Aromatic Side-Chain Interactions in Proteins. I. Main Structural Features

Annick Thomas,[1*] Rita Meurisse,[2] Benoit Charloteaux,[2] and Robert Brasseur[2]
[1]*Institut National de la Santé et de la Recherche Médicale (INSERM) Unité 410, Paris Cedex, France*
[2]*Centre de Biophysique Moléculaire Numérique (CBMN), Faculté Scientifique Agronomique de Gembloux (FSAGx), Gembloux, Belgium*

ABSTRACT **In a data set of 593 nonhomologous proteins from the PDB, we have analyzed the pairing of phenylalanine, tyrosine, tryptophan, and histidine residues with their closest aromatic partner. The frequency distribution of the shortest interatomic distance of partners is bimodal with a sharp peak at ~3.8 Å and a wider one at a longer distance. Only the 3.8 Å peak corresponds to direct ring–ring interactions thus aromatic pairs. The aromatic pairs were separated into two classes, near-sequence pairs and far-sequence pairs. Near sequence pairs stabilize local structure, and far-sequence pairs stabilize tertiary structure. Far-sequence pairs (74% of all pairs) mainly bridge two β-strands, followed by pairs that bridge a β-strand and a helix, and pairs that bridge a β-strand and a random coil structure. Pairs that bridge helices are rare. The secondary structure of the near-sequence pairs depends on the partner distance in the sequence. When the partners are 1, 3, or 4 residues apart in the sequence, pairs are mostly found in helical structures. When the partners are two apart, pairs are mostly found in the same β-strand. Analysis of the frequency of near sequence pairs supports the hypothesis that aromatic pairing occurs after, rather than before, the formation of secondary structures. Proteins 2002;48: 628–634.** © **2002 Wiley-Liss, Inc.**

Key words: aromatic amino acids; histidine; phenylalanine; tryptophan; tyrosine; aromatic pairs; secondary structures

INTRODUCTION

Aromatic residues are involved in the stabilization of protein structures, but the mechanism of their interactions is still unclear. Early analyses of aromatic pairs in proteins considered either a single residue, i.e., Trp,[1,2] three aromatic residues[3,4] or even the four homologous pairs Phe-Phe, Tyr-Tyr, Trp-Trp, and His-His.[5]

Because of the conjugated π electrons, aromatic rings should be capable of electromagnetic π-π stacking as in the on-line stacking of rings in benzene crystals. However, for benzene crystals, the cost in entropy was reported to be more important than the gain in enthalpy, suggesting that such structures are not the best solution for proteins.[8] However, this has been debated in the literature.[5] Nevertheless, in proteins, most pairs of aromatic rings adopt L- or T-shaped conformations,[3,9] which are more likely to involve electrostatic CH/π[8,10] interactions than electromagnetic π-π interactions. The CH/π interactions are low-energy H-bonds as are the NH/π interactions of the main-chain and side-chain NH with aromatic rings[4,11] and the aromatic NH/π and OH/π interactions,[10] in which histidine, tryptophan, and tyrosine could be involved all as H-donors. A growing body of evidence suggests that, besides the main-chain C═O/NH bonds, low-energy H-bonds are important in protein structures.[13–16] On the other hand, because the imidazolium group of histidine can be positively charged and the phenolic group of tyrosine and the indole group of tryptophan are electron-rich, interactions of aromatic side chains with charged residues should also be possible. In summary, the aromatic rings are capable of interacting in many different ways.

Given the diversity of aromatic interactions and their importance in protein structure and function, we have undertaken a statistical analysis of aromatic pairs in 593 proteins from the PDB. The tools recently developed for PDB analysis[6] enabled us to carry on an exhaustive analysis, and we have attempted to classify both the general and the specific properties of aromatic pairs. All aromatic residues were considered, and the shortest atomic distance between two side chains was used to identify the pairs. Those pairs were further characterized by the residue proximity in the sequence and by the secondary structure. We defined two major classes of pairs that correspond to different properties and structural function: the near-sequence and the far-sequence interactions of aromatic residues. This article should be followed by others that get deeper into the properties of each pair.

