J|A|C|S
ARTICLES
Published on Web 05/01/2002

Stacking and T-shape Competition in Aromatic−Aromatic Amino Acid Interactions

Riccardo Chelli, Francesco Luigi Gervasio, Piero Procacci,* and Vincenzo Schettino

Contribution from the Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, and European Laboratory for Nonlinear Spectroscopy (LENS), Via Nello Carrara 1, 50019 Sesto Fiorentino, Italy

Received September 12, 2001

Abstract: The potential of mean force of interacting aromatic amino acids is calculated using molecular dynamics simulations. The free energy surface is determined in order to study stacking and T-shape competition for phenylalanine−phenylalanine (Phe−Phe), phenylalanine−tyrosine (Phe−Tyr), and tyrosine−tyrosine (Tyr−Tyr) complexes in vacuo, water, carbon tetrachloride, and methanol. Stacked structures are favored in all solvents with the exception of the Tyr−Tyr complex in carbon tetrachloride, where T-shaped structures are also important. The effect of anchoring the two α-carbons (C_α) at selected distances is investigated. We find that short and large $C_\alpha-C_\alpha$ distances favor stacked and T-shaped structures, respectively. We analyze a set of 2396 protein structures resolved experimentally. Comparison of theoretical free energies for the complexes to the experimental analogue shows that Tyr−Tyr interaction occurs mainly at the protein surface, while Phe−Tyr and Phe−Phe interactions are more frequent in the hydrophobic protein core. This is confirmed by the Voronoi polyhedron analysis on the database protein structures. As found from the free energy calculation, analysis of the protein database has shown that proximal and distal interacting aromatic residues are predominantly stacked and T-shaped, respectively.

1. Introduction

Interactions between protein side chains (amino acid residues) are determining factors in the protein-folding processes,[1−5] and therefore, their understanding is a necessary step to gain new insights into the complex tertiary (and quaternary) structure of proteins. Among the various amino acid residues, the aromatics play a role of primary importance for determining the protein structure, mainly because of the possibility of $\pi-\pi$ interaction[2,6] or H-bonding between their polar groups. In fact, Burley and Petsko found[1] that ∼60% of aromatic side chains of a set of 34 surveyed proteins is involved in aromatic−aromatic interactions, and 80% of these interactions contributes to stabilize the tertiary structure by linking different elements of secondary structure. Quite recently, other statistical studies, principally aimed at probing differences in the free energy of pairs of aromatic residues, appeared in the literature.[3,4,7,8] Overall, these studies suggest the existence of a competition between stacked and T-shaped or crossed complexes and that this competition is strongly affected by the polarity of the chemical environment and by the possibility of forming H-bonds between the residues and the solvent. If on one hand, this statistical information can be useful to discover correlations and recurrences concerning aromatic−aromatic pair arrangements in proteins, on the other hand, this kind of analysis often lacks the necessary chemical−physical background to explain the observed recurrences or correlations.

Indeed, a systematic study on the specific interactions between amino acid residues is extremely difficult from the experimental standpoint. From the theoretical standpoint, the use of well-established computational techniques on simple molecular models can give valuable, albeit partial, information on this matter. Several studies have been undertaken in this direction. For example, computer simulations using empirical potentials were used to calculate the potential of mean force (PMF) of pairs of aromatic compounds in pure liquid phases[9] and in various solvents.[9−12] Many efforts were also devoted to investigate the potential energy surface (PES) of aromatic pairs using empirical and ab initio techniques.[10,13−16] Recently, we investigated the structural properties of a pair of aromatic amino

* To whom corresondence should be addressed. E-mail: procacci@chim.unifi.it.

(1) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23−28.
(2) Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837−846.
(3) Brocchieri, L.; Karlin, S. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9297−9301.
(4) Karlin, S.; Zuker, M.; Brocchieri, L. *J. Mol. Biol.* **1994**, *239*, 227−248.
(5) Mitchell, J. B. O.; Laskowski, R. A.; Thornton, J. M. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 370−380.
(6) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525−5534.
(7) Mitchell, J. B. O.; Nandi, C. L.; McDonald, J. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315−331.
(8) McGaughey, G. B.; Gagné, M.; Rappé, A. K. *J. Biol. Chem.* **1998**, *273*, 15458−15463.
(9) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768−4774.
(10) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 11217−11224.
(11) Stofer, E.; Chipot, C.; Lavery, R. *J. Am. Chem. Soc.* **1999**, *121*, 9503−9508.
(12) Gervasio, F. L.; Chelli, R.; Marchi, M.; Procacci, P.; Schettino, V. *J. Phys. Chem. B* **2001**, *105*, 7835−7846.
(13) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *14*, 117−135.

acids[12] (tryptophan, histidine) in various solvents (water, methanol, dimethyl sulfoxide, carbon tetrachloride) by using molecular dynamics simulations. By means of a Boltzmann sampling, we calculated the free energy surface of the Trp−His complex on the bidimensional space defined by the angle between the aromatic rings and by the distance between the centroids of the aromatic rings. Several remarkable features were observed: (i) T-shaped complexes, stabilized by strong H-bonds, are favored in nonpolar solvents such as CCl_4, while they are destabilized in solvents that can form H-bonds with the residues; (ii) stabilization of the stacked structures is observed when a distance constraint is imposed between the two C_α's of the aromatic residues and this $C_\alpha-C_\alpha$ distance is below a certain threshold; (iii) in water only stacked arrangements are stable irrespective of the presence of $C_\alpha-C_\alpha$ constraint; (iv) amphiphilic solvents, such as dimethyl sulfoxide and methanol (MeOH), destabilize both the stacked and the T-shaped complexes, thereby inhibiting aromatic−aromatic interactions. The present work is intended as an extension of the previous study[12] to other aromatic amino acid pairs. In particular, this study is devoted to further assessing the general validity of those findings,[12] by calculating, using well-established molecular dynamics (MD) techniques, the PMF for the tyrosine−tyrosine (Tyr−Tyr), tyrosine−phenylalanine (Phe−Tyr), and phenylalanine−phenylalanine (Phe−Phe) side-chain pairs in various solvents (water, CCl_4, MeOH). As was done for the Trp−His pair,[12] the models of the Tyr and Phe side chains are made from the corresponding amino acids by removing both the amino and carboxylic groups. These simplified models are far from being representative of the complex interplay of interactions occurring in a protein. However by taking into account the *protein side-chain interactions alone*, without interference from any other factor (e.g., polypeptidic chain or vicinal residues), we may hope to isolate the specific contribution of these side-chain interactions in a process as complex as the protein folding. The effect of different solvents and of distance constraints between the residues[12] has also been investigated. To further rationalize the solvent effect, calculations in vacuo were performed. For these interacting aromatic compounds, the investigation of the balance of electrostatic (H-bonding) and dispersive forces is important for understanding the competition between stacked and T-shaped arrangements. To this end, a thorough comparative study between the results obtained with the AMBER force field[17] and those obtained with different trial potential models (e.g., with attenuated dispersive forces) was undertaken.

The study of the Tyr−Tyr, Phe−Tyr, and Phe−Phe pairs is of general interest to understand the role of H-bonding in proteins, since Tyr and Phe are similar in structure and differ only by the presence of the hydroxy group that can be involved in H-bonding. In the discussion, the main differences of the present results from those obtained for the Trp−His pair[12] will be highlighted. This comparison is of interest not only for the reasons discussed before but also for studying the behavior of two different kinds of H-bond: a "rigid" kind typical of the Trp−His pair (in the sense that the H-bonded groups are rigidly connected to the body of the molecule) and a more "flexible" one in the Tyr−Tyr pair.

