Protein Engineering vol.15 no.2 pp.91–100, 2002

Aromatic–aromatic interactions in and around α-helices

Rajasri Bhattacharyya[1,2], Uttamkumar Samanta[2,3] and Pinak Chakrabarti[1,3,4]

[1]Department of Biochemistry and [3]Bioinformatics Centre, Bose Institute, P-1/12 CIT Scheme VIIM, Calcutta 700 054, India
[2]R.Bhattacharyya and U.Samanta contributed equally to this work
[4]To whom correspondence should be addressed.
E-mail: pinak@boseinst.ernet.in

To understand the role of aromatic–aromatic interactions in imparting specificity to the folding process, the geometries of four aromatic residues with different sequence spacing, located in α-helices or five residues from helical ends, interacting with each other have been elucidated. The geometry is found to depend on the sequence difference. Specific interactions (C–H···π and N–H···π) which result from this geometry may cause a given pair of residues (such as Phe–His) with a particular sequence difference to occur more than expected. The most conspicuous residue in an aromatic pair in the context of helix stability is His, which is found at the last (C1) position or the two positions (Ncap and Ccap) immediately flanking the helix. An α-helix and a contiguous 3_{10}-helix or two helices separated by a non-helical residue can have interacting aromatic pairs, the geometry of interaction and the relative orientation between the helices being rather fixed. Short helices can also have interacting residues from either side.

Keywords: α-helix/aromatic residues/interaction geometry/modelling/protein folding

Introduction

Weak interactions, although small in magnitude, being numerous, may contribute to the overall stability of protein molecules (Burley and Petsko, 1988; Samanta *et al.*, 2000) and also to the occurrence of special structural features, such as *cis* peptide bonds (Pal and Chakrabarti, 1999). The hallmark of such interactions is the specific geometry between the two interacting groups, usually comprising side chains (Singh and Thornton, 1992), but can also involve main-chain atoms (Pal and Chakrabarti, 1998). The distribution of interplanar angles between interacting side chains with well defined planar regions is non-random, smaller angles being found in excess (Mitchell *et al.*, 1997). However, to specify the relative orientation between two planar groups, additional parameters are needed (Brocchieri and Karlin, 1994; Samanta *et al.*, 1999). Additionally, it is not known if the geometry can also depend on the location of the residues in a given secondary structure. α-Helices, in which two side chains in the adjacent position or in the neighbouring turns can, in principle, be in contact with each other, offer a system where such a feature can be studied.

Interaction between aromatic residues (Phe, Tyr, His and Trp) can contribute to the stability of the native fold (Burley and Petsko, 1985) and it has been suggested that aromatic clusters could be a determinant of thermal stability in thermophilic proteins (Kannan and Vishveshwara, 2000). An aromatic residue has been implicated in the origin of DNA-binding specificity of Arc repressor (Schildbach *et al.*, 1999). For the proper understanding of the specificity of binding and stability, it is necessary to know if the aromatic rings have any preferred geometry of interaction and indeed the perpendicular (edge-to-face) orientation is generally favoured (Burley and Petsko, 1985; Singh and Thornton, 1985; Hunter *et al.*, 1991). However, none of these studies addressed the question of whether the geometry can vary depending on the relative location in a helix, if any particular pair is preferred among all the possible pairs of interacting aromatic residues and if within a pair a given residue has a preference to precede the other in the sequence.

Elucidation and quantification of different non-covalent interactions provide a foundation on which the modification of protein stability and the design of small proteins can be made in a rational way. It has been possible to dissect the contribution of local interactions to α-helix stability using synthetic model peptides (Chakrabartty and Baldwin, 1995; Serrano, 2000). For example, the interaction of Phe8 with His12 has been shown to account for the unexpected stability of the C-peptide helix (the N-terminal 13 residues of RNase A) (Dadlez *et al.*, 1988; Shoemaker *et al.*, 1990; Armstrong *et al.*, 1993). Likewise, Trp and His at positions i and $i + 4$, respectively, can stabilize the helix by 1 kcal/mol when the histidine is protonated (Fernandez-Recio *et al.*, 1997). For the design of supersecondary structures containing α-helix, it is also imperative to know the interactions that straddle the helix terminus, linking residues in helix to those beyond (Aurora and Rose, 1998). For a better understanding of the role of aromatic residues in providing stability to α-helices and their immediate neighbourhood, we analysed all known protein structures to identify pairs of interacting aromatic residues, both located within α-helices or one in a helix and the other within five residues of a helix-end, and then found the geometry of interaction. This led to the enumeration of patterns in sequence and structure involving the aromatic residues.

Materials and methods

Atomic coordinates of protein molecules were obtained from the Protein Data Bank (PDB), now operated by the Research Collaboratory for Structural Bioinformatics (RCSB; http://www.rcsb.org) (Berman *et al.*, 2000). The proteins selected were based on the January 2000 release of the representative list of structures (Hobohm and Sander, 1994) determined at a resolution of 2.0 Å or better and *R*-factor ⩽20%; the maximum sequence identity between any two of the polypeptide chains was ⩽25%.

The α-helices were located using the program DSSP (Kabsch and Sander, 1983) and all the aromatic residues within helices and five residues from either end of the helices were identified;

those with more than one conformation of the side chain were, however, excluded from further calculations. Distances between ring centroids of all aromatic residues in and around a given α-helix were calculated and any pair of residues within a distance of 7.5 Å [which was found to be optimal by Samanta *et al.* (Samanta *et al.*, 1999)] was assumed to be interacting. For the calculation of the geometry of interaction between the two rings, their centroids were calculated (for Trp, the mid-point of the CD2–CE2 bond was taken as the centroid) and a molecular axes system defined (with the origin at the centroid of the aromatic residue which precedes in the sequence and the *z*-axis perpendicular to its ring plane). Two parameters were calculated, theinterplanar angle (*P*) and the angle (θ) made by the *z*-axis with the line joining the origin to the centroid of the second ring.

The difference (Δ) in residue numbers between the two interacting residues, $[Ar_i]_1$ and $[Ar_j]_2$, where $Ar_{i/j}$ stands for any of the four aromatic residues whose relative positions in the sequence are indicated by the subscript 1 or 2 and the numbers ($O_{i,j}$) of the various combinations of $[Ar_i]_1$–$[Ar_j]_2$ pairs observed (for a given Δ) were found. The expected number ($E_{i,j}$) of the pairs is proportional to the product of the fractional abundances of the two residues in positions 1 and 2:

$$E_{i,j} = N\{([n_i]_1/N) \times ([n_j]_2/N)\}$$

where

$$n_i = \sum_{j=1}^{4} n_{i,j} \text{ and } N = \sum_{i=1}^{4} n_i$$

and n_i and n_j are the occurrences of each residue. All the pairs with the observed value of at least 5 and the observed to expected ratio of >1.2 or <0.8 were marked. To see if residues pair up in any sequential order (type *i* preceding *j* or vice versa) the differences between $n_{i,j}$ and $n_{j,i}$ were found. This difference is associated with an error, $(n_{i,j} + n_{j,i})^{1/2}$, based on counting statistics. The cases with the quantity $1n_{i,j} - n_{j,i}1/(n_{i,j} + n_{j,i})^{1/2}$ exceeding 1.5 were taken as significant.

The solvent accessible surface area (ASA) was computed using the program ACCESS (Hubbard, 1992), which is an implementation of the Lee and Richards (1971) algorithm. The relative accessibility of a residue (X) was obtained by dividing the observed ASA by its ASA in a model tri-peptide, Ala–X–Ala. Cartoon representations of molecules were generated using MOLSCRIPT (Kraulis, 1991).

The codes for the PDB files used are given in the Appendix.