MATERIALS AND METHODS

Pex

We used Pex, an automated programme that extracts an extensive set of protein parameters from PDB structures The PDB files were transformed into GF-Pex files,[6] which

Grant sponsor: Interuniversity Poles of Attraction Programme–Belgian State, Prime Minister's Office, Federal Office for Scientific, Technical, and Cultural Affairs; Grant number: PAI contract no. P4/03. Grant sponsor: Ministére de la Région Wallonne; Grant number: contract nos. 14540 and 991/4024 (Bioval).

*Correspondence to: A Thomas: CBMN, 2, passage des déportés, FSAGx, 5030 Gembloux. E-mail: thomas.a@fsagx.ac.be

Received 21 March 2002; Accepted 15 April 2002

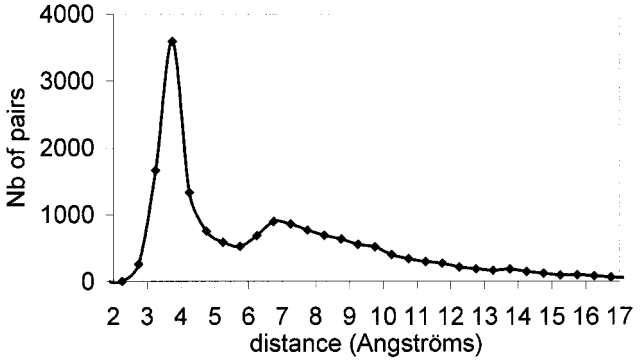

Fig. 1. Distribution of the minimal atomic distances between two aromatic residues of 593 proteins of the PDB. As explained in methods, for each aromatic residue, we looked on the C-side for the closest aromatic residue in the 3D structures with no limit of distance. Hence, each residue has a partner, but there is no pair redundancy. The distance between partners (in Angströms) was calculated as the shortest atomic interaction considering the C, N, and O atoms of side chains (CG and up); H were excluded because they are absent in many PDB.

tabulate the parameters in a standard spreadsheet format and can be downloaded from the Website of the Centre de Biophysique Moléculaire Numérique (http://www.fsagx.ac.be/bp/). We have used a dataset of 593 nonhomologous protein structures from the Brookhaven Protein Data Bank.[7]

Pair Definition

In Pex, pairs of aromatic residues are defined on a spatial proximity basis. The first residue, which we call the reference residue (Phe, His, Trp, and Tyr) is the first aromatic amino acid in the sequence (read from the N- to the C-end) and the partner (X) is defined as the closest aromatic residue of the C-side sequence in the 3D structure. As there is no limit of distance, all aromatic residues have a partner except the last aromatic residue of the sequence. Partners make a pair when the minimal distance between two non-H atoms (i.e., C, N, O) of their side chains is <5.5 Å (Fig.1). Contacts with the other 16 amino acids or with a cofactor, a substrate, or water molecules were not considered.

Duet Definition

The definition of aromatic pairs depends on the 3D structure. In contrast, the definition of aromatic duets refer to a distance in the sequence only. Aromatic duets are defined by the number of amino acids between them. Consequently, all pairs are duets, but not all duets are pairs.

Pair Structures

For simplicity, we have grouped all helices (alpha, 3,[10] and omega …) together under the symbol H, all beta structures (β-strands, antiparallel and parallel sheets) under B and, all turns under T. Otherwise, residues were said to be in random coil structures (C). In Pex, secondary structures are defined by ϕ/ψ values and the occurrence of main-chain H-bonds (O—HN distance < 3.5 Å). Residues in helices have ϕ/ψ values of $-57 \pm 45°$ and $-47 \pm 45°$, respectively, and are involved in a main-chain H-bond with a n-3 to n-6 neighbor in the sequence. Residues in B have ϕ/ψ values of $-129 \pm 90°$ and $123 \pm 90°$, respectively. We have used the definitions of the nine types of turns as defined by Srinivasan et al.[17]