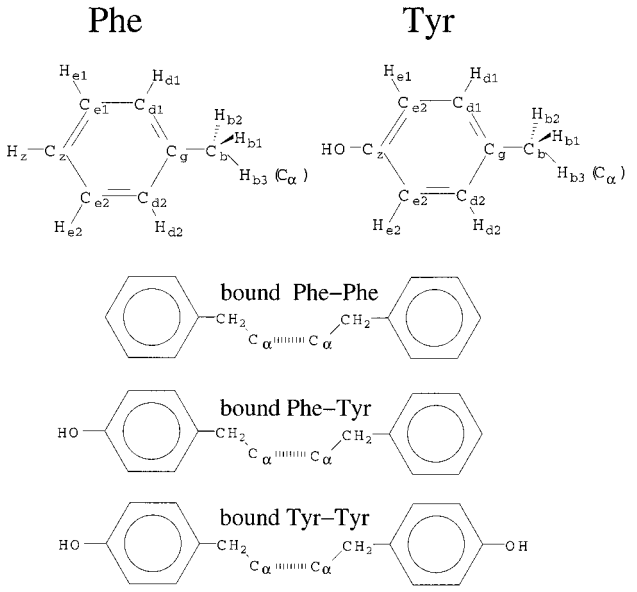

Figure 1. Unbound phenylalanine (Phe) and tyrosine (Tyr) modeled by toluene and *p*-cresol, respectively. The bound Phe−Phe, Phe−Tyr, and Tyr−Tyr pairs are modeled by imposing a harmonic potential between methylic hydrogen (H_{b3}) atoms (also labeled C_α in the picture).

To stress the relevance of this molecular modeling study in the biochemical field, in the final part of the paper, we compare the PMF-derived structural properties of the Tyr−Tyr, Phe−Tyr, and Phe−Phe complexes to those found from the analysis of a large database of experimentally resolved protein structures. For completeness, in this part, we also compare the experimental structures to the results obtained for the Trp−His complex.[12] We found that our simple model is able to explain many aspects of the interactions between aromatic side chains in real proteins, such as the polarity of the chemical environment and the effect of anchoring of the amino acid residues to the protein backbone.

The paper is organized as follows: in section 2, we discuss the potential model adopted for the residues and for the solvents as well as the methodology used in the simulations and for analyzing the data. In section 3, we discuss the structural properties of the bound and unbound complexes on the basis of the results obtained from the PES and from the PMF in vacuo, CCl_4, MeOH, and water. A discussion and interpretation of the Phe−Phe, Phe−Tyr, Tyr−Tyr, and Trp−His interactions, found in experimentally resolved protein structures, is given on the basis of the computational data in section 3. Finally, in section 4, conclusive remarks are presented.

2. Computational Models and Methods

2.1. Potential Model. Phe and Tyr residues are modeled by toluene and *p*-cresol, respectively. The choice of considering toluene as a better prototype than benzene for phenylalanine was addressed recently by Chipot et al.[10] Following Chipot et al.[10] and our own experience,[12] the choice of *p*-cresol as the prototype for tyrosine is thus well justified. In the MD simulations of the bound residues, a methylic hydrogen is substituted with a carbon (the C_α) and an additional harmonic potential between two C_α atoms is added (see Figure 1). In the previous work on the Trp−His pair,[12] the choice of the equilibrium distance of this additional constraining potential was thoroughly explained. To allow a direct comparison between the present calculations and those of ref

(14) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590−5596.
(15) Kratochvil, M.; Engkvist, O.; Sponer, J.; Jungwirth, P.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 6921−6926.
(16) Gervasio, F. L.; Procacci, P.; Cardini, G.; Guarna, A.; Giolitti, A.; Schettino, V. *J. Phys. Chem. B* **2000**, *104*, 1108−1114.
(17) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

Table 1. Potential Model for Unbound and Bound Complexes[a]

Phe			Tyr		
atom	type	charge	atom	type	charge
C_b	ct	−0.330 422	C_b	ct	−0.225 814
H_{b1}, H_{b2}	hc	0.096 465	H_{b1}, H_{b2}	hc	0.071 317
H_{b3} (C_α)	hc (ct)	0.096 465	H_{b3} (C_α)	hc (ct)	0.071 317
C_g	ca.	0.143 625	C_g	ca.	0.142 797
C_{d1}, C_{d2}	ca.	−0.115 342	C_{d1}, C_{d2}	ca.	−0.211 325
H_{d1}, H_{d2}	ha	0.109 987	H_{d1}, H_{d2}	ha	0.146 975
C_{e1}, C_{e2}	ca.	−0.210 331	C_{e1}, C_{e2}	ca.	−0.241 223
H_{e1}, H_{e2}	ha	0.139 573	H_{e1}, H_{e2}	ha	0.152 908
C_z	ca.	−0.056 016	C_z	c	0.377 992
H_z	ha	0.105 644	O	oh	−0.583 439
			H	ho	0.379 843

[a] Atom labels refer to those reported in Figure 1. The intra- and intermolecular potentials are obtained from the atom type (see "type" column) of the AMBER force field.[17] The atomic charges (in fractions of electron) are computed as described in the text. Chemically equivalent atoms are separated by a comma. In the bound complexes, the H_{b3} atom of Phe and Tyr is replaced by the C_α atom of ct type. The force constant of the harmonic potential between the C_α atoms of the bound complexes is 100 kcal mol^{-1} Å^{-2}.

12, we performed MD simulations using the same constraining distances used there, i.e., 6.5 and 3.8 Å. MD simulations of the bound and unbound Phe−Phe, Phe−Tyr, and Tyr−Tyr pairs were carried out in vacuo and in various solvents. Again, to be consistent with ref 12, water, CCl_4, and MeOH were used (dimethyl sulfoxide was omitted) to mimic polar, nonpolar, and amphiphilic environments.

The potential model adopted for the residues and the solvents is based on the AMBER force field.[17] In many instances, it was shown that the AMBER force field is able to reliably model the aromatic−aromatic interactions.[9,10,16] Though electronic effects, such as polarization or charge-transfer phenomena, are totally neglected in the AMBER approach, the good agreement between empirical and ab initio data for the intermolecular energy of aromatic−aromatic complexes[9,10,16,18] indicates that these electronic effects are not crucial for shaping the potential energy surface. In addition, it was also shown that the AMBER force field reproduces very satisfactorily both the dynamical and structural properties of H-bond-forming systems[19,20] (as remarked in the Introduction, modeling of H-bonding is important for the present study). As to solute−solvent interactions, polarization effects were recently accounted for in the calculation of the solvation free energy of amides[21] using MD simulations in combination with the fluctuating charge model.[22−24] The conventional electrostatic energy was found[21] to be by far the largest contribution to the total solvation free energy, while the polarization contribution was small and practically independent of the solute conformation, thereby equally affecting solute conformations with no or little effect in their free energy differences.

For the bound and unbound Tyr and Phe, the interaction potential is reported in Table 1, where the AMBER atomic types and the atomic charges are reported. On the basis of the force field database,[17] all constants referring to intra- and intermolecular potentials can be deduced from the knowledge of the molecular topology and of the atomic force field types. The atomic charges of Phe and Tyr were computed according to the prescriptions given in ref 17, i.e., by using a restrained electrostatic potential fit[25] with the gridded electrostatic potential evaluated ab initio at the MP2/6-31G* level of theory. Ab initio calculations were performed using the Gaussian98 program.[26] The atomic charges are the same in the bound and unbound complexes: in fact, to conserve the electroneutrality on each monomer of the pair, the same charge of a methylic hydrogen atom is assigned to the C_α atoms (see Figure 1 and Table 1).

Table 2. Summary of the Simulation Runs[a]

solvent	*D*	*L* (Phe–Phe)	*L* (Phe–Tyr)	*L* (Tyr–Tyr)
CCl_4	3.8	20.53	20.53	20.53
$N = 50$	6.5	20.54	20.54	20.53
$t = 9.6$ ns	uc	20.51	20.51	20.51
MeOH	3.8	20.43	20.42	20.42
$N = 121$	6.5	20.44	20.42	20.42
$t = 9.6$ ns	uc	20.40	20.40	20.39
water	3.8	20.36	20.36	20.35
$N = 250$	6.5	20.36	20.36	20.35
$t = 9.6$ ns	uc	20.33	20.33	20.32

[a] In the first column the solvent type, the number of solvent molecules (N), and the run length (t) are reported. D (in Å) is the $C_\alpha-C_\alpha$ constrained distance of the bound complex. The mean box side length L (in Å) for all the pairs (Phe−Phe, Phe−Tyr, Tyr−Tyr) are reported. uc refers to the unbound complex simulations. In vacuo simulations were done in the NVT ensemble for 14.4 ns in a cubic box of 30 Å side length.

The potential model for CCl_4 is taken from ref 27. The potential models for water and MeOH are reported in ref 12. Standard mixing rules[17] are used for the residue−residue, residue−solvent, and solvent−solvent Lennard-Jones interactions.

