.5 ± 5.4	60.5 ± 5.4	81	3.57
Phe	34.3 ± 2.2	4.9 ± 1.0	60.8 ± 2.3	51.2 ± 3.0	48.8 ± 3.0	285	0.17
Gln	24.4 ± 2.7	8.8 ± 1.8	66.8 ± 3.0	49.1 ± 3.9	50.9 ± 3.9	167	0.05
Glu	20.7 ± 2.1	11.9 ± 1.7	67.4 ± 2.5	52.9 ± 3.2	47.1 ± 3.2	244	0.80
Gly	22.9 ± 2.0	18.3 ± 1.8	58.8 ± 2.3	68.9 ± 2.8	31.1 ± 2.8	267	38.21
His	20.5 ± 3.0	14.6 ± 2.6	64.9 ± 3.5	58.3 ± 4.5	41.7 ± 4.5	120	3.33
Ile	31.3 ± 1.9	9.6 ± 1.2	59.1 ± 2.0	51.6 ± 2.6	48.4 ± 2.6	370	0.39
Leu	35.1 ± 1.7	12.5 ± 1.2	52.5 ± 1.8	47.8 ± 2.5	52.2 ± 2.5	395	0.73
Lys	24.8 ± 2.1	6.8 ± 1.2	68.4 ± 2.3	60.6 ± 2.9	39.4 ± 2.9	282	12.77
Met	33.2 ± 3.4	10.2 ± 2.2	56.7 ± 3.6	51.9 ± 4.9	48.1 ± 4.9	106	0.15
Pro	0.0 ± 0.0	19.2 ± 2.8	80.8 ± 2.8	0.0 ± 0.0	100.0 ± 0.0	164	164.00
Ser	22.0 ± 1.9	11.0 ± 1.4	67.0 ± 2.2	47.3 ± 2.8	52.7 ± 2.8	317	0.91
Thr	25.1 ± 1.7	7.8 ± 1.1	67.1 ± 1.9	46.0 ± 2.4	54.0 ± 2.4	430	2.69
Trp	33.5 ± 3.8	3.2 ± 1.4	63.3 ± 3.8	47.0 ± 5.0	53.0 ± 5.0	100	0.36
Tyr	38.5 ± 2.3	4.9 ± 1.0	56.6 ± 2.3	55.5 ± 3.1	44.5 ± 3.1	254	3.09
Val	31.3 ± 1.6	8.2 ± 0.9	60.5 ± 1.6	51.0 ± 2.2	49.0 ± 2.2	539	0.22

*The left-hand side of the table shows the location of each residue within antiparallel β-sheet. Edge residues are those which only have a β-sheet partner on one side. Interior residues have partners on both sides. Bulge residues are those classified by DSSP as not having (ϕ,ψ) angles in β conformation. They are always located on edges but the distribution of residues found in bulges is different to the edge residue distribution. Standard deviations quoted were calculated using $\sqrt{p(1-p)/N}$ where p is the fraction of each residue i found in a particular location and N is the total number of residue i in all locations. They are a measure of the accuracy of the percentages based on the number of counts. They do not reflect variations between proteins. The right-hand side of the table shows how the edge residues are distributed between the H-bonded and non-H-bonded sites. The results of a χ^2 test to determine whether the partitioning is nonrandom are shown in the last column. For a 50:50 partitioning of residues the expectation value would be half of the number of counts shown in the second last column. A χ^2 value greater than 3.84 indicates a nonrandom partitioning at the 95% confidence level. Proline, glycine, lysine, and arginine show a nonrandom partitioning at a high confidence level.

eral polar interactions or interactions that were important during the folding process.

Strong correlations: cystine pairs

Cys has an extraordinary preference to be its own partner in both the H-bonded and non-H-bonded sites (Tables III and IV). An examination of the structures reveals that this is largely due to pairs of disulfide-bonded cystines rather than free-thiol cysteines. The preference appears to be stronger in the non-H-bonded site, where it has the largest pair correlation (9.9). Of the 20 pairs, 19 are pairs of disulfide-bonded cystines. In five cases, the side chains are disulfide-bonded to each other, while in the remaining 14 cases the two cystines in the pair are distinct.

In the five cases where the two half-cystines form a single cystine, the residues adopt the gauche+ conformer. In order to accommodate the interstrand disulfide bond, each of the strands appears to shear in the direction of its C-terminus, as well as twisting in the normal direction. A similar shear coupled with the twist is observed for non-H-bonded branched residues (see below). Of the five examples, three are in influenza neuraminidase (5nn9) and the other two are in trypsin inhibitors (1tab and 1tie).

Of the separate cystines in the non-H-bonded site, four of the 14 pairs were in structures of high stereochemical quality. When these were superimposed, it was found that three of them adopted a similar close-packed conformation where a left-handed spiral cystine packed against a cystine in a right-handed spiral conformation (Fig. 2; nomenclature is from Richardson[16]). The "mirrored-spiral" arrangement allows three of the four sulfurs to pack tightly. Either one of the sulfur atoms in the right-handed cystine packs against the two sulfurs in the left-handed conformer. In all cases observed, the half-cystine in the left-handed spiral conformation adopts the gauche+ rotamer while the half-cystine in the right-handed spiral is trans.

In the H-bonded site, five of the seven Cys–Cys pairs are cystine pairs. Of these structures, three are of high stereochemical quality. Two of these three pairs also adopt the "mirrored-spiral" packing. The H-bonded and non-H-bonded Cys–Cys pairs differ from each other in that the non-H-bonded pairs are on more strongly twisted portions of β-sheet.

