Protein Science (1998), 7:2287–2300. Cambridge University Press. Printed in the USA.

Determinants of strand register in antiparallel β-sheets of proteins

E. GAIL HUTCHINSON,[1] RICHARD B. SESSIONS,[2] JANET M. THORNTON,[1] AND DEREK N. WOOLFSON[3]

[1]Biomolecular Structure and Modeling Unit, Department of Biochemistry and Molecular Biology, University College, Gower Street, London WC1E 6BT, United Kingdom
[2]The Department of Biochemistry, The School of Medical Sciences, University Walk, Bristol BS8 1TD, United Kingdom
[3]Centre for Biomolecular Design and Drug Development, The School of Biological Sciences, The University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom

(RECEIVED April 17, 1998; ACCEPTED July 20, 1998)

Abstract

Antiparallel β-sheets present two distinct environments to inter-strand residue pairs: $\beta_{A,HB}$ sites have two backbone hydrogen bonds; whereas at $\beta_{A,NHB}$ positions backbone hydrogen bonding is precluded. We used statistical methods to compare the frequencies of amino acid pairs at each site. Only ~10% of the 210 possible pairs showed occupancies that differed significantly between the two sites. Trends were clear in the preferred pairs, and these could be explained using stereochemical arguments. Cys-Cys, Aromatic-Pro, Thr-Thr, and Val-Val pairs all preferred the $\beta_{A,NHB}$ site. In each case, the residues usually adopted sterically favored χ_1 conformations, which facilitated intra-pair interactions: Cys-Cys pairs formed disulfide bonds; Thr-Thr pairs made hydrogen bonds; Aromatic-Pro and Val-Val pairs formed close van der Waals contacts. In contrast, to make intimate interactions at a $\beta_{A,HB}$ site, one or both residues had to adopt less favored χ_1 geometries. Nonetheless, pairs containing glycine and/or aromatic residues were favored at this site. Where glycine and aromatic side chains combined, the aromatic residue usually adopted the gauche$^-$ conformation, which promoted novel aromatic ring-peptide interactions. This work provides rules that link protein sequence and tertiary structure, which will be useful in protein modeling, redesign, and de novo design. Our findings are discussed in light of previous analyses and experimental studies.

Keywords: β-strand; β-structure; aromatic-aromatic interactions; aromatic-peptide interactions; profile analysis; protein design; protein folding; protein modeling; protein structure

The amino acid sequence of a protein dictates the acquisition and stabilization of its active state (Anfinsen, 1973). Therefore, to improve understanding of protein folding and design, it is critical to establish rules that link sequence and structure. Studies in this area have revealed that context is a key determinant of the preference of a residue for a particular secondary structure. At an elementary level, patterns of hydrophobic (*H*) and polar (*P*) residues play roles in the organization of protein structures (Lim, 1974a, 1974b; Dill, 1990; Huang et al., 1995; Sun et al., 1995; West & Hecht, 1995). For example, in soluble proteins, alternating, *HPHP*, patterns occur in β-strands, whereas α-helices show *PHPPHHP* and similar repeats. Such patterns produce amphipathic secondary structures, which can promote the organization of tertiary structure in globular proteins. However, there is no absolute requirement for such simple *HP* patterns in the secondary structures of proteins (West & Hecht, 1995). Moreover, it appears that for folding to unique structures, specific interactions superimposed on these binary patterns are required. This can be seen in different contexts within proteins.

For example, although many α-helical coiled-coil motifs are coded by the sequence repeat *HPPHPPP*, specific residue placements, particularly at the two *H* sites, are required to discriminate between dimer and trimer alternatives (Harbury et al., 1993; Woolfson & Alber, 1995; Gonzalez et al., 1996). Similarly, to obtain unique folded states in the design of four-helix-bundle proteins, DeGrado and colleagues have had to introduce complementary interactions into what began as an all-leucine hydrophobic core (Raleigh & DeGrado, 1992; Betz et al., 1993, 1995, 1996). One group has succeeded in designing a protein with the properties of a naturally folded molecule by experimentally selecting a sequence from a large pool based on simple *HP* patterns (Roy et al., 1997). However, a number of other researchers have chosen to develop methods to generate designer proteins with well-packed hydrophobic cores *in silico*, prior to the preparation and characterization of

Reprint requests to: Dek Woolfson, The School of Biological Sciences, University of Sussex, Falmer BN1 9QG, United Kingdom; e-mail: D.N.Woolfson@sussex.ac.uk.

a limited number of molecules (Munson et al., 1996; Dahiyat & Mayo, 1997; Lazar et al., 1997). More surprisingly, Tisi and Evans (1995) report interactions on the surface of proteins that appear to be important in specifying the active conformers of a family of β-sandwich structures. Thus, while structural space is undoubtedly reduced by the selection of certain *HP* patterns, additional specific interactions play pivotal roles in directing folding to a unique state. Here we describe interactions that the direct register of β-strands in antiparallel β-structures.

The rates of occurrence of the 20 amino acids in β-strands have been available for some time (Chou & Fasman, 1974; Levitt, 1978), and more recently, β-propensities have been determined experimentally (Kemp, 1990; Kemp et al., 1990; Kim & Berg, 1993; Minor & Kim, 1994a, 1994b; Smith et al., 1994; Otzen & Fersht, 1995). There is some consensus between the experimental and theoretical scales (Muñoz & Serrano, 1994; Finkelstein, 1995; Swindells et al., 1995); in particular, aromatic and β-branched hydrophobic residues are favored in all scales, while small hydrophilic residues are deemed destabilizing from experimental studies, and are less frequently observed in database analyses. However, in absolute terms, correlations between the two types of scale are less convincing. This has been reconciled to some extent by the work of Muñoz and Serrano (1994), who argue that inadequacies in the derivation of the original theoretical scales are largely at fault for the poor correlations. However, when the experimentally derived scales are compared, the rank orders of amino acids are not maintained either. One explanation is that these scales are subject to context dependence and that this differs between experiments. Consistent with this, Minor and Kim (1994b) show that β-propensities determined for an edge strand of a β-sheet differ from those for a central strand. As a central strand can make approximately twice as many interactions as an edge strand, interstrand contacts may influence the measured propensities. Indeed, Swindells et al. (1995) show that comparisons of β-propensities and database derived frequencies for an edge strand are best made with two correlations, one for the polar residues and another for the hydrophobic side chains. A step toward accounting for cooperative terms of this type is to establish potentials for pair-wise interactions in β-sheets. While the energetics and preferences for specific amino acid pairs have been studied in some detail in α-helices (Padmanabhan & Baldwin, 1994a, 1994b; Huyghues-Despointes et al., 1995; Stapley & Doig, 1997), analyses and experiments on side-chain to side-chain interactions in β-structure are less prolific.

Lifson and Sander have analyzed the frequencies of nearest-neighbor residues in parallel and antiparallel structures, and uncover a small number of trends in favored amino acid pairs (Lifson & Sander, 1980). More recently, two groups have focused on the pair-wise preferences of residues in antiparallel β-structure in general (Wouters & Curmi, 1995) and in the more restricted case of β-hairpins (Gunasekaran et al., 1997). For adjacent antiparallel β-strands two positions of closest approach can be distinguished, Figure 1: at $\beta_{A,HB}$ sites two residues are brought together and interact at the backbone via two hydrogen bonds; in contrast, at $\beta_{A,NHB}$ sites the backbone CO and NH groups of each residue point away from each other and intra-pair backbone hydrogen bonding is precluded. Wouters and Curmi (1995) have analyzed the amino acid occupancies at the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites. Their work reveals that the two sites have preferences for different residue pairs. Smith and Regan (1995) have taken some of these preferences and tested them experimentally. Correlations between

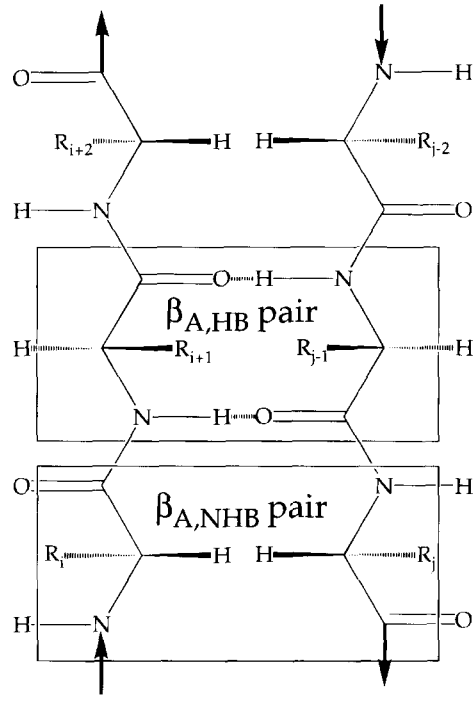

Fig. 1. Schematic diagram of an antiparallel β-sheet highlighting the two inter-strand sites of closest approach, $\beta_{A,HB}$ and $\beta_{A,NHB}$, and the intra-strand site, β_{II2}, taken as a "control" in this work.

the experimental scale and theoretical values are evident, although there are also discrepancies. Here we describe an in-depth analysis of the pair-wise preferences of the amino acids for antiparallel β-structure in proteins. Our work builds on previous studies in three important respects.

First, we applied statistical methods to compare the occupancies of residue pairs at the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites with those for a "control" data set. This was for the β_{II2} site, where residues were displaced i, $i + 2$ along the same β-strand (Fig. 1), and intimate side-chain to side-chain interactions were diminished. By using the β_{II2} data as a control, we were able to account for two leanings inherent in the raw data for the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites: (1) preferences for residue pairs in which both amino acids have high propensities for β-structure, and (2) skewing toward pairs of residues with similar chemistry. This second bias occurs because, in each case, the two residues at the β_{II2}, $\beta_{A,HB}$, and $\beta_{A,HB}$ sites fall on the same face of the β-sheet (Fig. 1). Thus, they will experience similar environments, and display similar side-chain chemistry; an exposed pair will often contain two polar side chains, whereas a buried pair is likely to have two hydrophobic residues. Indeed, Wouters and Curmi (1995) comment on such trends in their data, and note an indiscriminate preference for *PP*-type pairs and a dearth of *HP*-type interactions. Second, we compared the tables for the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites directly. As $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites alternate along β-strands and through β-sheets (Fig. 1), this direct comparison highlighted specific residue pairs that determine register in antiparallel β-structure. Third, and most importantly, for the majority of cases we were able to rationalize the observed preferences using straightforward stereochemical arguments and model building.

Results and discussion

Statistical analysis of the occupancies of residue pairs in antiparallel β-structure

The secondary-structure assignments introduced by Kabsch and Sander (1983) were used to locate the β_{II2}, $\beta_{A,HB}$, and $\beta_{A,NHB}$ pairs in 311 nonhomologous protein chains with structures determined to 2.5 Å or better. The occurrences of each of the 210 possible residue pairs were counted for each site and these data chelated in 20×20 contingency tables. In turn, these tables were analyzed and compared using statistical methods.

