Investigation of Aromatic-Backbone Amide Interactions in the Model Peptide Acetyl-Phe-Gly-Gly-*N*-Methyl Amide Using Molecular Dynamics Simulations and Protein Database Search

Gergely Tóth, Richard F. Murphy, and Sándor Lovas*

Contribution from the Department of Biomedical Sciences, School of Medicine, Creighton University, 2500 California Plaza, Omaha, Nebraska 68178

Received May 21, 2001

Abstract: Weakly polar interactions between the side-chain aromatic rings and hydrogens of backbone amides (Ar−HN) are found in unique conformational regions. To characterize these conformational regions and to elucidate factors that determine the conformation of the Ar−HN interactions, four 4-ns molecular dynamics simulations were performed using four different low-energy conformations obtained from simulated annealing and one extended conformation of the model tripeptide Ac-Phe-Gly-Gly-NH-CH_3 as starting structures. The Ar(*i*)−HN(*i*+1) interactions were 4 times more frequent than were Ar(*i*)−HN(*i*+2) interactions. Half of the conformations with Ar(*i*)−HN(*i*+2) interactions also contained an Ar(*i*)−HN(*i*+1) interaction. The solvent access surface area of the Phe side chain and of the amide groups of Phe1, Gly2, and Gly3 involved in Ar−HN interactions was significantly smaller than in residues not involved in such interactions. The number of hydrogen bonds between the solvent and Phe1, Gly2, and Gly3 amide groups was also lower in conformations with Ar−HN interactions. For each trajectory, structures that contained Ar(*i*)−HN(*i*), Ar(*i*)−HN(*i*+1), and Ar(*i*)−HN(*i*+2) interactions were clustered on the basis of similarity of selected torsion angles. Attraction energies between the aromatic ring and the backbone amide in representative conformations of the clusters ranged from −1.98 to −9.24 kJ mol^{-1} when an Ar−HN interaction was present. The most representative conformations from the largest clusters matched well with the conformations from the Protein Data Bank of Phe-Gly-Gly protein fragments containing Ar−HN interactions.

Introduction

The strength of the weakly polar interaction between the side-chain aromatic ring of an amino acid and a backbone amide of a polypeptide (Ar−HN interaction) can be as high as 16 kJ mol^{-1}.[1] This is comparable with the strength (8−29 kJ mol^{-1}) of a conventional hydrogen bond. The geometry of the Ar−NH interaction can be described using the angle α between the vector of the N−H bond and the plane of the aromatic ring. The Ar−HN interaction is regarded as perpendicular when α is larger than 30° and parallel when smaller. Ab initio values in a vacuum showed no significant difference between the maximum strength of the interactions in the two orientations.[1,2]

The conformation of polypeptide fragments containing Ar−HN interactions can depend on factors including the amino acid sequence, the hydrophobicity or hydrophilicity of the local environment, the degree of solvation, and the structural flexibility of the polypeptide fragment. Some of these factors were investigated[2−6] by data mining in the Protein Data Bank[7] (PDB), (a) The percentages of Ar(*i*)−HN(*i*+1), Ar(*i*)−HN(*i*+2), and Ar(*i*)−HN(*i*+3) interactions are 7.10, 2.08, and 0.54%, respectively, whereas the percentages of Ar(*i*)−HN(*i*−1), Ar(*i*)−HN(*i*−2), and Ar(*i*)−HN(*i*−3) interactions are 0.66, <0.1. and 0.18%, respectively.[6] (b) In Ar(i)−HN(*i*+2) interactions, the propensity for Gly to be in position *i*+2 is far higher than for other amino acids.[4] (c) Ar−HN interactions are mostly in parallel geometry in proteins because, in this orientation, the nitrogen of the amide is able to form an additional hydrogen bond with another residue, thereby achieving its maximal hydrogen-binding capacity.[2−4] (d) The aromatic side chain is constrained in either gauche+ and gauche− or trans and gauche+ orientations, depending on the type of Ar−HN interactions.[6] (e) Ar−HN interactions are found in a variety of secondary structures in which they could have structure-stabilizing roles.[6]

Ar−HN interactions have been the subject of several experimental and theoretical investigations. For example, NMR studies elucidated that the aromatic ring of Tyr10 and the backbone amide of Gly12 form an Ar(*i*)−HN(*i*+2) interaction in a bend conformation in bovine pancreatic trypsin inhibitor (BPTI).[8,9] Further NMR[10] and molecular dynamics[11−13] investigations of

* Corresponding author: (phone) (402) 280-5753; (fax) (402) 280-2690; (e-mail) vasz@bif1.creighton.edu.

(1) Duan, G.; Smith, V. H., Jr.; Weaver, D. *Chem. Phys. Lett.* **1999**, *310*, 323−332.
(2) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *239*, 315−331.
(3) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709−717.
(4) Worth, G. A.; Wade, R. C. *J. Phys. Chem.* **1995**, *99*, 17473−17482.
(5) Mitchell, J. B. O.; Nandi, C. L.; Ali, S.; McDonald, I. K.; Price, S. L.; Singh, J. *Nature* **1993**, *336*, 413
(6) Tóth, G.; Watts, C. R.; Murphy, R. F.; Lovas, S. *Proteins: Struct. Funct. Genet.* **2001**, *43*, 373−381.
(7) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.
(8) Kemmink, J.; van Mierlo, C. P. M.; Scheek, R. M.; Creighton, T. E. *J. Mol. Biol.* **1993**, *230*, 312−322.
(9) Kemmink, J.; Creighton, T. E. *J. Mol. Biol.* **1993**, *234*, 861−878.
(10) Kemmink, J.; Creighton, T. E. *J. Mol. Biol.* **1995**, *243*, 251−260.
(11) van der Spoel, D.; van Buuren, A. R.; Tieleman, D. P.; Berendsen, H. J. C. *J. Biomol. NMR* **1996**, *8*, 229−238.

10.1021/ja011245u CCC: $20.00
Published on Web 11/01/2001

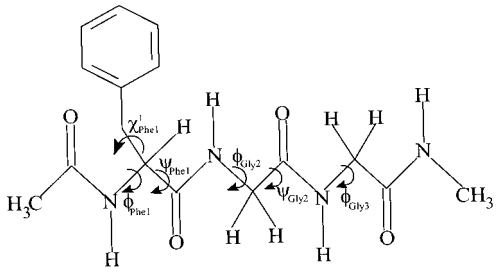

Figure 1. Torsion angles of Ac-Phe-Gly-Gly-*N*-methyl amide used in the cluster analysis of trajectories.

the 10−14 fragment (Tyr-Thr-Gly-Pro) of BPTI revealed local structures stabilized by the Ar(i)−HN(i+2) interactions.

The fundamental factors that determine the conformation of polypeptide fragments containing Ar−HN interactions can be investigated using model peptides in which no intramolecular hydrogen bonds constrain the structure. Thus, the Ar−HN interaction would be the strongest noncovalent force in the structure. Also, in such model peptides, no side chains should interfere with the formation of Ar−HN interactions. The tripeptide Ac-Phe-Gly-Gly-NME (FGG) is such an ideal model. Worth and Wade[4] performed a conformational search for low-energy structures of FGG containing Ar(i)−HN(i+2) interactions, using the CHARMM force field, by varying six torsion angles, χ^1_{Phe1}, χ^2_{Phe1}, ψ_{Phe1}, ϕ_{Gly2}, ψ_{Gly2}, and ϕ_{Gly3} (Figure 1). The conformational search yielded nine low-energy structures in simulated aqueous media and two in a vacuum. Tóth et al.[14] used these as starting structures for simulated annealing studies using three different force fields with the GB/SA solvation model.[15] The lowest energy structures obtained with the different force fields differed significantly, though all starting structures with the same force field resulted in the same lowest energy structure. The presence of Ar(i)−HN(i+2) interactions in the computed lowest energy structures in aqueous media implies that Ar−HN interactions take part in stabilizing the folded structure of a peptide.