All of these paired cystines are found in either

TABLE III. Pair Correlations in the H-bonded (Upper Right) and Non-H-bonded (Lower Left) Sites of Antiparallel β-Sheet*

				A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
A	g	1.4		1.1	0.5	1.0	0.8	0.9	0.9	0.9	1.3	0.6	1.5	0.8	0.6	0.0	0.7	1.1	0.7	0.9	1.3	0.3	1.0	A
A	%	50.0		50.0	75.0	50.0	50.0	50.0	50.0	50.0	75.0	90.0	97.5	50.0	50.0	50.0	50.0	50.0	75.0	50.0	90.0	90.0	50.0	A
A	C	11		11	4	8	11	17	17	6	34	10	43	6	5	0	6	15	11	22	49	2	21	A
C	g	0.5	9.9		4.9	1.0	0.2	1.5	1.0	2.3	1.5	1.2	0.6	1.0	1.0	0.0	0.6	0.9	0.5	0.5	0.9	1.2	0.7	C
C	%	75.0	99.5		99.5	50.0	90.0	75.0	50.0	95.0	90.0	50.0	75.0	50.0	50.0	50.0	50.0	50.0	75.0	75.0	50.0	50.0	50.0	C
C	C	4	20		7	3	1	11	7	6	15	8	7	3	3	0	2	5	3	4	13	3	5	C
D	g	0.5	0.5	1.2		1.8	1.1	0.0	1.3	1.1	0.9	2.1	0.6	0.0	2.4	0.0	1.1	1.9	1.2	1.4	0.7	0.4	0.5	D
D	%	75.0	50.0	50.0		50.0	50.0	99.5	50.0	50.0	50.0	99.5	75.0	90.0	99.0	50.0	50.0	95.0	50.0	50.0	50.0	50.0	75.0	D
D	C	4	2	2		3	6	0	10	3	9	15	7	0	8	0	4	11	8	13	11	1	4	D
E	g	0.4	0.0	1.0	1.0		0.4	0.5	0.6	1.5	0.8	3.4	0.6	0.2	1.1	0.0	0.3	3.4	1.0	0.8	0.8	1.2	0.7	E
E	%	90.0	97.5	50.0	50.0		75.0	90.0	75.0	50.0	50.0	99.5	90.0	90.0	50.0	50.0	90.0	99.5	50.0	50.0	50.0	50.0	75.0	E
E	C	5	0	5	4		2	7	7	7	13	40	12	1	6	0	2	32	11	13	21	5	9	E
F	g	1.2	1.0	1.3	0.6	0.7		2.4	1.0	1.6	1.0	0.3	1.0	0.8	0.9	0.0	0.7	0.5	0.7	0.7	1.4	1.1	1.2	F
F	%	50.0	50.0	50.0	75.0	50.0		99.5	50.0	75.0	50.0	99.5	50.0	50.0	50.0	50.0	50.0	90.0	75.0	75.0	97.5	50.0	50.0	F
F	C	21	9	10	7	7		24	19	11	25	5	29	6	7	0	6	7	11	17	52	7	24	F
G	g	1.3	1.3	2.2	0.9	0.8	1.6		0.9	0.8	1.1	0.3	0.9	1.1	1.5	0.0	0.8	0.3	1.0	1.0	1.5	1.6	1.4	G
G	%	50.0	50.0	99.0	50.0	50.0	75.0		50.0	50.0	50.0	99.5	50.0	50.0	75.0	50.0	50.0	97.5	50.0	50.0	99.0	75.0	90.0	G
G	C	14	7	11	7	9	6		8	5	26	5	24	8	11	0	7	4	16	23	51	10	27	G
H	g	1.2	0.4	3.0	0.8	0.5	1.4	0.0		2.5	0.4	0.7	0.4	1.5	1.8	0.0	0.3	0.6	0.9	1.1	1.2	1.3	1.0	H
H	%	50.0	50.0	99.5	50.0	50.0	50.0	50.0		75.0	90.0	50.0	95.0	50.0	75.0	50.0	75.0	50.0	50.0	50.0	50.0	50.0	50.0	H
H	C	6	1	7	3	3	5	0		3	4	4	4	4	5	0	1	3	5	9	15	3	7	H
I	g	1.4	1.0	0.5	1.0	1.0	0.9	0.4	1.2		1.0	0.7	1.3	1.6	0.3	0.0	0.8	0.8	0.7	0.7	1.0	1.4	1.4	I
I	%	90.0	50.0	75.0	50.0	50.0	50.0	75.0	50.0		50.0	75.0	95.0	95.0	97.5	50.0	50.0	50.0	75.0	75.0	50.0	75.0	90.0	I
I	C	31	11	5	16	24	13	3	19		17	15	52	16	3	0	9	15	14	22	48	12	36	I
K	g	0.9	0.3	1.7	3.2	0.5	0.6	0.0	1.0	0.2		1.5	0.6	1.6	1.7	0.0	2.1	0.5	0.8	1.5	0.8	0.7	0.7	K
K	%	50.0	90.0	90.0	99.5	90.0	75.0	95.0	50.0	90.0		75.0	90.0	90.0	90.0	50.0	99.5	90.0	50.0	95.0	50.0	50.0	75.0	K
K	C	11	2	10	29	7	5	0	18	1		12	17	11	12	0	17	6	12	30	27	4	12	K
L	g	1.0	0.6	0.6	0.7	1.7	1.4	0.3	1.3	0.6	1.7		1.3	1.1	0.4	0.0	0.7	0.5	0.9	0.8	1.2	1.8	1.3	L
L	%	50.0	75.0	75.0	75.0	99.5	90.0	90.0	75.0	95.0	99.5		75.0	50.0	95.0	50.0	50.0	95.0	50.0	75.0	75.0	97.5	75.0	L
L	C	28	9	7	14	51	27	3	48	12	40		29	13	5	0	10	11	23	27	66	17	38	L