RESULTS

In the dataset, there are 154,425 residues, and the four aromatic amino acids (histidine, phenylalanine, tryptophan, and tyrosine) account for 17,597 residues (11%). Phenylalanine and tyrosine residues account for 35% and 32% of the aromatic residues, respectively, whereas histidine and tryptophan account for only 19% and 14%, respectively. These ratios are similar to those obtained with a smaller database of 131 proteins.[6]

Aromatic Pairs

For each aromatic residue, we looked for the closest partner in the 3D structures defined by the shortest side-chain–side-chain interatomic distance. The frequency distribution of this distance for the 17,597 aromatic residues is bimodal, with a sharp peak at 3.8 Å and a wider one at ~6.5–8 Å (Fig. 1). The distribution was split at 5.5 Å and the interactions studied below have a minimal atomic distance <5.5 Å. This minimal distance does not allow a water molecule in between the two closest atoms. We call aromatic pairs two such residues with interatomic distance <5.5 Å.

In the initial Pex protocol, there was no pair redundancy: for instance, if the Phe30-Tyr80 pair existed, then Tyr80-Phe30 did not. However, one residue could be involved in several pairs, for instance, Tyr 80 could be in the Phe30-Tyr80, in Trp40-Tyr80, and in Tyr80-His90 pairs. Other residues could exist in only one pair; for instance, the histidine of the Tyr80-His90 pair could have no C-side partner at a distance <5.5Å. Using these restrictions, we collected 8200 unique pairs.

In order to quantify the number of aromatic residues involved in pairs, we also made a separate analysis in which pair redundancy was allowed. In this analysis, for each aromatic residue, we collected the closest partner in the 3D structure, regardless of whether this partner was on the N- or C-side in the sequence. This measures the number of residues found in pairs: 12,192 of 17,597 (69%) aromatic residues are in pairs accounting for 73, 68, 58, and 76% of all Phe, Tyr, His, and Trp residues, respectively.

Characterization of Aromatic Pairs

In most pairs, the aromatic rings are perpendicular to each other (Fig. 2). In a 0–180° scale, the median value for ring angles is 90°, but the large standard deviation (±42°) corresponds to a widely spread distribution in agreement with previous publications.[3,9]

Homologous pairs, Phe-Phe, Tyr-Tyr, Trp-Trp, and His-His, account for only 32% of all pairs. In Tables I and II, symmetric pairs such as Phe-Tyr and Tyr-Phe were gathered under the acronym Phe⟨⟩Tyr because we have found

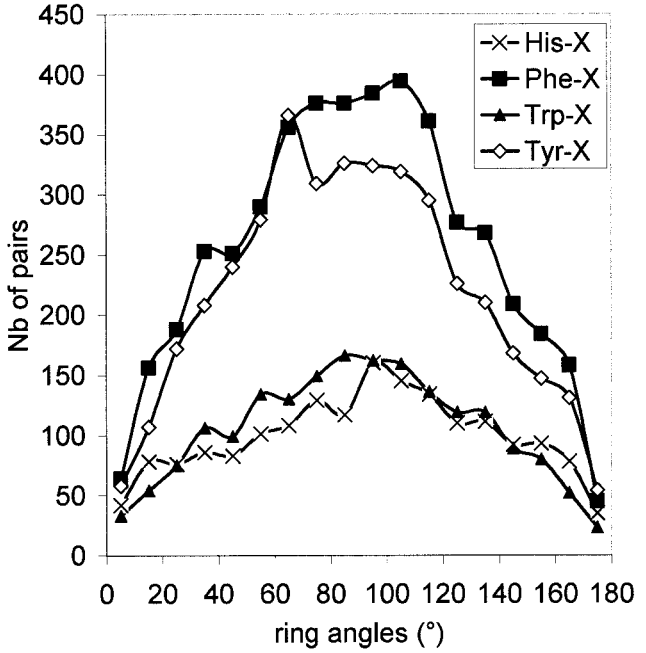

Fig. 2. Distribution of ring angle values in all aromatic pairs of the 593 protein structures. The pairs are clustered according to the name of the N-side partner, X; the C-side partner is either Phe, Tyr, His, or Trp.