2.2. Computational Details. All simulations, except those in vacuo, were done in the NPT ensemble (constant number of particles, pressure, and temperature) using a Parrinello−Rahman isotropic barostat[28] and a Nosé thermostat[29] with pressure set to 0.103 MPa and temperature to 300 K in a box of ∼20 Å side length. Simulations in vacuo were done in a cubic box of 30 Å side length in the NVT ensemble (constant number of particles, volume, and temperature) at 300 K. For each solution, including the in vacuo case, three simulations were performed, one for the unbound complex and two for the bound complex at different $C_\alpha-C_\alpha$ distances. The sample equilibration procedure is the same for all the MD simulations: from a starting arbitrary configuration, a simulation of 0.4 ns was performed with scaling of the atomic velocities. The simulation sample was then let to relax at the equilibrium density at constant temperature and pressure (constant volume and temperature for the in vacuo simulations) for ∼1.2 ns. The production stages, during which atomic coordinates were saved (every 60 fs), lasted for 9.6 ns for the solution samples and for 14.4 ns for the complex in vacuo. Simulation data are summarized in Table 2. As it can be observed in the table, the volume of the sample is practically independent of the $C_\alpha-C_\alpha$ distance and of the complex type.

Electrostatic interactions are calculated using the smooth particle mesh Ewald method.[30] For integrating the equations of motion the five time-steps r-RESPA algorithm, thoroughly described in ref 28, is used. Only the X−H bonds (with X being any non-hydrogen atom) are held

(18) Gervasio, F. L.; Chelli, R.; Procacci, P.; Schettino, V. *J. Phys. Chem. A* **2002**, *106*, 2945−2948.
(19) Chelli, R.; Procacci, P.; Cardini, G.; Della Valle, R. G.; Califano, S. *Phys. Chem. Chem. Phys.* **1999**, *1*, 871−877.
(20) Chelli, R.; Procacci, P.; Cardini, G.; Califano, S. *Phys. Chem. Chem. Phys.* **1999**, *1*, 879−885.
(21) Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1996**, *118*, 672−679.
(22) Rappé, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358−3363.
(23) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141−6156.
(24) Chelli, R.; Ciabatti, S.; Cardini, G.; Righini, R.; Procacci, P. *J. Chem. Phys.* **1999**, *111*, 4218−4229.
(25) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(26) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.5; Gaussian Inc.: Pittsburgh, PA, 1998.
(27) Fox, T.; Kollman, P. A. *J. Phys. Chem. B* **1998**, *102*, 8070−8079.
(28) Marchi, M.; Procacci, P. *J. Chem. Phys.* **1998**, *109*, 5194−5202.
(29) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511−519.
(30) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

fixed during the simulations, while all other intramolecular degrees of freedom, including stretchings involving non-hydrogen atoms, are explicitly integrated. All calculations were performed using the ORAC program.[31]

In the previous work on the Trp−His pair,[12] where the box side lengths were slightly smaller than those used in the present case and the residues were of comparable or even larger size, we convincingly showed that the periodicity of the solute, implicitly considered in the Ewald treatment of electrostatics, does not affect the results significantly, nor does the relatively small size of the simulation box. In fact, tests made by using a larger simulation box showed[12] that the box side length does not alter sensibly any feature of the PMF for distances 1−2 Å shorter than the threshold distance of half the box side length (i.e., ∼10 Å).

2.3. Potential of Mean Force: Theoretical Background. As previously stated, we are interested in assessing the free energy differences between the T-shaped and the stacked structures for the Phe−Phe, Phe−Tyr, and Tyr−Tyr complexes in different solvents. In this respect, the mutual arrangement of the two residues can be described by the distance R between the centroids of the rings and the angle θ between the normals to the ring planes. We define $\Delta n(R,\theta)$ as the number of system configurations in the element of volume $(R + \Delta R, \theta + \Delta\theta)$. For studying the relative stability of generic T-shaped and stacked structures, configurations at $\pi/2 - \theta$ and $\pi/2 + \theta$ need not be distinguished. Therefore, $\Delta n(R,\theta)$ is defined in the ranges $0 \le \theta \le \pi/2$ and $R > 0$. The distribution function in the R, θ space is given by

$$G(R,\theta) = \Delta n(R,\theta)/\Delta V \tag{1}$$

where $\Delta V = 2\pi R^2 \sin\theta\, \Delta R\, \Delta\theta$. The associated PMF[32] is hence given by

$$W(R,\theta) = -k_B T \ln[G(R,\theta)] \tag{2}$$

where k_B is the Boltzmann constant and T the temperature of the system. The function $W(R,\theta)$ corresponds to the reversible work needed to bring the two molecules from an infinite distance to the configuration (R,θ) in the solvent and in the given thermodynamic conditions. The minimum of the PMF, conventionally taken to be the zero of the PMF, corresponds to the most probable complex structure, i.e., to the state with the minimum free energy. Since only differences in the PMF can be obtained, the following function is actually calculated

$$W(R,\theta) = -k_B T \ln[G(R,\theta)/G_{max}] \tag{3}$$

where G_{max} is the maximum value of the $G(R,\theta)$ function. Hereafter the acronym PMF will refer to the function $W(R,\theta)$ of eq 3.

The PMF was computed using standard NPT (NVT in vacuo) molecular dynamics with conventional Boltzmann sampling. Convergence of the $W(R,\theta)$ function was checked by subdividing the simulation run into chunks of equal length and calculating the PMF for each of them.[12] The average value (in the R,θ space) of the standard deviation of the $\mathcal{N}$ chunks calculated PMF from the whole run calculated PMF is defined as follows:

$$\sigma_{\mathcal{N}} = \langle \{ \mathcal{N}^{-1} \sum_{i=1}^{\mathcal{N}} [W_i(R,\theta) - W_{tot}(R,\theta)]^2 \}^{1/2} \rangle \tag{4}$$

where $W_i(R,\theta)$ is the PMF at the point (R,θ) calculated in the ith chunk, $W_{tot}(R,\theta)$ is the PMF at the same point computed over the whole run, and the angular brackets indicate the average on the R,θ space. $\sigma_{\mathcal{N}}$ gives an estimate of the sampling error of the method (see section 3.2).

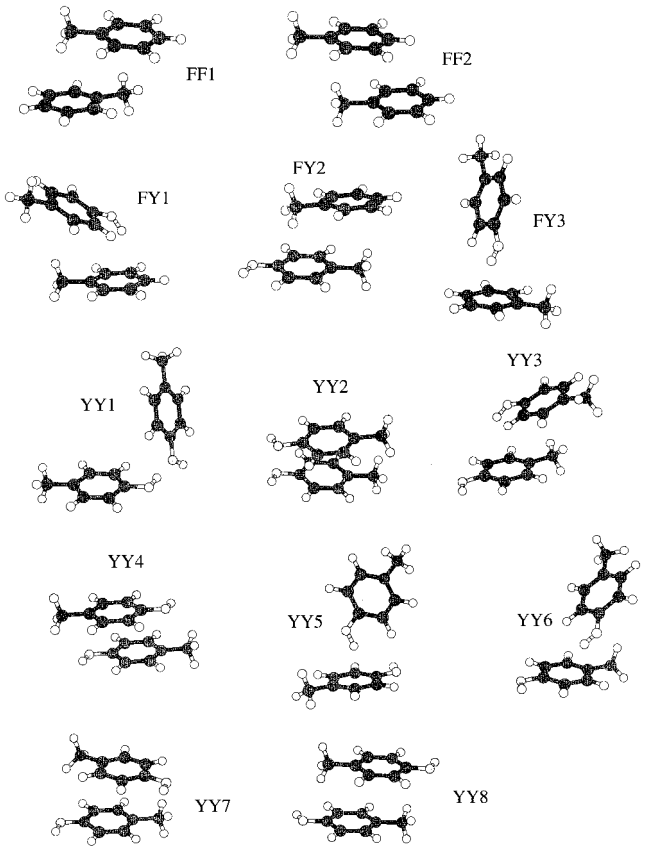

Figure 2. Minimum energy structures of the Phe−Phe (FF1, FF2), Phe−Tyr (FY1, FY2, FY3), and Tyr−Tyr (YY1, YY2, YY3, YY4, YY5, YY6, YY7, YY8) complexes. For the Tyr−Tyr complex, the eight most populated minimum energy structures are reported. The structural and energetical data of these structures are reported in Table 3.