M	g	1.1	0.8	1.2	0.6	1.3	0.6	0.4	1.1	0.4	1.5	1.3		0.0	1.0	0.0	0.9	1.3	1.1	1.2	0.8	0.8	0.8	M
M	%	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	75.0	90.0	50.0		75.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	M
M	C	8	3	4	3	10	3	1	11	2	19	2		0	3	0	3	7	7	11	11	2	6	M
N	g	0.8	0.6	1.5	1.2	0.8	1.0	1.3	0.7	2.1	0.6	1.9	2.5		0.0	0.0	0.8	0.7	1.8	2.0	0.7	0.4	0.9	N
N	%	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	99.0	75.0	75.0	90.0		75.0	50.0	50.0	50.0	95.0	99.5	50.0	50.0	50.0	N
N	C	6	2	5	6	6	5	3	7	12	7	6	4		0	0	3	4	12	19	11	1	7	N
P	g	0.8	1.3	0.2	0.6	1.3	1.7	2.0	0.8	0.9	0.6	1.4	1.4	0.7		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	P
P	%	50.0	50.0	75.0	50.0	50.0	90.0	90.0	50.0	50.0	75.0	50.0	50.0	50.0		50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	P
P	C	8	6	1	4	13	11	6	10	7	10	6	6	2		0	0	0	0	0	0	0	0	P
Q	g	0.6	0.2	0.7	1.0	1.6	0.2	1.4	0.7	1.0	0.5	0.5	0.8	1.1	0.8		1.9	2.5	0.4	1.8	0.9	1.3	0.8	Q
Q	%	75.0	75.0	50.0	50.0	90.0	95.0	50.0	50.0	50.0	95.0	50.0	50.0	50.0	50.0		75.0	99.5	90.0	97.5	50.0	50.0	50.0	Q
Q	C	5	1	3	6	15	1	4	9	7	7	2	3	6	2		4	16	3	19	15	4	7	Q
R	g	0.5	0.7	2.2	1.8	0.6	1.2	1.4	0.9	0.7	0.7	0.6	0.8	1.7	2.0	0.8		0.8	1.4	1.2	0.7	1.3	1.0	R
R	%	75.0	50.0	99.0	95.0	75.0	50.0	50.0	50.0	50.0	75.0	50.0	50.0	90.0	97.5	50.0		50.0	75.0	50.0	75.0	50.0	50.0	R
R	C	6	4	11	14	7	9	5	13	6	14	3	4	11	12	3		4	16	19	19	6	14	R
S	g	0.9	0.6	1.4	1.1	1.4	0.9	1.6	0.4	0.6	1.0	0.9	2.1	0.7	1.5	0.7	1.3		2.2	1.1	1.1	1.3	0.8	S
S	%	50.0	50.0	50.0	50.0	90.0	50.0	75.0	99.0	75.0	50.0	50.0	99.5	50.0	75.0	50.0	50.0		99.5	50.0	50.0	50.0	50.0	S
S	C	14	5	10	12	24	10	8	9	8	26	6	15	7	13	8	10		15	22	34	7	13	S
T	g	0.9	0.4	1.4	1.4	0.4	0.7	1.5	0.6	1.5	0.4	0.3	1.1	0.8	1.9	1.5	1.2	2.0		1.6	0.7	0.3	0.8	T
T	%	50.0	95.0	75.0	90.0	99.5	50.0	75.0	95.0	95.0	99.5	97.5	50.0	50.0	99.5	95.0	75.0	99.5		95.0	95.0	95.0	75.0	T
T	C	21	5	15	24	10	12	11	20	27	16	3	12	11	24	25	29	34		22	30	2	18	T
V	g	1.1	0.8	0.3	0.8	1.1	0.7	1.0	1.5	0.9	1.0	1.1	0.3	0.7	0.6	0.7	0.9	1.2	1.5		0.9	0.7	1.0	V
V	%	50.0	50.0	99.0	50.0	50.0	75.0	50.0	99.5	50.0	50.0	50.0	99.0	50.0	75.0	75.0	50.0	50.0	99.5		50.0	75.0	50.0	V
V	C	36	12	4	19	39	15	10	65	23	54	15	4	14	11	15	29	54	49		31	8	37	V
W	g	1.0	2.2	1.4	0.4	0.7	1.4	1.0	0.8	1.8	0.9	1.1	0.0	2.1	1.2	1.4	0.8	0.3	0.6	2.4		0.0	1.1	W
W	%	50.0	95.0	50.0	75.0	50.0	50.0	50.0	50.0	90.0	50.0	50.0	90.0	95.0	50.0	50.0	50.0	95.0	75.0	75.0		50.0	50.0	W
W	C	6	7	4	2	5	6	2	7	9	10	3	0	8	4	6	5	3	7	3		0	7	W
Y	g	1.2	1.0	0.6	1.1	0.9	0.7	1.3	1.0	1.4	1.3	1.4	1.0	1.6	1.6	1.0	0.9	0.7	0.8	1.3	0.2		1.2	Y
Y	%	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	75.0	75.0	75.0	50.0	90.0	90.0	50.0	50.0	75.0	75.0	50.0	97.5		50.0	Y
Y	C	21	9	5	14	16	8	7	24	19	38	11	8	16	15	12	15	18	27	9	2		12	Y
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y			

*Pairs which are favored or disfavored at the 99% confidence level are underscored. g_{ij} (top), confidence level (2nd line), pair counts n_{ij} (3rd line).

TABLE IV. High Scoring Pairs*

H-bonded		Non-H-bonded	
Cys-Cys	4.9	Cys-Cys	9.9
Glu-Lys	3.4	Glu-Lys	3.2
Glu-Arg	3.4	Asp-His	3.0
Gln-Arg	2.5	Ser-Asn	2.1
Phe-Phe	2.4	Thr-Thr	2.0
Ser-Ser	2.2		
Asp-Lys	2.1		
Gln-Lys	2.1		
Thr-Asn	2.0		

*The table lists the pair correlations g_{ij} in the H-bonded and non-H-bonded sites which have pair correlations greater than 2.0 at a confidence level of at least 99%.

toxins, protease inhibitors, or protein hormones except for those in wheat germ agglutinin (9wga). All of these proteins are small (usually about 50 residues) with only a few instances of α-helices (1gps, 1tfg, and 2crd) and mainly highly twisted β-sheet and irregular extended structures. They are generally held together by 3–4 disulfides. Wheat germ agglutinin is also composed of four such modules.