Results and discussion

Content of aromatic residues in α-helices

There are 2280 α-helices in 434 protein chains considered, of which 1553 contain 0 or 1 aromatic residue and thus have no pairwise contact. The number of residues which are aromatic as a function of helix length is plotted in Figure 1. The fitted straight line suggests that on average one aromatic residue is incorporated into the helix as its length increases by 10. This number is quite close to 11.4, the average percentage composition of aromatic residues in protein structures (Chakrabarti and Pal, 2001). One feature of the plot, for which there is no obvious explanation, is the dip in the number of aromatic residues when the length is 9 and 18. If one looks at the percentages of helices with different numbers of aromatic residues (Figure 2), one finds that the number of helices with

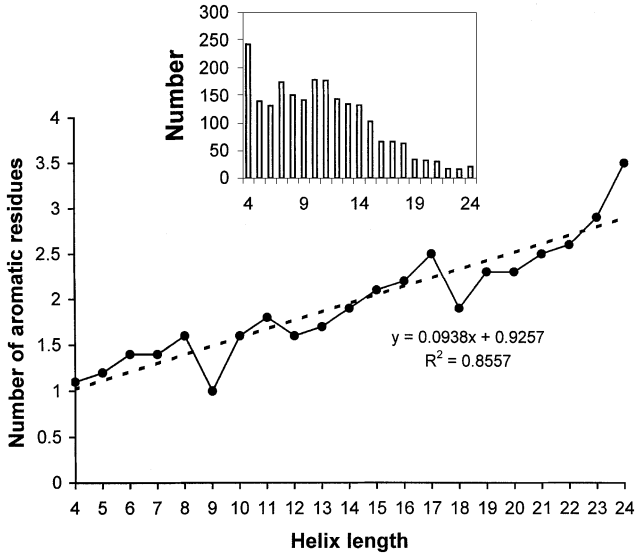

Fig. 1. The average number of aromatic residues in a helix containing a given number of residues (helix length). The dashed line shows the fitted equation. The inset shows the observed number of helices of a particular length.

no aromatic residue falls off monotonically as the length of the helix increases, but it is still possible to find long helices devoid of any aromatic residue. The numbers of helices with four or more aromatic residues show a reverse trend, appearing at a length of about 10. In between these two groups the percentages of helices with one or two aromatic residues remain fairly uniform over the helix length, being 30–40% for the former (for helices up to a length of 20) and 20–30% for the latter (when the helix is 9–20 residues long).

Interaction between helical residues and their sequence difference

Geometric considerations dictate that two helical residues can be in contact (centroid within 7.5 Å) only when there are specific sequence differences (Δ) between them. Table I shows that 50% of all residues interact when they are located one or three residues apart, whereas a much greater percentage (94%) can interact when they are four residues apart. This is also shown by a shorter contact distance which is possible for the latter relative position. The average values of the closest distance between any two atoms of the interacting pair for different Δ values are: 4.1(6) Å for 1; 4.3(7) Å for 3 and 3.9 (6) Å for 4. Interestingly enough, even when the two residues are one helical turn away (Δ = 7 or 8) there are examples having interaction.

Geometry and the side-chain conformation

As discussed by Samanta *et al.* (Samanta *et al.*, 1999), the relative orientation of the second ring with respect to the first (in a pair) can be idealized at three sets of discrete values of the two angular parameters *P* and Θ (Figure 3). Each relative orientation is designated by a two-letter code (*ff*, *ot*, *ef*, etc.). The first letter indicates if the first residue in a pair is interacting with its face (*f*), edge (*e*) or has the centroid of the second ring in an intermediate (offset or *o*) position. The second letter indicates if the second residue is tilted (*t*) with respect to the first or has its face (*f*) or edge (*e*) pointing towards the first.

The geometry observed when the two residues in an α-helix are interacting depends on the sequence difference and is very different when Δ = 1 than when it is 3 or 4 (Table II). In the

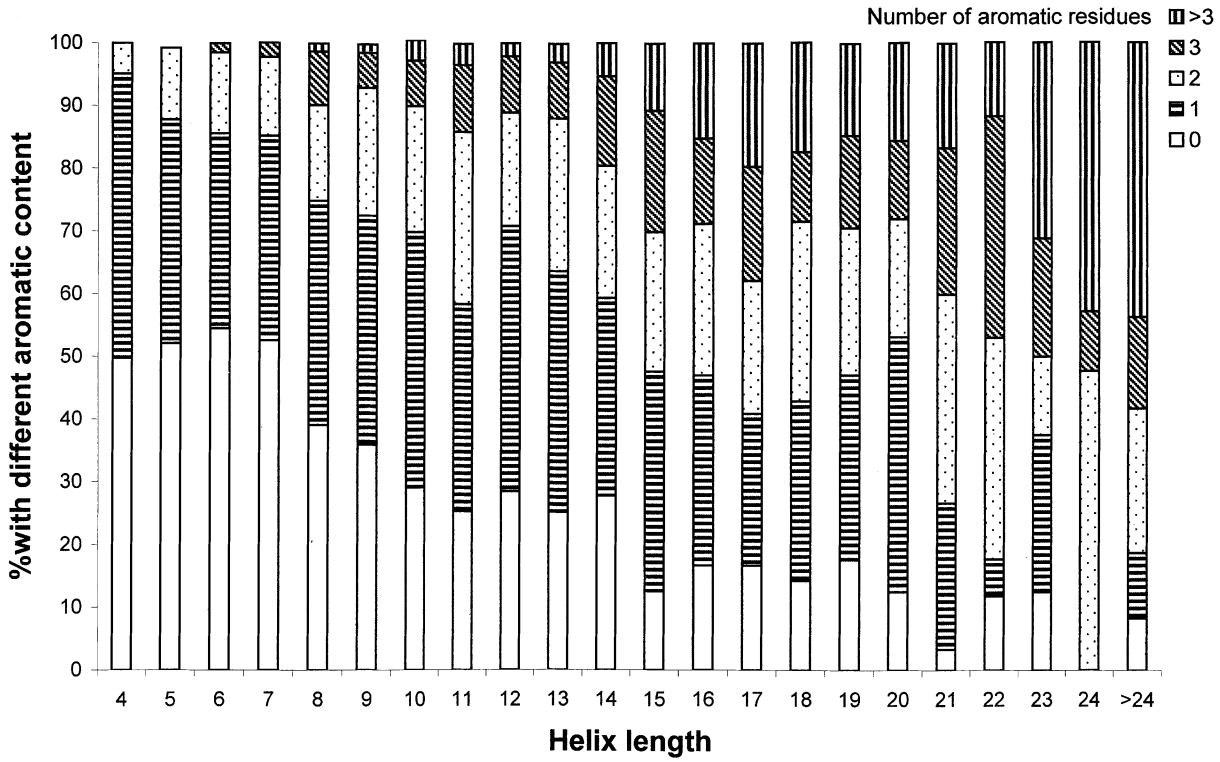

Fig. 2. Percentages of helices containing 0, 1, 2, 3 and ⩾4 aromatic residues plotted against helix length.

Table I. For a particular sequence difference (Δ), the numbers of residues observed and interacting in α-helices[a]

Δ	1	2	3	4	5	6	7	8	9	10
No. observed	300	178	236	177	89	64	60	42	20	20
No. interacting	147	0	125	167	4	0	1	3	0	0

[a]Only pairs of closest (along the sequence) aromatic residues are considered while defining Δ.

former, the predominant geometry is *ef (*shown by 68% of the cases). When Δ = 4, 49% of the geometries are of types *ot* and *of*; *ot* is the most preferred orientation (followed by *of* and *ee*) when Δ = 3. Of the four cases of interactions with Δ = 5, three have the geometry *oe* and one *ot*. Some examples of different geometries are given in Figure 4. When Δ = 8 (also see Table VI), there is a kink in the helix.

The geometry observed must be the most stabilizing one attainable under the constraint of the relative locations of the two residues in the helix and the possible side-chain conformations. Although the latter is characterized by three staggered values of χ_1 torsion angle, 180°, 60° and –60° (or 300°) (designated as *t*, g^- and g^+, respectively), because of the steric reason the g^- state is usually avoided in a helix (McGregor *et al.*, 1987; Chakrabarti and Pal, 2001). Figure 5 depicts the distribution of χ_1 angles for the two residues of the interacting pair. When Δ = 1, for the overwhelming majority, the first residue is in the *t* state and the second in g^+. For Δ = 3 and 4, in the majority of the cases both the χ_1s are in the *t* state. The percentages of various combinations of χ_1 angles for the two residues are as follows: when Δ = 3, *tt*: 59, g^+g^+: 13, tg^+: 2 and g^+t: 15; the corresponding values when Δ = 4 are 40, 17, 37 and 0, respectively (the rest being the outliers, i.e. residues not within 30° of the canonical angles for *t* and g^+ states).