The count data for the amino acid pairs at the β_{II2} sites were well represented; only 13 of the 210 possible pairs had fewer than five examples (see Supplementary material). Traditionally, skewed distributions in contingency tables of this type are gauged using chi-square tests. In this method the observed data are compared with an expected table, or null hypothesis (Lifson & Sander, 1980; Wouters & Curmi, 1995). The expected table is generated by distributing the total number of pairs in the observed table among the elements of the expected table considering only the total numbers of each amino acid; thus, all couplings between amino acids are ignored. A chi-square test on our β_{II2} data revealed that, taken together, the pairs statistically matched the expected values. In other words, the distribution of amino acid pairs at the β_{II2} positions was similar to that calculated assuming no coupling between amino acids. Most likely, this reflects an inability of side chains to interact at β_{II2} sites, because the C_α-C_α distance between the i and $i + 2$ residues in a β-strand is too great. Examination of molecular models confirmed that there was little possibility for intimate side chain interactions at the β_{II2} sites except in two special cases. First, two bulky aromatic side chains could contact each other in one orientation; with the i^{th} residue in a less favored trans (t) χ_1 conformation and the $i + 2^{th}$ residue in the favored gauche$^+$ (g^+) χ_1 conformation. Second, flexible and extended side chains, such as Arg and Lys, could contact each other, but again this demanded unusual side-chain conformations. Nevertheless, the distribution of residue pairs in the β_{II2} data set was relatively flat, and suggested that intra-pair side-chain interactions were rare. For this reason, we were confident in using the β_{II2} data set as an initial control for our statistical and stereochemical analyses of the pair-wise occupancies at the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites.

In chi-square tests, the $\beta_{A,HB}$ and $\beta_{A,NHB}$ count data differed from their calculated expected tables and the β_{II2} data above the 95% confidence limits. This indicated that, in contrast to the β_{II2} data, some pairs were favored at the expense of others, and suggested that some residue-residue pairs lead to favorable side-chain interactions. This stands to reason as C_α-C_α distances at $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites are shorter than those at β_{II2}: ≈4–5 Å compared with ≈7 Å. To determine which of the 210 elements in each of the $\beta_{A,HB}$ and $\beta_{A,NHB}$ tables were responsible for the divergence from the null hypotheses, we compared the individual elements of each table with corresponding elements of the β_{II2} data set. This was done using the statistical method of standard error of proportion, which we have used previously to establish amino acid differences that distinguish dimeric and trimeric coiled-coil structures (Woolfson & Alber, 1995). In the present work, a standard error was calculated for each element of the $\beta_{A,HB}$ and $\beta_{A,NHB}$ tables and used to calculate a z-score for the difference between that element and the corresponding one from the β_{II2} table.

Z-scores set confidence limits for calculated differences. In our analysis the pairs listed in Table 1 differed at, or above, the 99% confidence limit in the comparisons with β_{II2} data. It is striking that these subsets of residue pairs for the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites are very different; of the 28 pairs retrieved only one, Phe-Leu, is favored at both sites in comparison with β_{II2}. To probe the differences between the $\beta_{A,HB}$ and $\beta_{A,NHB}$ data sets further, we compared the $\beta_{A,HB}$ and $\beta_{A,NHB}$ data sets directly using the standard error of proportion (Table 2). This direct comparison carries the same normalizations afforded by the comparisons of the $\beta_{A,HB}$ and $\beta_{A,NHB}$ data with the $\beta_{A,HB}$ table; namely, the removal of biases toward *HH* and *PP* pairs, and those that contain residues with high individual preferences for β-structure. In addition, it should highlight pairs that discriminate between the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites, and help determine register in antiparallel β-sheets. Encouragingly, a number of indiscriminate *HP*-type pairs that were disfavored at the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites in comparison with the β_{II2} data did not reappear in the direct comparison. In total 15 amino acids pairs differed between the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites at, or above, the 99% confidence limit, and seven of these pairs were identical to pairs highlighted in the comparisons with the β_{II2} data (Table 1).

Similar trends were evident from the two different comparisons ($\beta_{A,HB}$ /$\beta_{A,NHB}$ vs. β_{II2}, and $\beta_{A,HB}$ vs. $\beta_{A,NHB}$):

Table 1. *Residue pairs that show strong preferences for either the $\beta_{A,HB}$ or $\beta_{A,NHB}$ in antiparallel β-structure in proteins*[a]

Favored site	Type of residue pair			
	PP	*HH*	Other	Total
$\beta_{A,HB}$ vs. β_{II2}	HH, KQ, KS, QR, RS	FL, FV, IY, VY, VW	GW	11
$\beta_{A,HB}$ vs. $\beta_{A,NHB}$	KS, RS	AW, FF, FV, VW, VY, IL	GA, GF, GR, GV	12
$\beta_{A,NHB}$ vs. β_{II2}	TT, ST, DH, DK, DR, EK, ER, ET, KN, KS, KY, RT	FA, FL	CC, FP, YP	17
$\beta_{A,NHB}$ vs. $\beta_{A,HB}$	TT, DR	VV	—	3

[a]Successive rows give pairs favored above the 99% confidence limits in statistical comparisons of the various data sets: $\beta_{A,HB}$ vs. β_{II2}, $\beta_{A,HB}$ vs. $\beta_{A,NHB}$, $\beta_{A,NHB}$ vs. β_{II2}, and $\beta_{A,NHB}$ vs. $\beta_{A,HB}$, respectively. The columns separate the favored pairs by type: *PP*, pairs of polar side chains; *HH*, pairs of hydrophobic residues; and *Other* refers to pairs that do not fit either of these categories. The final column gives the total number of favored pairs for each comparison.

Table 2. *Ratios ($\beta_{A,HB}/\beta_{A,NHB}$) of occupancies at the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites*[a]

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.71	1.49	0.99	1.49	0.61	**2.87**	0.85	0.96	1.98	1.13	0.85	1.49	0	1.24	2.03	1.22	0.92	1.28	**2.55**	1.29
C		0.15	2.97	0	1.06	2.23	5.95	0.69	0	0.64	1.49	4.46	0	0	0.74	0.50	0.59	1.22	0.74	1.12
D			1.49	5.95	0.50	0.42	0.62	1.06	0.80	0.85	0.50	3.72	0	0.45	**0.43**	0.59	0.87	2.60	0	0.17
E				0.89	0.65	1.78	1.49	0.77	0.96	0.87	0.25	0.37	0	0.34	1.03	0.25	0.65	1.29	0.89	1.12
F					**2.16**	**2.63**	2.38	1.15	1.22	1.03	1.35	0.12	0	0.56	0.74	0.70	1.27	**1.99**	0.89	1.49
G						1.08	1.98	0.94	0.74	1.55	1.12	0.68	0	1.49	**2.55**	1.19	0.66	**2.59**	5.95	1.39
H							10.41	0.25	1.49	1.19	0	0.74	0	0.74	1.49	0.30	0.74	1.73	0.74	1.73
I								0.89	1.39	**1.41**	1.49	0.25	0	1.86	0.69	0.71	1.32	0.78	0.95	1.53
K									0.99	1.22	2.97	0.68	0	2.04	1.19	**2.16**	0.92	1.20	0.54	0.69
L										0.99	0.93	0.99	0	0.99	1.26	0.71	1.30	1.05	0.67	1.36
M											0.37	0.59	0	2.48	1.12	1.86	1.49	0.79	0.74	1.16
N												1.49	0	0.89	0.45	0.79	0.58	0.74	2.48	0.42
P													0	0	0	0	0	0	0	0
Q														1.49	1.22	1.08	0.85	1.38	0.15	0.59
R															2.48	**2.30**	1.01	1.08	0.99	1.12
S																0.61	0.73	1.18	0.85	0.70
T																	**0.40**	0.94	0.33	0.69
V																		**0.66**	**2.68**	**2.53**
W																			1.49	0.68
Y																				2.38

[a]Differences that are significant above the 99 and 95% confidence levels are shown in bold and underlined, respectively.

residue pairs favored at the $\beta_{A,HB}$ sites contained aromatic residues and/or glycine, or the extended side chains of arginine and lysine; whereas, more specific pairs rather than general groupings were favored at $\beta_{A,NHB}$, the pairs of β-branched residues, Thr-Thr and Val-Val, and the charge-charge pairs. For the pairs that differed at statistically significant levels in the direct $\beta_{A,HB}$-$\beta_{A,NHB}$ comparison, we found the following themes: Gly-containing pairs, Gly-Ala, Gly-Phe, Gly-Arg, and Gly-Val were favored at the $\beta_{A,HB}$ sites, as were the aromatic-containing pairs, Ala-Trp, Phe-Phe, Phe-Val, Val-Trp, and Val-Tyr; at the $\beta_{A,NHB}$ sites, Thr-Thr, Val-Val, and the charge-charge pair Asp-Arg were all favored. These trends all continued in pairs whose occupancies differed between the two sites at the 95% significance level. For instance, more Gly- and aromatic-containing pairs were favored at the $\beta_{A,HB}$ sites, while Thr-containing pairs and the Ile-Val pairing were found for the $\beta_{A,NHB}$ sites. Below, we discuss the stereochemical origins of these observations. Pro-aromatic and Cys-Cys pairs are included in this analysis even though they were not revealed in the $\beta_{A,HB}$-$\beta_{A,NHB}$ comparison. This is because they showed significant differences in the $\beta_{A,NHB}$-β_{II2} comparison, which were lost in the direct comparison because Cys and Pro were under-represented in the $\beta_{A,HB}$ data; by definition, there were no Pro-containing pairs in the $\beta_{A,HB}$ data set and only two Cys-Cys pairs were retrieved for this site.

Favored interactions at the $\beta_{A,NHB}$ sites

Charge-charge pairs

In comparison to the β_{II2} data, almost all combinations of oppositely charged residues, Asp and Glu with His, Lys, and Arg, were favored at the $\beta_{A,NHB}$ sites (Table 1). However, this preference was found to be marginal as only the Asp-Arg pairing survived the direct $\beta_{A,HB}$-$\beta_{A,NHB}$ comparison (Tables 1, 2). This demonstrates an important aspect of making the direct comparison. Although charge-charge interactions show preferences for antiparallel β-structure (Table 1) (Lifson & Sander, 1980; Wouters & Curmi, 1995), our analysis showed that with proper normalization, or choice of controls, that this is limited to a general preference rather than a specific preference for either the $\beta_{A,HB}$ or the $\beta_{A,NHB}$ site. Nonetheless, charge-charge interactions may still contribute to register selection in antiparallel β-sheets: Smith and Regan (1995) show that charge-charge pairs placed at a $\beta_{A,HB}$ site in streptococcal protein G give large interaction energies (Smith & Regan, 1995). In addition, Pham et al. (1998) highlight Glu-Lys pairs as possible contributors to the structure and stability of a solvent-exposed single-layer β-sheet in OspA from *Borrelia burgdorferi*. This sheet has five inter-strand Glu-Lys pairs, which are split 2:3 between $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites consistent with our findings. Furthermore, the 7 and 24 examples of Asp-Arg that we located at $\beta_{A,HB}$ and $\beta_{A,NHB}$, respectively, had similar potential for ion-pair interactions; the (C_γ-C_ζ) distances between the charged moieties of the side chains were 5.61 ± 3.13 and 5.61 ± 2.04 Å, respectively. The probable root of the preference for the $\beta_{A,NHB}$ site is that side chains here could achieve these distances via favorable side-chain conformations [that is, g^+ and g^+t for Asp and Arg, respectively (Fig. 2A)], whereas at the $\beta_{A,HB}$ sites less common conformers were used because of constraints imposed by the backbone conformation. *N.b.*: For the 26 examples of Asp-Arg at β_{II2}, the C_γ-C_ζ distances were slightly greater (6.21 ± 2.46 Å) as expected.