Richardson[16] has identified these proteins as forming a separate class called "small irregular, cystine-rich proteins," and has also described the prevalence of cystine–cystine pairs in the non-H-bonded sites of such small proteins.[17] It is likely that double disulfides are necessary to stabilize these small proteins that have minimal hydrophobic cores. The reason for the pairing of cystines is unclear, although it may be to bring at least three sulfur atoms into close proximity. Previous studies have observed the tendency of sulfur containing residues to cluster.[18]

When these small, cystine-rich proteins are removed from the dataset, Cys–Cys pairs are still favored in the non-H-bonded site, although the numbers are not large enough to make the preference statistically significant. The differences between the usual "globular" proteins and the small, cystine-rich proteins are currently under investigation.

Six-membered ring aromatic side chains: Phe and Tyr

The six-membered ring aromatic sidechains favor the H-bonded site when pairing with themselves. This is reflected in the strong positive correlation of Phe–Phe in the H-bonded state and the negative correlation of Tyr–Tyr in the non-H-bonded site. As was noted by Richardson and colleagues,[1] phenylalanine has a strong preference for the gauche− rotamer within a β-sheet. In the H-bonded site, this allows the two side chains to interact via both hydrophobic and π–π interactions.[19] The less favorable interaction of Tyr-Tyr (cf. Phe–Phe) is likely to be due to the presence of the hydroxyl group interfering with these interactions.

Figure 3 shows an overlay of the six high quality Phe–Phe pairs. Most of the side chains adopt the gauche+ rotamer (five on the upper strand and three on the lower strand). Three of the residues adopt the gauche− rotamer which in two cases facilitates an offset-stacked π–π interaction with a gauche+ partner.

Branched residues: Val, Ile, Thr, and Leu

Many of the favored correlations with high statistical significance involve pairs of branched residues with matched properties (i.e., pairs of β- or γ-branched hydrophobic residues and β-branched hydrophilic residues). A comparison of the correlations for pairs of branched residues shows that they favor the non-H-bonded site (Table V). In β-sheets, β-branched side chains favor the rotamer that staggers the side chain with respect to the backbone (trans for Val, gauche+ for Ile and Thr). In the H-bonded site, this rotamer preference means that the side chains point away from each other. However, in the non-H-bonded site, the rotamer preference facilitates intersidechain interaction.

Val–Val pairs in the non-H-bonded site favor the trans, trans conformer state (9 of 12 high stereochemical quality structures). These pairs are shown in Figure 4a. The trans, trans rotamers allow the side chains to form a compact "nested packing," provided the backbones shear/twist slightly to one side or the other. There are two possible ways of nesting a Val–Val pair. However, only one of these is observed in the structures, i.e., the packing where the $C_{\gamma 2}$ carbon atoms nest in the fork of the partner residue. A superposition of trans, trans Val–Val pairs in the H-bonded site is shown for comparison (Fig. 4b).

Thirteen of the 20 Val–Ile pairs adopt the trans, gauche+ rotamers, respectively, and pack in a nested fashion in the non-H-bonded site. They show the same directionality in their packing interaction, with the valine C_γs packing around the C_δ-extended arm of isoleucine. This is also true for threonine (7 of 12 high stereochemical quality structures are in gauche+, gauche+). In this case, the favored packing direction results in the opposition of the two hydroxyl moieties (Fig. 5). This favors intersidechain hydrogen bonding.

Leucine is branched at the γ position but it shows the same site preferences and directional packing as the β-branched residues. An overlay of the high quality structures (Fig. 6) shows that one side chain is usually in the trans conformer (11 of 13), while the partner is distributed between the trans (four cases) and the gauche+ (seven cases) conformers. The pairs of side chains form compact structures, with the same directionality as the pairs of β-branched side chains.

The observed directionality of the nested packing between these pairs of branched side chains is consistent with the handedness of the twist in the

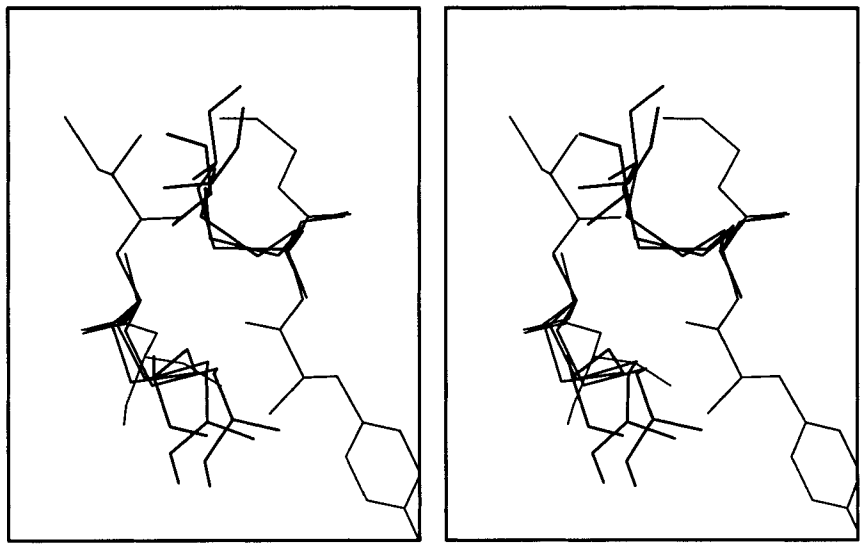

Fig. 2. A superposition of three high stereochemical quality "mirrored-spiral" cystine–cystine pairs in the non-H-bonded site (1tgf, 2sn3, 9wga). The cystine on the upper right adopts a left-handed spiral conformation packing both its sulfurs against either sulfur in the right-handed conformer (lower left). Their van der Waals radii are in contact.