To further characterize peptide conformations containing Ar−HN interactions and to investigate the solvation of these conformations, a molecular dynamics (MD) study of FGG structures using the modified GROMOS-87[16] force field was done. The starting structures for the MD were the lowest energy structures in aqueous media calculated by simulated annealing.[14] To characterize the conformational preference for Ar−HN interactions in the FGG peptide, the MD trajectories were clustered. Detailed examination of clusters revealed characteristic conformation regions with Ar−HN interactions. These were compared with conformation regions containing Ar−HN interactions in polypeptide fragments found in the PDB.

Methods

Molecular Dynamics Simulations. All molecular dynamics calculations for FGG were done using the modified GROMOS-87 force field as implemented in the GROMACS 1.6 program package.[16] Starting structures for trajectory **1**, **2**, and **3** were the lowest energy structures calculated using simulated annealing in aqueous media with the AMBER,[17] CHARMM,[18] and OPLSAA[19] force field, respectively, by Tóth et al.[14] The starting structure for trajectory **4** was an extended conformation of FGG in which χ^1_{Phe1} was 60°. Each starting structure was immersed in a cubic box (30 Å × 30 Å × 30 Å) of SPC/E water molecules so that all water molecules with oxygen atoms less than 2.8 Å or hydrogen atoms less than 2.0 Å from the peptide were removed. All systems were energy minimized using the steepest descent method until the difference between the total potential energy of the molecular system in two adjacent energy minimization steps was less than 0.001 kJ mol^{-1}. Then, NVT MD was performed for 20 ps by positionally restraining the peptide in the center of the box with a force of 1000 kJ mol^{-1} at 300 K to allow the solvent density to approach the equilibrium value. Finally, four separate 4040-ps molecular dynamics trajectories, at a constant temperature of 300 K and constant pressure of 1 bar, were generated. The first 40 ps was regarded as the equilibration period and was excluded from the trajectory analysis. The following parameters were used for the dynamics simulations: 2-fs time steps, a nonbonded interactions list updated in every 10 steps, 1.0 nm cutoff distance for evaluation of nonbonded interaction, the LINCS algorithm[20] to set bonds to their correct length with the warning angle of 30°, a constant dielectric of 1.0 for all Coulomb interactions, a cutoff of 1.0 nm, the peptide and solvent coupled to separate temperature baths with relaxation constant of 0.1 ps, and the peptide and solvent coupled to a pressure bath using isotopic and atomic scaling with a relaxation constant of 0.5 ps. The coordinates of the peptide were stored for evaluation after every 1000 steps to yield a total of 2000 sampled conformations for each trajectory.

The energies of the nonbonded interactions between the aromatic ring of Phe1 and the backbone amide of Phe1, Gly2, and Gly3 in sampled structures in the trajectory were also calculated with the modified GROMOS-87 force field as follows. Sampled structures from the trajectory having the average torsion angles of the clusters of the trajectories were energy minimized using the steepest descent algorithm. The maximum initial step size was 0.005 nm. The minimization converged when the maximum force was smaller than 0.001 kJ mol^{-1} nm^{-1}.

Trajectory Analysis. The trajectories were analyzed using the analysis suit of GROMACS 1.6 to determine the total energy, backbone RMSD, radius of gyration (R_g), torsion angles, number of hydrogen bonds between each backbone amide, and the solvent water molecules (NHB). Solvent-accessible surface area (SASA) of the peptide, the backbone amides, and the side-chain phenyl group was calculated with the NACCESS[21] program.

The Ar−HN interactions were assigned on the basis of the backbone amide hydrogen NMR ring shift[5] (δ_{ring}). δ_{ring} is the result of the change in the local magnetic field of the proton due to the nearby delocalized electrons of an aromatic ring of a side chain during an ^{1}H NMR experiment. The value of δ_{ring} is influenced by the interaction geometry of the Ar−HN interaction. An Ar−NH interaction was assigned when the δ_{ring} of the backbone amide hydrogen was −0.5 ppm or lower.[5] The Total[22] program was used to calculate the backbone amide hydrogen δ_{ring}.

Geometry of Ar−HN Interaction. Results from a protein database search[6] suggested the inverse perpendicular geometry of the Ar−HN interaction. This is the geometry of Ar−HN interactions when α is less than −30°.

(12) Nardi, F.; Worth, G. A.; Wade, R. C. *Folding Des.* **1997**, *2*, 62−68.
(13) Worth, G. A.; Nardi, F.; Wade, R. C. *J. Phys. Chem.* **1998**, *102*, 6260−6272.
(14) Toth, G.; Lovas, S.; Murphy, R. F. *Internet J. Chem.* **1999**, *2*, http://www.ijc.com/articles/1999v2/5/.
(15) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(16) van der Spoel, D.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; van Drunen, R.; Berendsen, H. J. C. *GROMACS User Manual*; University of Groningen, 1996. http://rugmd0.chem.rug.nl/~gmx.
(17) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(18) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(19) Jorgensen, W. L.; Maxwell, D. S.; Rives, J. T. *J. Am. Chem. Soc.* **1996**, *118*, 1125−1136.
(20) Hess, B.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(21) Hubbard, S. J.; Thornton, J. M. NACCESS program; Department of Biochemistry and Molecular Biology, University College, London, U.K., 1993.
(22) Williamson, P. W.; Asakura T. *J. Magn. Reson., Ser. B* **1993**, *101*, 67−71.

Cluster Analysis of Trajectories. Sampled conformations from each trajectory were collected in five groups on the basis of the presence of Ar(i)−HN(i), Ar(i)−HN(i+1), Ar(i)−HN(i+2), Ar(i)−HN(i+1,i+2), and no Ar−HN interactions. The four groups with Ar−HN interactions from each trajectory were clustered using the partitioning around medoids (PAM) clustering method.[23] This method of cluster analysis finds groups of related conformations on the basis of their pairwise dissimilarities. Dissimilarities between conformations were defined by calculating the torsion angle root-mean-square deviations for each pair of structures:

$$d_{ij} = \sqrt{\frac{1}{N}\sum_{k=1}^{N} \min[(\theta_k^{(i)} - \theta_k^{(j)})^2, (2\pi - \theta_k^{(i)} + \theta_k^{(j)})^2]} \qquad (1)$$

where N is the number of torsion angles, and $\theta_k^{(i)}$ and $\theta_k^{(j)}$, respectively are the torsion angle θ_k in structures i and j. χ^1_{Phe1} and ϕ_{Phe1} torsion angles were used to calculate the d_{ij} for the group with Ar(i)−HN(i) interactions, χ^1_{Phe1} and ψ_{Phe1} for the group with Ar(i)−HN(i+1) interactions, and χ^1_{Phe1}, ψ_{Phe1}, ϕ_{Gly2}, and ψ_{Gly2} for the two groups with Ar(i)−HN(i+2) and Ar(i)−HN(i+1,i+2) interactions. The dissimilarity matrix, constructed using a Perl script, was used as input file for the clustering program, PAM. Clustering was performed using a minimum of 2 and maximum of 10 clusters. The number of clusters representing the optimal clustering of the system was chosen on the basis of the highest average silhouette width of all clusters.[24,25]

Protein Database Search. A database of 560 coordinate files of proteins from the PDB, with less than 25% sequence similarity[26] and a resolution of 3 Å or better (the list of the redundant proteins was downloaded from EMBL file server: ftp.embl-heidelberg.de) was created using SYBYL 6.2.[27] The database was searched, using the SEARCH command of the Biopolymer module of SYBYL, for fragments containing either Phe, Tyr, or Trp at position i and any other residue, except Pro, at position i+1 and i+2. The resultant coordinate files for the protein fragments were stored in SYBYL databases. SYBYL script 1 was used to add amide hydrogen to the fragments and then to execute the Total program to calculate the amide hydrogen δ_{ring}. Next, fragments with a δ_{ring} of −0.5 ppm or less were selected using PERL script 1. SYBYL script 2 was used to measure selected torsion angles of fragments and the distance between the side-chain aromatic ring centroid and the amide hydrogen. PERL script 2 was used to identify Ar−HN interactions on the basis of the following criteria. The δ_{ring} of the amide hydrogen was −0.5 ppm or less, and the distance between the side-chain aromatic ring centroid and the amide hydrogen was less than 4.5 Å. Multiple copies of particular Ar−HN interactions, due to structural analogues of the same protein, were ruled out on the basis of similarities in amino acid sequence and secondary structure. The DSSP program[28] was used to determine the secondary structure of the fragments. Perl script 3 ruled out multiple copies of Ar−HN analogues and was used to tabulate the torsion angles of the selected protein fragments. Selected torsion angles of the protein fragments were clustered using PAM as described above.