Residues in the interacting pairs

It is of interest to know if any particular combination of aromatic residues is found to interact more than expected and data presented in Table III identify a few such pairs: His–His when Δ = 1 and 4 and Phe–His, Tyr–Trp and Trp–Phe pairs when Δ = 3. The smallest (His) and the largest (Trp) aromatic residues are found as one of the interacting partners. Calculations on relative ordering of the two residues in a pair (see Materials and methods) also suggest that the Phe–His pair has a much greater chance of occurrence than the His–Phe pair when Δ = 3 or 4 (particularly the latter).

Specific interaction between Phe (i) and His (i + 4) in α-helix

Using an alanine-based peptide system, Armstrong *et al.* have shown that the interaction between Phe and protonated His at relative position *i* / *i* + 4 (i.e. Δ = 4) is helix stabilizing and is independent of the exact location of the pair in the helix (Armstrong *et al.*, 1993). Likewise, Fernandez-Recio *et al.* studied the helical content of several peptides in which a Trp–His pair was placed at *i*, *i*+3 or *i*, *i* + 4 in either the N to C or the C to N orientation and found that the Trp–His pair at *i*, *i* + 4 gives rise to the highest helical content (i.e. the most stabilizing) when the histidine is protonated (Fernandez-Recio *et al.*, 1997). However, none of these dealt with the geometry between the two rings. As discussed in the previous section, our analysis shows that the Phe–His pair is observed significantly more than in the reverse order both when Δ =3 and 4. The geometry of His with respect to Phe is *ot* or *oe* (Figure 3) when Δ = 3, similar to other pairs with the same sequence difference (Table II). For the 12 cases observed with Δ = 4, the majority (seven) are in the geometry *ft* or *ot* (an

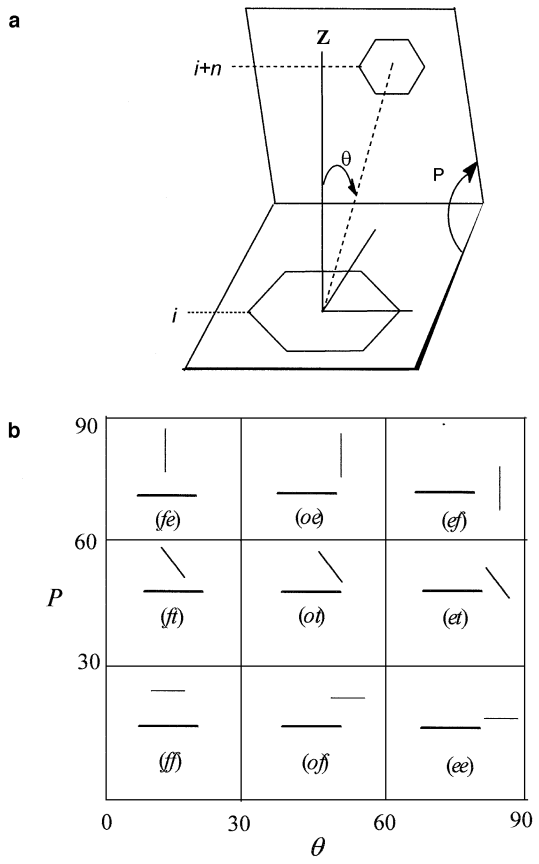

Fig. 3. (**a**) Definitions of *P*, the interplanar angle, and Θ, the angular location of the centroid of the aromatic ring of a helical residue (at $i + n$) from the *z*-axis of an axial system centred on the centroid of a preceding aromatic ring (at *i*). (**b**) Schematic representations and their nomenclature for ring orientations corresponding to various combinations of *P* and Θ values (in °). Lines signify aromatic planes (the longer one is for the residue with the smaller sequence number in an interacting pair) perpendicular to the paper.

Table II. Geometry of interaction of helical aromatic residues with three values of sequence-difference (Δ)[a]

(a) Δ = 1 (147 cases)			(b) Δ = 3 (125 cases)			(c) Δ = 4 (167 cases)		
4	9	100	9	16	14	7	19	6
1	8	23	3	33	9	16	41	16
0	2	0	0	21	20	8	40	14

[a]The numbers of observations at nine boxes corresponding to the geometry shown in Figure 3 are given.

additional two in *of* and the rest are scattered in other boxes of Figure 3). With the tilted orientation of the His ring relative to Phe (when Δ = 4), an X–H group (where X = C or N) usually at CD2 or ND1 of His is directed towards the π electron cloud of the Phe ring [with an average X···C distance of 3.7(5) Å] (Figure 6) resulting in a C–H···π or N–H···π interaction (Mitchell *et al.*, 1994; Nishio *et al.*, 1998; Samanta *et al.*, 2000). When the His ring is protonated, the N–H group at ND1 becomes available and the other C–H protons in the ring carry more partial positive charge, resulting in the occurrence or the strengthening of the X–H···π interaction (Samanta *et al.*, 1998). This explains the result from the solution studies showing an increase in the helical content of model peptides containing an aromatic–His pair (at *i*, $i + 4$)

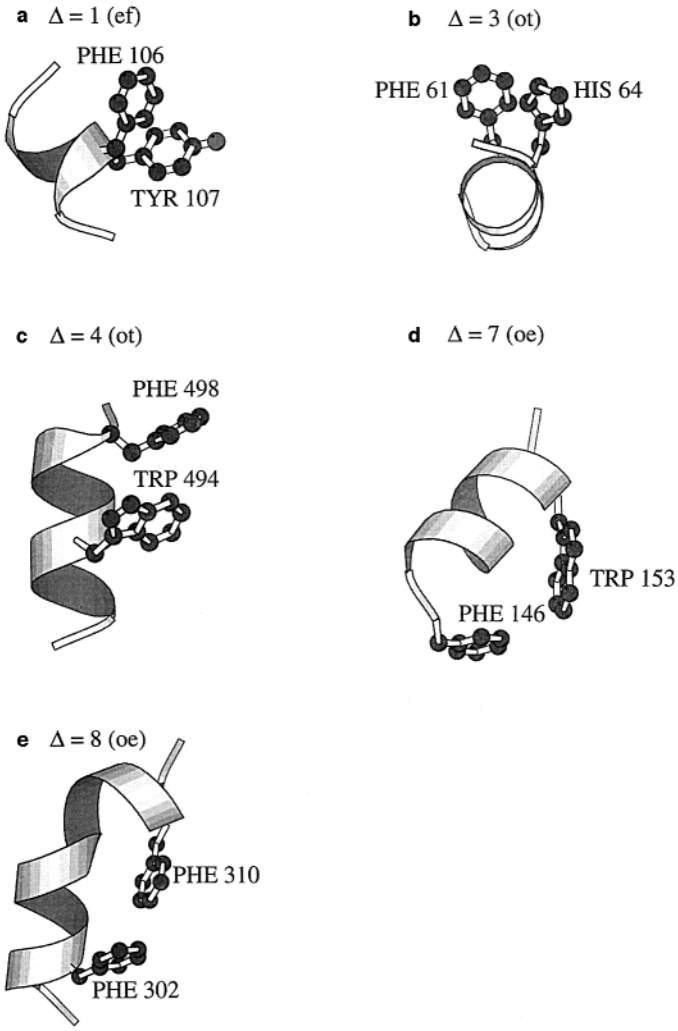

Fig. 4. Examples of interacting aromatic pairs (with different sequence differences, Δ) in α-helices. The PDB files and helix ranges (not all helices are shown in full) are (**a**) 1JDW, 104–110, (**b**) 1KP6, 58–67, (**c**) 2MYR, 490–499, (**d**) 1XNB, 146–153 and (**e**) 1HPM, 299–312. In each case the geometry is given in parentheses.