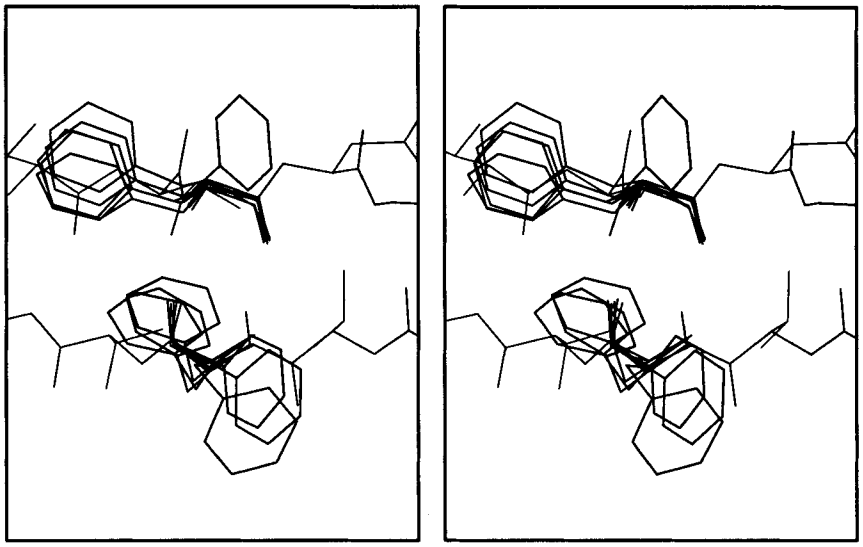

Fig. 3. A superposition of six high quality Phe–Phe pairs in the H-bonded site. Three of the six side chains on the lower strand are in the gauche− conformation. They are π–π interacting with partners on the upper strand. Two of the partners are in the gauche+ conformation. They form offset stacked pairs (1bop, 2er7). The third is the only residue on the top strand in the trans conformation. It forms an edge-on pair with its gauche− partner (1hle). The three structures which take up the gauche+ gauche+ conformation are not π interacting with each other but with other aromatics (1sha, 1shf, 2msb).

β-sheet (i.e., the unidirectional twisting of the sheet assists the unidirectional nesting of the branched side chains and vice versa). In the case of the side chains with a second chiral center, the observed enantiomer has a more favored interaction with its partner in the observed nested packing than it would if the nesting had the opposite directionality.

Finally, the backbone atoms also appear to slip relative to each other in β-branched non-H-bonded pairs (compare Val–Val pairs in the non-H-bonded and H-bonded sites, Fig. 4a and b, respectively). Although the sheets are antiparallel, the carbonyl groups appear to be aligned in non-H-bonded β-branched pairs. The hydrogen bonds on either side of the pair are not linear. CO..N angles and CO..C angles were measured for the backbone of pairs adjacent to all Val–Val pairs which are trans, trans in the non-H-bonded site. The average CO..C angle is 165±10° while the average CO..N angle is 157±7°. Thus the C=O . . . C=O vectors are closer to colin-

TABLE V. Comparison of Pair Correlations in the H-bonded and Non-H-bonded Sites for β-Branched Residues*

	H-bonded	Non-H-bonded
Val-Val	0.9	1.5
Val-Ile	1.0	1.5
Val-Thr	0.7	1.2
Ile-Ile	1.0	1.2
Ile-Thr	0.7	0.6
Thr-Thr	1.6	2.0

*The table lists the pair correlations g_{ij} for β-branched residues in the H-bonded and non-H-bonded sites. Note that for most pairs, the pair correlation in the non-H-bonded site is higher than in the H-bonded site.

earity than the C = O . . . H-N vectors. The slippage is always in the same direction, correlating again with the twist in the β-sheet.

Other favored pairs

Tryptophan is the only side chain capable of hydrogen bonding which does not have any favored interactions with other hydrogen bonding partners. This is probably because the side chain is too large to interact in such a way. The only favored interaction involving tryptophan is with proline in the non-H-bonded site. Proline is confined to the plane of the sheet and thus is only found on the edge in the non-H-bonded site. An overlay of the six medium resolution structures shows that tryptophan swings out over the proline (Fig. 7). This solves two problems: it allows the large tryptophan to be accommodated in the sheet as well as providing packing around the proline.

DISCUSSION

Dominant Factors Determining Pair Correlations

Electrostatic compatibility appears to be responsible for most of the strongest pair correlations. When it is violated, it results in the most significant negative correlations. Although these correlations are the strongest, an examination of the structures shows that these residues do not generally form specific intrapair interactions.

This conundrum has at least three possible explanations: (1) the high correlations are just due to a general polar/nonpolar partitioning of residues, (2) the observed structures do not truly represent the interacting pairs as they are on the protein surface and hence mobile, or (3) the compatible pairings are the predominant factors during folding that determine how β-strands align themselves. The specific nature of the strongly correlated pairs would argue against the first proposal. The second may be tenable in some cases where electron density maps are poorly defined and crystallographic *B*-factors high, however, many of the structures show well formed networks of polar side chains linked by salt bridges and hydrogen bonds in lieu of specific interactions within the favored pair. Given these arguments, we favor the third proposal, viz the electrostatically compatible pairings are crucial to sheet formation in the early stages of protein folding.

Relationship Between Branched Residues and β-Sheet Twist

The nested packing of branched residues always has the same directionality. When viewing a pair of residues with their side chains towards the observer, each side chain has slipped with respect to its partner towards its own C-terminal. This is in the same sense as the twist in β-sheets. Many have proposed reasons for the handedness of the twist in β-sheets[6,20–25] although none of these explanations is complete.

