Significance of Aromatic-Backbone Amide Interactions in Protein Structure

Gergely Tóth, Charles R. Watts, Richard F. Murphy, and Sándor Lovas*
Department of Biomedical Sciences, School of Medicine, Creighton University, Omaha, Nebraska

ABSTRACT **Weakly polar interactions between aromatic rings of amino acids and hydrogens of backbone amides (Ar-HN) have been shown to support local structures in proteins. Their role in secondary structures, however, has not been elucidated. To investigate the relationship between Ar-HN interaction and the stability of local and secondary structures of polypeptides and to improve the prediction of this interaction based on amino acid sequence, the structures of 560 nonhomologous proteins, from the Protein Data Bank, were searched for Ar–HN interactions between the aromatic ring of each Phe, Tyr, and Trp residue at position *i* and the backbone amide group of any residue, except Pro, at the positions *i*, *i* − 1, *i* − 2, *i* − 3, *i* + 1, *i* + 2, and *i* + 3. Ar-HN interactions were identified by calculating the chemical shift of the amide hydrogen caused by the proximal aromatic ring. Ar(*i*)–HN(*i* + 1, *i* + 2 and *i* + 3) interactions were more common (7.10%, 2.08%, and 0.54%, respectively) than were Ar(*i*)–HN(*i* − 1, *i* − 2, and *i* − 3) interactions (0.66%, <0.1%, and 0.18%, respectively). The value of the χ^1 torsion angle of the aromatic residue in position *i* depended on the direction of the Ar–HN interaction. The position of the aromatic ring in Ar(*i*)–HN(*i* + 1, *i* + 2, and *i* + 3) interactions was mostly *trans*, in Ar(*i*)–HN(*i* − 1, *i* − 2, and *i* − 3) interactions mainly *gauche*(−), and in Ar(*i*)–HN(*i*) interactions mostly *gauche*(+). The analyses of the secondary structures of the protein fragments containing Ar–HN interactions showed that Ar–HN interactions were in all types of secondary structures. Search results suggest that Ar–HN interactions have a stabilizing effect on all types of secondary structures. Proteins 2001;43:373–381.** © 2001 Wiley-Liss, Inc.

Key words: aromatic-amide interaction; peptide; protein; conformation; secondary structure; database search

INTRODUCTION

Attractive interactions between aromatic groups of amino acids and nearby amides, in polypeptides and proteins (Ar–HN), are weakly polar and have a quadrupole–dipole nature. These interactions can be effectively modeled by the electrostatic interaction between the partial negative charge of an aromatic ring and the partial positive charge of an amide hydrogen.[1] The strength of this interaction in vacuum (1–4 kcal mol^{-1}) is comparable with that of a conventional hydrogen bond[1–4] (2–7 kcal mol^{-1}). Therefore, the function and formation of Ar-HN interactions in protein structure are heavily dependent on the conformational consequences of stronger forces such as conventional hydrogen bond.

On the basis of the location of the HN group in a polypeptide, interactions between an aromatic ring and an HN can be characterized as either Ar–NH(side chain) and Ar–HN(backbone). Throughout this text, Ar–HN designates the Ar–HN(backbone) interaction. In case of Arg and Lys, the side chains are protonated under physiological conditions and, therefore, participate in a cation-aromatic interaction,[5] which can be much stronger than the Ar-HN interaction.

The geometry of the Ar–HN interaction can be described by using the angle α between the vector of the N—H bond and the plane of the aromatic ring (Fig. 1.). When α is larger than 30°, the Ar–HN interaction is regarded as perpendicular, and if smaller than 30°, parallel. When the N was closer to the aromatic ring than the amide H, the angle α was considered to be negative. Ab initio calculations in gas phase found that energetically the parallel and perpendicular Ar–HN interaction are similar.[4]

In solution, ^{1}H-NMR spectroscopy has been used to identify Ar–HN interactions in peptides and proteins.[6–11] The ^{1}H-NMR spectra of a disordered peptide is composed of random coil values of the protons. Local conformations change the chemical shift of the protons from their random coil values by the structural shift. The structural shift is composed from the effect of the electric field from polar groups, the effect of the delocalized electrons of the aromatic ring, and the effect to the anisotropy of the C═O and C—N bond. In the case of Ar–HN interactions, the aromatic ring is in the close proximity of the amide proton; thus, the delocalized electrons of the aromatic ring change the local magnetic field of the amide proton, thereby causing an anomalous shift called ring shift (δ_{ring}). The value of the ring shift can be related to the interaction geometry of the Ar–HN interaction.[11]

Weakly polar interactions in short peptides are able to constrain side-chain conformation and stabilize local struc-

The Supplementary Material referred to in this article can be viewed at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/43_4/v43_4.html

Grant sponsor: National Science Foundation; Grant number: 9720643; Grant sponsor: Carpenter Chair in Biochemistry, Creighton University.

*Correspondence to: Sándor Lovas, Department of Biomedical Sciences, School of Medicine, Creighton University, 2500 California Plaza, Omaha, NE 68178. E-mail: vasz@bif1.creighton.edu

Received 28 June 2000; Accepted 2 February 2001

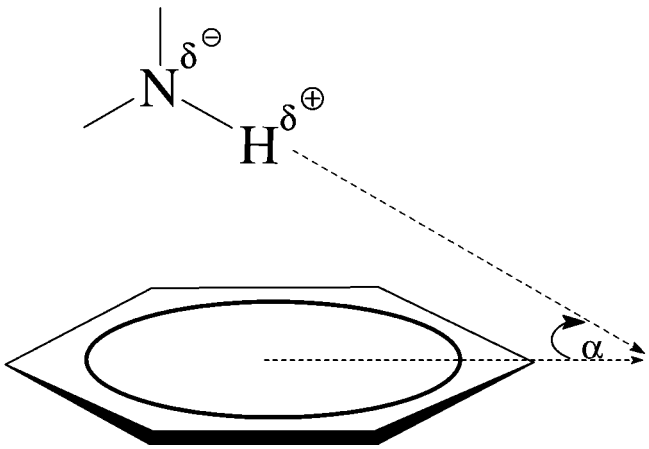

Fig. 1. Geometry of an Ar–HN interaction.

tures.[6–8,12] This is the case in the deltorfin 2 analogue, Hat-D-Ala-Phe-Glu-Ile-Ile-Gly-NH_2, in which an Ar(i)–HN(i + 2) interaction is between the aromatic ring of Phe3 and the backbone amide of Ile5 (Geza Toth, personal communication). Another example[13] is in crystalline Tyr-Tyr-Leu monohydrate where the amide group of Tyr2 interacts with the aromatic ring of Tyr1. The aromatic side chain of Tyr1 is in the *trans* position.