3. Results and Discussion

3.1. Potential Energy Surface. PMF in a defined phase space is a quantity that measures differences of probability in that space, as determined by energetic (i.e., the PES) and entropic factors. Entropy calculation, and hence PES, is actually a difficult task in systems of many particles such as ours. On the other hand, indirect information on the relative importance of the two contributions to the PMF can be obtained from the knowledge of the PES of the isolated complexes.

Following a well-established procedure,[16,33] all possible minimums for the intermolecular PES of the Phe−Phe, Phe−Tyr, and Tyr−Tyr complexes are determined by quenching, for each pair, 4000 structures regularly sampled from MD simulations in vacuo (for the simulation details see section 2.2 and Table 2). In addition to the minimum energy structures, this procedure also gives an idea of the extension of the energetical basin of each minimum, i.e., the probability that an instantaneous configuration sampled at high temperature has that minimum as inherent structure.[34,35] In Figure 2 the minimum energy structures of the Phe−Phe, Phe−Tyr, and Tyr−Tyr pairs are shown. In Table 3, some structural and energetical data of the

(31) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848−1862.
(32) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, 1987.
(33) Kratochvil, M.; Sponer, J.; Hobza, P. *J. Am. Chem. Soc.* **2000**, *122*, 3495−3499.
(34) Stillinger, F. H.; Weber, T. A. *Phys. Rev. A* **1982**, *25*, 978−989.
(35) Stillinger, F. H. *J. Phys. Chem.* **1984**, *88*, 6494−6499.

Table 3. Structural and Energetical Data of the Minimum Energy Structures Shown in Figure 2[a]

structure	Pop.	E_b	E_{el}	E_{LJ}	R	θ
FF1	91.0	−18.7	−0.7	−18.0	4.49	8.8
FF2	9.0	−15.2	2.1	−17.3	4.03	5.1
FY1	64.7	−23.8	−10.2	−13.6	4.62	31.7
FY2	31.4	−19.4	0.2	−19.6	4.29	7.3
FY3	3.9	−21.6	−12.3	−9.3	5.16	74.2
YY1	26.1	−27.7	−24.1	−3.6	5.90	69.8
YY2	15.3	−27.3	−6.2	−21.1	3.67	8.9
YY3	13.9	−22.8	−7.1	−15.7	4.54	21.4
YY4	10.3	−23.0	−2.5	−20.5	3.82	0.0
YY5	9.6	−23.2	−11.2	−12.0	4.91	53.4
YY6	6.4	−22.6	−10.7	−11.9	4.91	44.0
YY7	5.6	−20.5	3.3	−23.8	3.57	0.3
YY8	4.8	−20.4	2.7	−23.1	3.80	2.8

[a] Pop. is the percent population, namely, the ratio (× 100) between the number of occurrences of the minimum energy structure and the total number of quenched structures (see text for details). E_b (in kJ mol^{-1}) is the binding energy (energy of the complex subtracted out of the energies of the two monomers calculated separately). E_{el} and E_{LJ} are the electrostatic and Lennard-Jones contributions to E_b, respectively. R (in Å) is the distance between the two ring centroids. θ (in deg) is the angle between the two normals to the rings. For the Tyr−Tyr complex, the eight most populated minimum energy structures are reported.

minimums of Figure 2, namely, the population (i.e., the number of occurrences of the minimum energy structure), the distance between the centroids of the rings, the angle between the normals to the rings, the binding energies, and their electrostatic and Lennard-Jones contributions, are reported.[36]

Several features can be observed from Table 3: (i) Phe−Phe and Phe−Tyr pairs have only two and three minimum energy structures, respectively, while the Tyr−Tyr pair has eight minimum energy structures with significant population. (ii) The most populated minimum energy structures of the Phe−Phe and Phe−Tyr pairs, namely, FF1, FF2, FY1, and FY2, have stacked arrangements (see also Figure 2), only FY3 (population, 3.9%) being T-shaped. On the contrary, the Tyr−Tyr pair has two highly populated T-shaped minimum energy structures, namely, YY1 and YY5 (YY1, 26.1%; YY5, 9.6%). Nevertheless, the sum of the populations of all the stacked minimum energy structures (YY2, YY3, YY4, YY6, YY7, and YY8) remains greater than the sum of the occurrences of the T-shaped YY1 and YY5 minimums (56.3% versus 35.7%). (iii) The binding energies (E_b) of the minimum energy structures of the three pairs are quite different: the population-weighted binding energy for the Tyr−Tyr, Phe−Tyr, and Phe−Phe pairs is −24.7, −22.3, and −18.4 kJ mol^{-1}, respectively. (iv) The minimum energy structures are in general stabilized by dispersive forces.

There are, however, few minimum energy structures that are stabilized mostly by electrostatic interactions, namely, YY1 and FY3 and, to a lesser extent, YY5 and YY6. The YY1 structure is stabilized by a strong H-bond involving the two hydroxy groups, whereas FY3, YY5, and YY6 are characterized by having the hydroxy group of a Tyr monomer pointing toward the aromatic ring of the partner molecule in a not conventional H-bond.[37,38] Finally, we remark that these minimum energy structures, stabilized by electrostatics, have a θ angle characteristic of T-shaped (YY1, FY3) or oblique arrangements (YY5, YY6). This last feature confirms[12] that electrostatics, and H-bonding in particular (including that of the H-ring type[37,38] as observed in the FY3, YY5, and YY6 structures), is an important factor in driving side-chain orientation toward T-shaped arrangement.

3.2. Potential of Mean Force and Structural Properties. In Figures 3 and 4, the PMFs of the unbound and bound Phe−Phe, Tyr−Tyr, and Phe−Tyr complexes in vacuo and in solution are shown.[39] The sampling errors on the PMF, calculated as described in section 2.3 (eq 4), are reported in Table 4. For the bound complexes, the errors are, in general, of the order of 1 kJ mol^{-1} or less, whereas for the unbound ones, due to the extended configurational space, they are slightly larger.

For all pairs, the in vacuo PMF has a minimum corresponding to stacked arrangements ($\theta < 20°$). For Phe−Phe and Phe−Tyr, low free energy of stacked arrangements is consistent with the PES (see section 3.1), which has revealed that most of the minimum energy structures are of the stacked type. An analogous stabilization of energetic origin has been observed for the PMF of the Trp−His complex.[12] In fact, the Trp−His complex in vacuo at 300 K gives principally T-shaped structures stabilized by H-bonding. Correspondingly, the PES scanning, using a minimum energy search procedure identical to that used in the present work, yielded[16] the T-shaped structures as the most populated and energetically stable for the Trp−His complex. For Tyr−Tyr, the most stable and populated minimum in the PES, i.e., YY1, is T-shaped. In this case, therefore, the stabilization of the stacked structures has an entropic origin correlated with the greater number of accessible stacked configurations (see discussion in section 3.1). From the PMF in vacuo, the energetical stability of the complexes (coherently with the population-weighted PES binding energy reported in section 3.1) is in the order Tyr−Tyr > Phe−Tyr > Phe−Phe.

The solvated complexes (bound and unbound) have in general the same stability order as that observed in vacuo. As a general trend, and as observed for the Trp−His complex, the solvent has a destabilizing effect irrespective of the presence of the C_α−C_α constraint. From Figures 3 and 4 it can be observed that, while water has only a small destabilizing effect with respect to the in vacuo case, solubilization in CCl_4 results in a relatively strong destabilization of the complexes. An opposite behavior was found for the Trp−His complex, for which destabilization in water is larger than that observed in CCl_4 (see Figure 3 of ref 12). These findings can be explained by observing that, for Tyr−Tyr, Phe−Tyr, and Phe−Phe complexes, stabilization is due mostly to dispersive interaction (see Table 3). So, while the strong polar character of water favors destabilization of structures stabilized by electrostatic (H-bond) interactions, for example, those observed in the Trp−His PES,[16] CCl_4 weakens the complex stabilized by dispersive interactions, namely, Phe−Phe, Phe−Tyr, and Tyr−Tyr. It should be then clear why MeOH strongly destabilizes Phe−Phe, Phe−Tyr, and Tyr−Tyr: with its methyl group, MeOH can in fact inhibit, like CCl_4, residue−residue dispersive interactions, whereas with the hydroxy group, it can weaken interresidue electrostatical binding forces. This destabilizing power of MeOH due to its amphiphilic nature has also been observed in the case of the Trp−His complex.[12] An indeed remarkable effect, correlated to the suppression of the favorable residue−residue dispersive forces by apolar solvents,

(36) The atomic coordinates of the minimum energy structures are available upon request to the corresponding author.
(37) Levitt, M.; Perutz, M. F. *J. Mol. Biol.* **1988**, *201*, 751−754.
(38) Perutz, M. F. In *Pioneering Ideas for the Physical and Chemical Sciences*; Fleischhacker, W., Schönfeld, T., Eds.; Proceedings of the Josef Loschmidt Symposium, Vienna, Austria; Plenum Press: New York, 1995; pp 1−14.
(39) The color figures are available in PDF format upon request to the corresponding author.