Combinations of Ile and Val

In the direct comparison of the $\beta_{A,HB}$ and $\beta_{A,NHB}$ data (Table 2), the predominant *HH*-type pairs favored at $\beta_{A,HB}$ contained aromatic and/or glycine residues; whereas those favored at $\beta_{A,NHB}$ sites were restricted to combinations of β-branched aliphatic residues, Ile and Val. The comparisons with the β_{II2} data indicated that these particular preferences resulted from a slight dislike for

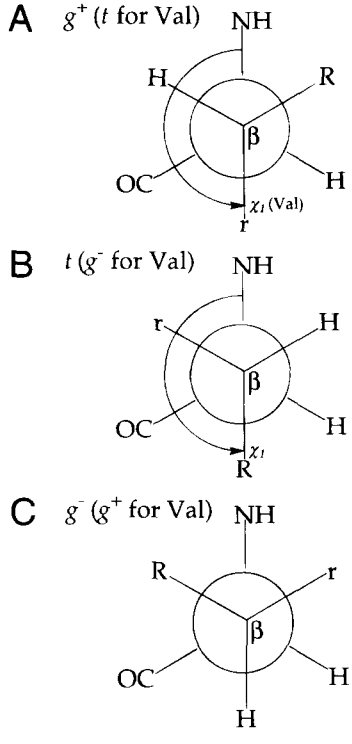

Fig. 2. Newman projections for residues other than Ala, Gly, and Pro. The views are down the C_β-C_α bond. χ_1 angles are measured anticlockwise on a 0–360° scale from the C_α-NH bond vector (**A**) to the C_β-r vector for Val, or (**B**) to the C_β-R vector for all other amino acids. Panel **A** shows the g^+ conformation (*t* for Val), which places R *trans* to CO and is the most favorable conformation sterically. Panel **B** shows the *t* conformation (g^- for Val). Panel **C** shows the g^- conformation (g^+ for Val), which places R *gauche* to both the CO and NH groups and is the least favorable conformation sterically. Key for relevant amino acids: r = CH_3 for Ile, Thr and Val, and r = H for all others; R = OH for Ser and Thr; R = CH_3 for Val; R = CH_2CH_3 for Ile; R = $CH(CH_3)_2$ for Leu; R = SH for Cys; R = C_6H_5 for Phe; R = C_6H_4.OH for Tyr; R = C_8H_6N for Trp; R = $C_3H_3N_2^+$ for His; R = $(CH_2)_3.NH_4^+$ for Lys; R = $(CH_2)_2.NH.C.(NH_2)_2^+$ for Arg; R = CO_2^- for Asp; R = $CH_2CO_2^-$ for Glu.

the $\beta_{A,HB}$ sites combined with preferences for the $\beta_{A,NHB}$ sites (Supplementary material). Moreover, a clear order in the preferences was apparent: the ratios of occupancy of the $\beta_{A,NHB}$ and $\beta_{A,HB}$ sites were Val-Val (1.52) > Ile-Val (1.28) > Ile-Ile (1.12); the values for Val-Val and Ile-Val were significantly above the 99% and 95% confidence limits, respectively (Table 2). The value for Ile-Ile was not statistically significant. To understand this trend we examined the structures of Val-Val pairs at the various sites.

The χ_1 angles of each valine in the Val-Val pair were plotted against each other for the three sites, β_{II2} (145 examples), $\beta_{A,HB}$ (36), and $\beta_{A,NHB}$ (81) (Fig. 3). In the control, β_{II2} set, the data were skewed toward the "middle row" and the "central column" of the plot with 90% of the data occupying this area. In this cross-shaped region, one or both of the side chains adopt the favored trans (*t*) χ_1 conformation. Furthermore, the majority of the pairs (57%) fell in the central region in which both residues were in the *t* χ_1 conformation. This preference is expected for valine, because this rotamer staggers all substituents on the C_β with respect to those on the C_α, and leaves the γ-methyl groups maximally displaced from both the backbone carbonyl and NH moieties (Fig. 2A). McGregor et al. (1987) and Ponder and Richards (1987) note that 67–69% of all Val residues in proteins occupy the *t* rotamer. The χ_1-χ_1 distribution for Val-Val pairs at the favored $\beta_{A,NHB}$ sites resembled that for the β_{II2} sites, but had a larger skew toward *t* χ_1 conformations; 95% of the data lay in the central cross-shaped region of the plot, and 63% of the pairs had both residues in the *t* χ_1 conformation (Fig. 3).

Inspection of structures of Val-Val pairs in "*tt*" conformation at a $\beta_{A,NHB}$ site revealed that this stereochemistry promoted good intra-pair van der Waals' contacts (Fig. 4A), referred to as "nested packing" (Wouters & Curmi, 1995). It is probable that this drives the preference for Val-Val pairs at these sites to some extent. However, as we noted above that a component of the difference in the occupancies at the $\beta_{A,NHB}$ and $\beta_{A,HB}$ positions by Val-Val pairs came from their preclusion from $\beta_{A,HB}$. The χ_1-χ_1 distribution for Val-Val at the $\beta_{A,HB}$ position was quite different from that for the β_{II2} sites (Fig. 3): First, fewer points (66%) fell in the cross-shaped region. Second, only 28% of the pairs had the paired *tt* conformation. This value was found to be significantly less (at the 99% confidence level) than the value from the β_{II2} plot. Thus, the χ_1-χ_1 distribution for the $\beta_{A,HB}$ sites was broader than for the $\beta_{A,NHB}$

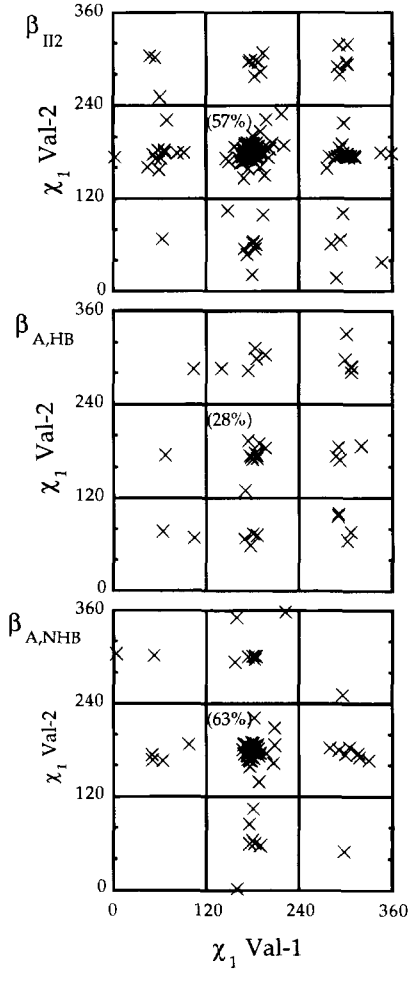

Fig. 3. χ_1-χ_1 distributions for Val-Val pairs at the β_{II2}, $\beta_{A,HB}$, and $\beta_{A,NHB}$ sites. Note the change in occupancy of the central regions of these plots, which house the pairs that have both residues in the sterically favored *t* conformations (Fig. 2A).

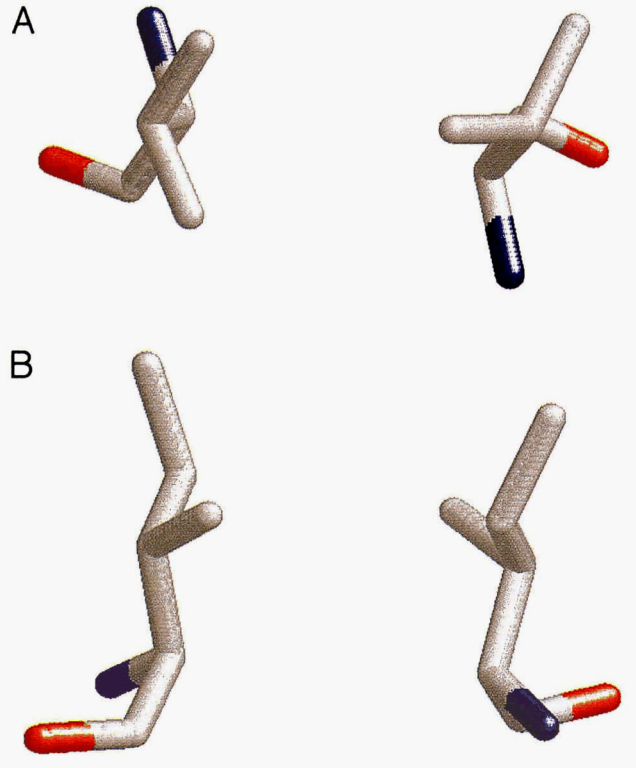

Fig. 4. Side-chain interactions in Val-Val and Ile-Ile pairs at $\beta_{A.NHB}$ sites. **A:** A Val-Val pair (residues I47 and I64 from 1CEW) viewed from "above"; that is, down the C_β-C_α bond vectors. Both side-chains are in the favored *t* conformation, which leads to a "nested" interaction involving the C_γ atoms (Wouters & Curmi, 1995). Sixty-three percent of the Val-Val pairs that we found at the $\beta_{A.NHB}$ sites took up this conformation (Fig. 3). **B:** An Ile-Ile pair (residues 71 and 96 from 1MAT) viewed side-on. Both residues are in the favored g^+t conformation, which we found for 71% of the Ile-Ile pairs at $\beta_{A.NHB}$. This arrangement led to an interaction similar to that shown in **A**, which does not involve the Ile C_δ atoms. These examples were chosen at random from our lists of Val-Val and Ile-Ile pairs at the $\beta_{A.NHB}$ sites. Key: blue, nitrogen; red, oxygen; grey, carbon.

sites (Fig. 3), which suggests that no particular conformation is preferred. This reflects an inability of Val-Val pairs to interact at $\beta_{A.HB}$ sites, which stands to reason as two valine side chains in the *t* conformation at a $\beta_{A.HB}$ site point away from each other and cannot interact (Figs. 1, 2). Two valine residues could interact at a $\beta_{A.HB}$ site with one or both of the residues in the unfavored g^+ or g^- χ_1 conformations, but these conformers are taken up by only 30% of the valine residues in the Protein Data Bank (McGregor et al., 1987; Ponder & Richards, 1987).