NMR measurement of reduced bovine pancreatic trypsin inhibitor (BPTI) revealed local nonrandom conformations.[14] These local structures are formed[6,9] by Ar(i)–HN(i + 2) interactions when the aromatic ring of Tyr10 interacts with the backbone amide of Gly12 and the aromatic ring of Tyr35 interacts with the backbone of Gly37. Further studies by Kemmink and Creighton[7,8] revealed that the interaction between Tyr35 and Gly37 assists in the folding of BPTI by stabilizing an intermediate structure along the folding pathway. The presence of Ar–HN interactions in the lowest-energy structures of Ac-Phe-Gly-Gly-$NHCH_3$, obtained by simulated annealing with and without an implicit GB/SA solvation model and using three different (Amber 5.0, CHARMM22, and OPLSAA) force fields,[10] also suggests a structure-stabilizing role of these interactions.

Several statistical surveys of sequence distribution and geometry of the Ar–HN interactions in proteins have been reported. Studies by Burley and Petsko[15,16] suggested involvement of Ar–HN(side chain) interactions in the stabilization of protein tertiary structures on the bases of their spatial distribution. Flocco and Mowbray[17] examined the occurrence and geometry of the interactions between the guanidinium group of Arg, the amide group of Asn or Gln, and the aromatic side chain of Phe, Tyr, and Trp in 83 nonhomologous protein structures. They found that the parallel arrangement of these groups is the most frequent geometry, because the parallel arrangement is sterically more favorable because the nitrogen is able to engage in hydrogen bonding with another residue. In this way the nitrogen can achieve its maximal hydrogen-binding capacity. Mitchell and coworkers[18] found that approximately 10% of the identified aromatic-amide, -amino, -guanido pairs of 55 nonhomologous protein structures are in either side chain-side chain or side chain-backbone interactions, mostly in parallel geometry. Worth and Wade[11] searched for Ar(i)–HN(i + 1 and i + 2) interactions in 297 nonhomologous structures and found that in Ar(i)–HN(i + 2) interaction the propensity for Gly to be in position i + 2 is by far the highest than for all other amino acids.

The Ar–HN(side chain) interactions stabilize tertiary and local structures and strengthen protein-ligand interactions. Conversely, the Ar–HN interactions have been mostly known to stabilize local structure. Secondary structures of proteins may create such three-dimensional space that could facilitate Ar–HN interactions, and these could then further stabilize these structures. To elucidate the role of Ar–HN interactions in secondary and local structures of polypeptides and to improve the prediction of the Ar–HN interactions on the basis of amino acid sequence, we searched 560 protein structures for Ar–HN interactions between the aromatic ring of Phe, Tyr, and Trp at position i and the backbone amide group of any residue, except Pro, at the positions $i - 3$, $i - 2$, $i - 1$, i, $i + 1$, $i + 2$, and $i + 3$. The secondary structures of fragments containing an Ar–HN interaction were determined.

MATERIALS AND METHODS

A database of 560 coordinate files of proteins with <25% sequence similarity and a resolution of 3 Å or better,[19] from the Protein Data Bank[20] was used. The protein coordinate files were downloaded from the EMBL file server (ftp.embl-heidelberg.de) and from them a database was generated using SYBYL 6.2.[21]

The database was searched for fragments, using the *SEARCH* command of the Biopolymer module of SYBYL, containing either Phe, Tyr, or Trp at position i and any others residue, except Pro at position $i - 1$ and $i + 1$, $i - 1$, $i + 2$, $i - 2$, $i + 3$, and $i - 3$. The resultant coordinate files for the protein fragments were stored in SYBYL databases.

SYBYL script #1 was used to add amide hydrogens to the fragments and then to execute the program Total,[22] which calculated the amide hydrogen ring shifts in protein fragments. Next, fragments with a ring shift of ≤0.5 ppm were selected using PERL script #1. The SYBYL script #2 was used to calculate the angle α, χ^1 and χ^2 torsion angles of the aromatic residue in these fragments as well as the distance between the side-chain aromatic ring centroid and the amide hydrogen.

PERL scripts #2 and #3 were used to tabulate distances, angles, and ring shifts, identified Ar–HN interactions and their geometry, and for final statistical analyses. A protein fragment was assigned to contain Ar–HN interaction if the ring shift of the amide hydrogen was ≤0.5 ppm.[9–11,23] and the distance between the side-chain aromatic ring centroid and the amide hydrogen was <4.5 Å. Multiple copies of particular Ar–HN interactions due to structural analogs in the same protein were ruled out on the basis of similarities in amino acid sequence and secondary structure. The geometry of Ar–HN interactions was defined according to angle α (Fig. 1.). If α is >30° than the interaction is perpendicular and parallel if α is <30°.

The secondary structure of the fragments was determined by the program DSSP.[24] To define the secondary structure of a protein fragment, the secondary structure of the whole

protein had to be determined. From the DSSP output files, PERL script #4 was used to select the secondary structure codes on the basis of chain and residue numbers of fragments. Four residues before and four residues after the fragments were also selected and analyzed.

The propensity (p) for an amino acid to be in one type of Ar–HN interaction was calculated by the formula: $p = n^{j,\ \mathrm{Ar}(i)\text{–}\mathrm{HN}(l)} \sum N^{\mathrm{all,\ proteins}} / n^{j,\ \mathrm{proteins}} \sum N^{\mathrm{Ar}(i)\text{–}\mathrm{HN}(l),\ \mathrm{fragments}}$,[25] where $n^{j,\ \mathrm{Ar}(i)\text{–}\mathrm{HN}(l)}$ is the occurrence of amino acid j in a certain Ar(i)–HN(l) interaction; $l = I - 3, i - 2, i - 1, i, i + 1, i + 2, i + 3$; $N^{\mathrm{all,\ proteins}}$ is the number of amino acids in all proteins in the database; $n^{j,\ \mathrm{proteins}}$ is the occurrence of amino acid j in all proteins in the database; $N^{\mathrm{Ar}(i)\text{–}\mathrm{HN}(l),\ \mathrm{fragments}}$ is the number of Ar(i)–HN(l) interactions in (i) − (l) fragments.