that far-sequence pairs are symmetrical. The most frequent pairs are Phe⟨⟩Tyr (22%) (Table I). Homologous pairs of phenylalanine (Phe-Phe) come next (16%), and they account for as much as 44% of all phenylalanine residues involved in pairs and thus are the main opportunity of that residue. Phenylalanine is the only aromatic residue that prefers an homologous to an heterologous interaction. Tyrosine is more frequently involved in Phe⟨⟩Tyr pairs, tryptophan in Phe⟨⟩Trp pairs, and histidine in His⟨⟩Tyr pairs.

If we further discriminate the pairs between different secondary structures, the same pattern of distribution for pairs is observed. The most common secondary structure of pairs is BB, followed by HH. Some exceptions occur for histidine and tryptophan pairs. The most noticeable exception occurs for turns, which are rarely observed: the His-His, His⟨⟩Trp, and Trp-Trp pairs are two to three times more frequent in turns than the Phe-Phe and Tyr-Tyr pairs.

Further analysis of partner frequencies shows that at least 36–47% of all pairs involve a B partner, the B-B pairs representing as much as 22–33% of all partners (Table I).

Near-Sequence and Far-Sequence Partners in Aromatic Pairs

A near-sequence pair bridges an aromatic residue (n) to a partner between ($n+1$) and ($n+5$) in the sequence; these pairs are often found within a single fragment of secondary structure. The far-sequence pairs bridge residues that are more than 5 apart in the sequence, where the partners come from different fragments of secondary structures. Consequently, the near-sequence pairs should be involved in the stabilization of secondary structures while the far-sequence pairs will stabilize tertiary structures such as β-sheets and helix bundles. There is a clear difference of structure pattern between the near-sequence and the far-sequence pairs: near-sequence pairs are abundant in helices (especially for Phe-X and Trp-X pairs) and the number of pairs abruptly drops when the distance in sequence increases to more than five residues (Fig.3). Near-sequence pairs are also abundant in B structures (especially for His-X and Tyr-X pairs) but the number of pairs decreases more smoothly when the distance in sequence between partners increases.

Far-sequence Aromatic Pairs

The far-sequence pairs are the most abundant since they account for 74% of all pairs. The difference between Table I (all pairs) and Table II (far-sequence pairs) highlights the specific properties of far-sequence pairs. The helix-helix configurations (HH) are less frequent: the frequency of far-sequence pairs is almost half that of all pairs. This is not due to a decrease of helix partners (25–38% in all pairs compared with 18–32% in the far-sequence pairs), but rather, when an aromatic residues is located in a helix, it interacts more frequently with a far-sequence partner found in a strand (B) (HB pairs are 14–25% in the far-sequence pairs compared with only 7–13% in all pairs). This trend increases with B partners (49–54% in far-sequence pairs compared with 36–47% in all pairs). One may finally notice that the secondary structure distribution of aromatic partners in far sequence pairs is similar to the secondary structure pattern of aromatic residues in proteins, paired or not (Table II compared with Table III). This suggests that far-sequence pairing is not specific of any aromatic residue conformation.

The rank order of secondary structures of far-sequence aromatic pairs is BB>HB>HH. Thus, aromatic pairs are more frequently found in β-sheets than in helix bundles, and for the aromatic residues in helices, the favorite partner will be in a strand rather than in another helix.

Near-Sequence Aromatic Pairs

Pairs of aromatic residues 1–5 apart in the sequence account for 26% of all aromatic pairs, and each sequence distance (positions 1–5) corresponds to a particular pattern of secondary structures (Table III, aromatic pairs). When the partners are 2 apart in the sequence, pairs are mainly in a β-strand (84%). When the partners are 1, 3, or 4 apart in the sequence, pairs are mainly in helices (53, 40, and 71%, respectively). The reason seems clear, because in all positions the most common secondary structure of pairs is the one that brings the side chains the closest: β-strand when the partners are 2 apart, helix when they are 1, 3, or 4 apart in the sequence.