when His is protonated. Again, the reason why the order in which the two residues occur is important becomes apparent from the consideration of geometry. The two cases of His–Phe (*i*, $i + 4$) pair observed have the geometry *et* (using a different convention, where the geometry of the first ring, i.e. His is expressed relative to the second, i.e. Phe), so that the edge of Phe is pointed towards the face of the His ring in a tilted fashion, which is not particularly stabilizing. The stabilizing nature of the edge of His interacting with the face of an aromatic ring is also exemplified by a tertiary interaction between His18 and Trp 94 in barnase, where the protonation of His increases the stability of the protein by 1 kcal/mol (Loewenthal *et al.*, 1992). The geometry of His with respect to Trp (based on the PDB file, 1B20) is *ft*, as observed for most of the Phe–His pairs (Δ = 4) in helix. When Δ = 3, the centroids of the two rings are further apart and an X–H···π interaction appears to exist only for a few pairs with geometry *ot*.

i, $i+3$ or *i*, $i + 4$ aromatic–His interactions are not dependent on the exact location in the helix (Armstrong *et al.*, 1993). In accordance with this, when Δ = 3, His is observed in the last helical turn in 33% of the cases only, while it is near the helix

Table III. Numbers ($O_{i,j}$) of interacting, helical aromatic residues ($[Ar_i]_1$ and $[Ar_j]_2$, 1 preceding 2 in sequence)[a]

		$[Ar_j]_2$			
		F	Y	H	W
(a) Δ = 1 (147 cases)					
$[Ar_i]_1$	F	23	16	9	12
	Y	18	10	5	7
	H	9	3	**7**	2
	W	13	7	2	4
(b) Δ = 3 (125 cases)					
		17	17	**9**	11
		13	9	2	**12**
		3	2	3	4
		13	8	0	2
(c) Δ = 4 (167 cases)					
		25	19	12	10
		18	14	10	10
		2	6	**10**	3
		10	8	6	4

[a]*Δ* corresponds to the difference in the sequence numbers; the one-letter amino acid code has been used. Cases with an observed value >5 and an observed to expected ratio of >1.2 are in bold.

centre in the remaining cases. However, when Δ = 4, in the majority (75%) of the cases His is located in the C-terminal turn, the last position of the helix being the most favoured position. In the interacting Phe–His pair with Δ = 4, for Phe $\chi_1 \approx 180°$ and $\chi_2 \approx 90°$, whereas for His, $\chi_1 \approx -60°$ (in nine cases) or 180° (three cases) and $\chi_2 \approx +90°$ or –90°. Lacroix *et al.* (1998) have explained the occurrence of His at helix C-termini based on the requirement of a $\chi_1 \approx -60°$, which is not easily attainable when His is in the middle of the helix. However, as mentioned above, His with a $\chi_1 \approx 180°$ are also found favourably positioned relative to a preceding Phe residue and moreover, aromatic residues with χ_1 around –60° are seen inside helices (Figure 5). It is plausible that for a charged His residue a position near the helix terminus offers a more solvent-exposed region in the structure, whereas being in the middle of the helix would, in general, mean a more buried location, not ideal for undergoing protonation. Indeed, assuming residues with a relative accessibility >5% to be exposed, all but one of the 12 cases of His residues in His–Phe (*i*, *i* + 4) pairs are on the surface, whereas the Phe residues are exposed to a lesser extent and five of them are completely buried.

Interaction of a helical aromatic residue with another one beyond the helix–helix capping

Helix capping motifs are specific patterns of hydrogen bonding and hydrophobic interactions found at or near the ends of helices and Aurora and Rose have systematically grouped the commonly observed motifs into seven classes, based on the positions of the two interacting hydrophobic residues across the helix termini (Aurora and Rose, 1998). As we enumerated all the interactions between an aromatic residue within a helix and another one within five residues from a helix end (Table IV), we wanted to see if any of the aromatic–aromatic pairs matched the patterns of hydrophobic–hydrophobic pairs. We could identify only two cases (when Δ = 5) with the already annotated capping motifs. Using the following nomenclature for helices and their flanking residues: ... –N″–N′–Ncap–N1–N2–N3 ... C3–C2–C1–Ccap–C′–C″– ... (where N1 = H_N– and C1 = –H_C in Table IV), the pattern with Δ = 5 at the helix N-terminus is equivalent to the N′→N4 motif of Aurora and Rose (Aurora and Rose, 1998) – here a Phe or Tyr residue one position prior to the Ncap interacts with a similar residue four positions after the Ncap. For about 37% of the 33 cases with Δ = 5 at the helix C-terminus the helix ends at X2 (i.e. C1 position) and these are equivalent to C″→C3 (or Schellman motif), with a Gly at the C′ position. Interestingly, with two aromatic residues interacting, the proportion of Trp is fairly high, especially at the C″ position. For other values of Δ, the motifs identified here are unique and can be broadly described as an aromatic interaction involving a His occupying the capping position and an α-helix capped by a 3_{10}-helix.

Interaction involving a His residue at the Ccap or Ncap position

Like the His occupying the C1 position in an α-helix interacting with a Phe ring four residues preceding it, a His at the Ccap position can interact in an analogous manner (geometry *ot* or *ft*), such that a C–H or N–H group of His is positioned over the face of the other aromatic ring. It can be seen in Table IV(b) that when Δ = 4, His is found in large number at the Ccap position (with a Phe or Trp at C4). Interestingly, even at the other end (Ncap), His is found with the maximum number of occurrences when Δ = 3 [Table IV(a)], and here too, the geometry is such that the edge of His can point towards the face of the partner aromatic ring. Even when Δ = 4, His is conspicuous as the first aromatic residue. Thus, although His is not favoured near the helix N-terminus (Richardson and Richardson, 1988), in association with another aromatic residue following it by three or four residues, His can offer stability to a helical structure.

Residues in α- and 3_{10}-helices

Short pieces of 3_{10}-helices occur fairly frequently in protein structures, especially at the termini of α-helices (more common at the C-terminal end) (Richardson, 1981; Baker and Hubbard, 1984; Barlow and Thornton, 1988; Doig *et al.*, 1997; Pal and Basu, 1999). However, it is not clear what factors are responsible for the tightening up of the last α-helical turn into the 3_{10} conformation. In this connection, it is interesting that aromatic residues (with a sequence difference of 3) from α- and 3_{10}-helices, which are contiguous along the chain, can interact [see Δ = 3 in Tables IV(b) and V] and possibly stabilize the structure. The geometry is very nearly identical (Figure 7a) with what has been observed when both the residues are in the same α-helix. One of the residues involved is usually Tyr (or Trp) with one of the intervening residues being hydrophilic. In 1YGE, there is a short hydrogen bond (2.73 Å) between the side chains of His and Tyr. Interestingly, there is an example (Figure 7b) where there is a kink between the two types of helices caused by a Pro residue and two residues with Δ = 7 interact with each other, as has also been seen within a kinked α-helix (Figure 4e).