Aromatic–aromatic interaction[26] was defined when the distance between aromatic ring centroid was <7 Å and the angle between the planes of the two rings between 60° and 120°. CH-π interaction[27] between an aromatic ring and the side chain of a Pro was assigned when the ring shift, calculated by the Total program, of a hydrogen bonded to a carbon was < -0.25 ppm.[28–30]

TABLE I. Number of Ar(i)—HN($i, i + 1, i + 2, i + 3, i - 1, i - 2$, and $i - 3$) Interactions[†]

Amide position	Phe	Tyr	Trp	Phe + Tyr + Trp
$i + 3$	12 (0.43)	18 (0.51)	14 (0.77)	44 (0.54)
$i + 2$	101 (1.46)	151 (2.42)	57 (3.47)	309 (2.08)
$i + 1$	469 (6.65)	465 (6.81)	258 (8.96)	1192 (7.10)
i	265 (3.33)	193 (2.68)	129 (4.32)	587 (3.24)
$i - 1$	59 (0.77)	21 (0.30)	35 (1.23)	115 (0.66)
$i - 2$	5 (<0.1)	4 (<0.1)	1 (<0.1)	10 (<0.1)
$i - 3$	10 (0.14)	7 (0.10)	13 (0.46)	30 (0.18)

[†]Relative frequencies (%) of the total number of fragments with Ar—HN interactions containing the given aromatic residue are in the brackets.

RESULTS AND DISCUSSION

Distribution of Ar–HN Interactions in Protein Structure

Ar(i)–HN($i + 1, i - 1, i + 2, i - 2, i + 3$, and $i - 3$) interactions

The numbers of the different Ar–HN interactions found in 560 coordinate files of the Brookhaven Protein Data Bank are listed in Table I. Ar(i)–HN($i + 1, i + 2$, and $i + 3$)

TABLE II. Number (N) of Ar(i)—HN($i + 1, i + 2, i + 3, i - 1, i - 2$, and $i - 3$) Interactions and Propensity (p) for a Residue (Xaa) to Be Therein

	$i + 1$			$i + 2$			$i + 3$		
Ar(i)	Xaa	p	N	Xaa	p	N	Xaa	p	N
Phe	Met	1.87	20	Gly	5.21	41	Tyr	4.60	2
	Cys	1.72	12	Cys	2.01	3	Gly	4.27	4
	Ser	1.56	43	Asn	1.31	6	Arg	3.44	2
	Asp	1.52	42	His	1.28	3	Ala	1.97	2
	Asn	1.40	30	Glu	1.12	7	Asn	1.83	1
Tyr	Gln	1.61	28	Gly	5.69	67	Gly	4.99	7
	Thr	1.49	41	Trp	1.74	4	Ser	2.83	3
	Ser	1.39	38	Tyr	1.27	7	Asp	1.89	2
	His	1.29	14	Arg	1.22	9	Glu	1.79	2
	Arg	1.24	28	Ser	1.01	9	Asn	1.22	1
Trp	Asp	1.77	27	Gly	4.50	20	Lys	5.10	4
	Gln	1.76	17	His	3.02	4	Trp	4.70	1
	Asn	1.53	18	Ser	1.49	5	His	3.07	1
	Thr	1.37	21	Lys	1.25	4	Arg	1.48	1
	Arg	1.35	17	Cys	1.18	1	Ile	1.25	1
	$i - 1$			$i - 2$			$i - 3$		
	Xaa	p	N	Xaa	p	N	Xaa	p	N
Phe	Cys	4.58	4	Gly		4	Gly		2
	Asn	2.23	6	Tyr		1	Asp		2
	Thr	1.53	7				Val		2
	Asp	2.01	6				Lys		1
	His	1.73	2				Ser		1
Tyr	Ser	2.43	3	His		1	Gly		4
	Thr	2.41	3	Phe		1	Asp		1
	Phe	2.37	2	Lys		1	Ser		1
	Met	2.10	1	Ala		1	Lys		1
	His	2.05	1						
Trp	Ser	2.92	6	Lys		1	Arg		3
	Tyr	1.58	2				Glu		2
	Gln	1.53	2				Ala		2
	Gly	1.47	4				Gly		1
	Asp	1.46	3				Ser		1

TABLE III. Number (*N*) of Ar(*i*)—HN(*i*) Interactions and Propensity (*p*) for an Amino Acid (Xaa) to Be Therein at Positions *i* − 1 and *i* + 1

	i − 1			*i* + 1		
Ar(*i*)	Xaa	*p*	*N*	Xaa	*p*	*N*
Phe	Pro	2.76	33	Phe	1.88	20
	Trp	2.61	10	Trp	1.74	7
	Tyr	2.08	19	Tyr	1.56	15
	Cys	1.61	6	Ser	1.48	23
Tyr	Trp	2.83	8	Pro	2.19	20
	Pro	2.16	19	Trp	2.05	6
	Tyr	1.78	12	Phe	1.68	13
	Phe	1.75	13	Thr	1.66	19
Trp	Pro	4.05	25	Ser	2.38	18
	Ser	2.62	20	Asp	2.11	16
	Cys	2.08	4	Glu	1.88	15
	Asp	1.57	12	Gln	1.66	8

interactions were more common than were Ar(*i*)–HN(*i* − 1, *i* − 2, and *i* − 3) interactions. More than three times more Ar(*i*)–HN(*i* + 1) than Ar(*i*)–HN(*i* + 2) interactions were found. Ar(*i*)–HN(*i*) interactions were almost twice as numerous as Ar(*i*)–HN(*i* + 2) interactions. Ar(*i*)–HN(*i* − 2, *i* − 3, and *i* + 3) interactions were rare.

More Ar(*i*)–HN(*i* + 1) interactions than Ar(*i*)–HN(*i* + 2) interactions, more Ar(*i*)–HN(*i* − 1) interactions than Ar(*i*)–HN(*i* − 2) interactions and more Ar(*i*)–HN(*i* + 2) interactions than Ar(*i*)–HN(*i* + 3) interactions were found. These findings suggest that the distance and number of separating residues between the backbone amide and the aromatic residue correlate with occurrence of the interaction. The difference in length seems to account for the predominance of Ar(*i*)–HN(*i* + 1, *i* + 2, and *i* + 3) interactions over Ar(*i*)–HN(*i* − 1, *i* − 2, *i* − 3) interactions. Only a C═O group exists between the aromatic side chain and the *i* + 1 backbone amide, whereas, in Ar(*i*)–HN(*i* − 1) interactions, the HN(*i* − 1) is further away because of the $C_\alpha(i-1)$ atom. The side chain of the *i* − 1 residue my also interfere with the Ar–HN interaction.

The probability of the involvement of Trp in Ar–HN interactions is the highest among aromatic amino acids. This can be attributed to its larger size compared with Phe and Tyr. The incidence of Tyr in Ar(*i*)–HN (*i* + 1, *i*+ 2, *i* + 3) interactions was higher, whereas in Ar(*i*)–(*i* and *i* − 1)HN interactions lower, than that of Phe.