Turning to the order of preference Val-Val > Ile-Val > Ile-Ile at $\beta_{A.NHB}$ sites, one might expect that inferences drawn for the Val-Val pairs should apply to Ile-Val and Ile-Ile pairs; Ile is β-branched and has a strongly preferred χ_1 rotamer, g^+, and the side chains of two Ile residues paired (g^+g^+) at $\beta_{A.NHB}$ sites will point toward each other and interact. However, Ile has an additional C_δ atom to Val, which introduces new considerations. First, Ile can adopt more side-chain rotamers than Val, and selection of one conformer will incur an entropic penalty in folding. Second, some of the side-chain conformations may lead to side-chain to side-chain clashes at $\beta_{A.NHB}$ sites. There is evidence for this in our data. The Ile side chains in the Ile-Ile pairs at $\beta_{A.NHB}$ take up predominantly one conformation, namely $\chi_1 = g^+$ and $\chi_2 = t$; 71% of the Ile residues in Ile-Ile pairs at $\beta_{A.NHB}$ sites have this conformation in comparison to 60% at the β_{II2} sites. With two Ile residues in the g^+t conformation at a $\beta_{A.NHB}$ site intimate, non-clashing nested interactions do occur (Fig. 4B). However, these involve only the C_γ atoms and not the C_δ methyl. Thus, we propose that Ile "weakens" interactions in Ile/Val pairs at $\beta_{A.NHB}$ because one specific side-chain rotamer must be frozen out without a concomitant improvement in van der Waals' interactions. Hence the order of preference: Val-Val > Ile-Val > Ile-Ile at $\beta_{A.NHB}$.

Combinations of Thr and Ser

Wouters and Curmi (1995) find that the Thr-Thr pair is favored at both the $\beta_{A.HB}$ and $\beta_{A.NHB}$ sites in antiparallel β-structure with frequencies of 1.6 and 2, respectively. In contrast, we found that this pair was favored only at the $\beta_{A.NHB}$ position, and that there were 2.5 times as many examples at this site compared with $\beta_{A.HB}$. We attribute this discrepancy to (1) a greater number of proteins in our survey, and (2) the additional normalization that we perform to remove biases toward residue pairs with high intrinsic (individual) preferences for β-structure. Moreover, we were able to rationalize the preference for Thr-Thr at the $\beta_{A.NHB}$ site by inspection of protein crystal structures.

χ_1-χ_1 Distributions for the Thr-Thr pairs found at β_{II2}, $\beta_{A.HB}$, and $\beta_{A.NHB}$ are plotted in Figure 5. Like Ile and Val, Thr is β-branched and has a strong preference for the g^+ χ_1 conformation. Correspondingly, 49% of the Thr-Thr pairs at the β_{II2} sites had the paired g^+g^+ conformation. In comparison, the χ_1-χ_1 distribution for the $\beta_{A.NHB}$ site was tighter with more pairs (59%) in the g^+g^+ region. This g^+g^+ arrangement at $\beta_{A.NHB}$ directs the O_γ atoms toward each other and promotes side-chain hydrogen bonding (Fig. 6A). Indeed, half of the Thr-Thr pairs in this region made O_γ-O_γ hydrogen bonds (Fig. 5). At a $\beta_{A.HB}$ site, side-chain hydrogen bonds are only possible with both Thr residue in the least favored g^- χ_1 conformation. Indeed, the single pair at a $\beta_{A.HB}$ site with this conformation did form an O_γ-O_γ hydrogen bond (Fig. 5). Presumably, the steric penalty of placing two threonine residues in the g^- conformation is not balanced by intra-pair hydrogen bonding. Taken together, these observations suggest that the preference of Thr-Thr for the $\beta_{A.NHB}$ positions is driven, at least in part, by the formation of side-chain hydrogen bonds with the residues in their sterically favored g^+ conformations. However, given the chemical similarity of serine and threonine, it is odd that the Ser-Ser and Ser-Thr pairs were not also favored at $\beta_{A.NHB}$; the $\beta_{A.NHB}/\beta_{A.HB}$ ratios for these pairs were 1.6 and 1.4, respectively, and neither was statistically significant.

To explore this, we compared the χ_1-χ_1 distributions at $\beta_{A.NHB}$ for the series Thr-Thr, Ser-Thr, Ser-Ser. In this series the occupancy of the g^+g^+ region dropped from 59% to 34% to 24%, but the proportions of pairs that made intra-pair side-chain hydrogen bonds remained steady at ~1/3 of the total numbers of pairs; 11/37, 13/45, and 6/17, respectively. Thus, although there was increased restriction of conformational space in the series Ser-Ser to Ser-Thr to Thr-Thr, this did not lead to increased hydrogen bonding. Thus, rotamer restriction in the Thr-Thr pair must play an important role in driving its preference for $\beta_{A.NHB}$, otherwise equally high preferences would be expected for Ser-Ser and Ser-Thr. Further support for this argument comes from the $\beta_{A.NHB}$-β_{II2} comparison (Table 1), in which both Thr-Thr and Ser-Thr, but not Ser-Ser, were favored.

The experimental studies of Smith and Regan (1995) also support our finding that Thr-Thr is not favored at both sites in anti-

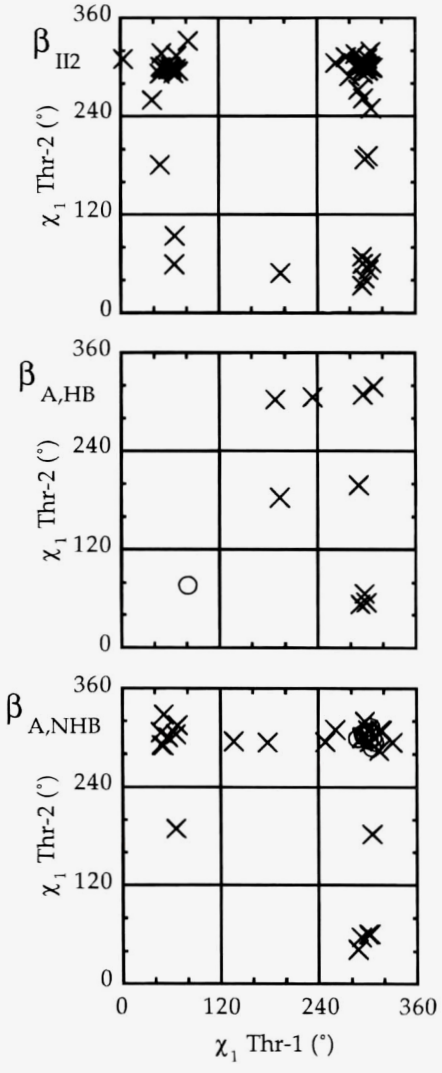

Fig. 5. χ_1-χ_1 distributions for Thr-Thr pairs at the β_{II2}, $\beta_{A,HB}$, and $\beta_{A,NHB}$ sites. Circles in the plots for the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites are for pairs that make side-chain to side-chain hydrogen bonds (O_γ-O_γ distances < 3.5 Å), all other pairs are represented by crosses. Comparison of the plots reveals the skew toward the g^+g^+ conformation at the $\beta_{A,NHB}$ positions.

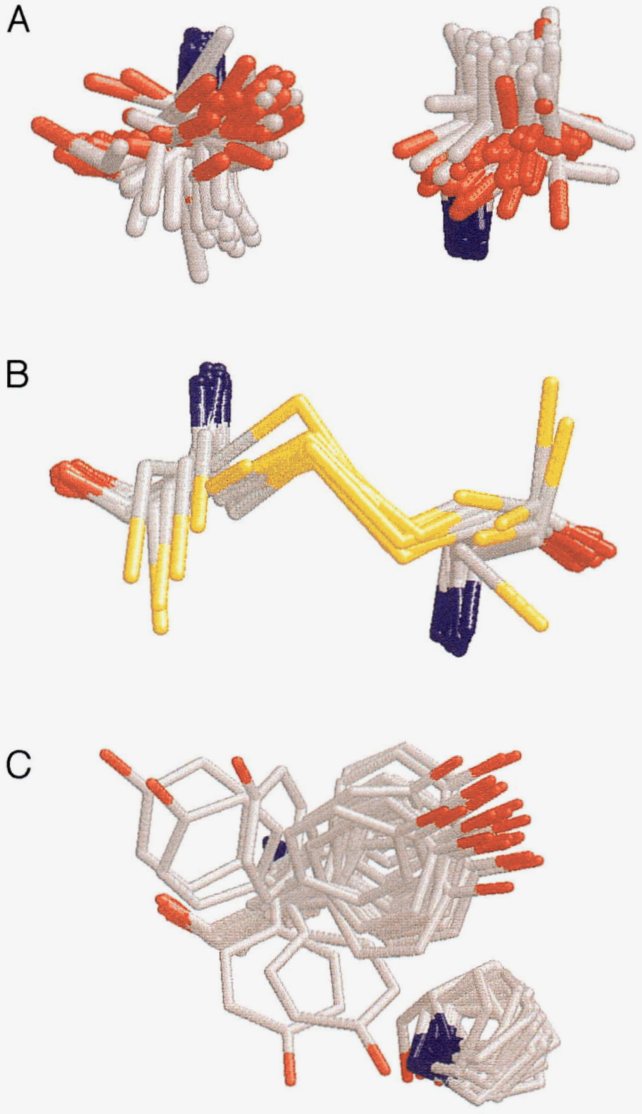

Fig. 6. Superposition of certain favored residue pairs found at $\beta_{A,NHB}$ sites in antiparallel β-sheets. Intimate side-chain interactions are evident. **A:** The 37 Thr-Thr pairs, showing the preference for the g^+g^+ conformation, which brings O_γ atoms within hydrogen bonding range. **B:** The 20 examples of Cys-Cys pairs. In this case, the g^+g^+ conformation brings the S_γ atoms close and accommodates intra-pair disulfide bonding. **C:** The 28 Pro-Tyr pairs. The aromatic residues lie predominantly in the most favored g^+ conformation, which brings the aromatic rings and the constrained proline side chains into van der Waals contact. All superpositions were made by aligning the two corresponding sets of backbone heavy atoms (N, C_α, C, and O). All structures are viewed along the C_β-C_α bond vectors. Key: blue, nitrogen; red, oxygen; grey, carbon; yellow, sulfur.

parallel β-sheets. These workers measure interaction energies between pairs of amino acids substituted at an inter-strand site of protein G. They find that the Thr-Thr pair has a low interaction energy and state that this is not consistent with foregoing statistical work of Wouters and Curmi (1995). The experimental result is, however, consistent with our analysis. This is because the Thr-Thr pair investigated by Smith and Regan (1995) was placed at a $\beta_{A,HB}$ site. Therefore, we would not expect an interaction between the two side chains. However, we predict a strong interaction between two Thr residues substituted at a $\beta_{A,NHB}$ position.