Interaction between aromatic residues from two helices separated by a non-helical residue

The existence of an interacting pair of aromatic residues located on two halves of a kinked helix (Figures 4e and 7b) made us look for a situation when such residues are from two different helices separated by a non-helical residue. The results presented in Table VI reveal a few interesting points. When Δ = 7, the geometry is fixed with the edge

of the first residue interacting with the face of the second. The angle between the two helix-axes is also fairly constant, being in the range 102–130°, and the intervening residue does not have any secondary structure. When Δ = 8, however, the residue is in the turn conformation, the geometry is *et* and the angle is in the range 66–86°. The near constancy of the interhelix angle and the geometry of packing of the aromatic residues against each other suggests that such aromatic pairs can act as a prop for maintaining the proper relative orientation of the helices

Table IV. Residues involved and patterns in secondary structure when an aromatic residue in an α-helix interacts with another one within five residues from the helix termini[a]

Δ	No. of cases[b]	Ar_1	X_1	X_2	X_3	X_4	Ar_2	Geometry
(a) Ar_1 before the helix N-terminus, Ar_2 in helix								
3	27 (33)	**H**-10, F-8, W-6 [C(12), T(5)]	Ch-8, Po-6, P-4 [**H_N**–]	Ch-7, Po-6, Ho-6 [**H**]			F-13, Y-6 [**H**]	*ef*(8), *oe*(6)
4	31	**H**-9, F-9, Y-8 [C(16), S(7)]	Ho-10, Ch-6, Po-6 [H(16), C(10)]	Ch-11, Po-5, P-5 [H(24), T(6)]	Ch-12, Po-7 [H(26), T(5)]		Y-10, F-9, H-7 [H]	*ef*(8), *oe*(7)
5	22 (39)	F-10, Y-9 [C(14), S(4)]	Po-12, Ch-9 [C(14)]	Ch-8, Ar-6 [**H_N**-]	Ch-9, Ho-6 [**H**]	Ch-14 [**H**]	F-10, Y-7 [**H**]	*et*(7), *oe*(5)
(b) Ar_1 in helix, Ar_2 after helix C-terminus								
3	20	F-9, Y-6, H-5 [H]	Ch-8, G-4 [H(10), G(4)]	Po-6, Ch-5 [H(6), T(6), G(4)]			**Y-**10, W-4, F-4 [T(8), **G**(6)]	*oe*(6)
4	44 (73)	F-21, Y-10, W-9 [**H**]	Ho-18, Po-11, Ch-9 [**H**]	Ch-20, Po-10 [**H**]	Ho-14, Ch-12, Po-11 [**–H_C**]		Y-17, F-14, **H-**13 [T(28), C(13)]	*ot*(11), *of*(9), *ft*(8)
5	33	F-12, Y-11, W-6 [H]	Ch-13, Ho-9 [H(31)]	Po-14, Ch-9 [H(26)]	Ch-16, Ho-7 [H(15), T(14)]	**G-**12, Ch-10 [T(18), H(8)]	Y-10, **W**-9, F-9 [T(9), H(5)]	*oe*(8),*ot*(3), *of*(3),*et*(3)

[a]Δ = sequence difference between the two interacting aromatic residues and X_1, X_2, etc. are the intervening residues. Prominent residues occupying these positions (with the number of occupancies indicated after a hypen) are given in one-letter amino acid code or two-letter class code. Classes are Ar (aromatic = F, Y, H and W); Ch (charged = D, E, K and R); Po (polar = S, T, N and Q); and Ho (hydrophobic = A,C, M, V, I and L). The secondary structure at each position is indicated in square brackets [H, α-helix; H_N– and –H_C are N/C-terminal positions of α-helix; G, 3_{10}-helix; T, hydrogen-bonded turn; C, non-regular structure], with number of occurrences of each element given in parentheses. Significant residues and secondary structures observed for a given *Δ* are indicated in bold. The geometry of interaction is defined in Figure 3.

[b]In cases where two numbers are given, the one in parentheses is the total number of cases for a particular *Δ*, of which the cases corresponding to the other number have a definite pattern of secondary structure and the statistics given for the rest of the entry correspond to this.

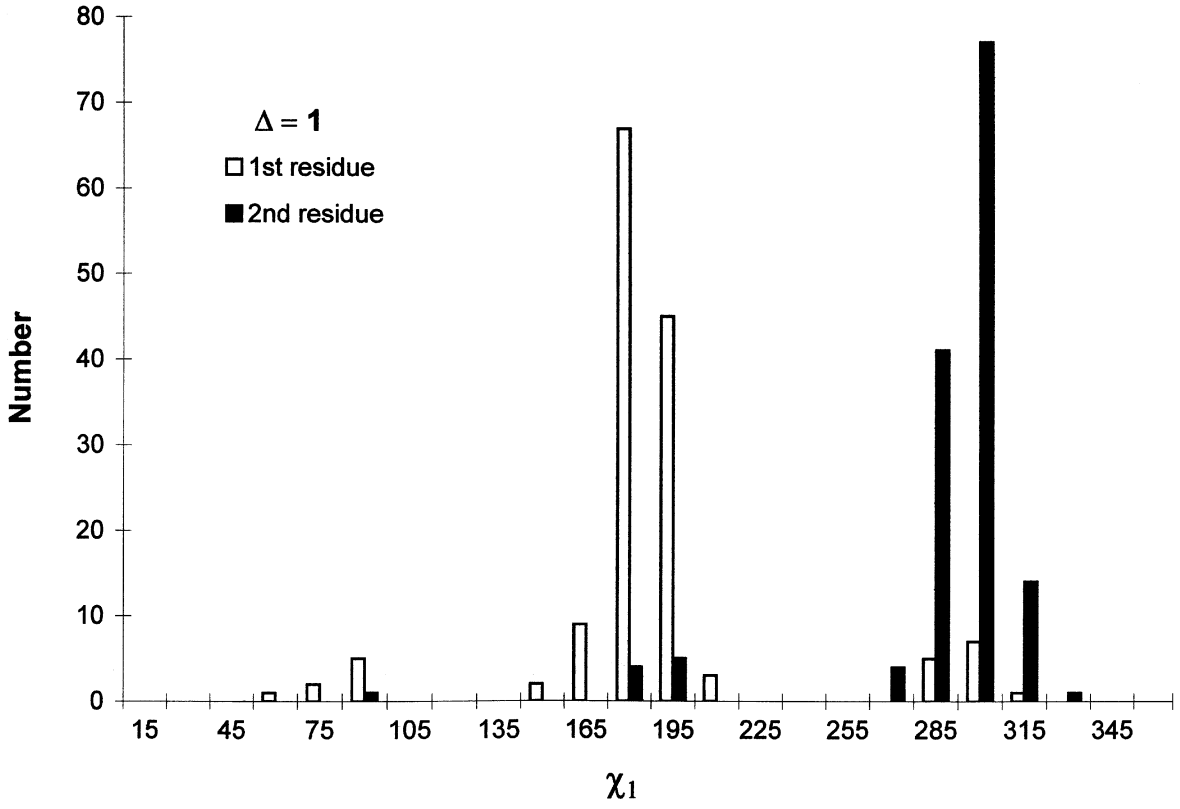

Fig. 5. continued.

and their interaction energy may be enough to force a residue out of what would otherwise have become a single helix into a non-helical conformation. Thus, the specific geometry of aromatic–aromatic interaction can confer specificity (uniqueness) to the packing of secondary structural elements.

Interaction between aromatic residues from two sides of a helix

The possibility of aromatic residues located within five residues of the two helical ends was also considered. As expected, of the 27 cases the majority involve short helices. There are four cases with $\Delta = 10$ and the central helix of length 4, with the two