Involvement of amino acids in Ar(*i*)–HN(*i* + 1, *i* + 2, *i* + 3, *i* − 1, *i* − 2, and *i* − 3) interactions are listed in Table II and in the Supplementary Online Material section. Ser, Thr, Asp, Asn, and Gln promoted Ar(*i*)–(*i* + 1)HN interactions, whereas Leu and Ala at the *i* + 1 position had a tendency to hinder the interactions. More than 41% of all Ar(*i*)–HN(*i* + 2) interactions and 9.1% of Phe-Xaa^1-Xaa^2 fragments contained Gly at the *i* + 2 position. Gly was an outstanding promoter of Ar(*i*)–HN(*i* + 2) interactions, and this finding agrees well with the results of the database search by Worth and Wade.[11] In Ar(*i*)–HN(*i* + 3) interactions, Gly was the most frequent at position (*i* + 3) when Phe and Tyr was in position (*i*), whereas Trp in position (*i*) interacted often with the backbone amide of Lys. Ser and Thr seemed to promote Ar(*i*)–HN(*i* − 1) interactions. The incidence of Ar(*i*)-HN(*i* − 2 and *i* − 3) interactions was too low to group participating residues as either promoting or hindering the interactions.

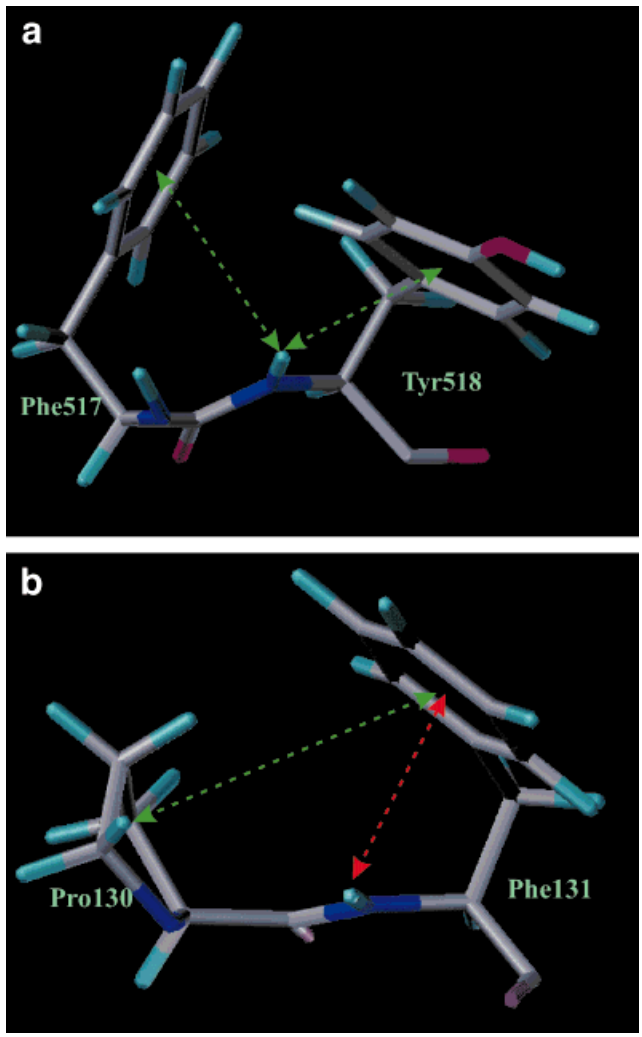

Fig. 2. **a:** Ar(*i*)–HN(*i*) interaction (green dotted line) in Tyr518 and Ar(*i*)–HN(*i* + 1) interaction between the aromatic ring of Phe517 and the amide of Tyr518, and aromatic–aromatic interaction between the two aromatic rings of Phe517 and Tyr518 in Phosphoinoitide-Specific Phospholipase C (pdb code 1djx). **b:** Ar(*i*)–HN(*i*) interaction in Phe131 and CH-π interaction (red dotted line) between the aromatic ring of Phe131 and the side chain C^{δ}-$H^{\delta1}$ of Pro130 in Ascorbate Oxidase (pdb code 1aoz).

Ar(i)–HN(i) interactions

The number of Ar(*i*)–HN(*i*) interactions and propensity for residues to be at positions *i* + 1 and *i* − 1 are shown in Table III. When Trp was at the position *i* in Ar(*i*)–HN(*i*) interactions, the propensity for Ser to be in both *i* + 1 and *i* − 1 positions was high, and for aromatic residues normal to low. Pro was the most frequent residue found in the *i* − 1 position. In Ar(*i*)–HN(*i*) interactions, the propensities for Phe and Tyr to be in positions *i* − 1 and *i* + 1 was exceptionally high. To examine whether these high propensities were by chance or due to other effects, the possibility of interactions between the aromatic residue at the *i*

TABLE IV. Relative Frequency (%) of Parallel (par) and Perpendicular (per) Ar(*i*)—HN Interactions

Amide position	Phe		Tyr		Trp		Phe + Tyr + Trp	
	par	per	par	per	par	per	par	per
$i + 3$	50.00	50.00	61.11	38.89	71.43	28.57	61.36	38.64
$i + 2$	47.52	52.48	23.18	76.82	42.11	57.89	34.63	65.37
$i + 1$	97.87	2.13	98.28	1.72	96.12	3.88	97.65	2.35
i	74.72	25.28	77.72	22.28	76.74	23.25	76.15	23.85
$i - 1$	65.71	34.29	84.75	15.25	66.67	33.33	75.65	24.35
$i - 2$	60.00	40.00	50.00	50.00	100.00	0.00	60.00	40.00
$i - 3$	50.00	50.00	42.86	57.14	61.54	38.46	53.33	46.67

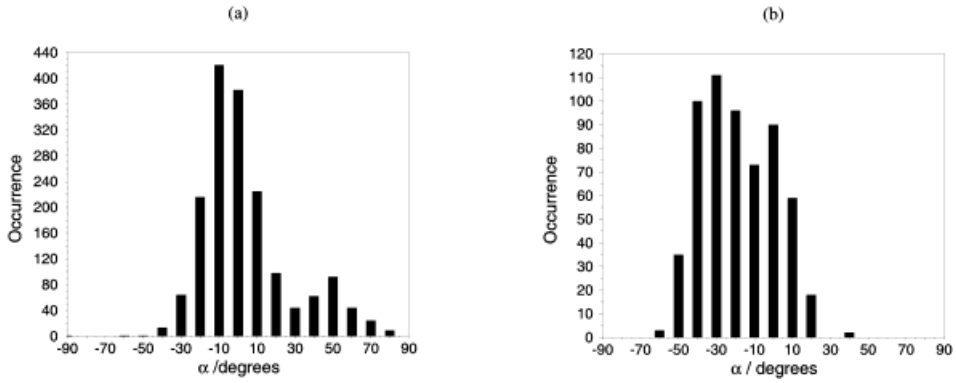

Fig. 3. **a:** Distribution of α angles in Ar–HN interactions except Ar(*i*)–HN(*i*) interactions. **b:** Distribution of α angles in Ar(*i*)–HN(*i*) interactions.