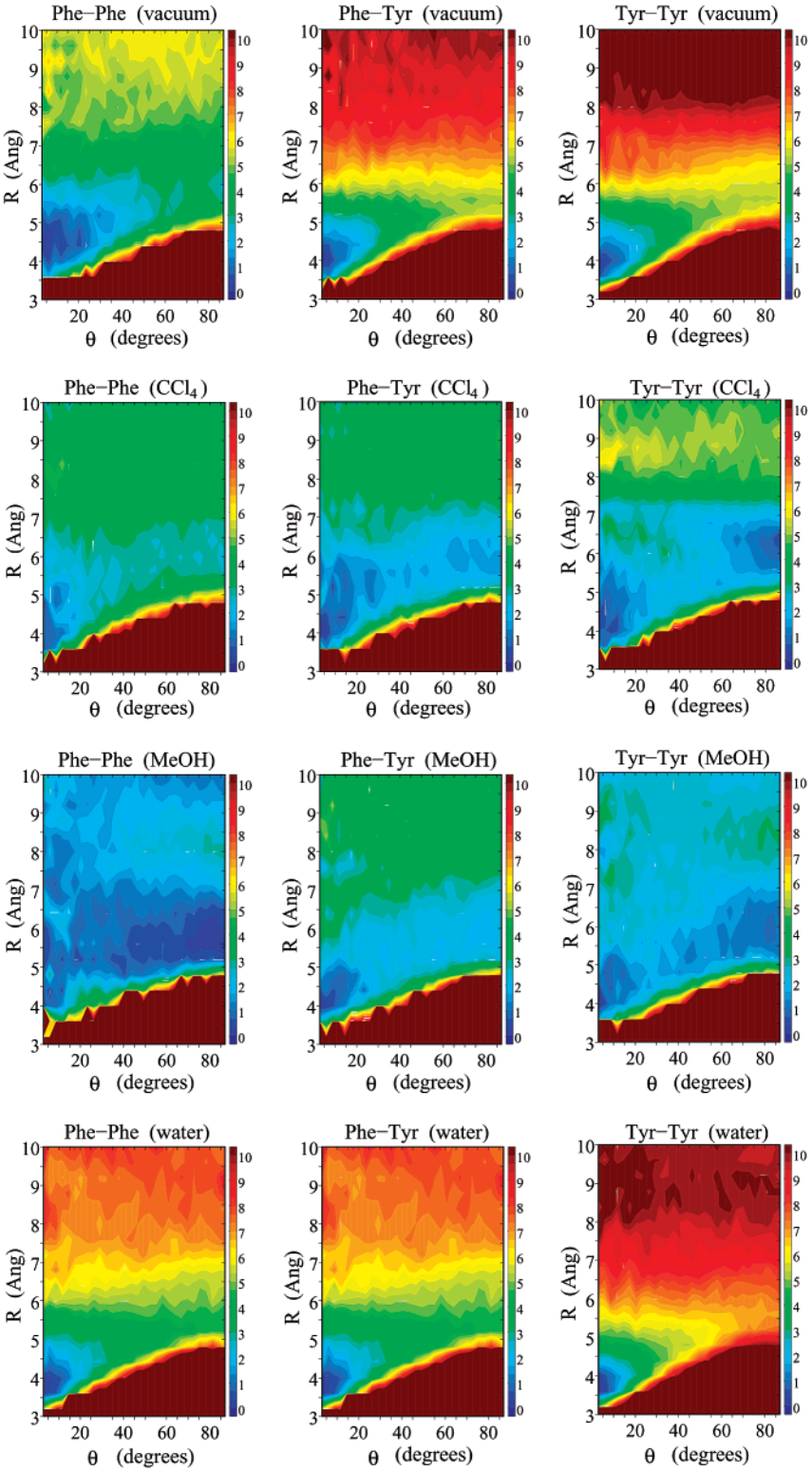

Figure 3. PMF for the Phe−Phe, Phe−Tyr, and Tyr−Tyr unbound complexes in vacuo and in CCl_4, MeOH, and water (see top of the pictures) as a function of the ring centroid distance, R, and of the angle between the normals to the rings, θ. On the right side of each picture, the chromatic energy scale is shown in units of kJ mol^{-1}. The brown regions correspond to free energies greater than 10 kJ mol^{-1}.

is observed for the Tyr−Tyr complex in CCl_4 (see Figure 3). As expected, the complex is less stable in CCl_4 than in vacuo (bound states are from 2 to 4 kJ mol^{-1} lower than unbound states), but in the binding region (R < 7.3 Å), T-shaped arrangements emerge (note PMF for R < 7.3 Å and θ > 60° in Figure 3) along with stacked ones. In fact, while CCl_4 has no significant effect on electrostatics, it weakens the Tyr−Tyr dispersive interactions, thereby reversing the balance between stacked and T-shaped structures toward the latter.

On the basis of the previous discussion, the rather different behavior of Tyr−Tyr and Trp−His complexes, though both these molecules are able to form quite strong intermolecular H-bonds, is not surprising. In fact, as revealed by the PES, the dispersive interactions for the Tyr−Tyr pair are more important