Cys-Cys pairs

One of the largest differences between $\beta_{A,HB}$ and $\beta_{A,NHB}$ data sets was for the Cys-Cys pairs. After normalization for the different sizes of the databases, we found nearly 10 times as many Cys-Cys pairs at the $\beta_{A,NHB}$ sites (Table 2). However, it was not possible to gauge the significance of this result because of a low occupancy at the $\beta_{A,HB}$ sites. Wouters and Curmi (1995) comment on an overall preference for Cys-Cys pairs in β-sheets, but do not see the large and contrasting difference between the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites that we observed. More recently and in agreement with our analysis, Gunasekaran et al. (1997) report a preference for Cys-Cys at $\beta_{A,NHB}$ sites in a database limited to β-hairpin structures. We show here that this preference of Cys-Cys for $\beta_{A,NHB}$ can be rationalized by considering the stereochemical requirements for disulfide links at the different sites of antiparallel β-structure.

We identified six Cys-Cys pairs at β_{II2} sites. None of these made intra-pair disulfide links. This was expected because the inter-

residue distances at this site are too big for covalent bonding. At $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites, however, the distances could be bridged comfortably by a disulfide bond. Indeed, of the 20 examples of the pair at the $\beta_{A,NHB}$ sites 14 formed an intra-pair disulfide bond (Fig. 6B). However, neither of the 2 Cys-Cys pairs at the $\beta_{A,HB}$ sites were cystines. Thus, cystine rather than cysteine-cysteine pairs was favored at the $\beta_{A,NHB}$ sites of antiparallel β-structure and precluded from the $\beta_{A,HB}$.

A rationale for this finding followed from molecular-dynamics studies conducted in water on a series of model cyclic peptides. The parent molecule was the cyclic dimer {AA-AAAAA-AG}$_2$. The two central stretches of five alanine residues were modeled as a pair of hydrogen bonded antiparallel β-strands connected by the Ala-Gly-Ala-Ala sequences configured as type-II′ reverse turns. This model was robust in minimization (in vacuo) and during 200 ps of dynamics at 300 K in water. In the simulation some local breathing motions were observed, otherwise hydrogen-bond-distance and backbone-torsion-angle parameters remained stable (data not shown). Simulations were performed on related peptides, in which cystines were placed at $\beta_{A,NHB}$ and $\beta_{A,HB}$ sites.

The torsion-angle and hydrogen-bond-length trajectories for peptide {AA-AACAA-AG}$_2$, which had a cystine at a $\beta_{A,NHB}$ site, are shown in Figures 7A and 7B, respectively. Apart from some breathing motions between 80 and 110 ps into the run (Fig. 7B), both the structure and the parameters were stable over the course of the simulation. In contrast, for peptide {AA-ACAAA-AG}$_2$, where the cystine was at $\beta_{A,HB}$, the trajectories were not stable (Fig. 7C,D). First, we observed that, after minimization, this and similar starting structures with cystine at $\beta_{A,NHB}$ sites, had energies, on average, 10 kcal mol^{-1} greater than those with cystine at $\beta_{A,NHB}$. Much of this strain was likely to be in the S-S dihedral, which in the starting structures lay away from the preferred 100°. Indeed, early (0–5 ps) into the simulation of the "$\beta_{A,HB}$" peptide the S-S dihedral relaxed (Fig. 7C). This was followed closely by changes in the χ_1 and χ_2 angles of the half cystines (Fig. 7C), which then led to a loss of backbone hydrogen bonding around the disulfide (Fig. 7D). Later in the trajectory, near 60 ps, there was further relaxation of the χ_1 and χ_2 torsion angles of Cys4 and widespread loss of the hydrogen bond network (Fig. 7C,D).

The contrasting behavior of the two model peptides reflects the different environments that cystine bonds experience at the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites. At the $\beta_{A,HB}$ position the two half-cystines are required to adopt sterically hindered g^- conformations [an angle of 60° (Fig. 7C)] and the S-S bond is strained. During our simulations this strain was lost, but at the expense of the hydrogen bond network. In comparison, the preferred g^+ conformation of cysteine

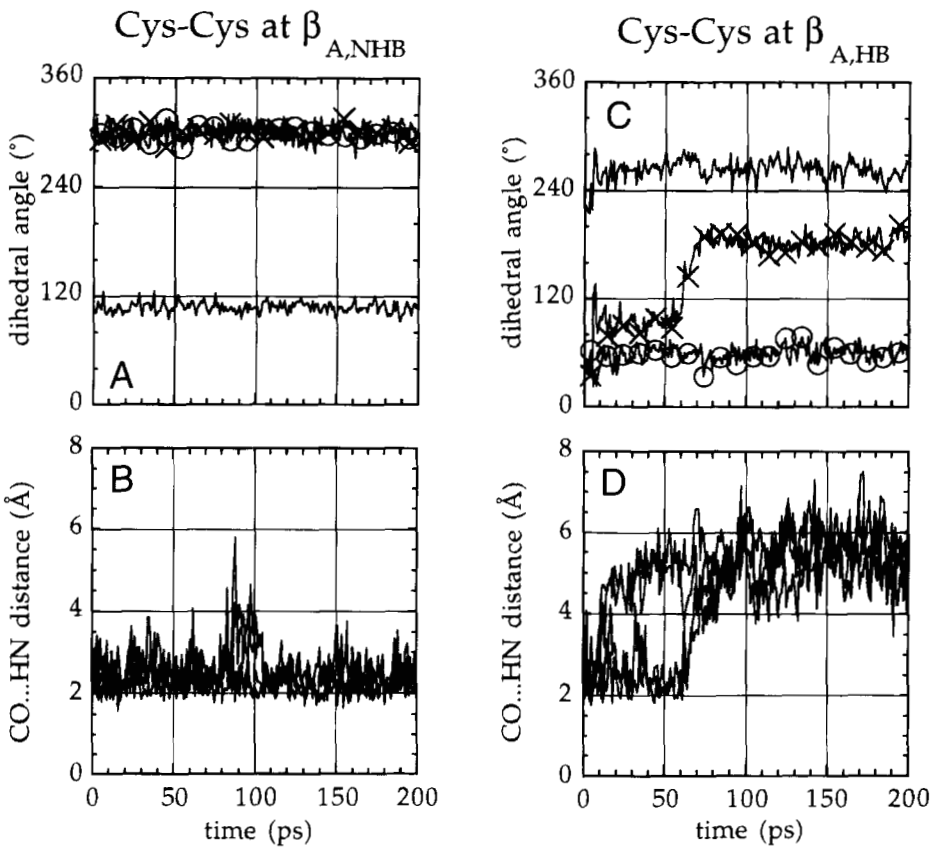

Fig. 7. Summary of the molecular-dynamics studies on the cystine-containing cyclic peptides. **A:** The evolution of dihedral angles during the simulation of the peptide with cystine placed at a $\beta_{A,NHB}$ site. Key: solid trace, C_β-S-S-C_β torsion angles; traces highlighted by circles and crosses, χ_1 torsion angles for the half cystines. **B:** Changes in hydrogen bonding distances (backbone CO to NH) for the four inter-strand hydrogen bonds closest to the Cys-Cys pair. The parameters shown in **A** and **B** were stable over the 200 ps simulation in water. This contrasts with the data shown in panels **C** and **D**, which are the corresponding plots for the trajectory of the model peptide with a disulfide bridge at a $\beta_{A,HB}$ site. Kinemages for the initial and final structures of these simulations can be found in the Electronic supplementary material.

at the $\beta_{A,NHB}$ site promotes intra-pair disulfide bonds. With two cysteine residues in the g^+ conformation, the two S_γ atoms point toward each other and can form an unstrained covalent bond (Fig. 6B), which survives the molecular-dynamics simulations.

Pro-aromatic ring interactions

Within proteins proline residues do not carry a proton. Therefore, proline cannot occupy $\beta_{A,HB}$ sites in antiparallel β-sheets, and we were unable to make a direct comparison of frequencies of proline-containing pairs at $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites. However, in comparison with the β_{II2} data, combinations of proline and aromatic residues (collectively labeled Ar here) were found to be favored at the $\beta_{A,NHB}$. The $\beta_{A,NHB}/\beta_{II2}$ ratios were Pro-Tyr (6.5) > Pro-His (3.5) > Pro-Phe (2.5) > Pro-Trp (1.6) with the values for Pro-Tyr and Pro-Phe significantly above the 99% confidence limit. The correlations of Wouters and Curmi (1995) show similar trends, with the Pro-Trp correlation being significant. Inspection of the Pro-Ar pairs in our data set revealed a common theme. The conformation of the proline ring was restricted and the predominant rotamer for the aromatic side chains was the sterically most favored g^+ conformation. This gave good van der Waals' contacts over the faces of the aromatic and proline rings (Fig. 6C). Presumably, these contacts drive this preference for Pro-Ar combinations at the $\beta_{A,NHB}$ site.

Favored interactions at the $\beta_{A,HB}$ sites

The above analyses revealed a theme in residue-residue interactions at $\beta_{A,NHB}$ sites: all the preferred pairings exhibited intimate side-chain contacts with both residues in the most favored χ_1 conformations (g^+ for Cys, Thr, and Ar residues and t for Val). In order for two residues to be directed toward one another and interact at a $\beta_{A,HB}$ site, one or both must adopt a sterically less favored conformation. This is a consequence of different main-chain geometries at the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites (Figs. 1, 2). Nevertheless, we found that a number of residue pairs were favored at the $\beta_{A,HB}$ sites (Tables 1, 2). Surprisingly, one-half of these pairs contained the bulky aromatic side chains, Phe, Trp, or Tyr. By contrast, one-third of the favored pairs contained glycine, which has only a proton for a side chain and is widely held to be a breaker of secondary structure. Although aromatic residues are known to have a high propensity for β-structure, and a preference for Gly-containing pairs at the $\beta_{A,HB}$ sites has been noted previously (Wouters & Curmi, 1995), no strong physical argument has been presented to explain these preferences. We grouped the favored pairs—as Gly-Ar, Ar-Val, and a smaller set of Ar-Ar pairs—and found structural trends within these groupings that helped rationalize the observed preferences.