position and side chains of Phe, Tyr, Trp, and Pro at the $i - 1$ position was examined. The aromatic–aromatic interactions involving the two side chains could additionally influence the formation of a high number of Ar(*i*)–HN(*i*) interactions. About 51% of fragments that had Ar(*i*)–HN(*i*) interactions had aromatic-aromatic interactions as well. In addition, about 30% of the same fragments simultaneously had Ar(*i*)–HN(*i*), Ar($i - 1$)–HN(*i*), and aromatic-aromatic interactions. The Ar(*i*) and Pro($i - 1$) were frequently involved in CH-π interaction. Such simultaneous interactions are illustrated in Figure 2. CH-π and/or aromatic-aromatic interactions may be responsible for forcing the aromatic ring into a position that could favor the Ar(*i*)–HN(*i*) interaction. Weakly polar interactions should, thus, be considered collectively. The existence of one type of interaction is likely to be responsible for the formation of another and vice versa. These individual interactions are weak, but together they may have a much stronger influence on the rigidity of local structures in proteins.

Geometry of Ar–HN Interactions

Table IV lists the relative frequencies of the parallel and perpendicular geometries from the database search in this study. As the distance of the backbone amide from the aromatic residue increased, the occurrence of perpendicular interactions increased. The perpendicular geometry was only favored in Ar(*i*)–HN($i + 2$) interactions. The predominance of the perpendicular geometry in Ar(*i*)–HN($i + 2$) interactions could be attributed to the presence of Gly in the $i + 2$ position. When Gly was in position $i + 2$ in Ar(*i*)–HN($i + 2$) interactions, 79% of fragments had perpendicular and 21% had parallel geometry. All other residues were 55% in perpendicular and 45% in parallel geometry. In low-energy structures of the model tripeptide Ac-Phe-Gly-Gly-$NHCH_3$ the geometry of Ar(*i*)–HN($i + 2$) interactions, in agreement with present results, were perpendicular.[10] By not following the general geometric trends for Ar–HN interactions, Ar(*i*)–HN($i + 2$) interactions with Gly at position $i + 2$ are special cases for unknown reasons. Further investigation of this phenomenon by molecular dynamics simulations is under way.

The distribution of the α angle in parallel geometry in Ar(*i*)–HN(*i*) interactions was well below 0°, whereas all other parallel Ar-HN interactions had a majority at 0° (Fig. 3.). In this arrangement, the amide nitrogen was closer to the centroid of the aromatic ring than was the amide hydrogen. Another common orientation in this case was when the amide nitrogen and the amide hydrogen was almost equally distanced from the centroid of the aromatic ring, and the amide nitrogen was close to selected hydrogens of the aromatic ring.

To examine whether the side chains of the aromatic residues at position *i* prefer certain conformations, χ^1 and χ^2 torsion angles were measured. Figure 4 illustrates the distributions of the χ^1 angle in various Ar–HN interactions. Side chains of the aromatic residues in almost all Ar(*i*)–HN($i + 1$, $i + 2$, $i + 3$) interactions were mostly in *trans* and substantially less in *gauche*(+) or *gauche*(−) position. Ar(*i*)–HN(*i*) interactions favored *gauche*(+) and Ar(*i*)–HN($i - 1$, $i - 2$, $i - 3$) *gauche*(−) orientation. In all cases, the Ar–HN interactions supported different rotamer