Results

Ar−HN Interactions in FGG during the Trajectories. The percentages of Ar−HN interactions in FGG during the simulations are summarized in Table 1. The occurrence of Ar(i)−HN(i+1), Ar(i)−HN(i+2), and Ar(i)−HN(i+1,i+2) interactions was similar in the trajectories except in trajectory **4**, in which the occurrence of Ar(i)−HN(i+1) was much lower than in trajectories **1**−**3**. No Ar(i)−HN(i) interaction existed in trajectory **1**.

Table 1. Percentage of Sampled Conformations with Ar−HN Interactions in the Trajectories

	trajectories			
Ar-HN interaction	**1**	**2**	**3**	**4**
(i)−(i)	0.00	0.35	3.00	4.44
(i)−(i+1)	37.88	41.58	39.43	18.94
(i)−(i+2)	13.20	16.40	9.75	11.34
(i)−(i,i+1)	0.00	0.00	0.06	0.00
(i)−(i+1,i+2)	6.00	8.00	5.50	3.60

Dynamics of FGG. During the simulations, FGG showed moderate flexibility, even though it contained no intramolecular hydrogen bonds. Figure 2 shows values of selected torsion angles, the radius of gyration (R_g), and Ar−HN interactions for the simulation of FGG structures in trajectories **1** and **4**. Results for two of the four trajectories are shown because similar trends were observed in all four trajectories. The extent of folding/unfolding is represented by the low/high values of R_g. In trajectory **1**, the conformation of the tripeptide was stable from 300 to 3300 ps. In this period, the R_g was low and had small fluctuations (Figure 2A), Ar(i)−HN(i+1) and Ar(i)−HN(i+2) interactions were present continuously (Figure 2B), χ^1_{Phe1} was around 180° and torsion angles ψ_{Phe1} and ϕ_{Gly2} fluctuated between 120° and 150° and ⁻80° and ⁻150°, respectively. The average δ_{ring} of Gly2 and Gly3 amide hydrogens for this period were −0.44 ± 0.41 and −0.28 ± 0.31, respectively. FGG was more flexible in trajectory **4** than in trajectory **1**, since R_g maximums were higher and fluctuated more. For the first 1400 ps of this trajectory, Ar(i)−HN(i+1) and Ar(i)−HN(i+2) interactions were continuously present. In this period, the torsion angles and δ_{ring} averages for the Gly2 and Gly3 amide hydrogens were similar to those in the 300−3300-ps period of trajectory **1**. In trajectory **4** at 1400 ps, a conformation change took place as values of ψ_{Phe1} and ϕ_{Gly2} shifted. From about 1400 to 2850 ps, only occasional Ar(i)−HN(i+2) interactions occurred. From 2850 to 3000 ps, the same regions of ψ_{Phe1} and ϕ_{Gly2} torsional phase space and similar frequencies of Ar(i)−HN(i+1), Ar−HN(i+2), and Ar−HN(i+1,i+2) interactions were observed as for the first 1400 ps of the trajectory. At 3000 ps, a large maximum in R_g was indicative of unfolding of the peptide which could be attributed to a change in the orientation of the aromatic side chain from trans to gauche− (Figure 2H). Eventually, the R_g decreased when the aromatic side chain moved to gauche+ orientation and formed an Ar(i)−HN(i) interaction.

Clustering the Conformations in the Trajectories. To examine the conformational characteristics of Ar−HN interactions in FGG, a multistep clustering procedure was performed. The clusters are summarized in Table 2. Ar(i)−HN(i) interactions took place when facilitated by appropriately complementary values of χ^1_{Phe1} and ϕ_{Phe1}. The number of the sampled structures with Ar(i)−HN(i) interactions was low, and so, the selected average torsion angles characterizing the clusters (Table 2) may be misleading. Conformers in cl1_t_i and cl2_t_i had similar ϕ_{Phe1}, while the orientation of the aromatic side chain was either gauche+ or gauche−. The formation of Ar(i)−HN(i+1) interactions depended on the values of torsion angles χ^1_{Phe1} and ψ_{Phe1}. Conformers in cl1_t_i+1 and cl2_t_i+1 were sampled in all four trajectories. The formation of Ar(i)−HN(i+2) and Ar(i)−HN(i+1,i+2) interactions depended on suitable combinations of χ^1_{Phe1}, ψ_{Phe1}, ϕ_{Gly2}, and ψ_{Gly2} values. Conformers in cl2_t_i+2, cl2_t_i+1,i+2, cl3_t_i+2, and cl3_t_i+1,i+2 were present in all four trajectories.

Database Search for Protein Fragments Containing Ar−HN Interactions. A total of 5.45, 7.24, 7.24, and 1.81% of

(23) Kaufman, L.; Rousseeuw, P. J. Wiley: New York, 1990.
(24) Watts, C. R.; Tóth, G.; Murphy, R. F.; Lovas, S. *J. Mol. Struct. (THEOCHEM)* **2001**, *535*, 171−182.
(25) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(26) Hobohm, U.; Scharf, M.; Schneider, R.; Sander, C. *Protein Sci.* **1992**, *1*, 409−417.
(27) Tripos Inc. *Sybyl Users Manual*; St. Louis, MO, 63144.
(28) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.