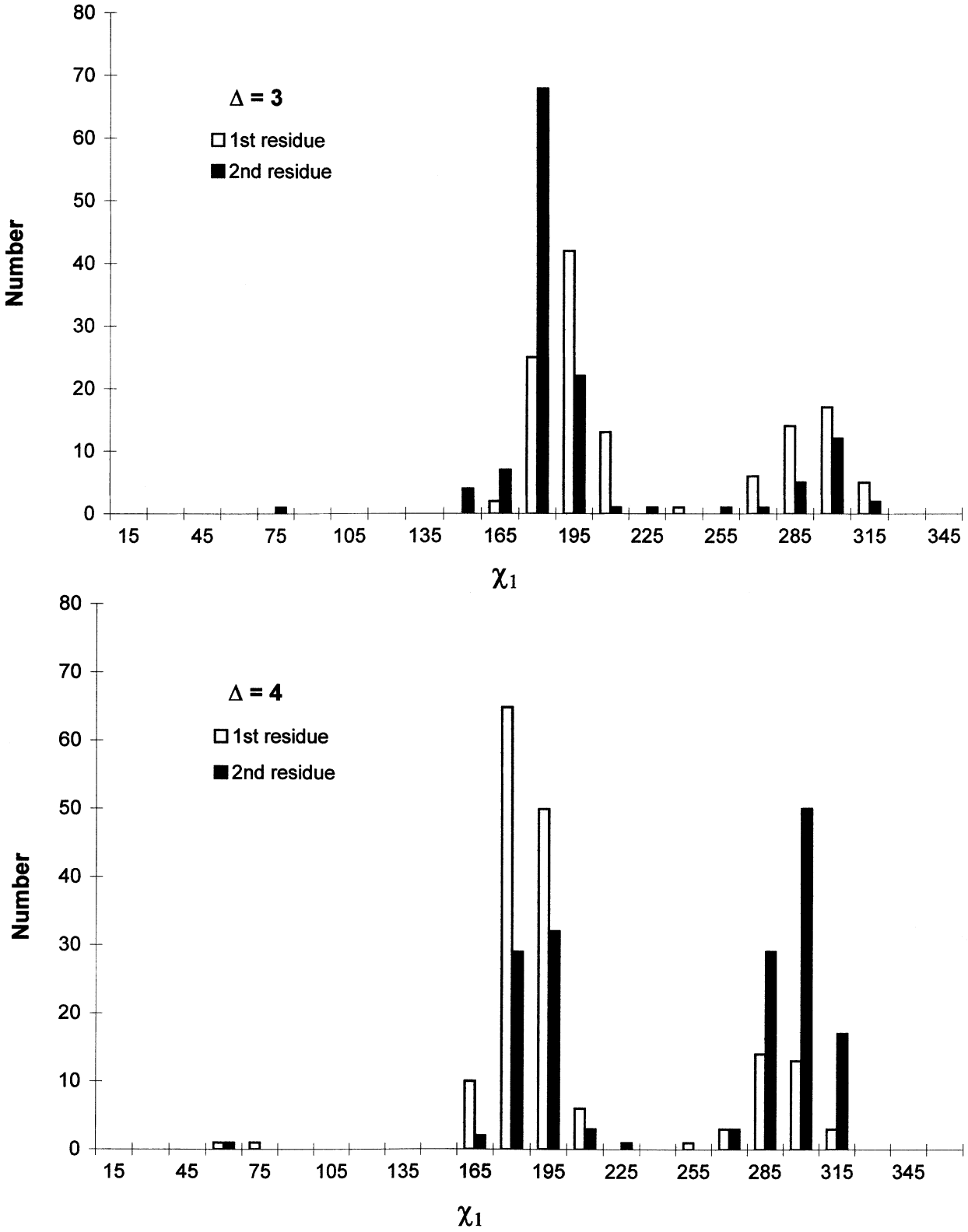

Fig. 5. Histogram showing the distribution of χ_1 angles (°) of the interacting pairs of aromatic residues (with different sequence differences) in α-helices.

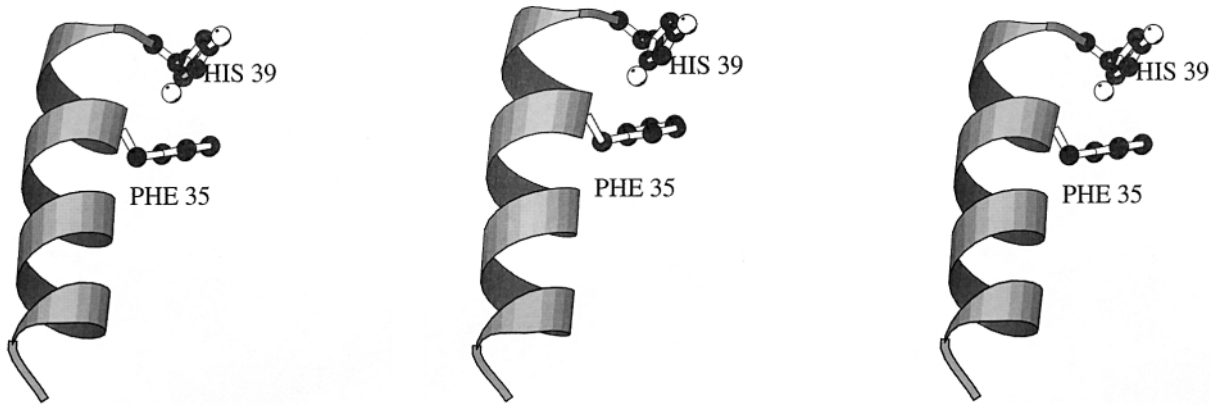

Fig. 6. Stereo plot showing the interaction between Phe and His residues with Δ = 4 (and geometry *ft*) in a helix (residues 25–39 in the structure 1MTY; the two NH hydrogen atoms in His are shown).

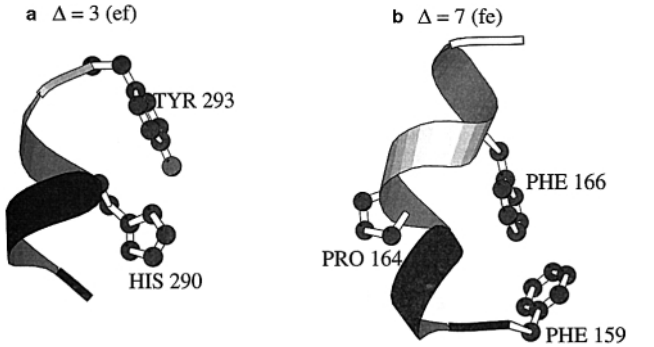

Fig. 7. Examples of interacting aromatic pairs with the first residue in α-helix (darker in shade) and the second in a contiguous 3_{10}-helix. The sequence difference, Δ, and the interaction geometry are indicated. The PDB files and (helix ranges) are (**a**) 1YGE (286–290, 291–293) and (**b**) 1BAM (159–162, 163–169).

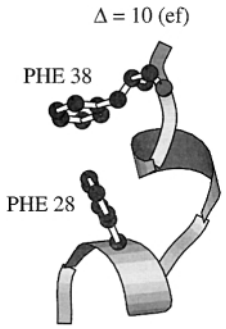

Fig. 8. A pair of interacting aromatic residues from two sides of an α-helix (residues 32–35 in the structure, 1ECD), with Δ and geometry indicated.

aromatic residues being three and four residues away from the two ends of the helix (Figure 8) – in all the cases the geometry is *ef* and one of the residues is in another helix, while the other residue has a non-regular conformation. When residues from the two ends interact they are mostly (63%) Phe.

Implications and summary

In this paper we have looked into the geometry of interaction between aromatic residues located in and around α-helices. The geometry is found to depend on the sequence difference (Δ) between the two residues (Table II). When Δ = 1, the preferred geometry puts the edge of the first ring against the face of the second, whereas when Δ = 4 (which gives the closest contact), the second ring usually interacts in a tilted

Table V. Sequence of the polypeptide chain containing interacting aromatic residues belonging to α- and 3_{10}-helices[a]

PDB	Sequence	Geometry
1CLE	**HLI**W	*ot*
1FCE	**Y**LRY	*et*
1GPE	**F**GDY	*ot*
1YGE	**H**DLY	*ef*
7A3H	**Y**GQF	*et*
1A7S	F**RDW**	*ef*
1BAM	**FEEL**EPYF	*fe*

[a]One-letter amino acid code is used. The two interacting residues are at the two ends of the sequence. α-Helical residues are in bold. Except for 1A76, in other cases the α-helix leads into a 3_{10}-helix. The sequence difference is 7 in the last case and 3 in the rest.

fashion, offset from the centre of the first. Based on the position of the two residues in the helix, preferences for the side-chain χ_1 angles are fairly strong, especially when Δ = 1 (Figure 5). If a given relative location offers the chance of a specific interaction taking place, it is observed more frequently. For example, a Phe–His pair (Δ = 4) with the His residue occupying the C1 or Ccap position in the helix is found more frequently than expected because an N–H or C–H group from His can interact with the p electron cloud of the other ring—a geometry which is not attainable if the order of the two aromatic residues is reversed. The stability of aromatic–His pair was known from solution studies (Armstrong *et al.*, 1993; Fernandez-Recio *et al.*, 1997); however, only an analysis of the geometry could reveal the origin of the stability. Although the residues involved can be broadly categorized as hydrophobic, the interaction (NH–CH···π) between them is more like a hydrogen bond.