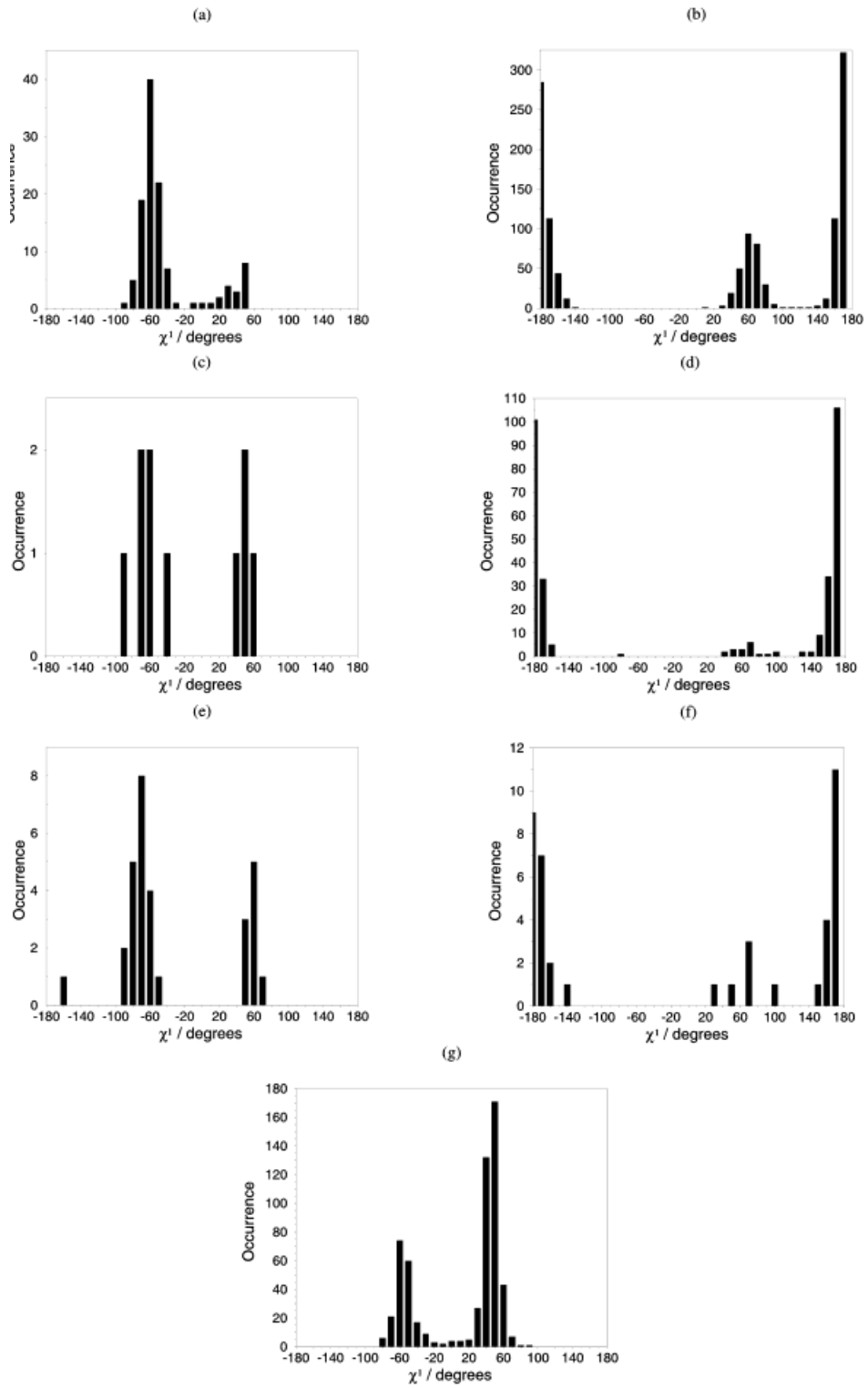

Fig. 4. Distribution of χ^1 torsion angles of the aromatic side chain of the residue at position (i) in **(a)** Ar(i)–HN($i - 1$), **(b)** Ar(i)–HN($i + 1$), **(c)** Ar(i)–HN($i - 2$), **(d)** Ar(i)–HN($i + 2$), **(e)** Ar(i)–HN($i - 3$), **(f)** Ar(i)–HN($i + 3$), and **(g)** Ar(i)–HN(i) interactions.

preferences for aromatic side chains compared with the distribution of χ^1 in all protein sequences.[25] Values of the χ^2 torsion angle were in good agreement with random distributions of aromatic side chains.

On the basis of the distribution of the χ^1 rotamer in Ar–HN interactions, the freedom of rotation of the aromatic ring is restricted, whereas rotamer distributions usually depend on backbone conformation.[31]

TABLE V. Relative Frequency (%) of Secondary Structures of Protein Fragments Containing Ar—HN Interactions

Amide position	β-sheet	Turn and bend	Helix	Random structure	Other[a]
$i + 1$	30.32	24.19	14.55	15.88	9.06
$i - 1$	70.70	9.60	3.54	12.12	4.04
$i + 2$	13.39	38.39	10.72	14.58	22.92
$i - 2$	0.00	53.85	30.77	15.38	0.00
$i + 3$	0.00	64.71	35.29	0.00	0.00
$i - 3$	0.00	75.00	25.00	0.00	0.00

[a]Protein fragments with Ar—HN interactions containing more than one type of secondary structure.

The rotation restriction in Ar–HN interactions can be responsible for side-chain configurational entropy loss.

Ar–HN Interactions in Protein Secondary Structures

The relative frequency of the secondary structures of fragments containing Ar–HN interactions is summarized in Table V. Ar(i)–HN(i) interactions were omitted from Table V, because the role of this interactions was only to control the orientation of the aromatic side chain as illustrated in Figure 4g. Furthermore, the effect of this interaction on the conformation of the side chain should be independent of the current secondary structure of the polypeptide in which it is present. Ar(i)–HN($i - 1$ and $i + 1$) interactions were in all major categories of secondary structure; fragments mostly were part of β-sheets. Ar(i)–HN($i + 2$ and $i - 2$) interactions were predominantly in turns and bends, and they were present only in the N-caps of helices. No Ar(i)–HN($i - 2$) interactions were in β-sheets, whereas some Ar(i)–HN($i + 2$) interactions were in β-sheets. Ar(i)–HN($i + 3$ and $i - 3$) interactions were almost exclusively in turns, bends, and helical structures, and no fragments therewith were found in β-sheets. Ar(i)–HN($i - 2$, $i + 3$, and $i - 3$) interactions were not formed in β-sheets, because the backbone amide and the aromatic side chains were too far away. This is due to the extended conformation of the backbone in β-sheets.

Ar–HN interactions in random meander structures

Random meander structures contain no hydrogen bonding between the backbone of the residues, which could define ordered secondary structures. As a consequence, the existence of an Ar–HN interaction could shape the local conformation dramatically. In agreement with previous

