Cation−π Interactions in Proteins: Can Simple Models Provide an Accurate Description?

Hervé Minoux and Christophe Chipot*

Contribution from the Laboratoire de Chimie Théorique, Unité Mixte de Recherche CNRS 7565, Université Henri Poincaré, Nancy I, B.P. 239, 54506 Vandœuvre-lès-Nancy, France

Received March 22, 1999. Revised Manuscript Received June 4, 1999

Abstract: It has been suggested that cation−π interactions constitute a strong, specific driving force that plays a key role in molecular recognition. The importance of such interactions in biological systems is explored here via two complementary approaches. The first one relies on an analysis of the association of phenylalanine, tyrosine, and tryptophan with arginine and lysine in 1718 representative protein structures, highlighting orientational preferences in cation−π complexes. The second one consists of an MP2/6-311++G**//MP2/6-31G** ab initio investigation of the dimers formed by relevant models of the amino acid side chains that are engaged in cation−π interactions. The estimated induction contribution to the binding energies confirms that polarization effects are significant. The ability of commercial, two-body potential energy functions to describe cation−π interactions accurately is also investigated, and the inclusion of correcting parameters in the force field is discussed. Put together, these results provide new insights into the nature of cation−π association in proteins.

Introduction

Cation−π interactions in biological systems have recently emerged as one of the driving forces in molecular recognition processes.[1−3] These interactions are strong enough to compete with the solvation of hydrophilic, charged moieties and allow ligand−receptor binding in a hydrophobic pocket constituted of aromatic residues, within the core of the receptor.[2] Such noncovalent interactions have been hypothesized to be responsible, among others, for the activity and the selectivity of the potassium channel, as well as for the binding of acetylcholine to acetylcholinesterase.[2,4−6] Furthermore, numerous cation−π interactions have been observed in experimentally determined protein structures. Whereas it is generally assumed that their function is not crucial for the stability of the protein,[7,8] they could, nevertheless, participate in its folding.[2]

From an electrostatic point of view, the dominating component in cation−π interactions is the attraction of the charge toward the quadrupole created by the π-electron cloud of the aromatic ring.[2,6,9−12] The significant polarizability of the latter combined with the polarizing nature of the positively charged ion makes cation−π interactions difficult to describe using simplified models. Such models, however, are desirable for exploring complex biological systems and can only be built from a better understanding of the contributions governing the binding in small and simple cation−π assemblies. In this article, the interaction between biologically relevant cations and aromatic compounds is studied via two different, yet synergistic, approaches that make use of both experiment and theory.

First, cation−π interactions formed by the side chain of phenylalanine, tyrosine, and tryptophan, on one hand, and lysine and arginine, on the other hand, are analyzed from a representative list of protein structures[13] of the Brookhaven protein data bank (PDB).[14,15] Orientational preferences with respect to the distance separating the positively charged side chain from the aromatic one are discussed, offering a direct quantification of cation−π interactions in systems of biological interest.

Next, models of cation−π complexes, constructed from toluene, *p*-cresol, methyl−indole, ammonium, and guanidinium, are investigated using higher level ab initio calculations than have been employed hitherto to study such systems. The optimized geometries are compared to those observed in the related biological structures, and the induction contribution to the total quantum mechanical interaction energies is estimated.

Last, because they virtually represent the most cost-effective solution for handling large molecular assemblies, the adequacy of additive potential energy functions to describe cation−π interactions accurately is investigated. The ability of three different, commercial force fields to reproduce the energetics of the model systems explored quantum mechanically is discussed. Corrective parameters for improving, in an average sense, the description of induction phenomena are proposed,

(1) Burley, S. K.; Petsko, G. A. *FEBS Lett.* **1986**, *203*, 139−143.
(2) Dougherty, D. A. *Science* **1996**, *271*, 163−168.
(3) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303−1324.
(4) Dougherty, D. A.; Stauffer, D. A., *Science* **1990**, *250*, 1558−1560.
(5) Scrutton, N. S.; Raine, A. R. C. *Biochem. J.* **1996**, *319*, 1−8.
(6) Kumpf, R. A.; Dougherty, D. A. *Science* **1993**, *261*, 1708−1710.
(7) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709−717.
(8) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315−331.
(9) Stauffer, D. A.; Barrans, R. E.; Dougherty, D. A. *J. Org. Chem.* **1990**, *55*, 2762−2767.
(10) Luhmer, M.; Bartik, K.; Dejaegere, A.; Bovy, P.; Reisse, J. *Bull. Soc. Chim. Fr.* **1994**, *131*, 605−606.
(11) Mecozzi, S.; West, A. P., Jr.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10566−10571.
(12) Mecozzi, S.; West, A. P., Jr.; Dougherty, D. A. *J. Am. Chem. Soc.* **1996**, *118*, 2307−2308.
(13) Hobohm, U.; Scharf, M.; Schneider, R.; Sander, C. *Protein Sci.* **1992**, *1*, 409−417.
(14) Abola, E. E.; Sussman, J. L.; Prilusky, J.; Manning, N. O. Protein Data Bank Archives of Three−Dimensional Macromolecular Structures. In *Methods in Enzymology*; Carter, C. W., Sweet, R. M., Eds.; Academic Press: San Diego, CA, 1997; Vol. 277, pp 556−571.
(15) Sussman, J. L.; Lin, D.; Jiang, J.; Manning, N. O.; Prilusky, J.; Ritter, E. E. *Acta Crystallogr., Sect. D* **1998**, *54*, 1078−1084.