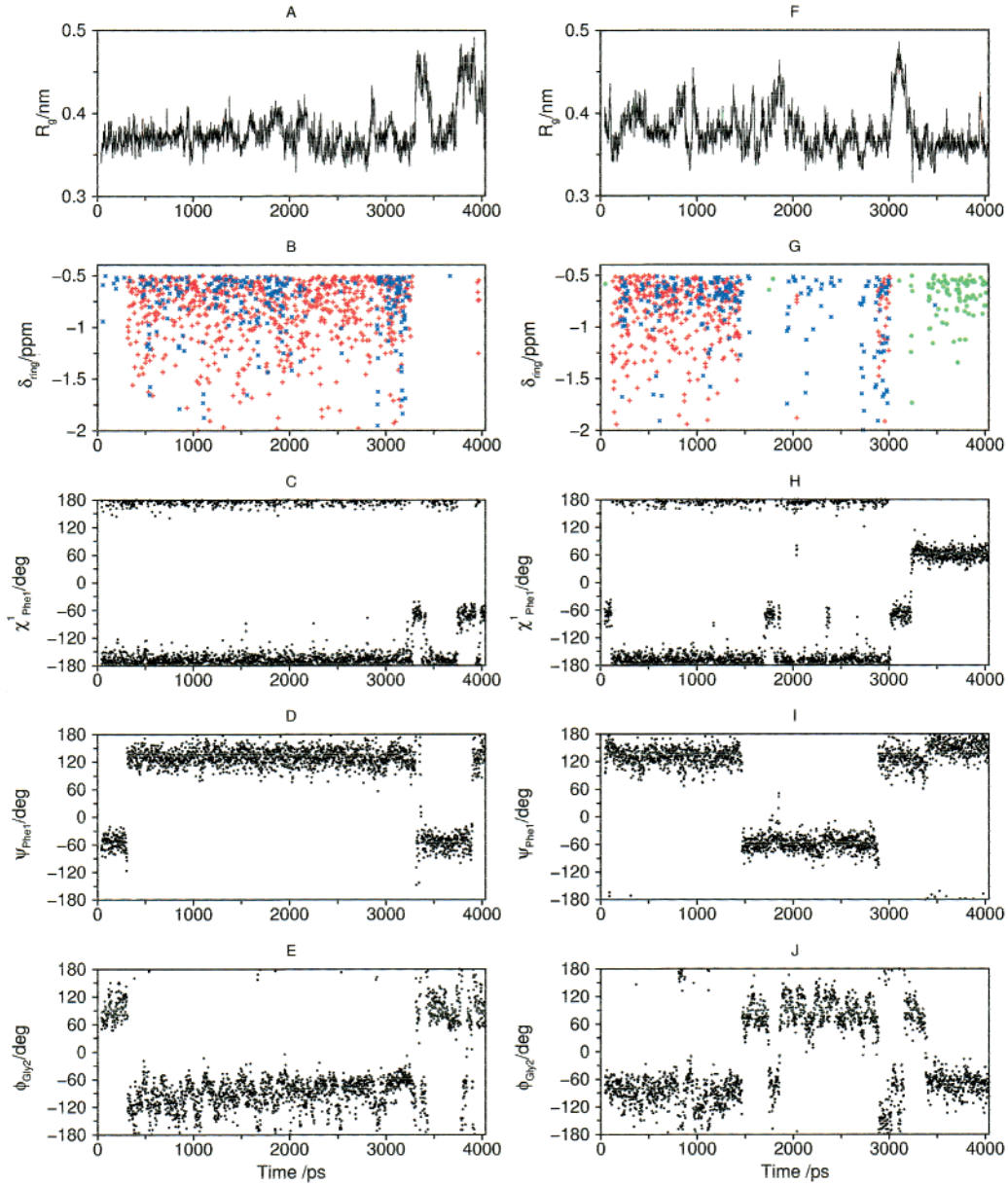

Figure 2. Evolution of radius of gyration (R_g) in trajectories **1** (A) and **4** (F), δ_{ring} in trajectories **1** (B) and **4** (G), χ^1_{Phe1} in trajectories **1** (C) and **4** (H), ψ_{Phe1} in trajectories **1** (D) and **4** (I), and ϕ_{Gly2} in trajectories **1** (E) and **4** (J).

FGG protein fragments contained Ar(i)−HN(i), Ar(i)−HN(i+1), Ar(i)−HN(i+2), and Ar(i)−HN(i+1,i+2) interactions, respectively (Table 3). Ar(i)−(i)HN and Ar(i)−HN(i+1) were found in turns, α-helices and β-sheets, while Ar(i)−HN(i+2) and Ar(i)−HN(i+1,i+2) interactions were present only in random meander. FGG fragments were mostly at the surface of the investigated proteins, except fragments from proteins with PDB access code *1cxs* and *2er7*. Torsion angles of FGG protein fragments with Ar−HN interactions were in the torsion angle regions described by the clusters from the trajectories (Table 2). Torsion angles of FGG fragments with Ar(i)−HN(i+1) interactions from proteins with PDB access codes *2er7* and *2pcd* were similar to those in cl2_t_i+1, *2kau* to cl1_t_i+1 and *1gof* to cl3_t_i+1. Torsion angles of FGG fragments with Ar(i)−HN(i+2) interactions in proteins with PDB access codes *1cxs* and *1djx* were similar to those in cl3_t_i+2 and cl2_t_i+2, respectively, and torsion angles of FGG fragments with Ar(i)−HN(i+1,i+2) interactions were similar to those in cl1_t_i+1,i+2.

To examine whether the conformation of Ar−HN interactions defined by the clusters from the trajectories are characteristic either of FGG or of all amino acid sequences, a database scan of torsion angles of protein fragments containing Ar(i)−HN-(i+1), Ar(i)−HN(i+2), and Ar(i)−HN(i+1,i+2) interactions was done. The collected torsion angles were subjected to clustering using the PAM method (Table 4). The values of ψ_{Phe1} in structures of cl1_d_i+1 were contained within both cl1_t_i+1 and cl2_t_i+1, while only part of cl2_d_i+1 was present in cl3_t_i+1. Structures in cl1_d_i+2 and cl3_d_i+2 had similar χ^1_{Phe1}, ψ_{Phe1}, and ϕ_{Gly2} torsion angle values, while their ψ_{Gly2} were different. Structures in cl1_d_i+2, cl2_d_i+2, and cl3_d_i+2 included the torsion angle regions defined by structures in cl2_t_i+2. The values of torsion angles χ^1_{Phe1}, ψ_{Phe1}, ϕ_{Gly2}, and ψ_{Gly2} in structures in cl4_d_i+2 and cl2_t_i+2 were similar. Structures in cl5_d_i+2 did not sample any of the conformational-phase space in the trajectories. Structures in cl1_d_i+1,i+2, cl3_d_i+1,i+2, and cl4_d_i+1,i+2 had similar χ^1_{Phe1}, ψ_{Phe1}, and ϕ_{Gly2} torsion angles, while their ψ_{Gly2} angles were different. The χ^1_{Phe1}, ψ_{Phe1}, amd ϕ_{Gly2} torsion angles of structures in cl1_d_i+1,i+2 and cl4_d_i+1,i+2 were similar

Table 2. Selected Average Torsion Angles (in deg) of Structures with (**a**) Ar(*i*)−HN(*i*), (**b**) Ar(*i*)−HN(*i*+1), (**c**) Ar(*i*)−HN(*i*+2), and (**d**) Ar(*i*)−HN(*i*+1,*i*+2) Interactions in the Clusters from the Trajectories

clusters	tr[a]	χ^1_{Phe1}[b]	ϕ_{Phe1}	frequency (%)
			(**a**) Ar(*i*)−HN(*i*)	
cl1_t_i	**3**	g−	−62.99 ± 19.11	9.0
cl1_t_i	**4**	g−	−62.77 ± 20.25	6.8
cl2_t_i	**4**	g+	−68.41 ± 14.69	22.7
cl3_t_i	**3**	g+	−115.33 ± 24.94	91.0
cl3_t_i	**4**	g+	−39.73 ± 8.29	1.1
clusters	tr[a]	χ^1_{Phe1}[b]	ψ_{Phe1}	frequency (%)
			(**b**) Ar(*i*)−HN(*i*+1)	
cl1_t_i+1	**1**	t	151.57 ± 8.42	47.6
cl1_t_i+1	**2**	t	148.58 ± 9.92	49.3
cl1_t_i+1	**3**	t	152.67 ± 9.01	42.2
cl1_t_i+1	**4**	t	151.24 ± 11.34	48.5
cl2_t_i+1	**1**	t	128.14 ± 10.95	52.4
cl2_t_i+1	**2**	t	129.29 ± 9.36	50.7
cl2_t_i+1	**3**	t	131.66 ± 8.53	50.3
cl2_t_i+1	**4**	t	130.61 ± 7.16	50.4
cl3_t_i+1	**3**	g+	−42.02 ± 14.60	7.5
cl4_t_i+1	**4**	g+	−39.73 ± 8.29	1.1