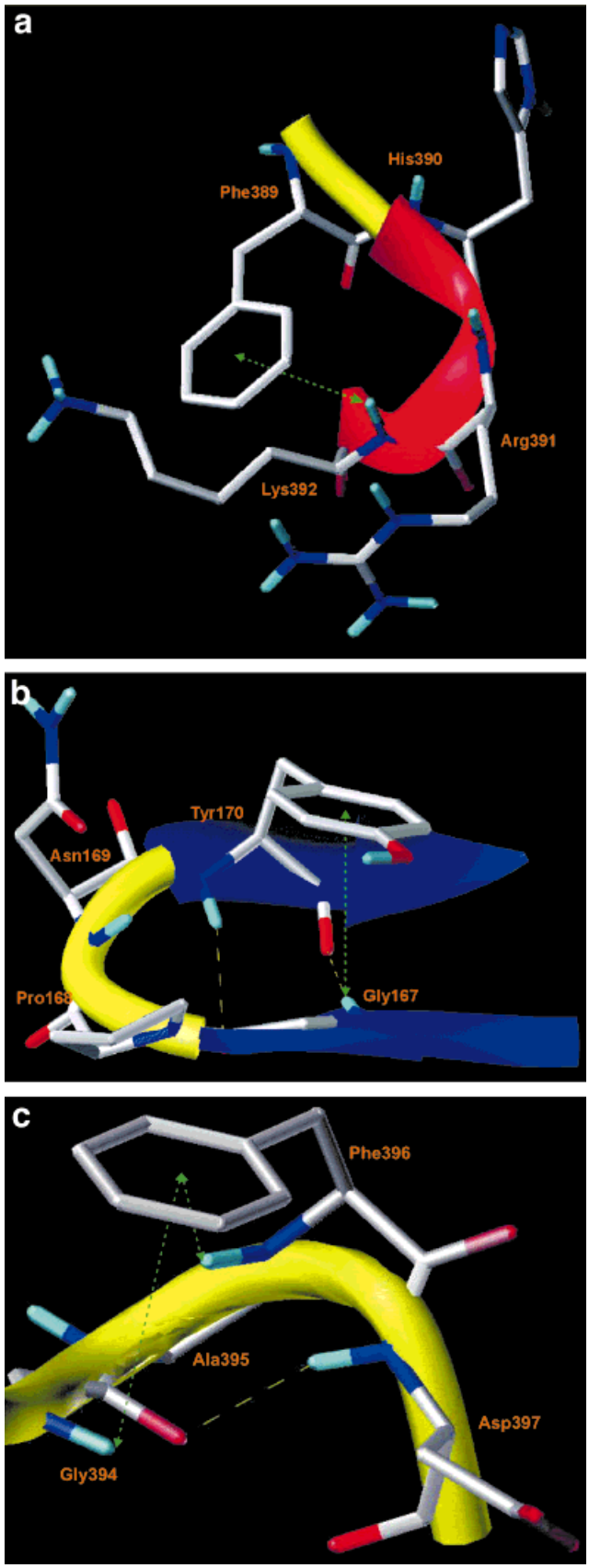

Fig. 5. Ar–HN interactions (green dotted lines) in protein secondary structures. For clarity, only amide and hydroxyl hydrogens are shown. Hydrogen bonds are indicated by yellow dotted lines. Red ribbons, blue arrows, and yellow tubes indicate helices, β-sheets, and turns, respectively. **a:** The Ar(i)–HN($i + 3$) interaction between the aromatic ring of Phe389 and the backbone amide of Lys392 in the N-terminal region of a helix in pyruvate kinase (pdb code: 1pkm) **b:** The Ar(i)–HN($i - 3$) interaction between the aromatic ring of Tyr170 and the backbone amide of Gly167 in the two strands of a β-sheet in papain (pdb code: 9pap) **c:** The Ar(i)–HN(i and $i - 2$) interactions between the aromatic ring of Phe396 and its own backbone amide and the backbone amide of Gly394 in position 1 and 3 of a turn in chitinase A (pdb code: 1ctn).

data,[6–9,23] Ar–HN interactions formed local structures. Ar(i)–HN($i - 3$ and $i + 3$) interactions were never found in random meander structures.

Ar–HN interactions in helices

Ar(i)–HN($i - 2$, $i + 2$, and $i + 3$) and many of the Ar(i)–HN($i - 3$) interactions, located in helices, were in the N-terminal region of helices. In the interior of the helix, only Ar(i)–HN($i - 1$ and $i + 1$) interactions could be found. The higher occurrence of Ar–HN interactions in the N-terminal than in the inner helix can be explained by the fact that the first three amide hydrogens of a helix do not form hydrogen bonds with the backbone carbonyl oxygen, so the helical geometry could be frayed (Fig. 5a). In the interior of the helix, formation of Ar–HN interactions is also possible, because in parallel geometry the aromatic side chains do not compete with the backbone carbonyl groups for the amide hydrogen; thus, both contacts are possible simultaneously.

The propensity of aromatic residues to be in helices is low, partly because of the loss of side-chain configurational entropy.[32] Ar–HN interactions are additional restrictions on the freedom of rotation of the aromatic side chains in helices. This suggests that the enthalpic component of Ar–HN interaction may offset somewhat the entropic cost of fixing an aromatic side chain in the core of a protein.

Ar–HN interactions in β-structures

The high propensity of aromatic amino acids to be in β-sheets[33] is partly explained by their tendency to form interstrand side-chain contacts, thereby shielding the polar backbone atoms from solvent molecules and ultimately promoting hydrogen bonding between the backbone of the β-strands.[34,35] Investigation of antiparallel β-sheets revealed interstrand contacts between the side chain of an aromatic amino acids and amide backbone of a Gly, which depend heavily on the interstrand proximity of the residues in the β-sheets and on the position of their side chain rotamers.[36] The high occurrence of Ar(i)–HN(i, $i + 1$ and $i - 1$) interactions in β-sheets (Fig. 5b) suggests that these interactions may contribute to the stability of β-sheets by not only favorable entalpies but also by their ability to hinder solvent access to the backbone amide. Hydrogen bonding of the amide to the carbonyl group of the other strand is still possible, because the majority of Ar(i)–HN(i, $i + 1$, and $i - 1$) interactions had parallel arrangement (Table IV).

Ar–HN interactions in turns and bends

Ar(i)–HN($i - 1$ and $i + 1$) interactions were mostly in positions 2 and 3 of a turn. Approximately half of Ar(i)–HN($i - 2$) interaction were in positions 1 and 3, and half in positions 2 and 4. Ar(i)–HN($i + 2$) interactions were almost entirely in positions 1 and 3. As many Ar(i)–HN($i + 2$) interactions were also found in positions −1 and 2 as in positions 1 and 3. All Ar(i)–HN($i - 3$ and $i + 3$) were in positions 1 and 4.

Turns in proteins are usually found near the surface, where they have direct contact with solvent molecules; thus, the occurrence of hydrophobic amino acids in turns is not favored. In model tetrapeptides AA-Ar-Pro-Ar-Hp, where AA represents any amino acid, Ar an aromatic residue, and Hp a small hydrophilic residue, aromatic side chains packing against Pro stabilize the β-turn by partial exclusion of hydration water from the Pro ring.[37–39] The interaction of the aromatic ring with the backbone amide could play a similar role as the Pro-Ar contacts in turns, but in such cases the aromatic ring would protect the backbone amide from solvent access. Ar(i)–HN($i - 2$, $i - 3$, $i + 2$, and $i + 3$) interactions may also further stabilize the turn structure (Fig. 5b and c).

CONCLUSION

Structures of 560 nonhomologous proteins, from the Protein Data Bank, were searched for Ar–HN interactions between the aromatic ring of the residue Phe, Tyr, and Trp at position i and the backbone amide group of any residue, except Pro, at the positions i, $i - 1$, $i - 2$, $i - 3$, $i + 1$, $i + 2$, and $i + 3$. The results showed the presence of Ar–HN interactions in different secondary structures. This suggests that Ar–HN interactions could further support the hydrogen-bonded secondary structures. The geometry of the Ar–NH interactions was mostly parallel except in Ar(i)–HN($i + 2$) interactions. In the N-terminal of helices, where side-chain backbone amide interaction become important for stability, several Ar–HN interactions were found. Being consistent with the general notion of protein folding, in which apolar side chains are packed tightly to bury the backbone from the solvent to assist secondary structure formation, the results of this study suggest that Ar–HN interactions may have such a stabilizing effect on all types of secondary structures.

Only a few residues that could promote Ar–HN interaction were identified. In Ar(i)–HN(i) interactions Pro, Phe, Tyr, and Trp most often preceded the aromatic residues. The propensity for Gly to be at positions $i + 2$ and $i - 2$ to form Ar(i)–HN($i + 2$ and $i - 2$) interactions was much greater than for the other residues.

Ar–HN interactions had an additional role in shaping local structures. Evidence for this is that the direction of Ar–HN interaction, for example (i)–($i - 1$) or (i)–($i + 1$), defines the χ^1 torsional angle of the aromatic side chain. In Ar(i)–HN(i and $i + 1$) interaction, two local interactions that involve weakly polar interactions other than Ar–NH interaction were identified: (1) CH-π interaction between Pro and the aromatic ring and (2) Ar–Ar interactions between Phe, Tyr, and Trp. Either CH-π or Ar–Ar interactions together with Ar–HN interactions can constrain the aromatic ring into a position that favors both types of interaction. Local conformations, supported by such weakly polar interactions combined, may be more stable than expected.