From there, one can ask whether the possibility of a side-chain pairing has favored the occurrence of a certain type of secondary structure during protein folding. In other words, does the proximity of two aromatic residues in the sequence correlate with a particular type of secondary structure? To answer the question, we

TABLE I. Distribution of the Secondary Structure Pattern of All Aromatic Pairs in 593 Proteins

All = 8,200	Phe-Phe 16%	Tyr-Tyr 10%	His-His 4%	Trp-Trp 3%	Phe\/Tyr 22%	Phe\His 9%	Phe\/Trp 11%	Tyr\/His 11%	Tyr\/Trp 10%	His\/Trp 5%
All	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
B-B %	33	30	30	22	29	26	29	29	27	24
H-H %	27	23	15	21	25	20	26	15	22	16
C-C %	4	4	9	6	4	5	4	8	8	9
T-T %	0	0	1	0	0	0	0	0	0	0
B-H %	13	13	8	12	13	11	13	12	11	7
B-C %	12	14	19	13	13	18	10	16	12	17
B-T %	2	2	2	3	2	2	0	2	2	1
H-C %	6	9	7	9	10	9	11	11	9	10
H-T %	0	1	0	1	0	0	0	1	0	1
T-C %	2	4	9	13	4	8	6	7	8	15
All = 16,400	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
BX %	47	45	45	36	43	42	41	44	40	36
HX %	37	34	23	32	36	31	38	27	32	25
CX %	14	17	26	23	18	23	18	25	23	30
TX %	2	3	6	9	3	5	4	5	5	9

The pairs with a minimal distance less than 5.5 Å (see Fig. 1) were analyzed for partners and secondary structures. Symmetric pairs were clustered: for example Phe-Tyr and Tyr-Phe are together as Phe\/Tyr. The convention for secondary structure symbol is explained in Methods. B is a beta conformation (including strand, antiparallel, and parallel sheets), H is helix (alpha, 3_{10}, ω, π), T is one of the nine turns previously described by Srinivasan et al.,[15] and C gathers all other conformations. In the first line, the total number of pairs is given (left) before the relative percentages of each pair (right). The percentages of all secondary structures are listed. In the four bottom lines, all pairs are analyzed for the secondary structure frequency of both partners; thus, the 8,200 pairs correspond to 16,400 partners.

TABLE II. Pattern of the Secondary Structure of Far-Sequence Pairs in 593 Proteins (Partners Are More Than 5 apart in the Sequence)

All = 6,108	Phe-Phe 18%	Tyr-Tyr 10%	His-His 4%	Trp-Trp 3%	Phe\/Tyr 23%	Phe\His 8%	Phe\/Trp 11%	Tyr\/Trp 10%	Tyr\/His 10%	His\/Trp 4%
All	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
B-B %	35	34	39	28	33	33	36	34	34	32
H-H %	14	12	8	11	12	9	13	12	8	9
C-C %	2	2	6	6	3	5	2	5	6	9
B-H %	25	22	14	21	23	19	23	22	18	14
B-C %	12	16	23	15	14	21	12	13	19	21
B-T %	1	2	3	5	1	2	1	2	2	2
H-C %	9	10	5	12	11	11	12	10	10	12
H-T %	1	1	0	2	1	1	1	1	1	1
C-T %	1	1	0	1	1	1	0	1	1	0
All = 12,216	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
B-X %	54	54	59	49	53	53	54	53	53	51
H-X %	32	29	18	28	30	24	31	29	23	22
C-X %	13	16	21	20	16	21	15	17	22	26
T-X %	2	2	2	4	2	2	1	2	2	2