clusters	tr[a]	χ^1_{Phe1}[b]	ψ_{Phe1}	ϕ_{Gly2}	ψ_{Gly2}	frequency (%)
				(**c**) Ar(*i*)−HN(*i*+2)		
cl1_t_i+2	**1**	t	−62.25 ± 17.39	91.20 ± 20.37	−69.36 ± 42.49	3.4
cl1_t_i+2	**2**	t	−66.24 ± 18.56	84.26 ± 12.85	−62.67 ± 17.42	6.4
cl1_t_i+2	**4**	t	−67.47 ± 15.92	97.93 ± 24.82	−69.62 ± 22.82	14.1
cl2_t_i+2	**1**	t	132.32 ± 16.47	−76.15 ± 27.35	−73.10 ± 22.72	72.4
cl2_t_i+2	**2**	t	133.83 ± 16.66	−83.62 ± 28.25	−72.02 ± 22.91	76.2
cl2_t_i+2	**3**	t	136.66 ± 14.84	−78.72 ± 29.38	−70.08 ± 19.66	65.1
cl2_t_i+2	**4**	t	134.41 ± 20.14	−82.99 ± 28.07	−73.80 ± 21.60	50.2
cl3_t_i+2	**1**	t	123.21 ± 18.37	−134.98 ± 31.72	71.70 ± 25.09	24.2
cl3_t_i+2	**2**	t	129.08 ± 16.80	−135.93 ± 33.08	78.46 ± 23.46	17.4
cl3_t_i+2	**3**	t	133.99 ± 23.38	−132.29 ± 35.54	70.23 ± 29.94	24.6
cl3_t_i+2	**4**	t	123.75 ± 16.19	−138.50 ± 24.34	69.99 ± 25.23	23.8
cl4_t_i+2	**3**	g+	−42.91 ± 12.20	99.66 ± 26.19	73.92 ± 24.63	10.3
cl4_t_i+2	**4**	t	−75.26 ± 14.57	43.67 ± 18.34	56.86 ± 22.10	11.9
				(**d**) Ar(*i*)−HN(*i*+1,*i*+2)		
cl1_t_i+1,i+2	**1**	t	139.36 ± 13.75	−85.49 ± 27.41	−77.14 ± 18.01	82.5
cl1_t_i+1,i+2	**2**	t	137.51 ± 30.17	−92.69 ± 29.32	−75.59 ± 21.59	83.1
cl1_t_i+1,i+2	**3**	t	142.88 ± 12.98	−88.67 ± 30.24	−72.22 ± 20.86	69.7
cl1_t_i+1,i+2	**4**	t	144.06 ± 11.44	−93.52 ± 25.90	−79.99 ± 20.22	75.9
cl2_t_i+1,i+2	**1**	t	133.89 ± 14.71	−156.42 ± 25.51	65.80 ± 19.30	17.5
cl2_t_i+1,i+2	**2**	t	138.38 ± 10.79	−155.91 ± 23.94	82.07 ± 23.41	16.9
cl2_t_i+1,i+2	**3**	t	139.99 ± 6.05	−152.54 ± 27.11	73.09 ± 21.53	13.8
cl2_t_i+1,i+2	**4**	t	135.75 ± 11.61	−158.73 ± 16.91	71.19 ± 22.11	24.1
cl3_t_i+1,i+2	**3**	g+	−45.48 ± 9.89	101.43 ± 26.71	76.37 ± 23.72	16.5

[a] Trajectory. [b] Orientation of the phenyl side chain: g−, gauche−; g+, gauche+; t, trans.

to those in cl1_t_i+1,i+2. Structures in cl3_d_i+1,i+2 were similar to those in cl2_t_i+1,i+2.

Effect of the Ar−HN Interactions on the Solvation of the Backbone Amide. The averaged data from structures in each cluster (Table 6) was compared to averaged data from structures without Ar−HN interactions (Table 5). A linear correlation, with 0.92 correlation coefficient, between corresponding NHB and $SASA_{HN}$ of each cluster of the trajectories (Table 6) was observed (Figure 3.). Ar(*i*)−HN(*i*) interactions and cl3_t_i+1 were not sampled statistically significantly, and so, they were not included in this analysis. The average NHB between a backbone amide of FGG and solvent water molecules was 0.43, 0.15, and 0.23 higher for the Phe1, Gly2, and Gly3 amide groups, respectively, when the amide was not involved in an Ar−HN interaction than when it was. The average solvent access surface area of Phe1 amide, Gly2 amide, Gly3 amide, and the Phe1 phenyl side chain was 6.49, 2.18, 5.71, and 6.64 Å^2 lower, respectively, when the backbone amide was interacting with the aromatic ring than when it was not. The difference in $SASA_{Phe}$ (6.0%) was almost negligible, while the differences in $SASA_{HN}$ of Phe1 (59.7%), Gly2 (19.6%), and Gly3 (43.4%) were distinct. The average values of $SASA_{Total}$ of FGG with and without Ar−HN were statistically identical.

Table 6 lists the geometrical features of the Ar−HN interactions. In all clusters, the parallel Ar−HN interactions were predominant. They were least frequent in Ar(*i*)−HN(*i*+2) interactions, while the frequency of inverse perpendicular interactions was correspondingly higher. When the Ar−HN interactions were in parallel geometry, the backbone amide generally formed more hydrogen bonds with the water molecules than when the Ar−HN interactions were in either perpendicular or inverse perpendicular geometry. Furthermore, at high NHB values, the Ar−HN interactions were mostly in parallel geometry. As the value of NHB decreased, the frequency of parallel Ar−HN interactions also decreased, while the frequency of the perpendicular and inverse perpendicular interactions increased.

Characterization of the Ar−HN Interaction Energy. The energies of the nonbonded interactions between the aromatic ring of Phe1 and the backbone amide of Phe1, Gly2, and Gly3 in structures derived from the average torsion angles in the

Table 3. Torsion Angles (in deg) of FGG Protein Fragments with Ar−HN Interactions from the PDB

PDB Access Code	Structure[a]	χ^1_{Phe1}	ϕ_{Phe1}	ψ_{Phe1}	ϕ_{Gly2}	ψ_{Gly2}	δ_{ring}	α
Ar(i)-HN(i)								
1gof	TFGGLA -TT---	45.16	-163.34	147.12	75.32	-162.69	-0.55	-46.70
1kve	TFGGSP EE--BS	54.04	-87.98	170.45	51.73	36.25	-0.54	4.43
3eng	QFGGL HHS--	45.27	-136.80	12.17	110.88	-3.13	-0.75	-21.00
Ar(i)-HN(i+1)								
1gof	NDAFGGSPG SS--TT-S	76.10	-126.70	-26.20	-66.10	132.20	-1.18	-0.87
2er7	SSCFGGOQS SEEEESEEE	-168.20	-93.70	137.40	-86.10	158.30	-0.80	7.77
2kau	EVKFGGKV ----STTSS	171.90	-103.50	143.70	146.50	-179.50	-1.02	9.99
2pcd	DPNFGGVGR -TT---EEE	-166.30	-137.43	127.00	-99.90	-50.60	-0.75	1.86
Ar(i)-HN(i+2)								
1arv	AGQFGGGGA TTS----SS	178.90	-70.67	128.70	-101.30	3.30	-2.15	58.65
1cxs	TGTFGGSYG GGGE-----	175.90	-102.22	123.20	-132.50	39.60	-1.99	63.99
1djx	SVJFGGFSS EE-------	-179.20	-57.87	125.10	-91.70	-60.10	-0.74	18.61
2pcd	DPNFGGVGR -TT---EEE	-166.30	-137.43	127.00	-99.90	-50.60	-0.98	1.83

[a] The upper line is the primary structure and the bottom line is its secondary structure of protein fragment: T, turn, S,bend B, β-bridge; E, β-sheet; H, α-helix; G, 3_{10} -helix; -, random coil.

Table 4. Selected Average Torsion Angles (in deg) of Structures with (**a**) Ar(i)−HN(i+1), (**b**) Ar(i)−HN(i+2), and (**c**) Ar(i)−HN(i+1,i+2) Interactions in Clusters from the Protein Database

interactions	χ^1_{Phe1}[a]	ϕ_{Phe1}	frequency (%)
		(**a**) Ar(i)−HN(i+1)	
cl1_d_i+1	t	139.04 ± 14.06	74.7
cl2_d_i+1	g+	−26.84 ± 10.87	25.3

	χ^1_{Phe1}	ψ_{Phe1}	ϕ_{Gly2}	ψ_{Gly2}	frequency (%)
		(**b**) Ar(i)−HN(i+2)			
cl1_d_i+2	t	133.49 ± 7.45	−110.87 ± 10.35	−45.77 ± 13.43	13.7
cl2_d_i+2	t	128.82 ± 23.18	−64.52 ± 13.50	−25.96 ± 15.01	28.3
cl3_d_i+2	t	123.42 ± 15.45	−114.20 ± 17.39	−16.00 ± 17.65	31.7
cl4_d_i+2	t	132.25 ± 17.38	−124.04 ± 22.46	104.21 ± 18.14	23.6
cl5_d_i+2	t and g+[b]	−29.29 ± 37.99	78.48 ± 22.74	−1.76 ± 31.80	2.7
		(**c**) Ar(i)−HN(i+1,i+2)			
cl1_d_i+1,i+2	t	135.67 ± 6.68	−115.37 ± 10.41	−48.41 ± 9.713	25.8
cl2_d_i+1,i+2	t	137.34 ± 10.85	−65.24 ± 15.05	−35.50 ± 16.60	20.4
cl3_d_i+1,i+2	t	141.16± 7.57	−141.28 ± 14.48	119.72 ± 18.28	23.7
cl4_d_i+1,i+2	t	131.69 ± 10.37	−126.83 ± 14.47	14.27 ± 13.39	30.1

[a] Orientation of the phenyl side chain: g−, gauche−; g+, gauche+; t, trans. [b] 38% gauche+, 62% trans.