Ar-Gly pairs

The direct $\beta_{A,HB}$-$\beta_{A,NHB}$ comparison showed pairings of Gly with Phe, His, Trp, or Tyr favored the $\beta_{A,HB}$ site (Table 2). Although only the value for Gly-Phe was statistically significant, the trend was repeated in the $\beta_{A,HB}$-β_{II2} comparison (Table 2 and Supplementary material). Taken together, the Gly-Ar data show that Ar residues at the β_{II2} sites adopted the sterically preferred g^+ rotamer (Fig. 8). A similar profile was observed for Ar-Gly pairs at $\beta_{A,NHB}$ sites (not shown). However, at $\beta_{A,HB}$ the dominant conformation for the aromatic rings switched to g^- (Fig. 8). This fits with our prediction that the g^- conformer is required to promote

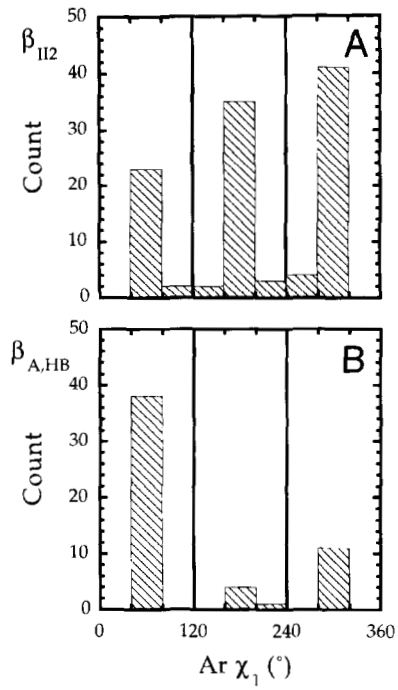

Fig. 8. Comparison of the χ_1 distributions of the aromatic side chains in Gly-Ar pairs at **(A)** β_{II2} and **(B)** $\beta_{A,HB}$ sites. Note the switch in preferred rotamer from the sterically favored g^+ at β_{II2} to g^- at the $\beta_{A,HB}$ sites.

side-chain interactions at $\beta_{A,HB}$ sites, but Gly does not have a side chain with which to interact.

Figures 9A and 9B show overlays for all 15 Gly-Tyr pairs found at $\beta_{A,HB}$ sites in our search. The superposition illustrates how all but two of the aromatic rings adopt the unusual g^- conformation. The χ_2 conformation of these residues all lay around the sterically preferred ±90°. Overall, this conformation placed the aromatic rings directly above the Gly partners and the hydrogen bond systems of the $\beta_{A,HB}$ positions. Similar structures were found for Gly-Phe and Gly-Trp, but there were too few examples of Gly-His to make any confident assertions. We note that one-fourth or less of the aromatic side chains would be expected to adopt the g^- conformation by chance (McGregor et al., 1987; Ponder & Richards, 1987). We argue therefore that this unusual Ar-Gly interaction is driven, and that the driving force is a hitherto unreported aromatic-peptide π-π stacking interaction. In support of this notion, we found that for the next most simple case, the Ar-Ala pairs, all correlations were lost. The Ala-Trp pair was favored at $\beta_{A,HB}$, but Ala-His and Ala-Tyr pairs were evenly distributed between the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites, and the Ala-Phe pair was favored at the $\beta_{A,NHB}$ site above the 95% confidence level. The 12 Ala-Trp pairs found at $\beta_{A,HB}$ sites did not show a strong bias for the g^- (six examples), t (two examples), or g^+ (four examples) rotamers. Furthermore, comparing the Gly-Phe and Ala-Phe pairs at $\beta_{A,HB}$, we found that the preference for Phe in the g^- conformer was lost with the additional methyl; whereas 13 of the 23 Gly-Phe pairs took the g^- conformation, only 4 of the 16 Ala-Phe pairs did so. It is probable that the extra bulk of alanine forces the aromatic ring away from the backbone π-system, and with nothing to gain from adopting the disfavored g^- conformation the aromatic ring takes up g^+.

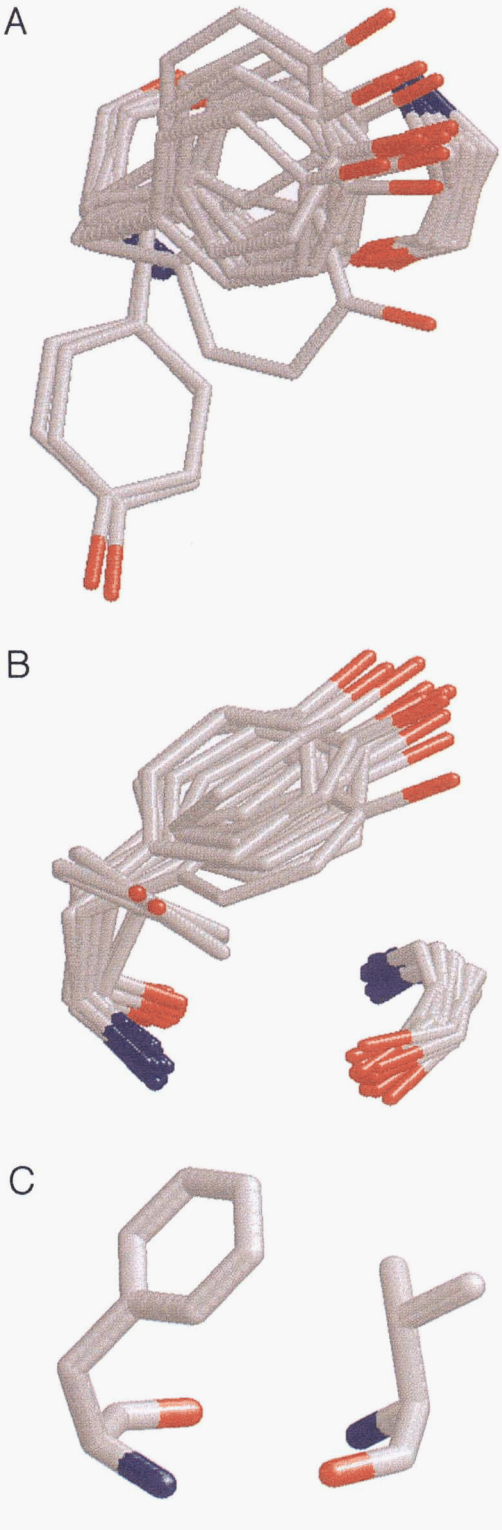

Fig. 9. Examples of interactions made by residue pairs favored at the $\beta_{A,HB}$ sites. **A, B:** Top and side views, respectively, of the superposition of the 15 Tyr-Gly pairs located by our search. In all but two of these cases, the aromatic side chain adopts the sterically least favored g^- conformation, which places the aromatic ring above the glycine backbone and promotes a novel π-π interaction. **C:** A Phe-Val pair (A495 and A427 of 1AOZ). In this case, Phe is in the g^- conformation and Val in t. This leads to good ring-to-C_β van der Waals' interactions. Note how the aromatic ring is forced further away from the partnering backbone than in **B**. Key: blue, nitrogen; red, oxygen; grey, carbon.

Electrostatic interactions involving aromatic residues in proteins are well documented (Burley & Petsko, 1988; Hunter et al., 1991). However, we are unaware of any report of Ar-peptide interactions of the type noted above. Some reference to related interactions in small-molecule studies has been made previously (Hatton & Richards, 1962a, 1962b; Bhacca & Williams, 1964; Holroyd et al., 1993), but little attention has been paid with regard to protein structure and stability (Mitchell et al., 1994). Specific Ar-Gly interactions have been treated (Kemmink et al., 1993; Kemmink & Creighton, 1995; Worth & Wade, 1995; Nardi et al., 1997), but these studies focus on peptide fragments examined in solution where solvent hydrogen bonding dominates, or on Ar-Gly pairs that fall in specific and irregular regions of protein structures.

Ar-Val pairs

All pairings between valine and the hydrophobic aromatic residues, Phe, Trp, and Tyr, were favored above the 99% confidence limit at the $\beta_{A,HB}$ site. This was the case in both the comparison with the β_{II2} data and in the direct $\beta_{A,HB}$-$\beta_{A,NHB}$ comparison (Tables 1, 2). Analysis of the χ_1-χ_1 distributions for Ar-Val pairs at β_{II2}. $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites showed that the dominant valine conformer at all three sites was t, with populations of 68%, 70%, and 69%, respectively. The distributions of Ar conformers were also alike for the three sites, with populations for the g^-, t, and g^+ rotamers of 21 ± 1, 24 ± 3, and 55 ± 2%, respectively. When the possible paired conformations were considered, there was slight preference for the g^-t conformer at the $\beta_{A,HB}$ sites; 19% of the pairs adopted this conformation at $\beta_{A,HB}$ in comparison with 14% at β_{II2}, and 13% at $\beta_{A,NHB}$ (these differences were small, but reliable because the data sets were large, with 238, 144, and 94 Ar-Val pairs at the β_{II2}, $\beta_{A,HB}$, and $\beta_{A,NHB}$ sites, respectively). The g^-t conformation at the $\beta_{A,HB}$ site led to a good packing interaction between the face of the aromatic ring and the C_β atom of the valine (Fig. 9C). We note that, as discussed above for the Ile/Val-Ile/Val pairings at $\beta_{A,NHB}$ sites, the substitution of Ile for Val in Ar-Val pairs is not likely to lead to additional intra-pair van der Waals' contacts. It is possible that this is the reason why Ar-Ile pairs are not generally favored at the $\beta_{A,HB}$ sites like Ar-Val.

Ar-Ar pairs

Possible interactions between aromatic side chains paired across β-sheets have been discussed recently (Wouters & Curmi, 1995; Smith & Regan, 1995). We found that the Phe-Phe pair was favored at the $\beta_{A,HB}$ site in the direct $\beta_{A,HB}$-$\beta_{A,NHB}$ comparison (Table 1). However, a more general preference for Ar-Ar pairs at this site was not evident; in addition to the Phe-Phe pair, only Phe-His and Tyr-Tyr showed differences between the two sites (at the 95% confidence level) (Table 2). Moreover, in the comparisons with the β_{II2} data, of all the possible Ar-Ar pairs, only the His-His pair showed a significant difference from the expected rates (Table 1). Inspection of molecular models and analysis of the χ_1-χ_1 distributions for Ar-Ar pairs at the β_{II2}, $\beta_{A,HB}$, and $\beta_{A,NHB}$ sites offered some explanation for this. At the $\beta_{A,NHB}$ sites the g^+g^+ conformation led to a side-chain clash between the aromatic rings. Indeed, only 2 (3%) of the 78 Ar-Ar pairs that we found at this site adopted this paired conformation, which was a 12-fold lower occupancy than for the corresponding conformation at the control β_{II2} sites. Balanced against this, in the tg^+ conformations, Ar-Ar pairs at the $\beta_{A,NHB}$ sites can achieve archetypal off-set stacked aromatic interaction (Hunter & Sanders, 1990; Hunter et al., 1991), which explains our finding that 52% of these pairs are in tg^+

conformations. It is probable that these factors contribute to an indifference of Ar-Ar pairs for the $\beta_{A,NHB}$ sites that we observe. As discussed above, for most residue pairs, including those with Ar residues, the g^+g^+ conformation does not foster any side-chain interactions at the $\beta_{A,HB}$ positions. In addition, Ar-Ar pairs at this site are precluded from the g^-g^- conformation because of a side-chain clash. Thus, most of the Ar-Ar pairs at the $\beta_{A,HB}$ sites (55%) take up g^-t or g^-g^+ conformations. In these conformers reasonable side-chain interactions were possible, but involved the face, or edge of one ring interacting with the C_β atom of its partner, rather than being intimate aromatic-aromatic interactions.