The pairs with a minimal distance less than 5.5 Å (see Fig. 1) were analyzed for partners and secondary structures. Symmetric pairs were clustered: for example Phe-Tyr and Tyr-Phe are together as Phe\/Tyr. The convention for secondary structure symbol is explained in Methods. B is a beta conformation (including strand, antiparallel, and parallel sheets), H is helix (alpha, 3_{10}, ω, π), T is one of the nine turns previously described by Srinivasan et al.,[15] and C gathers all other conformations. In the first line, the total number of pairs is given (left) before the relative percentages of each pair (right). The percentages of all secondary structures are listed. In the four bottom lines, all pairs are analyzed for the secondary structure frequency of both partners; thus, the 6,108 pairs correspond to 12,216 partners.

compared the numbers of aromatic duets and of aromatic pairs corresponding to sequence partners $n+1$, $n+2$, $n+3$, $n+4$ and $n+5$ apart in the dataset (Table III duets compared with pairs). First, the numbers of duets are similar at all positions. This supports the idea that evolution has not selected any peculiar motif of sequence with respect to that sequence separation. The most frequent secondary structure of all aromatic residues in proteins is B (an average of 50% is found) and, only 29% are in helices (Table III, All). Similar patterns of secondary structure distribution are observed in duets irrespective of the presence of a second aromatic residue nearby (Table III, duets, positions 1–5). Therefore, there is no significant increase of the H frequency when a second aromatic stands $n+1$, $n+3$, or $n+4$ apart in the sequence, and no significant increase of B structure frequency when a second aromatic is at $n+2$. Thus, the pairing of near-sequence aromatic residues does not