We have observed a correlation between the directionality of the branched side chain packing, the handedness of the β-sheet twist and the handedness of the enantiomers of amino acids which have a second chiral center. These observations lead us to speculate that the packing of branched sidechains provides the energetic reason as to why the β-sheets are twisted. However, this is not in itself directional. The observed directionality is most likely to arise from the asymmetry of the left-handed amino acids. Thus, the packing of branched side chains destabilizes the flat β-sheet, favoring either the left- or right-handed twisted sheet. Backbone interactions of L-amino acids then select the observed twist handedness. Finally, the observation that the naturally occurring enantiomers of the two amino acids with second chiral centers are the best choice for branched amino acid interactions within a non-H-bonded pair is likely to be an evolutionary selective effect.

Salemme and Weatherford[6] pointed out that due to the dyad symmetry, hydrogen bonds between antiparallel β-strands could be uniformly deformed. We have observed such a deformation in the relative shearing of neighboring strands. The direction of the shear and the handedness of the twist appear to be coupled. The direction of shearing results in a more even spacing between C = O groups by widening the separation between those in the H-bonded site and narrowing it in the non-H-bonded site. Shearing of the strands in the opposite direction would bring C = O groups into closer proximity in the H-bonded site, producing unfavorable steric and coulombic effects. These differences may be significant in favoring the observed shear direction.

Thus, it is possible that the preferred shear direction selects the handedness of the twist in antiparallel β-sheets. If this is correct, then the origin of the handedness in the twist in parallel β-sheets would be different, as these sheets do not have the same local dyad symmetry as antiparallel ones.

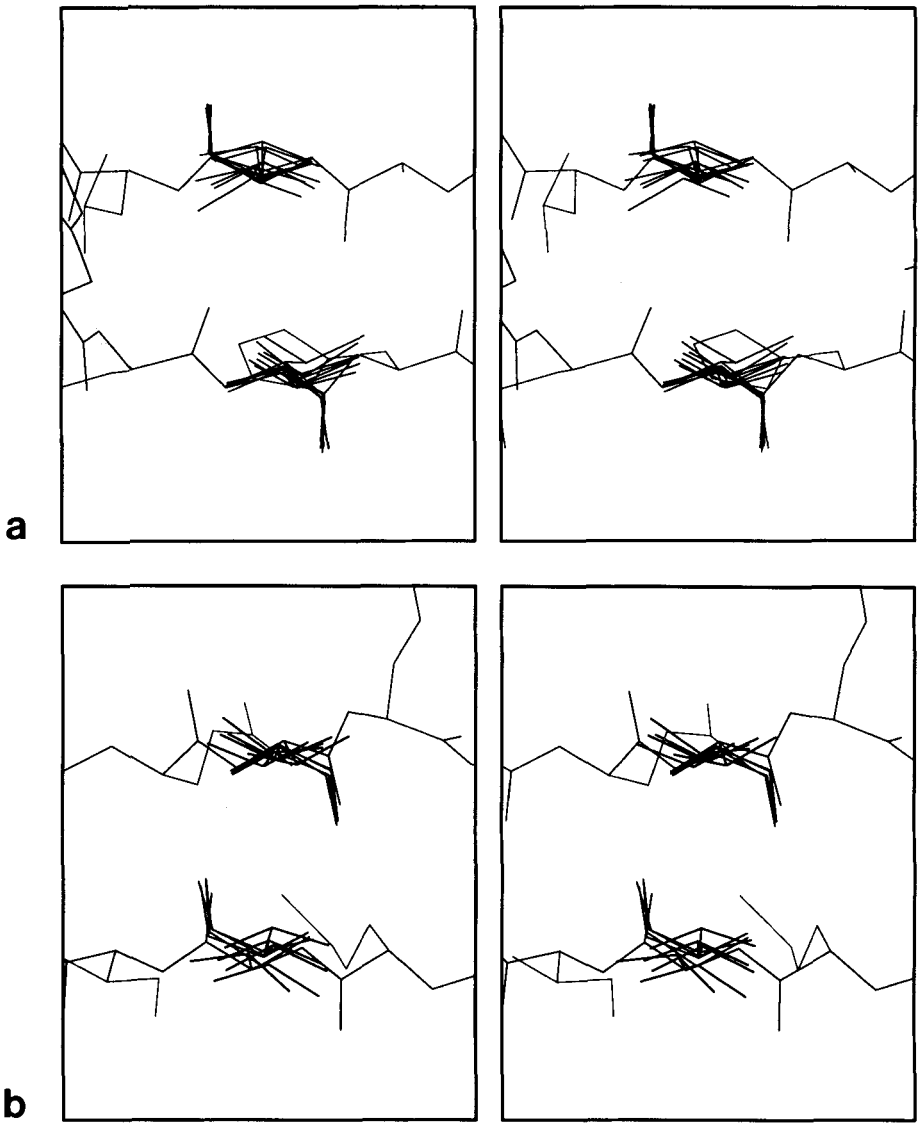

Fig. 4. (**a**) A superposition of seven Val–Val pairs in the non-H-bonded site (1aba, 1caj, 1mdc, 1ppf, 1ttb, 3cox, 3grs). The pairs adopt a nested packing. Note that each residue appears to move with respect to its partner. The direction of this movement is parallel to its strand and towards its C-terminal end. (**b**) A superposition of five high quality Val–Val pairs in the H-bonded site (1ifc, 1lts, 1ttb, 2er7, 2mnr). In general, both residues adopt the low energy trans conformation (the same conformation is called gauche+ for threonine.)

Protein Folding and β-Sheet Formation

The picture that emerges from this study is that the docking of extended segments of polypeptide chain to form an antiparallel β-sheet is determined by electrostatic compatibility. This makes intuitive sense, since in the early phases of folding (the molten globule state), intimate side chain packing has not occurred. Thus, only long-range effects (forces) can be responsible for early-phase interaction patterns. Once the interstrand register is set by these effects, the side chain packing commences. In this phase, the packing of branched side chains forces the sheet to shear and twist. The direction of this twist and shear is dictated by asymmetric factors associated with the interstrand backbone interactions. The twist appears to be stabilized by the side chain interactions of branched residues.