Solvent accessibilities of the favored pairs

For each residue in our database, we calculated a solvent-accessible surface relative to that of the corresponding residue in a model Gly-Xaa-Gly peptide. These relative accessibilities were used to assess the extent of burial for pairs highlighted by our analysis. Trends in theses data were largely as might be expected on the basis of the side-chain chemistry in the different pairs. For example, for the favored *HH*-type pairs—Val-Val and Cys-Cys at $\beta_{A,NHB}$, and Ar-Gly and Ar-Val at $\beta_{A,HB}$—between 66 and 77% of residues had relative accessibilities of ≤10%. A contribution to this burial could come from intra-pair contacts, but it is most likely that the bulk of it comes from the fact that these pairs will tend to lie in the interiors of proteins. At the $\beta_{A,NHB}$ sites, the Ar-Pro pair and the *PP*-type pairs, Thr-Thr and Asp-Arg, showed bimodal distributions in solvent accessibility. A proportion of the residues in these pairs (42, 19, and 23%, respectively) were almost totally inaccessible to solvent with relative accessibilities of ≤10%; again a contribution to this inaccessibility could come from close side-chain contacts. However, the remainder of the residues showed high relative accessibilities of up to 80% with median values in the range 20–40%. This is consistent with the majority of these pairs falling on the surfaces of protein structures. In the case of the Ar-Pro pairs, which are strictly speaking *HH*-type pairs, the high accessibility is presumably linked to the requirement for Pro to be in a strand on the edge of a β-sheet.

Conclusions

Implications for protein-structure prediction and design

We have described a number of residue pairs that have significantly different preferences for the two types of inter-strand sites that can be distinguished in antiparallel β-structure. In the majority of cases, these preferences are driven by specific interactions that can only be accommodated at the preferred site. Unlike residues that are brought close together within α-helices, there is no restriction on the length of sequence that intervenes between the residues brought together by the formation of a β-sheet; these structures are better considered as tertiary elements rather than local, secondary structures. It is appreciated that the structures of proteins are directed and stabilized by large cooperative networks of interactions of many noncovalent forces. For example, the register of β-strands in antiparallel β-hairpins is influenced by the turns linking them (Searle et al., 1995; de Alba et al., 1997a, 1997b; Ramirez-Alvarado et al., 1997). However, given the strength of the correlations that we observe, the repeated trends and the simplicity of the stereochemical arguments that explain them, it is probable that the interactions highlighted by our analysis exert considerable influence on the selection of the "correct" register in antiparallel β-sheets. It is envisaged that the results of this study will be of use in predicting protein structure from sequence and in the rational design and redesign of peptides and proteins. Indeed, studies using related information are underway both in protein-structure prediction (Hubbard & Park, 1995; Valencia et al., 1995; Frishman & Argos, 1996), and in rational protein design (Smith & Regan, 1995; de Alba et al., 1996; Ramirez-Alvarado et al., 1996).

Methods

Databases and classification of inter- and intra-strand pairs

All data were derived from a set of 311 nonhomologous chains, which were selected from the September 1996 version of the Brookhaven Protein Data Bank (Bernstein et al., 1977) as follows: first, only structures determined by X-ray crystallography to a resolution of 2.5 Å, or better were used; second, structures with >20% of residues outside the core ϕ-ψ regions as judged by PROCHECK (Laskowski et al., 1993) were rejected; third, protein chains were chosen such that no two had more than 25% sequence identity; finally, multiple copies of structural analogues, identified by the structural-alignment program SSAP (Orengo et al., 1992), were eliminated. Secondary structure assignments for the 311 protein chains were made using a modified version (D.K. Smith, unpubl.) of the DSSP algorithm (Kabsch & Sander, 1983). These assignments were used to identify one type of intra-strand pair (β_{II2}) in all β-strands, and two types of inter-strand residue pairs ($\beta_{A,HB}$ and $\beta_{A,NHB}$ sites) in antiparallel β-structure. The intra-strand pair was for residues displaced i to $i + 2$ from one another. In this case, residues from $i - 1$ through to $i + 3$ were all required to be in β-conformations; i.e., in EEEEE, eEEEE, EEEEe, and eEEEe units. For the inter-strand sites, pairs were selected if two residues involved in a bridge (Kabsch & Sander, 1983), and flanked on either side by at least one residue in a β-conformation, i.e., EEE, eEE, EEe, and eEe units. These pairs were classified on the basis of the number of hydrogen bonds between the bridge partners to distinguish (1) antiparallel bridges with two hydrogen bonds ($\beta_{A,HB}$ sites) and (2) antiparallel bridges with no hydrogen bonds ($\beta_{A,NHB}$ sites). The total number of pairs (T) found at each of the sites, β_{II2}, $\beta_{A,HB}$, and $\beta_{A,NHB}$, were 7,076, 2,061, and 3,065, respectively. For each case, the number of occurrences of each of the 210 possible combinations of the 20 amino acids were counted and collected in 20 × 20 contingency tables (Supplementary data).

Statistical methods

Contingency tables for the β_{II2}, $\beta_{A,HB}$, and $\beta_{A,NHB}$ were analyzed as a whole in the traditional manner by comparing observed (O_{rc}) and expected (E_{rc}) occupancy for each element (rc) in chi-square tests. Expected values were calculated as follows:

$$E_{rc} = \frac{\sum_c O_r \cdot \sum_r O_c}{T}. \quad (1)$$

These values were used in chi-square analyses:

$$\chi^2 \equiv \sum_{r,c} \frac{(Orc - Erc)^2}{Erc}. \quad (2)$$

Significance levels were calculated using standard tables assuming 190 degrees of freedom for each table.

In addition, individual elements of the $\beta_{A,HB}$ and $\beta_{A,NHB}$ tables were compared with corresponding elements of β_{II2} data using the statistical method of standard error of proportion. We have used this method previously to tackle a similar problem (Woolfson & Alber, 1995). In this procedure the proportions of residues at a particular element are compared for the two data sets and a z-score determined. For example, the comparison of the $\beta_{A,HB}$ and β_{II2} data used the following equations:

$$Z_{rc} = \frac{P_{rc,\beta A,HB} - P_{rc,\beta II2}}{\sigma_{rc}} \quad (3)$$

where

$$P_{rc,\beta A,HB} = \frac{O_{rc,\beta A,HB}}{T_{\beta A,HB}} \quad (4)$$

$$P_{rc,\beta II2} = \frac{O_{rc,\beta II2}}{T_{\beta II2}} \quad (5)$$

and

$$\sigma_{rc} = \sqrt{\frac{(P_{rc,\beta A,HB})\cdot(1 - P_{rc,\beta A,HB})}{T_{\beta A,HB}} + \frac{(P_{rc,\beta II2})\cdot(1 - P_{rc,\beta II2})}{T_{\beta II2}}}. \quad (6)$$

Values of z_{rc} lying outside the range -2.58 to 2.58 were taken to indicate that the two elements under scrutiny differed above the 99% confidence limit. The 95% confidence limits were taken as ± 1.96. z-Scores were only calculated when both of the observed counts were 5.

The $\beta_{A,HB}$ and $\beta_{A,NHB}$ data sets were compared directly with z-scores calculated from the standard errors of proportion using equations similar to those given above, but with appropriate substitutions for the P and T values. In addition, the ratios (R) of the occupancies at these sites were obtained thus:

$$R_{rc} = \frac{P_{rc,\beta A,HB}}{P_{rc,\beta A,NHB}}. \quad (7)$$

Manipulation and visualization of molecular structures and structural data

Protein structures were visualized and manipulated in INSIGHT II (MSI, San Diego, California) and using CPK models (Harvard Apparatus, Holliston, Massachusetts). Protein structure figures were created using RasMol V2.6 (Sayle & Milner-White, 1995). Dihedral-angle data were manipulated and visualized using KaleidaGraph V3.0.2 (Synergy Software, Reading, Massachusetts). Newman projections were created in ChemDraw Plus V3.0.2 (Cambridge Soft Corporation, Cambridge, Massachusetts).

Molecular-dynamics simulations

Parent peptide

A cyclic 18-residue peptide was constructed using molecular graphics on an SGI Indigo^2 workstation. The structure described by the notation {AA-AAAAA-AG}$_2$ was energy minimized (Discover V2.95) (MSI, San Diego, California) in vacuum. This structure was soaked in a box of water molecules $40 \times 25 \times 25$ Å, minimum image periodic boundary conditions were applied, with a nonbonded energy cut-off of 11 Å switched to 0 between 9 and 11 Å by a smoothing function. This system was energy minimized for 1,000 cycles of conjugate gradient minimization then subjected to 200 ps of molecular dynamics at 300 K using the leapfrog algorithm with an integration step of 1 fs. Structures were saved every 1 ps for further analysis.

Cystine-containing peptides

Four cystine linked cyclic peptides were constructed from the parent peptide by substituting two alanines with cysteine in each case. The cysteine χ_1 torsion angles were altered by hand, to best approximate a disulfide bond distance between the two sulfur atoms, and the cystine created. This resulted in the two peptides with cystine at $\beta_{A,NHB}$ sites, {AA-AACAA-AG}$_2$ and {AA-CAAAA-AG}$_2$, which had g^+ χ_1 angles for the half cystines; and two peptides with cystines at $\beta_{A,HB}$ sites, {AA-ACAAA-AG}$_2$ and {AC-AAAAA-AG}$_2$, which had half-cystines with g^- χ_1 angles. Each structure was minimized, soaked, and subjected to 200 ps molecular dynamics as described for the parent peptide. Every tenth dynamics frame (equivalent to 10 ps), (peptide plus water) was energy minimized to a maximum derivative less than 0.5 kcal/mol to give 21 local minima, for each system, close to the structure of the peptide at that point in the dynamics trajectory. This produced a set of representative peptide structures whose internal energies could be compared.

Analysis of simulations

The trajectories of the eight β-sheet hydrogen bonds and eighteen χ angles for each of the five dynamics simulations were extracted from the history files using FOCUS (Sessions et al., 1989) and displayed graphically. The cystine side-chain torsion angles and bond angles were analyzed likewise. Representative examples of these data are shown in Figure 7, and initial and final structures from the simulations can be found in Kinemages 1–6.

Electronic supplementary material

Supplementary data include: Three tables of raw count data for the β_{II2} (Table S1), $\beta_{A,HB}$ (Table S2), and $\beta_{A,NHB}$ (Table S3) positions. Tables S2 and S3 also give the propensities, and associated statistical significances, for all residue pairs at the $\beta_{A,HB}$ and $\beta_{A,NHB}$ sites, respectively. These were measured relative to frequencies of occurrence of the corresponding pairs at the β_{II2} sites. Six kinemages for the initial and final model peptide structures from the molecular-dynamics studies; Kinemages 1 and 2 are for the parent peptide, Kinemages 3 and 4 are for a peptide with cystine at a $\beta_{A,NHB}$ site, and Kinemages 5 and 6 are for a peptide with cystine placed at $\beta_{A,HB}$.