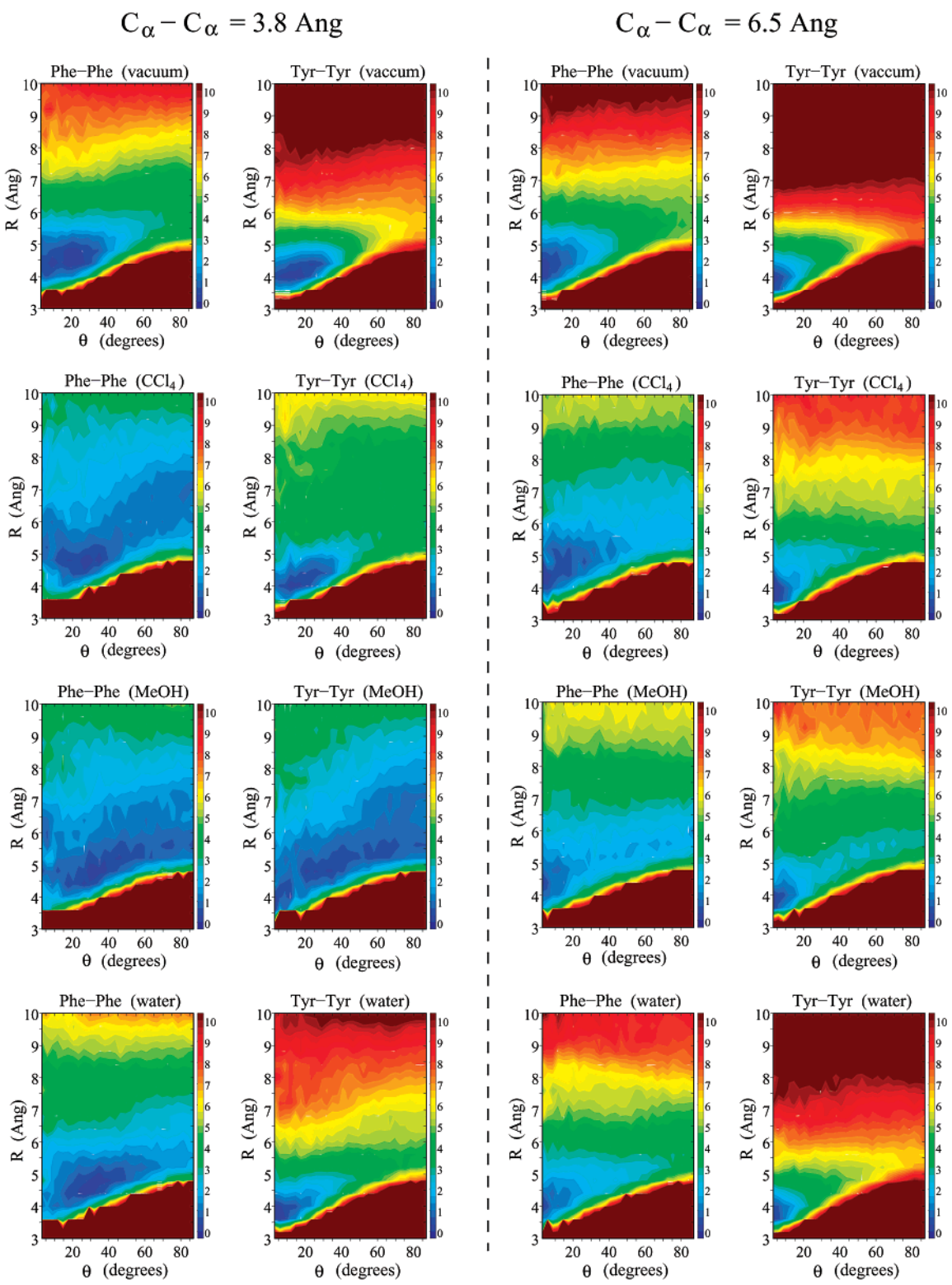

Figure 4. PMF for the Phe−Phe and Tyr−Tyr bound complexes in vacuo and in CCl_4, MeOH, and water (see top of the pictures) as a function of the ring centroid distance, *R*, and of the angle between the normals to the rings, *θ*. The PMFs for the bound complexes having the constrained $C_\alpha - C_\alpha$ distance equal to 3.8 and 6.5 Å are reported on the left and on the right, respectively. On the right side of each picture, the chromatic energy scale is shown in units of kJ mol^{-1}. The brown regions correspond to free energies greater than 10 kJ mol^{-1}. PMFs for the Phe−Tyr bound complex (not reported) are similar to the Phe−Phe ones.

with respect to electrostatics (see discussion in section 3.1) than in Trp−His; correspondingly, hydrophobic environment destabilizes more efficiently the Tyr−Tyr complex. This can be clearly shown by performing MD simulations of dummy models with quenched Lennard-Jones interactions. In Figure 5 the PMF for three different model systems are reported. Models A and B refer to the Tyr−Tyr unbound complex in vacuo where the ϵ Lennard-Jones constants of all the atoms, except that of the oxygen, are divided by a factor of 2 and 5, respectively. In model C, all the ϵ Lennard-Jones constants, including those of the

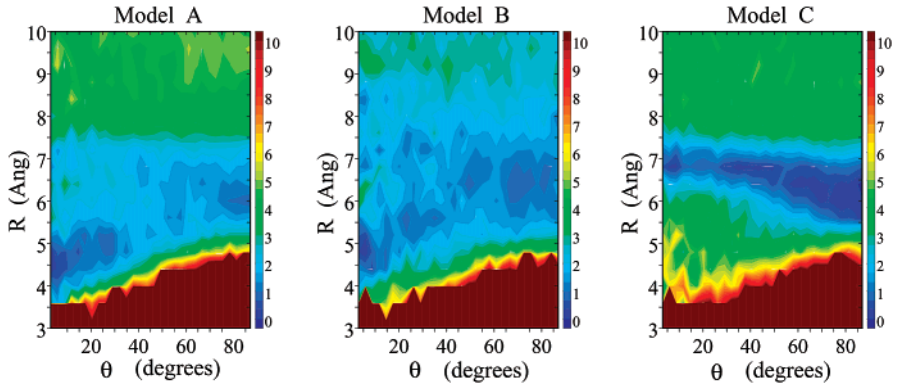

Figure 5. PMF for three models of Tyr−Tyr unbound complex in vacuo with decreasing Lennard-Jones interactions (see text for the definition of the models).

Table 4. Sampling Error of the PMF for Each Simulation[a]

	Phe–Phe			Tyr–Tyr			Phe–Tyr
D	3.8	6.5	uc	3.8	6.5	uc	uc
vacuum	0.49	0.86	1.19	0.42	0.46	1.14	1.92
CCl_4	1.11	1.29	1.33	0.89	1.40	1.27	1.87
MeOH	0.98	0.83	1.83	1.35	0.84	1.64	1.52
water	0.56	0.89	1.09	1.00	1.30	0.72	1.24

[a] $\sigma_{\mathcal{N}}$ values for $\mathcal{N} = 5$ are reported, $\mathcal{N}$ being the number of simulation chunks (see text for details). *D* (in Å) is the $C_\alpha - C_\alpha$ constrained distance of the bound complexes. uc refers to the unbound complexes. Values are given in kJ mol^{-1}.

oxygen atom, are divided by a factor of 5. The MD simulations of the A, B, and C models were performed using the same method (thermodynamic ensemble, equilibration procedure, box size, etc.) of the other in vacuo simulations. In going from model A to model C, we could imagine to have solubilized the Tyr−Tyr complex in apolar solvents of increasing solubilization power (with regard to residue−residue dispersive forces). The PMF of model A, because of the reduced dispersive forces, results in a less stable complex, but now T-shaped arrangements appear: by using a simple scaling factor for the different interactions in vacuo, we have reproduced the effect on the PMF of solubilizing the complex in CCl_4 (compare PMF of model A with PMF of the Tyr−Tyr unbound complex in CCl_4 shown in Figure 3). The PMF of model B shows a large destabilization with respect to model A, with stacked and T-shaped configurations barely distinguishable. Finally, although model C differs from model B only for the ϵ Lennard-Jones constant of the oxygen atom, the differences between the PMF of the two models are spectacular. The lowering of the ϵ Lennard-Jones constant of the oxygen atom restores the possibility of forming a strong H-bond any time the two hydroxy groups come in contact. Because of the enhanced strength of H-bond, a larger stabilization of the complex for model C with respect to model B is observed, and correspondingly, we observe the appearance of stable T-shaped electrostatically favored structures.

As seen previously, the H-bond plays an important role in shaping the PMF of aromatic amino acid residues by selectively favoring T-shaped arrangements. In this scenario, the solvent effect acts as a regulatory mechanism for H-bond association. In fact, for Trp−His (see Figure 4 of ref 12), we observed that H-bond-forming solvents (water, MeOH, dimethyl sulfoxide) strongly inhibit H-bond interaction between the two aromatic residues. An almost complementary behavior was observed for CCl_4, where the effect of the hydrophobic solvent is to restore H-bond-stabilized structures of the complex.

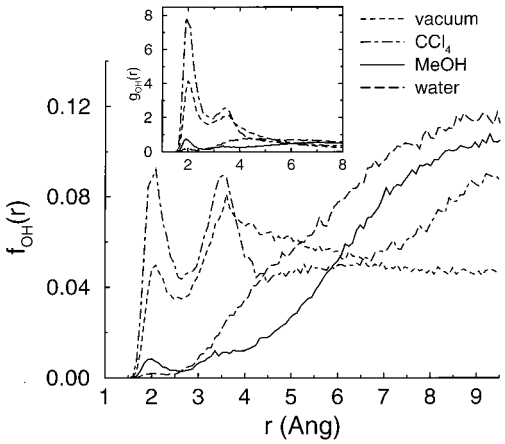

Figure 6. Radial distribution functions $f_{OH}(r)$ and $g_{OH}(r)$ (inset) of the *oxygen−[hydroxy]hydrogen* pair for the Tyr−Tyr unbound complex in vacuo and in various solvents.