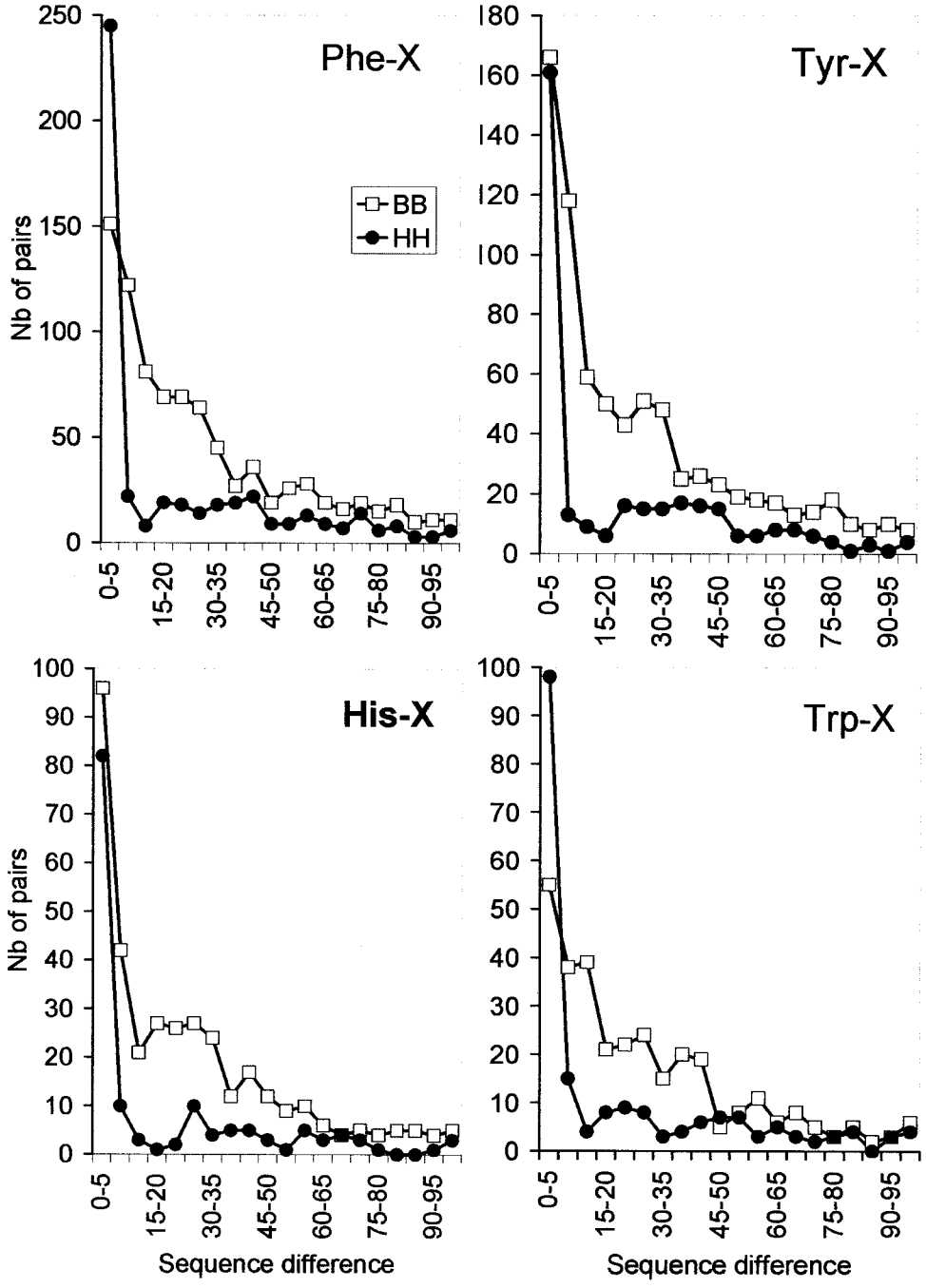

Fig. 3. Frequency distribution of the beta-beta (BB) and the helix-helix (HH) pairs with the sequence length between partners. The four figures stand for all aromatic pairs with a minimal distance <5.5 Å, X being the C-side partner. The secondary structures of pairs were attributed as explained in Methods, and in all instances beta structures are squares and helices are triangles. The pairs were clustered, assuming a sequence window of five residues. At >100 residues, the plots continue to decay, with no particular feature.

enforce secondary structure but, rather is a consequence of secondary structure.

DISCUSSION

In this article, we have analyzed the aromatic interactions in a large set of proteins. The minimal distance between ring atoms clearly demonstrates the existence of a population of aromatic residues that are in direct interaction, with no intermediate even water molecules; those were defined as pairs. As previously reported, the rings of most aromatic pairs stand perpendicular one to the other, and L and T shape geometries are favored.[3,9] However the distribution of ring angles is wide, and it is more correctly described by saying that few rings are stacked in parallel.

TABLE III. Analysis of Near-Sequence Pairs

	Aromatic aa	Aromatic duets					Aromatic pairs				
	All (n = 17,597)	Pos1 (n = 2,094)	Pos2 (n = 2,073)	Pos3 (n = 2,148)	Pos4 (n = 2,169)	Pos5 (n = 2,039)	Pos1 (n = 496)	Pos2 (n = 546)	Pos3 (n = 337)	Pos4 (n = 480)	Pos5 (n = 233)
ss of the N residue (%)											
B	50	48	56	47	47	51	10	84	33	15	56
H	29	32	25	30	33	28	53	3	40	71	30
C	18	19	17	21	18	19	35	9	24	13	11
T	2	1	2	2	2	2	2	4	2	1	2

Comparison of aromatic pairs and duets of near-sequence interactions in 593 proteins. Column "All", for the 17,597 aromatic residues of the 593 protein structures, we calculated the relative percentage of secondary structures (as defined in Methods). Columns "Aromatic duets": in the same proteins, we counted the occurrence of aromatic duets at sequence positions +1 to +4 and calculated the relative percentage of secondary structures of their N-side partner. Columns "Aromatic pairs": Among the duets, some were involved in pairs, we counted the pairs and calculated the relative percentages of secondary structures of their N-side partner.

Parallel stacking has been reported to be specific to isolated pairs.[5]

Previous articles analyzed the geometry of aromatic pairs, but few considered their dependence on secondary structures,[6] and, to our knowledge, none has clearly differentiated the near-sequence from the far-sequence interactions. Indeed, Burley and Petsko[6] reported that 22% of the pairs are between partners <7 apart in the sequence, but their dataset was too small to further study these near-sequence pairs. Finally, they ignored the near-sequence pairs when they concluded that 64% of all aromatic pairs stabilized tertiary structures and 16% quaternary structures. We find that 26% of all aromatic pairs stabilize short fragments of sequences, because they occur between residues that are 1–5 apart in the sequence. These near-sequence interactions are frequently found in helix, but the structure actually varies with the sequence distance between pair partners.