Table 5. Number of Hydrogen Bonds between Solvent Water Molecules and Each Backbone Amide (NHB) and the SASA (in Å^2) of Each Backbone Amide ($SASA_{HN}$) and of the Phe Side Chain ($SASA_{Phe}$) in FGG Structures with No Ar−HN Interaction

	Phe1			Gly2		Gly3	
tr[a]	NBH	$SASA_{NH}$	$SASA_{Phe}$	NHB	$SASA_{HN}$	NHB	$SASA_{HN}$
1	1.06 ± 0.52	11.32 ± 2.61	110.03 ± 11.21	1.03 ± 0.49	10.29 ± 3.03	1.03 ± 0.55	13.38 ± 4.06
2	1.05 ± 0.53	11.56 ± 2.84	109.47 ± 11.43	1.05 ± 0.51	10.83 ± 3.10	1.07 ± 0.56	13.39 ± 3.81
3	1.03 ± 0.56	10.17 ± 3.33	112.74 ± 10.75	1.03 ± 0.54	11.07 ± 3.60	0.97 ± 0.57	12.95 ± 4.00
4	1.01 ± 0.54	10.50 ± 3.30	111.61 ± 10.83	1.01 ± 0.53	12.71 ± 3.29	0.97 ± 0.58	12.71 ± 3.29

[a] Trajectory.

clusters from the trajectories are summarized in Table 7. When an Ar−HN interaction occurred, interaction energies ranged from −1.98 to −9.24 kJ mol^{-1}. When Ar(i)−HN(i+1) and Ar(i)−HN(i+2) interactions occurred simultaneously, the sum of the nonbonded interaction energies was less than −8 kJ mol^{-1}. Either the Coulomb or the Lennard-Jones energies were predominant energies of the Ar−HN interactions (Table 7.)

Discussion

Conformations of clusters cl1_t_i+1, cl2_t_i+1, cl2_t_i+2, cl3_t_i+2, cl1_t_i+1,i+2, and cl2_t_i+1,i+2 sampled from 300 to 3300 ps in trajectory **1**, belong to the folded state of the FGG peptide on the basis of their low R_g values. During this period, Ar−HN interactions did not occur in every sampled conforma-

Table 6. Average Number of Hydrogen Bonds between Solvent Water Molecules and the Backbone Amide (NHB) and the Average SASA (in Å^2) of the Backbone Amide ($SASA_{HN}$) and of the Phe Side Chain ($SASA_{Phe}$) Involved in (**a**) Ar(*i*)−HN(*i*), (**b**) Ar(*i*)−HN(*i*+1), (**c**) Ar(*i*)−HN(*i*+2), and (**d**) Ar(*i*)−HN(*i*+1,*i*+2) Interactions in FGG Structures in Clusters from the Trajectories (tr)

tr	$P_{\|\|}$[a]	$P_{\perp}$[b]	P_i[c]	NHB	$H_{b\|\|}$[d]	$H_{b\perp}$[e]	H_{bi}[f]	$SASA_{HN}$	$SASA_{Phe}$
					(**a**) Ar(*i*)−HN(*i*)				
					cl1_t_i				
3	100	0	0	0.83 ± 0.37	0.83	0.00	0.00	5.37 ± 3.35	125.69 ± 3.18
4	100	0	0	0.74 ± 0.47	0.33	0.00	0.00	4.40 ± 1.46	126.15 ± 6.25
					cl2_t_i				
4	100	0	0	0.55 ± 0.59	0.55	-	-	3.83 ± 1.36	118.09 ± 11.32
					cl3_t_i				
3	90	2	8	0.83 ± 0.61	1.00	0.80	1.20	4.35 ± 2.17	114.39 ± 6.91
4	94	6	0	0.74 ± 0.47	0.72	1.00	-	4.51 ± 1.86	106.41 ± 7.71
					(**b**) Ar(*i*)−HN(*i*+1)				
					cl1_t_i+1				
1	78	14	8	0.81 ± 0.54	0.81	0.78	0.80	8.15 ± 2.67	103.66 ± 6.24
2	84	3	13	0.80 ± 0.57	0.86	0.27	0.50	8.11 ± 2.70	102.50 ± 6.31
3	84	11	5	0.75 ± 0.55	0.83	0.53	0.28	8.23 ± 2.66	105.92 ± 7.68
4	91	0	1	0.79 ± 0.48	0.82	-	0.47	8.44 ± 2.71	104.76 ± 6.88
					cl2t_i+1				
1	78	15	7	0.96 ± 0.44	0.93	0.94	1.00	9.65 ± 2.44	105.15 ± 6.63
2	97	1	2	0.97 ± 0.49	0.98	0.66	0.75	9.70 ± 2.63	104.52 ± 7.17
3	97	2	1	0.92 ± 0.51	0.93	0.83	0.50	9.90 ± 2.51	105.91 ± 7.34
4	96	3	1	0.93 ± 0.52	0.93	1.00	0.83	9.71 ± 0.52	105.56 ± 7.95
					cl3_t_i+1				
3	85	15	0	0.95 ± 0.34	1.00	0.67	-	4.53 ± 2.21	112.11 ± 6.40
4	100	0	0	1.00 ± 0.00	1.00	-	-	3.46 ± 1.28	108.49 ± 1.70
					(**c**) Ar(*i*)−HN(*i*+2)				
					cl1_t_i+2				
1	44	0	56	0.44 ± 0.49	0.25	-	0.60	7.21 ± 2.31	104.46 ± 6.44
2	86	5	9	0.19 ± 0.39	0.17	1.00	0.00	4.21 ± 2.44	98.41 ± 6.51
4	88	0	12	0.53 ± 0.66	0.57	-	0.25	5.72 ± 3.82	101.42 ± 7.10
					cl2t_i+2				
1	53	11	36	0.90 ± 0.45	0.93	0.66	0.91	8.28 ± 2.97	102.14 ± 6.39
2	90	1	9	0.92 ± 0.50	0.98	1.00	0.35	8.38 ± 3.30	99.93 ± 5.53
3	82	7	11	0.87 ± 0.53	0.93	0.67	0.57	8.31 ± 3.15	101.97 ± 5.48
4	88	3	9	0.80 ± 0.48	0.84	0.33	0.50	7.92 ± 2.77	99.49 ± 5.76
					cl3_i+2				
1	59	10	31	0.38 ± 0.48	0.34	0.67	0.35	5.40 ± 3.77	99.43 ± 4.74
2	39	5	56	0.67 ± 0.57	0.91	1.00	0.46	5.83 ± 3.58	100.61 ± 5.25
3	42	8	50	0.52 ± 0.50	0.65	0.50	0.41	5.53 ± 2.73	97.92 ± 4.32
4	26	0	74	0.54 ± 0.53	0.86	-	0.43	6.19 ± 3.57	97.11 ± 5.02
					cl4_i+2				
3	90	10	0	0.85 ± 0.48	0.94	0.00	-	8.45 ± 3.98	108.79 ± 6.17
					cl5i+2				
4	7	0	93	0.37 ± 0.48	0.50	-	0.36	4.66 ± 2.99	97.54 ± 5.29
					(**d**) Ar(*i*)−HN(*i*+1,*i*+2)				
					cl1_t_i+1,i+2 − Ar(*i*)−HN(*i*+1)				
1	79	16	5	0.95 ± 0.48	0.92	1.00	1.20	8.28 ± 2.10	102.61 ± 5.95
2	92	0	8	0.93 ± 0.51	0.96	-	0.63	8.12 ± 2.39	100.66 ± 5.08
3	86	4	10	0.86 ± 0.53	0.89	1	0.5	8.04 ± 2.33	102.28 ± 5.56
4	93	0	7	0.77 ± 0.56	0.79	-	0.50	8.66 ± 2.20	99.38 ± 5.51
					cl1_t_i+1, i+2 − Ar(*i*)−HN(*i*+2)				
1	54	9	37	0.98 ± 0.37	1.01	0.77	0.97	8.34 ± 2.63	
2	92	0	8	0.94 ± 0.51	1.00	-	0.30	8.63 ± 3.18	
3	87	7	6	0.86 ± 0.55	0.90	0.60	0.60	8.43 ± 3.15	
4	92	2	6	0.82 ± 0.43	0.84	0.00	0.75	8.62 ± 2.43	
					cl2_t_i+1,i+2 − Ar(*i*)−HN(*i*+1)				
1	70	30	0	0.90 ± 0.43	1.00	0.83	-	7.40 ± 2.49	99.82 ± 4.83
2	86	0	14	0.86 ± 0.44	0.96	-	0.25	6.94 ± 1.53	102.00 ± 4.07
3	93	0	7	0.93 ± 0.25	0.93	-	1.00	7.17 ± 2.00	99.65 ± 4.02
4	94	0	6	0.84 ± 0.49	0.88	-	-	7.69 ± 2.57	99.68 ± 4.24
					cl2_t_i+1, i+2 − Ar(*i*)−HN(*i*+2)				
1	62	10	28	0.29 ± 0.46	0.31	0.50	0.17	4.17 ± 2.19	
2	32	4	64	0.75 ± 0.57	1.00	1.00	0.61	6.38 ± 3.72	
3	33	13	54	0.66 ± 0.47	0.80	0.50	0.63	5.41 ± 2.38	
4	21	0	79	0.42 ± 0.49	1.00	-	0.27	5.25 ± 3.60	
					cl3t_i+1,i+2 − Ar(*i*)−HN(*i*+1)				
3	83	17	0	0.89 ± 0.31	0.93	0.67	-	3.55 ± 1.92	109.11 ± 6.23
					cl3t_i+1, i+2 − Ar(*i*)−HN(*i*+2)				
3	94	6	0	0.89 ± 0.45	0.94	-	-	8.91 ± 3.77	