To compare the H-bond behavior for the Tyr−Tyr pair with that observed for Trp−His, the radial distribution function of the *oxygen−[hydroxy]hydrogen* pair was calculated as

$$f_{OH}(r) = n\langle\delta(r - R_{OH})\rangle \qquad (5)$$

In eq 5, the angular brackets indicate an average on the configurations and R_{OH} is the distance between the oxygen of a residue and the hydroxy hydrogen of the partner residue; *n* is a normalization factor such that $\int_0^{R_m} f_{OH}(r)\, dr = 1$, where R_m is the maximum distance between two particles in the simulation box. $f_{OH}(r)$'s for Tyr−Tyr unbound complex in various solvents are reported in Figure 6. The standard radial distribution functions, $g_{OH}(r)$, normalized with respect to the radial particle population,[40] are reported in the inset of Figure 6 to allow a direct comparison with the Trp−His case. H-bonding is evident from the first peak of $f_{OH}(r)$ and $g_{OH}(r)$ at 2 Å, while the other peak at ~3.5 Å (in CCl_4 and in vacuo) is due to the second OH pair of H-bonded configurations. As expected, H-bond association is strongly suppressed for Tyr−Tyr with respect to Trp−His (compare inset of Figure 6 with Figure 4 of ref 12). The Tyr−Tyr complex, like Trp−His, is mostly H-bonded in vacuo and in CCl_4, while the other (H-bond-forming) solvents, especially water, destabilize the residue−residue H-bond. The value of $f_{OH}(r)$ at distances greater than 7 Å reflects the different

(40) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U.K., 1987.

stability of the complex (see PMF of the unbound complexes in Figure 3). A significant exception to this trend is given by the case of water. In fact, contrarily to the $f_{OH}(r)$ behavior that indicates less Tyr−Tyr association in water with respect to all other solvents, the PMF reveals a relatively strong Tyr−Tyr association. The behavior of $f_{OH}(r)$ can be explained as follows: the water molecules form strong H-bonds with the OH group of the Tyr monomers; the resulting solvation shells around the hydroxy groups of the two Tyr monomers tend to repel each other, yielding $f_{OH}(r)$ values larger (at large distances) with respect, for example, to the in vacuo case.

The effect of constraining the distance between two C_α atoms (see models in Figure 1) is shown in Figure 4. As observed for the Trp−His case, the presence of the constraint generally enhances the stability of the complexes by favoring stacking. Stabilization of stacked structures due to the constraint can be clearly observed, for example, for Tyr−Tyr in CCl_4 and for both Phe−Phe and Tyr−Tyr in MeOH. Finally, it can be noted that the constraint at 6.5 Å has a more efficient stabilization effect than the constraint at 3.8 Å, indeed remarkable in some cases (see for instance the bound complexes in MeOH and CCl_4).

3.3. Can the PMF Teach Us Something of the Aromatic−Aromatic Interactions in Proteins? The computational investigation of the PMF provides a great deal of clear and surprisingly rich chemical−physical information. Furthermore, such a technique is very versatile as it could in principle be applied to model any kind of residue−residue interaction in environments of different polarity. On the other hand, the modelistic nature of the PMF calculation poses several questions. Can the PMF analysis really be related to experimental findings on proteins? If this is the case, what can we learn about the protein structures from PMF analysis? Are the interactions between aromatic amino acids in proteins really *mainly* regulated by their aromatic side chains as the PMF analysis implicitly assumes? Does the polarity of different protein regions (core or surface) play a role in the aromatic−aromatic side-chain arrangements?

Before addressing these questions, we summarize the PMF results we believe important for rationalizing the aromatic−aromatic interactions in proteins: (i) In a hydrophilic environment, such as that experienced by amino acid side chains on the protein surface, stacked arrangements are strongly favored for both the unbound and bound complexes. (ii) In a hydrophobic environment, such as the protein core, the free energy surface for the Phe−Phe and Phe−Tyr complexes is rather flat with respect to the Tyr−Tyr case, with only a slight prevalence of stacking. The behavior of the Trp−His pair in a hydrophobic environment (CCl_4) is more uncertain because the pair arrangement is strongly dependent on the C_α−C_α distance. In fact, the structure of the complex is T-shaped for C_α−C_α = 6.5 Å, whereas it is stacked when C_α−C_α = 3.8 Å. (iii) In a hydrophobic environment, the probability of stacking is enhanced by the C_α−C_α constraint. Assuming the bound complexes as a plausible model for aromatic−aromatic side-chain interactions in proteins, on the basis of the previous observations, one should expect the following: (i) In the core and in the surface of the proteins, the stacking is the most probable pair arrangement for the Phe−Phe, Phe−Tyr, and Tyr−Tyr side-chain interactions. (ii) The stacked interactions are more

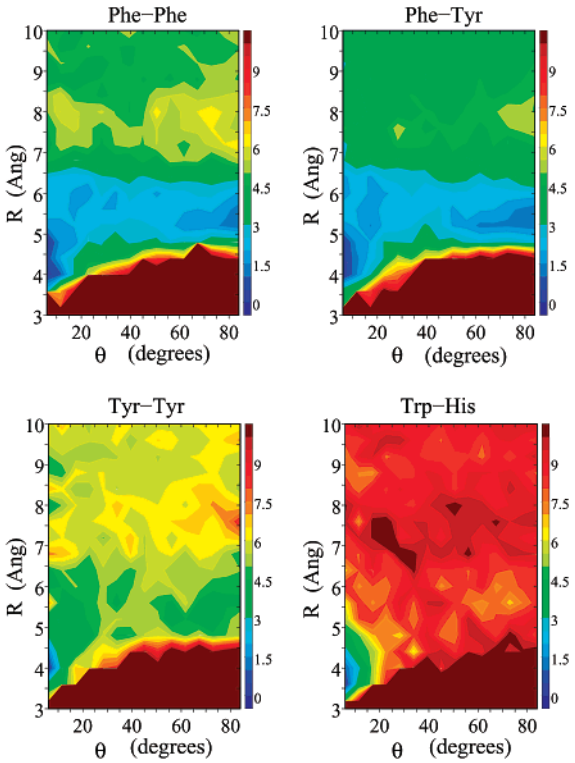

Figure 7. PMF_{pdb} of Phe−Phe, Phe−Tyr, Tyr−Tyr, and Trp−His residue pairs for protein structures resolved experimentally.

stabilized for Tyr−Tyr than they are for Phe−Phe and Phe−Tyr. (iii) In the hydrophobic core, Phe−Tyr and Phe−Phe interactions are less specific with little difference of probability between stacking and T-shape. (iv) For the Trp−His pair, one should expect a stacked arrangement on the protein surface, while a not-well-defined situation is expected in the core.

To verify these predictions based on the PMF analysis, we have analyzed aromatic−aromatic interactions in a database of 2396 nonhomologous proteins extracted from the Brookhaven Protein Data Bank[41,42] (PDB). The protein can be considered as a *special* solvent where several aromatic side chains, rigidly anchored to the protein backbone, are solvated. Many interacting (neighboring) and noninteracting aromatic residue pairs can be identified in the protein data set, and all these pairs can be considered as a statistical ensemble. We can therefore apply to this ensemble the same procedure described in section 2.3 and calculate the PMF in the R, θ space (hereafter indicated as PMF_{pdb}) of the Tyr−Tyr, Phe−Tyr, Phe−Phe, and Trp−His pairs in the "protein solvent". As all the structures in the PDB refer to proteins existing in physiological conditions, for the calculation of PMF_{pdb}, we set the temperature at 300 K. PMF_{pdb}'s for the Phe−Phe, Phe−Tyr, Tyr−Tyr, and Trp−His pairs are reported in Figure 7.

In the R, θ phase space of Figure 7, i.e., for $0 < R < 10$ Å and $0 \le \theta \le 90°$, we found 15064, 20293, 8577, and 4081 pairs of kind Phe−Phe, Phe−Tyr, Tyr−Tyr, and Trp−His, respectively. The statistical error on the PMF_{pdb} was evaluated by calculating σ_2 (see eq 4), which was 0.96, 0.55, 0.50, and 0.63 kJ mol^{-1} for Phe−Phe, Phe−Tyr, Tyr−Tyr, and Trp−

(41) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(42) The PDB names of the analyzed proteins are available upon request to the corresponding author.