When the two aromatic residues are far apart in the sequence (74% of all pairs), they stabilize a tertiary structure. The most frequent tertiary structure is the β-sheet (BB) and the stabilization of helix bundles is less frequently found.

The existence of near-sequence pairs and their requirement for a specific secondary structure allows us to ask whether aromatic residue pairing influences the secondary structure. The answer is no. There is no major change of secondary structure patterns when two aromatic residues are close in the sequence, and pairing mostly results from the opportunities created by secondary structures. This leads us to conclude that, in the course of protein folding, pairing of aromatic residues should occur after the secondary structures.

ACKNOWLEDGMENTS

The authors acknowledge Dr. B. Ho for his help in the manuscript. R. Brasseur is Research Director at the National Funds for Scientific Research of Belgium (FNRS). R. Meurisse and Benoit Charloteaux are Scientific Research Workers at the National Funds for Scientific Research of Belgium (FNRS).

REFERENCES

1. Samanta U, Debnath P, Chakrabarti P. Packing of aromatic rings against tryptophan residues in proteins. Acta Cryst 1999;D55: 1421–1427.
2. Samanta U, Debnath P, Chakrabarti P. Environment of tryptophan side chains in proteins. Proteins Struct Funct Genet 2000;38: 288–300.
3. Burley SK, Petsko GA. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. Science 1985;229:23–28.
4. Tóth G, Watts CR, Murphy RF, Lovas S. Significance of aromatic-backbone amide interactions in protein structure. Proteins Struct Funct Genet 2001;43:373–381.
5. McGaughey GB, Gagne M, Rappe AK. pi-Stacking interactions. Alive and well in proteins. J Biol Chem 1998;273:15458–15463.
6. Thomas A, Bouffioux O, Geeurickx D, Brasseur R. PEX, analytical tools for PDB files. I. Pex, analytical tools for PDB files. I. GF-Pex: the basic file to describe a protein. Proteins Struct Funct Genet 2001;43:28–36.
7. Liu W-m, Chou K-C. Prediction of protein secondary structure content. Protein Eng 1999;12:1041–1050.
8. Nishio M, Hirota M, Umezawa Y. The CH/π interaction—Evidence, nature, and consequences. In: Marchand AP, editor.

Series: Methods in stereochemical analysis. New York: Wiley-VCH Inc.; 1998. p 46.
9. Singh J, Thornton JM. The interaction between phenylalanine rings in proteins. FEBS Lett 1985;191:1–6.
10. Hunter C, Singh J, Thornton J. π-π Interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. J Mol Biol 1991;218:837–846.
11. Mitchell JBO, Nandi L, McDonald IK, Thornton J, Price SL. Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? J Mol Biol 1994;239:315–331.
12. Ma JC, Dougherty DA. The cation-π interaction. Chem Rev 1997;97:1303–1324.
13. Derewenda ZS, Lee L, Derewenda U. The occurrence of C—HO hydrogen bonds in proteins. J Mol Biol 1995;252:248–262.
14. Martin TW, Derewenda ZS. The name is bond, H-bond. Nat Struct Biol 1999;6(5):403–406.
15. Thomas A, Benhabiles N, Meurisse R, Ngwabije R, Brasseur R. PEX, analytical tools for PDB files. II. H-Pex: noncanonical H-bonds in alpha helices. Proteins Struct Funct Genet 2001;43:37–44.
16. Scatena LF, Brown MG, Richmond GL. Water at hydrophobic surfaces: weak hydrogen bonding and strong orientation effects. Science 2001;292:908–912.
17. Srinivasan R, Rose GD. LINUS: a hierarchic procedure to predict the fold of a protein. Proteins Struct Funct Genet 1995;22:81–99.