Percentage of [a]parallel, [b]perpendicular, and [c]inverse perpendicular Ar−HN interaction in structures in the cluster. [d]The number of hydrogen bonds between the solvent water molecule and the backbone amide when the amide is involved in a parallel, [e]perpendicular, and [f]inverse perpendicular Ar−HN interaction.

Table 7. Nonbonded Energies (in kJ mol^{-1}) between the Phenyl Group and the Backbone Amides

	Phe(i)−NH(i)				Phe(i)−NH(i+1)			Phe(i)−NH(i+2)		
cluster	E_c[a]	E_{LJ}[b]	E_{1-4LJ}[c]	$\sum E_{NB}$[d]	E_c	E_{LJ}	$\sum E_{NB}$	E_c	E_{LJ}	$\sum E_{NB}$
cl1_t_i	−2.48	−2.42	0.77	−4.13	0.35	−0.62	−0.27	−0.04	−0.08	−0.12
cl2_t_i	−4.5	−2.23	4.46	−2.27	−0.23	−0.82	−1.05	−0.09	−0.25	−0.37
cl3_t_i	−2.88	−2.04	0.6	−4.32	−1.11	−1.47	−1.58	−0.85	−0.93	−1.42
cl1_t_i+1	0.5	−0.98	−0.53	−1.01	−2.69	−2.69	−5.38	0.25	−1.28	−1.03
cl2_t_i+1	0.36	−0.97	−0.52	−1.13	−2.14	−2.65	−4.79	−0.48	−0.65	−1.13
cl3_t_i+1	−0.29	−0.99	−0.07	−1.35	−0.16	−3.21	−3.37	1.41	−2.68	−1.27
cl1_t_i+2	0.96	−0.98	−0.5	−0.52	1.66	−2.52	−0.86	−1.42	−1.18	−2.6
cl2_t_i+2	0.59	−0.93	−0.51	−0.85	−0.60	−3.14	−3.74	0.87	−2.97	−2.1
cl3_t_i+2	0.04	−0.91	−0.52	−1.39	0.60	−2.58	−1.98	−5.85	−3.39	−9.24
cl4_t_i+2	1.62	−1.68	−0.13	−0.19	−3.54	−2.57	−6.11	−1.92	−2.86	−4.78
cl5_t_i+2	0.31	−0.92	−0.52	−1.13	2.18	−2.29	−0.11	−2.59	−3.67	−6.26
cl1_t_i+1+2	0.31	−0.93	−0.52	−1.14	−1.62	−2.92	−4.54	−0.62	−2.85	−3.47
cl2_t_i+1+2	0.43	−0.92	−0.51	−1	−1.90	−2.67	−4.57	−3.05	−3.03	−6.08
cl3_t_i+1+2	1.62	−1.68	−0.13	−0.19	−3.54	−2.57	−6.11	−1.92	−2.86	−4.78

[a] Coulomb interaction energy. [b] Lennard-Jones interaction energy. [c] 1−4 Lennard-Jones interaction energy. [d] Sum of nonbonded energies.

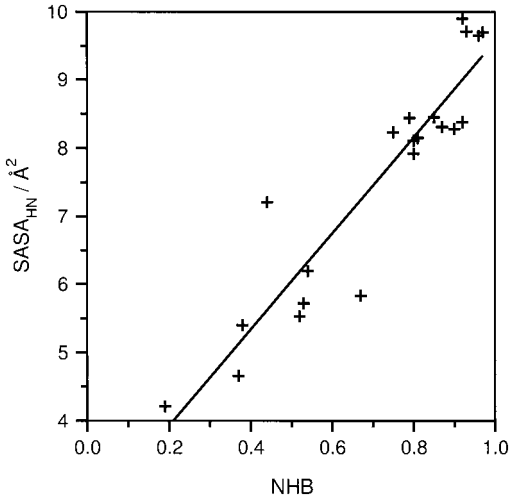

Figure 3. Correlation between number of hydrogen bonds between each backbone amide and the solvent water molecules (NHB) and $SASA_{HN}$ in clusters from trajectories.

tion even though the conformation of FGG was moderately stable. The net effect of the torsion angle fluctuation resulted in constant movement of the aromatic ring relative to the backbone amide, which caused the amide hydrogen δ_{ring} to be larger than −0.5 ppm. The averaged δ_{ring} of the Gly2 and Gly3 amide hydrogen for the 300−3300-ps period showed that the aromatic ring was always close to the amides, particularly the Gly2 amide. The formation of Ar(i)−HN(i+2) interactions depended on the ψ_{Gly2} torsion angle, which intensely fluctuated during this period, having optimal value for the Ar(i)−HN(i+2) interaction occasionally. Therefore, the stability of Ar−HN interactions in peptides should be regarded as a function of their conformational flexibility.

The conformation regions of the i and i+1 residues from the trajectories were similar to those of Tyr-Thr-Gly-Pro identified by Worth and associates.[13] The conformation of the i and i+1 residues in cluster cl4_t_i+2 were similar to those of clusters ArHN1 and ArHN2 in ref. 13. Conformations in cluster cl2_t_i+2 were similar to those in clusters ArHN3, ArHN5, ArHN7, and ArHN8, and conformations in cluster cl3_t_i+2 were similar to those of clusters ArHN4, ArHN6, and ArHN9. While cluster cl2_t_i+2 contained the highest number of conformations, clusters ArHN1 and ArHN2 had the highest number of conformations in ref. 13. This difference suggests that the side chain of residue i+1 strongly influences the conformation of residues i and i+1.

The environment of Ar−HN interaction, in particular the availability of hydrogen bond acceptors, affects the conformation of the resultant local structure.[4] The finding that FGG fragments were located mostly at the surface of the proteins suggests that the solvent accessibility of these fragments could be similar to that of the FGG peptide. This could be a reason for such good agreement between the conformations of FGG fragments from the database and the averaged conformations of the clusters from the trajectories. Also, the structure of FGG protein fragments with Ar(i)−HN(i+2) was random meander. In such fragments, an Ar(i)−HN(i+2) interaction may be the predominant structural influence on the formation or stabilization of the polypeptide conformation, because no other forces, such as hydrogen bonds, are present.