REFERENCES

1. Levitt M, Perutz MF. Aromatic rings act as hydrogen bond acceptors. J Mol Biol 1988;201:751–754.
2. Cheney J, Cheney BV, Richard G. Calculation of NH…π hydrogen bond energies in basic pancreatic trypsin inhibitor. Biochim Biophys Acta 1988;954:137–139.
3. Rodham DA, Suzuki S, Suenram DR, Lovas FJ, Dasgupta S, Goddard WA III, Blake GA. Hydrogen bonding in the benzene-ammonia dimer. Nature 1993;362:735–737.

4. Duan G, Smith VH Jr, Weaver DF. An ab initio and data mining study on aromatic-amide interactions. Chem Phys Lett 1999;310: 323–332.
5. Gallivan J, Dougherty DA. Cation-π interactions in structural biology. Proc Natl Acad Sci USA 1999;96:9459–9464.
6. Kemmink J, van Mierlo CPM, Scheek RM, Creighton TE. Local structure due to an aromatic-amide interaction observed by H-nuclear magnetic resonance spectroscopy: peptides related to the N terminus of bovine pancreatic trypsin inhibitor. J Mol Biol 1993;230:312–322.
7. Kemmink J, Creighton TE. Local conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor and their roles in folding. J Mol Biol 1993;234:861–878.
8. Kemmink J, Creighton TE. The physical properties of local interactions of tyrosine residues in peptides and unfolded proteins. J Mol Biol 1995;243:251–260.
9. Nardi F, Worth GA, Wade RC. Local interactions of aromatic residues in short peptides in aqueous solution: a combined database and energetic analysis. Fold Design 1997;2:62–68.
10. Tóth G, Lovas S, Murphy RF. Simulated annealing studies on aromatic-amide interaction in Phe-Gly-Gly tripeptide using different force fields. Internet J Chem 1998;2:http://www.ijc.com/articles/1999v2/5/
11. Worth GA, Wade RC. The aromatic–(i + 2) amine interaction in peptides. J Phys Chem 1995;99:17473–17482.
12. Shimohigash Y, Nose T, Yamauchi Y, Maeda I. Design of serine protease inhibitors with conformation restricted by amino acid side-chain-side-chain CH/π interaction. Biopolymers 1999;51:9–17.
13. Steiner T. Structural evidence for the aromatic-(i + 1) amine hydrogen bond in peptides: L-Tyr-L-Tyr-L-Leu monohydrate. Acta Crystal Sec D 1999;D54:584–588.
14. van Mierlo CPM, Darby NJ, Keeler J, Neuhaus D, Creighton TE. The partially folded conformation of the (30–51) intermediate in the disulphide folding pathway of bovine pancreatic trypsin inhibitor: ^{1}H and ^{15}N resonance assignments and determinataiton of the backbone dynamics from ^{15}N relaxation measurements. J Mol Biol 1993;229:1125–1146.
15. Burley SK, Petsko GA. Weakly polar interactions in proteins. Adv Protein Chem 1988;39:125–189.
16. Burley SK, Petsko GA. Amino-aromatic interactions in proteins. FEBS Lett 1986;203:139–143.
17. Flocco MM, Mowbray SL. Planar stacking interactions of arginine and aromatic side-chains in proteins. J Mol Biol 1994;235:709–717.
18. Mitchell JBO, Nandi CL, McDonald IK, Thornton JM. Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? J Mol Biol 1994;239:315–331.
19. Hobohm U, Scharf M, Schneider R, Sander C. Selection of representative protein data sets. Protein Sci 1992;1:409–417.
20. Bernstein FC, Koetzle TF, Williams GJ, Meyer EE Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
21. Tripos Inc. Sybyl Users Manual. St. Louis: MO 63144, USA.
22. Williamson PW, Asakura T. Empirical comparisons of models for chemical-shift calculation in proteins. J Mag Res Ser B 1993;101: 67–71.
23. Worth GA, Nardi F, Wade RC. Use of multiple molecular dynamics trajectories to study biomolecules in solution: the YTGP peptide. J Phys Chem 1998;102:6260–6272.
24. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
25. Penel S, Hughes E, Doig AJ. Side-chain structures in the first turn of the α-helix. J Mol Biol 1999;287:127–143.
26. Burley SK, Petsko GA. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. Science 1985;229:23–28.
27. Nishio M, Umezawa Y, Hirota M, Takeuchi Y. The CH/π interaction: significance in molecular recognition. Tetrahedron 1995;51: 8665–8701.
28. Wu W-J, Raleigh DP. Local control of peptide conformation: stabilization of cis proline peptide bonds by aromatic proline interactions. Biopolymers 1998;45:381–394.
29. Shimohigashi Y, Nose T, Yamauchi Y, Maeda I. Design of serine protease inhibitors with conformation restricted by amino acid side-chain-side-chain CH/π interaction. Biopolymers 1999;51:9–17.
30. Nardi F, Kemmink J, Sattler M, Wade RC. The cisproline(i − 1)-aromatic(i) interaction: folding of the Ala-cisPro-Tyr peptide characterized by NMR and theoretical approaches J Biomol NMR 2000;17:63–77.
31. Dunbrack RL, Karplus M. Backbone-dependent rotamer library for proteins. J Mol Biol 1993;230:543–547.
32. Stapley BJ, Doig AJ Free energies of amino acid side-chain rotamers in α-helices, β-sheets and α-helix N-caps. J Mol Biol 1999;272:456–464.
33. Chou PY, Fasman GD. Conformational parameters for amino acids in helical, beta-sheet and random coil regions calculated from proteins. Biochem 1974;13:11–22.
34. Kortemme T, Ramirez-Alvarado M, Serrano L. Design of a 20-amino acid, three-stranded β-sheet protein. Science 1998;281:253–256.
35. Merkel SJ, Regan L. Aromatic rescue of glycine in β sheets. Fold Design 1998;3:449–455.
36. Hutchinson EG, Sessions BS, Thornton JM, Woolfson DN. Determinants of strand register in antiparallel β-sheets of proteins. Protein Sci 1998;7:2287–2300.
37. Yao J, Dyson HJ, Wright PE. Three-dimensional structure of a type VI turn in a linear peptide in water solution. J Mol Biol 1994;243:754–766.
38. Yao J, Feher VA, Espejo BF, Reymond MT, Wright PE, Dyson HJ. Stabilization of a type VI turn in a family of linear peptides in water solution. J Mol Biol 1994;243:736–753.
39. Demchuk E, Bashford D, Gippert GP, Case DA. Thermodynamics of a reverse turn motif: solvent effects and side-chain packing. J Mol Biol 1997;270:305–317.