Aurora and Rose named capping motifs based on the closest pair of interacting hydrophobic residues that straddles the helix terminus (Aurora and Rose, 1998). We identified some new capping motifs (Table IV) where a His is at the capping position. Although His is not found near the helix N-terminus (Richardson and Richardson, 1988), this bias can be overcome and His can occupy the Ncap position in association with another aromatic residue following it by three or four positions (Table IV). Hence there is scope for improvement of secondary structure prediction if information on pairwise structural correlations in α-helices, as observed here and also by others (Klingler and Brutlag, 1994; Walther and Argos, 1996), is used in conjunction with traditional secondary structural propensity

Table VI. Interacting aromatic residues (with Δ = 7 or 8) from an α-helix or two helices with a single intervening non-helical residue

Δ	PDB code	Helix range[a]	Interacting aromatic residues[b]	Geometry[c]	Angle between helix axes (°)[d]
(a) With no intervening non-helical residue					
7	1XNB	146–155	F146, W153	*oe*	180
8	1HPM	299–312	F302, F310	*oe*	111
	1QHF	97–114	F104, Y112	*fe*	56[e]
	2LIS	44–74	Y57, Y65	*et*	129
(b) With one intervening non-helical residue					
7	1PHF	253–276 (S267,C)	F263, H270	*ef*	103
	1U9A	131–154 (N140,C)	Y137, Y144	*ef*	130
	1YAC	180–204 (I191,C)	F187, Y194	*ef*	124
	1YCC	61–74 (N70,C)	Y67, Y74	*ef*	114
	1QH4	81–96 (F85,T)	Y82, F89	*ot*	102
8	1CTJ	4–19 (C15,T)	F11, H19	*ot*	73
	1QKS	51–69 (C65,T)	Y61, H69	*et*	66
	2EBN	64–76 (N72,T)	H68, Y76	*et*	86

[a]When there are two helices the range encompasses both the helices and the non-helical residue, along with its secondary structure (T = turn, C = non-regular structure), is specified in parentheses.
[b]One-letter amino acid code, followed by the residue number, is used.
[c]As shown in Figure 3.
[d]Calculated as discussed by Chakrabarti and Chakrabarti (Chakrabarti and Chakrabarti, 1998).
[e]It may be better to consider this as made up of two different helices.

values, which are derived assuming that amino acids have no effect on each other. Additionally, interactions in helices are assumed to be local confined to consecutive turns of the helix (Aurora *et al.*, 1997). Here we identified a few cases of aromatic residues which interact even though they are separated by a helical turn (Table VI, Figures 4e and 7b). Specificity of folding is generally ascribed to the presence of buried polar interactions (Lumb and Kim, 1995). However, the near constancy of the geometry of the His–Phe interaction and the order in which the residues are to be placed within a helix for stability (Figure 6) and the relative orientation between two helices when they have a pair of interacting aromatic residues (Table VI) suggest that the specificity can also be controlled by aromatic–aromatic interaction. Interacting aromatic pairs can be used for the design and modelling of proteins. Finally, the functional role of His is well known, this study also points to a structural role for the residue.

Acknowledgements

This work was supported by the Department of Biotechnology and the Council of Scientific and Industrial Research.

References

Armstrong,K.M., Fairman,R. and Baldwin,R.L. (1993) *J. Mol. Biol*., **230**, 284–291.
Aurora,R. and Rose,G.D. (1998) *Protein Sci*., **7**, 21–38.
Aurora,R., Creamer,T.P., Srinivasan,R. and Rose,G.D. (1997) *J. Biol. Chem*., **272**, 1413–1416.
Baker,E.N. and Hubbard,R.E. (1984) *Prog. Biophys. Mol. Biol*., **44**, 97–179.
Barlow,D.J. and Thornton,J.M. (1988) *J. Mol. Biol*., **201**, 601–619.
Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) *Nucleic Acids Res.*, **28**, 235–242.
Brocchieri,L. and Karlin,S. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 9297–9301.
Burley,S.K. and Petsko,G.A. (1985) *Science*, **229**, 23–28.
Burley,S.K. and Petsko,G.A. (1988) *Adv. Protein Chem*., **39**, 125–189.
Chakrabarti,P. and Chakrabarti,S. (1998*) J. Mol. Biol*., **284**, 867–873.
Chakrabarti,P. and Pal,D. (2001) *Prog. Biophys. Mol. Biol*., **76**, 1–102.
Chakrabartty,A. and Baldwin,R.L. (1995) *Adv. Protein Chem.*, **46**, 141–176.
Dadlez,M., Bierzynski,A., Godzik,A., Sobocinska,M. and Kupryszewski,G. (1988) *Biophys. Chem*., **31**, 175–181.
Doig,A.J., MacArthur,M.W., Stapley,B.J. and Thornton,J.M. (1997) *Protein Sci*., **6**, 147–155.
Fernandez-Recio,J., Vazquez,A., Civera,C., Sevilla,P. and Sancho,J. (1997) *J. Mol. Biol*., **267**, 184–197.
Hobohm,U. and Sander,C. (1994) *Protein Sci*., **3**, 522–524.
Hubbard,S. (1992) *ACCESS: a Program for Calculating Accessibilities.* Department of Biochemistry and Molecular Biology, University College London, London.
Hunter,C.A., Singh,J. and Thornton,J.M. (1991) *J. Mol. Biol*., **218**, 837–846.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Kannan,N. and Vishveshwara,S. (2000) *Protein Eng.*, **13**, 753–761.
Klingler,T.M. and Brutlag,D.L. (1994) *Protein Sci*., **3**, 1847–1857.
Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
Lacroix,E., Viguera,A.R. and Serrano,L. (1998) *J. Mol. Biol.*, **284**, 173–191.
Lee,B. and Richards,F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.
Loewenthal,R., Sancho,J. and Fersht,A.R. (1992) *J. Mol. Biol.*, **224**, 759–770.
Lumb,K.J. and Kim,P.S. (1995) *Biochemistry*, **34**, 8642–8648.
McGregor,M.J., Islam,S.A. and Sternberg,M.J.E., (1987) *J. Mol. Biol*., **198**, 295–310.
Mitchell,J.B.O., Nandi,C.L., McDonald,I.K. and Thornton,J.M. (1994) *J. Mol. Biol.*, **239**, 315–331.
Mitchell,J.B.O., Laskowski,R.A. and Thornton,J.M. (1997) *Proteins: Struct. Funct. Genet.*, **29**, 370–380.
Nishio,M., Hirota,M. and Umezawa,Y. (1998) *The CH/π Interaction. Evidence, Nature and Consequences*. Wiley-VCH, New York.
Pal,D. and Chakrabarti,P. (1998) *J. Biomol. Struct. Dyn*., **15**, 1059–1072.
Pal,D. and Chakrabarti,P. (1999) *J. Mol. Biol*., **294**, 271–288.
Pal,L. and Basu,G. (1999) *Protein Eng*., **12**, 811–814.
Richardson,J.S. (1981) *Adv. Protein Chem.*, **34**, 167–339.
Richardson,J.S. and Richardson,D.C. (1988) *Science*, **240**, 1648–1652.
Samanta,U., Chakrabarti,P. and Chandrasekhar,J. (1998) *J. Phys. Chem. A*, **102**, 8964–8969.
Samanta,U., Pal,D. and Chakrabarti,P. (1999) *Acta Crystallogr.*, **D55**, 1421–1427.
Samanta,U., Pal,D. and Chakrabarti,P. (2000) *Proteins: Struct. Funct. Genet.*, **38**, 288–300.
Schildbach,J.F., Karzai,A.W., Raumann,B.E. and Sauer,R.T. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 811–817.
Serrano,L. (2000) *Adv. Protein Chem*., **53**, 49–85.
Shoemaker,K.R., Fairman,R., Schultz,D.A., Robertson,A.D., York,E.J., Stewart,J.M. and Baldwin,R.L. (1990) *Biopolymers*, **29**, 1–11.
Singh,J. and Thornton,J.M. (1985) *FEBS Lett.*, **191**, 1–6.
Singh,J. and Thornton,J.M. (1992) *Atlas of Protein Side-Chain Interactions*, Vols I and II. IRL Press, Oxford.
Walther,D. and Argos,P. (1996) *Protein Eng*., **9**, 471–478.