The orientation of the side chain aromatic ring in position i and the backbone amide in position i+1 depends only on torsion angles χ_i^1 and ψ_{i+1}. The calculated averages of these two torsion angles in clusters with Ar(i)−HN(i+1) interactions from the trajectories were similar to those in the PDB. Therefore, the side chain in position i+1 may not influence the formation of Ar(i)−HN(i+1) interactions. This view is also supported by the random incidence of residues in position i+1 in protein fragments with Ar(i)−HN(i+1) interactions.[6]

Comparison of the clusters containing structures with Ar(i)−HN(i+2) interactions from the database search of Worth and Wade [4] to the clusters of this study revealed some similarities, as CL4 was similar to cl1_d_i+2, CL3 to cl2_d_i+2, CL2 to cl3_d_i+2, and CL1 to cl4_d_i+2. Average torsion angle values of structures with Ar−HN(i+2) interactions in clusters from the trajectories were not always the same as in the clusters from the PDB (Table 4), possibly because the side chains at position i+2 affects the orientation of the aromatic side chain at position i. The values could also be affected by the local environment (solvent accessibility and hydrogen bond acceptors) of the protein fragments, depending on their location in the protein.

Clusters from the simulations and clusters from the database search revealed a similar trend in the conformation of structures with Ar(i)−HN(i+1,i+2) interactions. Over 90% of the sampled conformations from the trajectories with Ar(i)−HN(i+2) interactions (clusters cl2_t_i+2, cl3_t_i+2, and cl4_t_i+2) met the conformational requirements for Ar(i)−HN(i+1,i+2) interactions (clusters cl1_t_i+1,i+2, cl2_t_i+1,i+2, and cl2_t_i+1,i+2), while over 97% of the conformations of protein fragments with Ar(i)−HN(i+2) interactions (clusters cl1_d_i+2, cl2_d_i+2, cl3_d_i+2, and cl4_d_i+2) met the conformational requirements for Ar(i)−HN(i+1,i+2) interactions (clusters cl1_d_i+1,i+2,

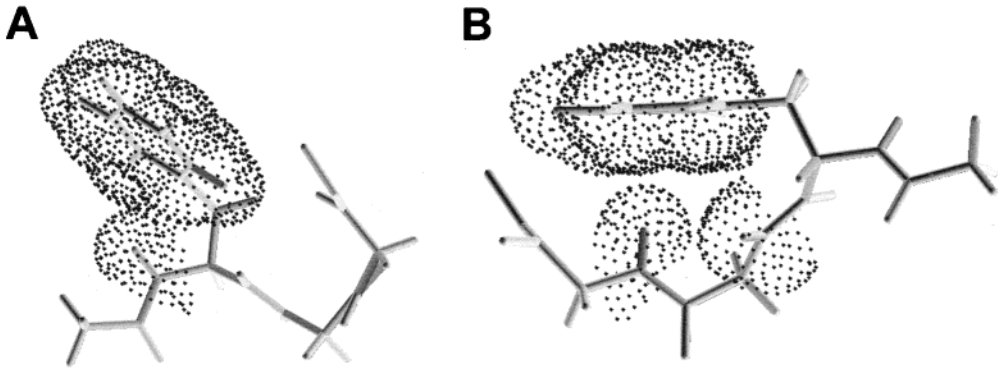

Figure 4. van der Waals surface of the phenyl side chain and backbone amide groups in the average structure of (A) cl2_t_i and (B) cl2_t_i+1, i+2.

cl2_d_i+1,i+2, cl3_d_i+1,i+2, and cl4_d_i+1,i+2). Furthermore, the conformation of the Zaa-Xaa and Zaa-Xaa-Yaa (Zaa = Phe, Tyr, or Trp; Xaa or Yaa = any residue) protein fragments containing Ar−HN interactions can be represented by the clusters identified in MD simulations. Thus, Ar(*i*)−HN(*i*+1) interactions in polypeptides can be characterized by two pairs of $\chi_{Zaa}{}^1$, ψ_{Zaa} torsion angles and Ar(*i*)−HN(*i*+2) interactions by three different sets of $\chi_{Zaa}{}^1$, ψ_{Zaa}, ϕ_{Xaa}, and ψ_{Xaa} torsion angles in which two sets also include the conformations of Ar(i)−(*i*+1 and *i*+2) interactions.

The amide backbone involved in Ar−HN interactions is less accessible to the water molecules because its solvation is hindered by the side-chain aromatic ring (Figure 4.). This phenomenon is reflected by the loss of the $SASA_{HN}$, which should be the predominant determinant of the increase of the free energy of solvation. Therefore, the free energy of the Ar−HN interaction must compensate the loss of the free energy of solvation in thermodynamically stable conformations.

It can be regarded that the difference in the strength of the ideal perpendicular and parallel Ar−HN interaction in a vacuum is negligible.[1] If external hydrogen bond acceptors are available to the backbone amide in a solution, Ar−HN interactions should be in parallel geometry.[2,4] This is supported by the present observation that the number of hydrogen bonds between the backbone amide and the solvent water molecules is higher in conformations with parallel than with perpendicular Ar−HN interactions in the trajectories.

The force field dependence of the simulation of structures of peptides with Ar−HN interactions is clear from the results of our previous study[14] and others.[11,13] Therefore, the validity of conclusions of any molecular mechanics study of such problems could be questionable. Nevertheless, van der Spoel and colleagues found that, of several combinations of force field and explicit water models, the SPC/E water model together with the revised GROMOS-87 force field gives the closest agreement with NMR experimental data.[11] The present studies based on the similarities of the conformations in the clusters from the trajectories and from the database suggest that this combination of force field and explicit water model can be used to simulate Ar−HN interactions. Since most clusters were sampled in each trajectory, 4-ns simulation time was enough of to sample all major conformational states, even when starting from an extended conformation (trajectory **4**). The course of dynamics of the peptide in each trajectory, however, did depended on the starting structures.

Conclusion

Conformation of the residues involved in Ar(*i*)−HN(*i*+1), −HN(*i*+2), and −HN(*i*+1,*i*+2) interactions was characterized by molecular dynamics simulations and by PDB search. Selected average torsion angles of FGG structures in clusters from the trajcectories were similar to those in protein fragments. The conformational requirements for Ar(*i*)−HN(*i*+2) interactions were almost always the same as those for the Ar(*i*)−HN(*i*+1) interactions, so that two conformational regions in which the formation of either Ar(*i*)−HN(*i*+1), Ar(*i*)−HN(*i*+2) or Ar−HN(*i*+1+2) interactions were possible.

The decrease in accessibility of water to the backbone, due to Ar−HN interactions, caused the partial burial of the backbone. The extent of burial determined the geometry of the Ar−HN interaction. Thus, the predominance of the parallel over the perpendicular geometry is determined by solvation thermodynamics.

Ar−HN interactions were found in folded structures of FGG which were stable throughout the simulations. Thus, the attractive force between the backbone amide and the side-chain aromatic ring is strong enough to outweigh any free energy losses due to entropic costs of backbone and side-chain stabilization and to solvation.

Abbreviations

NME, *N*-methyl amide; Ac, acetyl; cl, cluster; _t, trajectory; _d, database; _i, Ar(*i*)−HN(*i*) interaction; _i+1, Ar(*i*)−HN(*i*+1) interaction; _i+2, Ar(*i*)−HN(*i*+2) interaction; _i+1,i+2, Ar-(*i*)−HN(*i*+1,*i*+2) interaction; FGG, Ac-Phe-Gly-Gly-*N*-methyl amide; NHB, number of hydrogen bonds between each backbone amide and the solvent water molecules.

Acknowledgment. This work was supported by grants from the National Science Foundation (EPS-9720643), and the Carpenter Chair in Biochemistry, Creighton University.

JA011245U