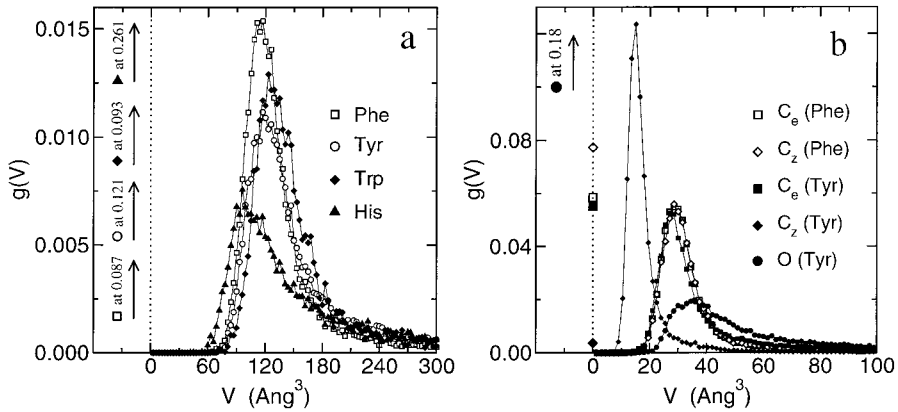

Figure 8. (a) Voronoi polyhedron volume distribution for the centroids of the Phe, Tyr, Trp, and His residues. In the first channel, the fraction of residues with volume greater than 1000 Å^3 is reported. (b) Voronoi polyhedron volume distribution for the C_e, C_z, and oxygen atoms of the Phe and Tyr residues (see text for details and Figure 1 for the atom labels). C_e was calculated by averaging on C_{e1} and C_{e2}. C_z of the Tyr residue has a small volume being surrounded by three covalently bonded atoms, namely, C_{e1}, C_{e2}, and oxygen. In the first channel, the fraction of residues having an atom with volume greater than 1000 Å^3 is reported.

His, respectively. The experimental PMF_{pdb}'s are strongly reminiscent of the calculated PMFs (see Figure 7 of ref 12 and Figure 4), suggesting the side-chain interactions as a determining factor for the reciprocal arrangement of neighbor aromatic amino acids in proteins. The well-defined free energy minimum (corresponding to stacked arrangements) in the PMF_{pdb} of the Tyr−Tyr pair suggests (on the basis of the PMF of Figure 4) that these contacts can mainly occur in a hydrophilic environment. The same conclusion can be deduced for the Trp−His pair (compare PMF_{pdb} and PMF of Figure 7 of ref 12). From the comparison of the PMF_{pdb} and PMF, we can also state that the Phe−Phe (and Phe−Tyr) interactions are mainly present in the hydrophobic core of proteins.

To verify these statements, we evaluated the solvent exposure of the Phe, Tyr, Trp, and His protein side chains by means of the Voronoi polyhedron analysis.[43] Side-chain polyhedra were computed by representing each side chain as a single point at its centroid. Closure of the Voronoi polyhedra for the exposed residues was achieved by adding eight distant dummy atoms (1000 Å away from the protein center of mass) at the vertexes of a cube. In this fashion, large Voronoi polyhedron volumes correspond to exposed side chains. Using the same closure procedure, the Voronoi polyhedron volumes of selected atoms of the Phe and Tyr side chains were also computed. The distribution of the Voronoi polyhedron volumes of the Phe, Tyr, Trp, and His centroids and that corresponding to several side-chain atoms of Phe and Tyr are reported in Figure 8a and b, respectively. In both cases, hydrogen atoms have not been considered in the volume calculation. In the first channel of the volume distributions, volumes greater than 1000 Å^3 (signature of solvent exposure) are reported. From the value of the distribution at the first channel and from its persistent tail at large volumes, we deduce that the Tyr residue is more exposed to the solvent than the Phe residue is. In particular, Tyr is exposed by means of its oxygen atom, which shows a very broad distribution extending far toward large volumes. Solvent exposure is even larger for His residues. All these findings are perfectly consistent with PMF analysis that strongly suggests a correspondence between stacking and exposure to hydrophilic environment.

(43) Rapaport, D. C. *The Art of Molecular Dynamics Simulation*; Cambridge University Press: New York, 1995, and references therein.

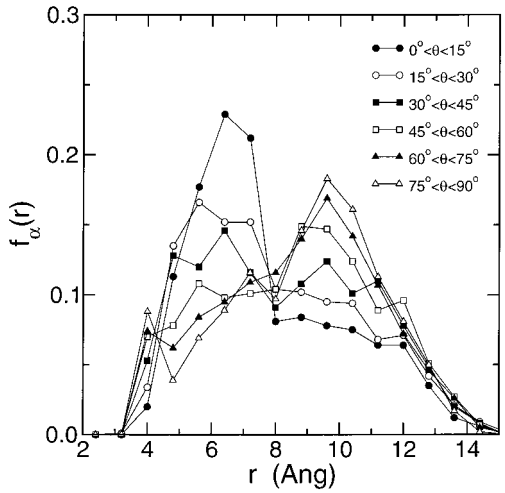

Figure 9. Pair radial distribution function, $f_\alpha(r)$, of the Phe C_α's for various ranges of the angle (θ) between the normals to the rings. Phe residues are from PDB.

An interesting feature is the presence, in the PMF_{pdb} of the Phe−Phe, Phe−Tyr, and, to a less extent, Tyr−Tyr pairs, of a local free energy minimum corresponding to T-shaped arrangements. Such a minimum is yielded by pairs of aromatic residues whose C_α's are at large distances, larger than the constraint distance of 6.5 Å used in our PMF analysis. This hypothesis is confirmed by the pair radial distribution function, $f_\alpha(r)$ (see eq 5), of the Phe C_α's calculated from the PDB data and reported in Figure 9 (similar results have been obtained for the other pairs). Six $f_\alpha(r)$'s were calculated, corresponding to six different ranges for the θ angle defined by the normals to the rings. These angular ranges are as follows: [0°, 15°], (15°, 30°], (30°, 45°], (45°, 60°], (60°, 75°], and (75°, 90°]. The figure clearly shows that stacked conformations are in general favored by short $C_\alpha-C_\alpha$ distances, while a large $C_\alpha-C_\alpha$ distance (8−10 Å) favors T-shaped conformations. Incidentally and remarkably, in agreement with the PMF of the bound complexes, the most favorable $C_{\alpha\prime}$-C_α distance for stacked structures is slightly greater than 6 Å. The fact that a large $C_\alpha-C_\alpha$ distance for interacting aromatic residues favors T-shaped arrangements was also verified by performing a simulation of 14.4 ns at 300 K of the Phe−Phe complex in vacuo with the $C_\alpha-C_\alpha$ distance constrained at 10 Å. The obtained PMF (data not shown) yields a broad but well-

defined T-shaped local minimum in agreement with experimental findings.

4. Conclusions

A study of the structural properties of interacting pairs of aromatic amino acid residues (Tyr−Tyr, Phe−Tyr, Phe−Phe) has been performed using MD simulations. Stacking and T-shape competition is investigated as a function of the chemical environment (solvent) and of a constraint between the α-carbons by determining the potential energy surface and the potential of mean force of the pairs.

In general, stacking is found to be the more common arrangement, being strongly favored by the constraint between the α-carbons. In water, stacking is favored irrespective of the α-carbon distance. In apolar environments, where residue−solvent dispersive interactions destabilize stacking, the possibility of forming an H-bond between the residues results in a stabilization of T-shaped structures. These findings are confirmed by a model of the Tyr−Tyr unbound complex in vacuo, where dispersive interactions are progressively switched off, resulting in a progressive stabilization of T-shaped structures. MeOH has in general a destabilizing effect on the complexes because is able to solvate both the T-shaped (Tyr−Tyr) structures by forming H-bonds with the two residues and the stacked structures by solvating the aromatic rings with the methyl group. These findings are consistent with previous theoretical studies on the Trp−His pair.

The PMF analysis as a function of solvent polarity and distance constraint provides an effective tool for rationalizing structural data obtained from a large database of experimentally resolved proteins. By comparing MD computed PMF and the experimental analogue calculated from a large protein database we can state the following: (i) The Tyr−Tyr and Trp−His interactions occur mainly in regions of the protein exposed to the solvent, and the possibility of H-bond formation between the residues is not important for their arrangement. (ii) Phe−Phe and Phe−Tyr arrangements are determined from a competition between stacking, favored by dispersive interactions, and T-shape, favored by a large distance between the α-carbons. Hence, interacting distal Phe−Phe and Phe−Tyr are prevalently T-shaped, while interacting proximal Phe−Phe and Phe−Tyr are prevalently stacked.

PES, PMF, and PMF_{pdb} calculations give indications that protein folding is a process where energetics of the aromatic protein side chains plays a determinant role. In this respect, studies on pair interactions between apolar protein side chains found in the protein core, such as valine, leucine, or isoleucine, could be important to understand the interplay of different amino acids in the protein folding.

Acknowledgment. This work was supported by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), by the Italian Consiglio Nazionale delle Ricerche (CNR), and by the European Union (Contract HPRI-CT-1999-00111).

JA0121639