CONCLUSION

The results of this study show that there are statistically significant correlations between cross-strand neighbors within antiparallel β-sheets. They also show that the two types of pair in the antiparallel sheet behave differently. Consideration of these pair correlations will be important in protein design, engineering, and structure prediction. There

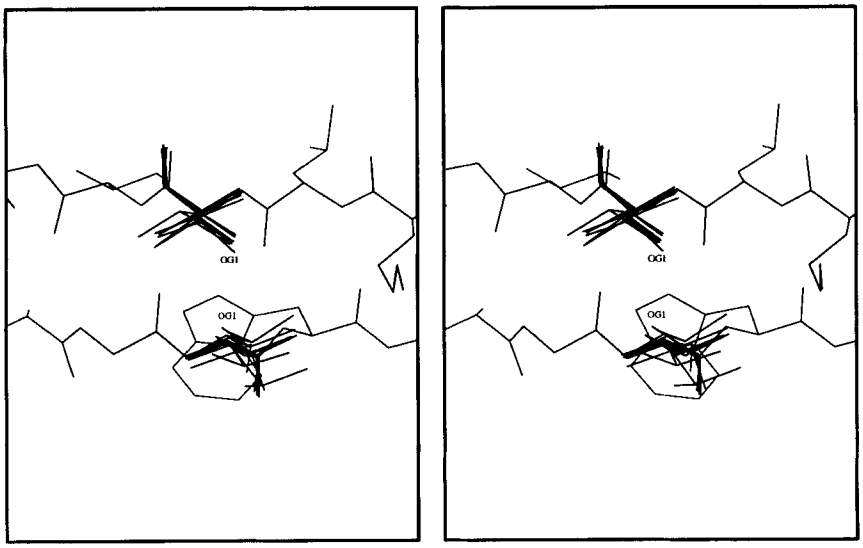

Fig. 5. A superposition of the six high quality Thr–Thr pairs in the non-H-bonded site in the data set (1bbp, 1hil, 1ifc, 2er7, 3dfr, 7aat). All pairs are gauche+, gauche+. The side chains move with respect to each other towards the C-terminal end of their strand, opposing their hydroxyl groups (marked "OG1"). Side chains are hydrogen bonded in two of the four high quality structures. The enantiomer of threonine found in proteins allows cross-strand hydrogen bonding of threonine pairs in the non-H-bonded site of twisted sheets with normal handedness.

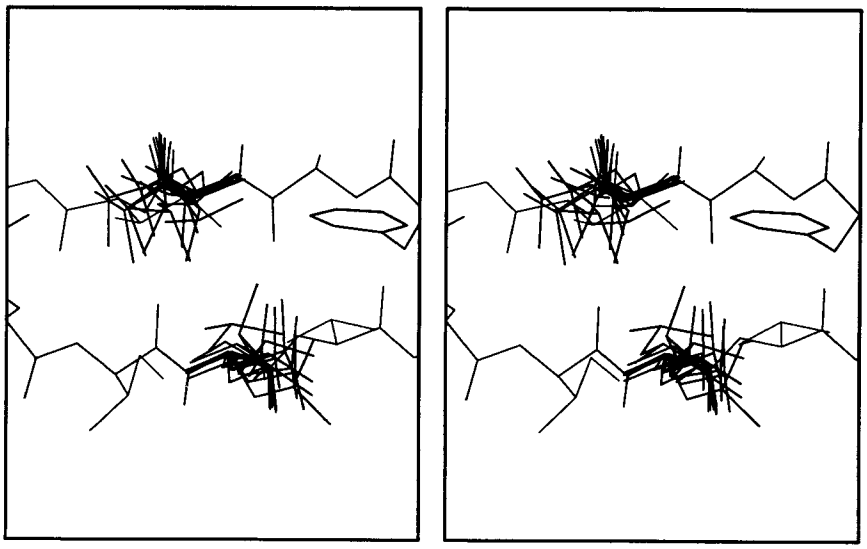

Fig. 6. Leu–Leu non-H-bonded pairs adopt a nested packing with the same handedness as Val–Val and Thr–Thr pairs in the same site. The superposition is of 11 high quality structures (1arb, 1bop, 1hiv, 1paz, 1sgt, 1snc, 1ten, 2er7, 2mnr, 2pia, 4gcr). One side chain (lower strand) is usually in the trans conformer (11 out of 13) while its partner is distributed between the trans (four cases) and the gauche+ (seven cases) conformers. The two other high resolution structures are gauche+, gauche+. They are not included in the diagram for simplicity.

have been several recent instances where peptides have been generated to inhibit protein function by preventing protein–protein interaction via β-sheet extension (HIV protease dimerization[26]; PapD chaperone action[27]). Our results should provide a basis for improving such peptide-based therapeutics by facilitating de novo peptide design. We are currently conducting a similar study of parallel β-sheets. By combining the correlations for both parallel and antiparallel sheets it should be possible to construct a novel algorithm for the prediction of β-sheets, their connectivities and topology.