Received March 5, 2001; revised October 15, 2001; accepted November 11, 2001

Appendix

The codes for the PDB files used are as follows (the subunit identifier, if present, is given as the fifth letter):
1A1IA, 1A1YI, 1A28B, 1A2PA, 1A2ZA, 1A34A, 1A3C_, 1A48_, 1A4IB, 1A6M_, 1A7S_, 1A8D_, 1A8E_, 1A9XB, 1ABA_, 1ADOA, 1ADS_, 1AE9B, 1AFWA, 1AGQA, 1AHO_, 1AIE_, 1ALVA, 1AMF_, 1AMM_, 1AMX_, 1AOCA, 1AOHB, 1APYA, 1AQB_, 1ARV_, 1ATLA, 1AUN_, 1AVWB, 1AXN_, 1AY7B, 1AYFA, 1AYL_, 1AYOA, 1AZO_, 1B0NA, 1B0NB, 1B0UA, 1B0YA, 1B16A, 1B2VA, 1B3AA, 1B4KB, 1B5EA, 1B65A, 1B67A, 1B6A_, 1B6G_, 1B7CA, 1B8OA, 1B93A, 1BA8A, 1BABB, 1BAM_, 1BBHA, 1BBPA, 1BDO_, 1BE9A, 1BEA_, 1BEC_, 1BENB, 1BF6A, 1BFG_, 1BFTA, 1BG6_, 1BGF_, 1BI5A, 1BJ7_, 1BK0_, 1BK7A, 1BKRA, 1BQCA, 1BRT_, 1BS4A, 1BS9_, 1BSMA, 1BTN_, 1BU7A, 1BX4A, 1BX7_, 1BXAA, 1BXOA, 1BY2_, 1BYI_, 1BYQA, 1BYRA, 1C24A, 1C2AA, 1C3D_, 1C3MA, 1C3WA, 1C52_, 1CBN_, 1CC8A, 1CCZA, 1CEQA, 1CEWI, 1CEX_, 1CF9A, 1CFB_, 1CG6A, 1CKAA, 1CLEA, 1CMBA, 1CNV_, 1COZA, 1CPO_, 1CPQ_, 1CQYA, 1CS1A, 1CTJ_, 1CTQA, 1CV8_, 1CVL_, 1CXQA, 1CXYA, 1CY5A, 1CYDA, 1CYO_, 1CZFA, 1CZPA, 1D3VA, 1D7PM, 1D9CB, 1DBWB, 1DCIA, 1DCS_, 1DF4A, 1DFNA, 1DG9A, 1DGWY, 1DHN_, 1DI6A, 1DIN_, 1DLFH, 1DLFL, 1DOKA, 1DOSA, 1DOZA, 1DPSD, 1DPTA, 1DUN_, 1DXGA, 1ECD_, 1ECPA, 1EDG_, 1EDMB, 1EGPA, 1EUS_, 1EXTB, 1EZM_, 1FCE_, 1FIPA, 1FIT_, 1FLEI, 1FLTV, 1FLTY, 1FNA_, 1FRPA, 1FUS_, 1FVKA, 1G3P_, 1GCI_, 1GDOB, 1GOF_, 1GP1A, 1GPEA, 1GSA_, 1GUQA, 1HFC_, 1HFES, 1HKA_, 1HLEB, 1HOE_, 1HPM_, 1HTRP, 1HUUA, 1HXN_, 1IAB_, 1ICFI, 1IDAA, 1IFC_, 1IIBA, 1ISUA, 1IXH_, 1JDW_, 1JER_, 1JHGA, 1KNB_, 1KOE_, 1KP6A, 1KPTA, 1KVEA, 1KVEB, 1LAM_, 1LATA, 1LBU_, 1LCL_, 1LKFA, 1LKKA, 1LOUA, 1LTSA, 1LTSC, 1LUCA, 1MAI_, 1MDC_, 1MFMA, 1MGTA, 1MKAA, 1MLA_, 1MML_, 1MOF_, 1MOLA, 1MOQ_, 1MPGA, 1MRJ_, 1MROA, 1MROB, 1MROC, 1MSI_, 1MSK_, 1MTYB, 1MTYG, 1MUGA, 1MUN_, 1NAR_, 1NBCA, 1NCOA, 1NIF_, 1NKD_, 1NKR_, 1NLS_, 1NOX_, 1NP4A, 1NPK_, 1NULB, 1OAA_, 1OBWA, 1OPD_, 1OPY_, 1ORC_, 1OTFA, 1PBE_, 1PCFA, 1PDO_, 1PGS_, 1PHF_, 1PLC_, 1PNE_, 1POA_, 1POC_, 1PPN_, 1PSRA, 1PTQ_, 1PTY_, 1PYMB, 1QB7A, 1QCXA, 1QCZA, 1QD1A, 1QDDA, 1QFMA, 1QFOA, 1QGIA, 1QGWB, 1QGWD, 1QH4A, 1QH5A, 1QH8A, 1QH8B, 1QHFA, 1QJ4A, 1QJ8A, 1QKSA, 1QMPD, 1QQ4A, 1QQ5A, 1QQP1, 1QQP2, 1QQP4, 1QREA, 1QRRA, 1QSGA, 1QTSA, 1QTWA, 1QU9A, 1RB9_, 1RCF_, 1REC_, 1REGY, 1RGEA, 1RHS_, 1RIE_, 1RZL_, 1SCJB, 1SFP_, 1SGPI, 1SLUA, 1SMD_, 1SMLA, 1SRA_, 1SUR_, 1SVFA, 1SVFB, 1SVPA, 1SVY_, 1SWUB, 1TAFA, 1TAXA, 1TC1A, 1TEN_, 1TGXA, 1TIB_, 1TIF_, 1TL2A, 1TML_, 1TOAA, 1TTBA, 1TVXB, 1U9AA, 1UBPA, 1UBPB, 1UNKA, 1UOX_, 1VCAA, 1VFRA, 1VFYA, 1VHH_, 1VID_, 1VIE_, 1VLS_, 1VNS_, 1VSRA, 1WAB_, 1WAPB, 1WDCA, 1WHI_, 1WHO_, 1WWCA, 1XNB_, 1YACA, 1YAGG, 1YCC_, 1YGE_, 1YTBA, 2A0B_, 2ABK_, 2ACY_, 2AHJC, 2ARCB, 2AYH_, 2BC2A, 2BOPA, 2BOSA, 2CBP_, 2CCYA, 2CHSA, 2CPGA, 2CTC_, 2DRI_, 2DTR_, 2EBN_, 2EBOA, 2END_, 2ERL_, 2FDN_, 2GAR_, 2GDM_, 2HBG_, 2HDDB, 2HFT_, 2HMZA, 2IGD_, 2ILK_, 2KNT_, 2LISA, 2MSBB, 2MYR_, 2NLRA, 2PII_, 2PSPA, 2PTH_, 2PVBA, 2QWC_, 2RN2_, 2SAK_, 2SICI, 2SN3_, 2SNS_, 2SPCA, 2TNFA, 2TPSA, 2TRXA, 2TYSB, 2UBPC, 3CHBD, 3CHY_, 3CLA_, 3CYR_, 3ENG_, 3EZMA, 3GRS_, 3LZT_, 3PTE_, 3PVIA, 3PYP_, 3SDHA, 3SEB_, 3SIL_, 3STDA, 3TDT_, 3TSS_, 3VUB_, 4EUGA, 4MT2_, 5HPGA, 5PTI_, 6CEL_, 6GSVA, 7A3HA, 7RSA_, 8ABP_, 8PRKA, 9WGAA, 16PK_, 19HCA, 153L_, 256BA, 451C_.