Acknowledgments

We thank Ian Badcoe, Tony Clarke, and Dudley Williams for useful discussions on various aspects of this work, and Zoltan Dienes for extremely helpful advice on the statistical analysis. We are grateful to the BBSRC of the UK, the Royal Society, and the Wellcome Trust for financial support.

References

Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science 181*:223–230.

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Betz SF, Bryson JW, Degrado WF. 1995. Native-like and structurally characterized designed alpha-helical bundles. *Curr Opin Struct Biol 5*:457–463.

Betz SF, Raleigh DP, Degrado WF. 1993. De-novo protein design—From molten globules to native-like states. *Curr Opin Struct Biol 3*:601–610.

Betz SF, Raleigh DP, DeGrado WF, Lovejoy B, Anderson D, Ogihara N, Eisenberg D. 1996. Crystallization of a designed peptide from a molten-globule ensemble. *Fold Design 1*:57–64.

Bhacca NS, Williams DH. 1964. *Applications of NMR spectroscopy in organic chemistry*. San Francisco, London, Amsterdam: Holden-Day.

Burley SK, Petsko GA. 1988. Weakly polar interactions in proteins. *Adv Protein Chem 39*:125–189.

Chou PY, Fasman GD. 1974. Conformational parameters for amino acids in helical, beta-sheet and random coil regions calculated from proteins. *Biochemistry 13*:211–222.

Dahiyat BI, Mayo SL. 1997. De novo protein design: Fully automated sequence selection. *Science 278*:82–87.

de Alba E, Angeles-Jiménez M, Rico M. 1997a. Turn residue sequence determines beta-hairpin conformation in designed peptides. *J Am Chem Soc 119*:175–183.

de Alba E, Angeles-Jimémez M, Rico M, Nieto JL. 1996. Conformational investigation of designed short linear peptides able to fold into beta-hairpin structures in aqueous solution. *Fold Design 1*:133–144.

de Alba E, Rico M, Angeles-Jiménez M. 1997b. Cross-strand side-chain interactions versus turn conformation in β-hairpins. *Protein Sci 6*:2548–2560.

Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*:7133–7155.

Finkelstein AV. 1995. Predicted beta-structure stability parameters under experimental test. *Protein Eng 8*:207–209.

Frishman D, Argos P. 1996. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng 9*:133–142.

Gonzalez LJ, Woolfson DN, Alber T. 1996. Buried polar residues and structural specificity in the GCN4 leucine-zipper. *Nature Struct Biol 3*:1011–1018.

Gunasekaran K, Ramakrishnan C, Balaram P. 1997. Beta-hairpins in proteins revisited: Lessons for de novo design. *Protein Eng 10*:1131–1141.

Harbury PB, Zhang T, Kim PS, Alber T. 1993. A switch between 2-stranded, 3-stranded and 4-stranded coiled coils in GCN4 leucine-zipper mutants. *Science 262*:1401–1407.

Hatton JV, Richards RE. 1962a. NMR shifts in aromatic solvents. *Mol Phys 5*:153–159.

Hatton JV, Richards RE. 1962b. Solvent effects in NMR spectra of amide solutions. *Mol Phys 5*:139–152.

Holroyd SE, Groves P, Searle MS, Gerhard U, Williams DH. 1993. Rational design and binding of modified cell-wall peptides to vancomycin-group antibiotics-factorizing free-energy contributions to binding. *Tetrahedron 49*:9171–9182.

Huang ES, Subbiah S, Levitt M. 1995. Recognizing native folds by the arrangement of hydrophobic and polar residues. *J Mol Biol 252*:709–720.

Hubbard TJ, Park J. 1995. Fold recognition and ab-initio structure predictions using hidden markov-models and beta-strand pair potentials. *Proteins Struct Funct Genet 23*:398–402.

Hunter CA, Sanders JKM. 1990. The nature of π-π interactions. *J Am Chem Soc 112*:5525–5534.

Hunter CA, Singh J, Thornton JM. 1991. π-π-interactions—The geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J Mol Biol 218*:837–846.

Huyghues-Despointes BMP, Klingler TM, Baldwin RL. 1995. Measuring the strength of side-chain hydrogen bonds in peptide helices—The Gln-center-dot-Asp-(i,i + 4) interaction. *Biochemistry 34*:13267–13271.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure—Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers 22*:2577–2637.

Kemmink J, Creighton TE. 1995. The physical properties of local interactions of tyrosine residues in peptides and unfolded proteins. *J Mol Biol 245*:251–260.

Kemmink J, Vanmierlo CPM, Scheek RM, Creighton TE. 1993. Local-structure due to an aromatic amide interaction observed by H-1-nuclear magnetic-resonance spectroscopy in peptides related to the n terminus of bovine pancreatic trypsin-inhibitor. *J Mol Biol 230*:312–322.

Kemp DS. 1990. Peptidomimetics and the template approach to nucleation of beta-sheets and alpha-helices in peptides. *Trends in Biotechnology 8*:249–255.

Kemp DS, Bowen BR, Muendel CC. 1990. Synthesis and conformational-analysis of epindolidione-derived peptide models for beta-sheet formation. *J Organic Chem 55*:4650–4657.

Kim CA, Berg JM. 1993. Thermodynamic beta-sheet propensities measured using a zinc finger host peptide. *Nature 362*:267–270.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: Program to check the stereochemical quality of protein structures. *J App Cryst 26*:283–291.

Lazar GA, Desjarlais JR, Handel TM. 1997. De novo design of the hydrophobic core of ubiquitin. *Protein Sci 6*:1167–1178.

Levitt M. 1978. Conformational preferences for amino acids in globular proteins. *Biochemistry 17*:4277–4285.

Lifson S, Sander C. 1980. Specific recognition in the tertiary structure of β-sheets of proteins. *J Mol Biol 139*:627–639.

Lim VI. 1974a. Algorithms for prediction of α-helical and β-structural regions of globular proteins. *J Mol Biol 88*:873–894.

Lim VI. 1974b. Structural principles of the globular organization of protein chains: A stereochemical theory of globular protein secondary structure. *J Mol Biol 88*:857–872.

McGregor MJ, Islam SA, Sternberg MJ. 1987. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J Mol Biol 198*:295–310.

Minor DL, Kim PS. 1994a. Measurement of the beta-sheet-forming propensities of amino acids. *Nature 367*:660–663.

Minor DL, Kim PS. 1994b. Context is a major determinant of beta-sheet propensity. *Nature 371*:264–267.

Mitchell JBO, Nandi CL, McDonald IK, Thornton JM, Price SL. 1994. Amino/aromatic interactions in proteins—Is the evidence stacked against hydrogen-bonding? *J Mol Biol 239*:315–331.

Muñoz V, Serrano L. 1994. Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices—Comparison with experimental scales. *Proteins Struct Funct Genet 20*:301–311.

Munson M, Balasubramanian S, Fleming KG, Nagi AD, Obrien R, Sturtevant JM, Regan L. 1996. What makes a protein a protein—Hydrophobic core designs that specify stability and structural properties. *Protein Sci 5*:1584–1593.

Nardi F, Worth GA, Wade RC. 1997. Local interactions of aromatic residues in short peptides in aqueous solution: A combined database and energetic analysis. *Fold Design 2*:S62–S68.

Orengo CA, Brown NP, Taylor WR. 1992. Fast structure alignment for protein databank searching. *Proteins 14*:139–167.

Otzen DE, Fersht AR. 1995. Side-chain determinants of beta-sheet stability. *Biochemistry 34*:5718–5724.

Padmanabhan S, Baldwin RL. 1994a. Helix-stabilizing interaction between tyrosine and leucine or valine when the spacing is i,i + 4. *J Mol Biol 241*:706–713.

Padmanabhan S, Baldwin RL. 1994b. Tests for helix-stabilizing interactions between various nonpolar side-chains in alanine-based peptides. *Protein Sci 3*:1992–1997.

Pham TN, Koide A, Koide S. 1998. A stable single layer beta-sheet without a hydrophobic core. *Nature Struct Biol 5*:115–119.

Ponder JW, Richards FM. 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.

Raleigh DP, DeGrado WF. 1992. A de novo designed protein shows a thermally induced transition from a native to a molten globule-like state. *J Am Chem Soc 114*:10079–10081.

Ramirez-Alvarado M, Blanco FJ, Niemann H, Serrano L. 1997. Role of beta-turn residues in beta-hairpin formation and stability in designed peptides. *J Mol Biol 273*:898–912.

Ramirez-Alvarado M, Blanco FJ, Serrano L. 1996. De-novo design and structural-analysis of a model beta-hairpin peptide system. *Nature Struct Biol 3*:604–612.

Roy S, Ratnaswamy G, Boice JA, Fairman R, McLendon G, Hecht MH. 1997. A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. *J Am Chem Soc 119*:5302–5306.

Sayle RA, Milner-White EJ. 1995. Rasmol—Biomolecular graphics for all. *Trends Biochem Sci 20*:374–376.

Searle MS, Williams DH, Packman LC. 1995. A short linear peptide derived from the N-terminal sequence of ubiquitin folds into a water-stable non-native beta-hairpin. *Nature Struct Biol 2*:999–1006.

Sessions RB, Dauber-Osguthorpe P, Osguthorpe DJ. 1989. Filtering molecular-dynamics trajectories to reveal low-frequency collective motions: Phospholipase A_2. *J Mol Biol 210*:617–633.

Smith CK, Regan L. 1995. Guidelines for protein design—The energetics of beta-sheet side-chain interactions. *Science 270*:980–982.

Smith CK, Withka JM, Regan L. 1994. A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry 33*:5510–5517.

Stapley BJ, Doig AJ. 1997. Hydrogen bonding interactions between glutamine and asparagine in α-helical peptides. *J Mol Biol 272*:465–473.

Sun SJ, Thomas PD, Dill KA. 1995. A simple protein-folding algorithm using a binary code and secondary structure constraints. *Protein Eng 8*:769–778.

Swindells MB, MacArthur MW, Thornton JM. 1995. Intrinsic phi,psi propensities of amino acids, derived from the coil regions of known structures. *Nature Struct Biol 2*:596–603.

Tisi LC, Evans PA. 1995. Conserved structural features on protein surfaces—Small exterior hydrophobic clusters. *J Mol Biol 249*:251–258.

Valencia A, Hubbard TJ, Muga A, Banuelos S, Llorca O, Carrascosa JL, Valpuesta JM. 1995. Prediction of the structure of GroES and its interaction with GroEL. *Proteins Struct Funct Genet 22*:199–209.

West MW, Hecht MH. 1995. Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. *Protein Sci 4*:2032–2039.

Woolfson DN, Alber T. 1995. Predicting oligomerization states of coiled coils. *Protein Sci 4*:1596–1607.

Worth GA, Wade RC. 1995. The aromatic-(i + 2) amine interaction in peptides. *J Phys Chem 99*:17473–17482.

Wouters MA, Curmi PMG. 1995. An analysis of side-chain interactions and pair correlations within antiparallel beta-sheets: The differences between backbone hydrogen bonded and non-hydrogen-bonded residue pairs. *Proteins 22*:119–131.