ACKNOWLEDGMENTS

We thank David Smith, Andrew Fisher, and Bridget Mabbutt for fruitful discussions and sugges-

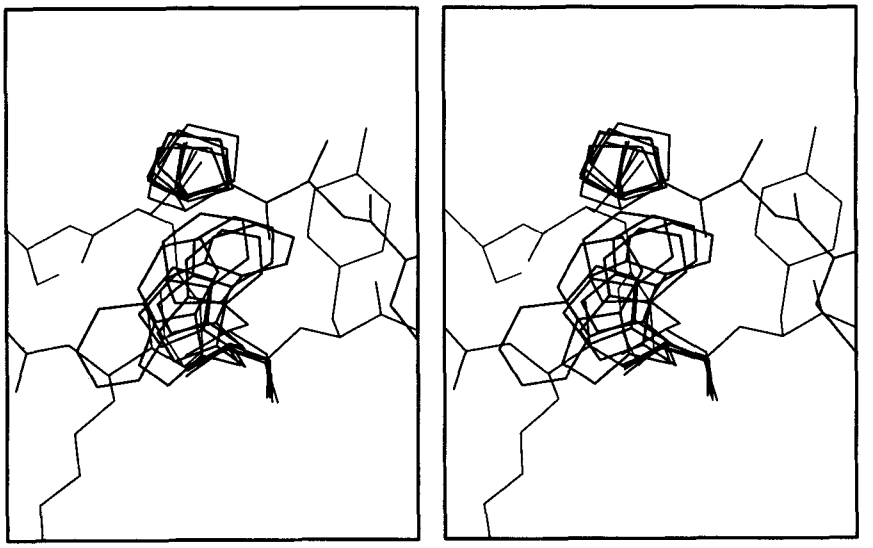

Fig. 7. Six medium resolution Trp–Pro pairs in the non-H-bonded site (1aak, 1aap, 1gbp, 1shf, 3sc2, 5nn9). Proline is always on the edge in the non-H-bonded site since in any other position it would interfere with the neighboring strand. The tryptophan residue does not fit comfortably into all sites in β-sheet because of its large size. In all six structures, the tryptophan swings out over the proline (gauche+ χ_1). Five of the six structures also have the same χ_2.

tions. This research was facilitated by grants from the University of New South Wales and the Clive and Vera Ramaciotti Foundation.

REFERENCES

1. Richardson, J.S., Richardson, D.C., Tweedy, N.B., Gernert, K.M., Quinn, T.P., Hecht, M.H., Erickson, B.W., Yan, Y., McClain, R.D., Donlan, M.E., Surles, M.C. Looking at proteins: Representations, folding, packing, and design. Biophys. J. 63:1186–1209, 1992.
2. Lifson, S., Sander, C. On the mutual recognition of strands in beta-sheets. In: "Molecular Mechanisms of Biological Recognition." Balaban, M., ed. Amsterdam: Elsevier-North Holland Medical Press, 1979: 145–156.
3. von Heijne, G., Blomberg, C. The β structure: Inter-strand correlations J. Mol. Biol. 117:821–824, 1979.
4. Lifson, S., Sander, C. Specific recognition in the tertiary structure of β-sheets of proteins. J. Mol. Biol. 139:627–639, 1980.
5. Hobohm, U., Scharf, M., Schneider, R., Sander, C. Selection of representative protein data sets. Protein Sci. 1:409–417, 1992.
6. Salemme, F.R., Weatherford, D.W. Conformational and geometrical properties of β-sheets in proteins II. Antiparallel and misced β-sheets. J. Mol. Biol. 146:119–141, 1981.
7. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
8. Hobohm, U., Sander, C. Enlarged representative set of protein structures. Protein Sci. 3:522–524, 1994.
9. Kabsch, W., Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637, 1983.
10. Morris, A.L., MacArthur, M.W., Hutchinson, E.G., Thornton, J.M. Stereochemical quality of protein structure coordinates. Proteins 12:345–364, 1992.
11. Schreuder, H. personal communication.
12. Kabsch, W. A solution for the best rotation to relate two sets of vectors. Acta Cryst. A32:922–923, 1976.
13. InsightII is a program by Biosym Technologies of San Diego.
14. Kraulis, P.J. Molscript: A program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. 24:946–950, 1991.
15. Lifson, S., Sander, C. Antiparallel and parallel β-strands differ in amino acid residue preferences. Nature (London) 282:109–111, 1979.
16. Richardson, J.S. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34:167–339, 1981.
17. Richardson, J.S. Protein Tourist #4: A survey of proteins in the "small irregular" category of tertiary structure. Protein Sci. 1, Kinemage diskette 7 (July), 1992.
18. Warme, P.K., Morgan, R.S. A survey of atomic interactions in 21 proteins. J. Mol. Biol. 118:273–287, 1978.
19. Hunter, C.A., Singh, J., Thornton, J.M. π–π interactions: The geometry and energetics of phenylalanine-phenylalanine interactions in proteins. J. Mol. Biol. 218:837–846, 1991.
20. Chothia, C. Conformation of twisted β-pleated sheets in proteins. J. Mol. Biol. 75:295–302, 1973.
21. Chou, K.C., Scheraga, H.A. Origin of the right-handed twist of β-sheets of poly(LVal) chains. Proc. Natl. Acad. Sci. U.S.A. 79:7047–7051, 1982.
22. Chou, K.C., Pottle, M., Nemethy, G. Ueda, Y., Scheraga, H.A. Structure of β-sheets. Origin of the right-handed twist and of the increased stability of antiparallel over parallel sheets. J. Mol. Biol. 162:89–112, 1982.
23. Chou, K.C., Nemethy, G., Scheraga, H.A. Role of interchain interactions in the stabilization of the right-handed twist of β-sheets. J. Mol. Biol. 168:389–407, 1983.
24. Salemme, F.R. Structural properties of protein β-sheets. Prog. Biophys. Mol. Biol 42:95–133, 1983.
25. Vtyurin, N. The role of local tight packing of hydrophobic groups in β-structure. Proteins 15:62–70, 1993.
26. Babe, L.M., Rosé, J., Craik, C.S. Synthetic "interface" peptides alter dimeric assembly of the HIV1 and 2 proteases. Protein Sci. 1:1244–1253, 1992.
27. Kuehn, M.J., Ogg, D.J., Kihlberg, J., Slonim, L.N., Flemmer, K., Bergfors, T., Hultgren, S.J. Structural basis of pilus subunit recognition by the PapD chaperone. Science 262:1234–1241, 1993.