10.1021/ja990914p CCC: $18.00
Published on Web 10/22/1999

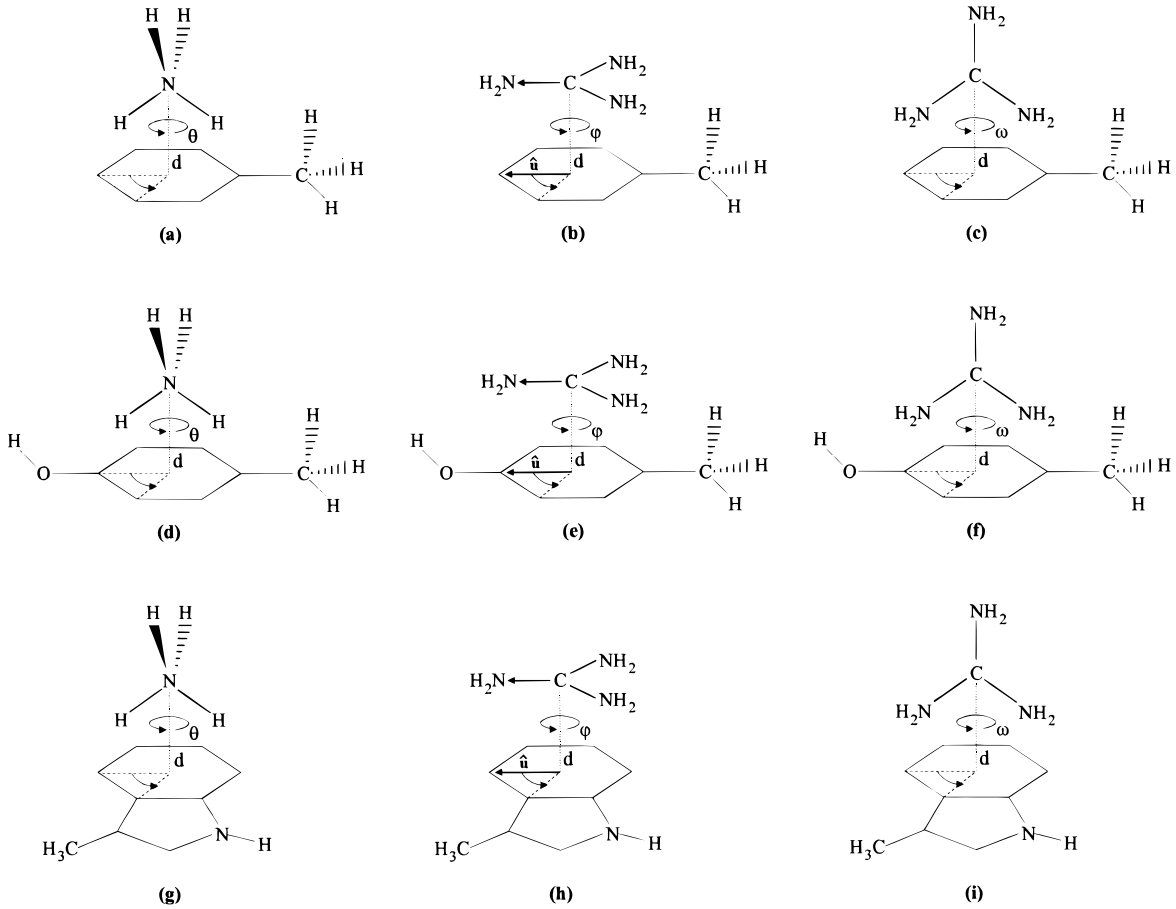

Figure 1. Structure of the cation−π complexes investigated at the ab initio, BSSE-corrected MP2/6-311++G**//MP2-6-31G** level of approximation. In (a), (d), and (g), $\theta = 0°$. In (b), (e), and (h), $\varphi = 0°$. In (c), (f), and (i), $\omega = 0°$.

and the limitations of additive molecular mechanical models are assessed.

Methods

The analysis of cation−π interactions in the Brookhaven PDB[14,15] was performed over a representative list of protein chains.[13] From the 1718 nonredundant protein structures selected in the March 1997 version of the list defined by Hobohm and Sander,[16] all possible interactions of the side chain of phenylalanine, tyrosine, or tryptophan with that of lysine or arginine were considered. In the case of lysine, together with the distance separating the centroid of the six-membered aromatic ring, the angle of approach, χ, of the onium group toward the π −electron cloud was determined. χ corresponds to the angle between the normal to the aromatic ring and the unitary vector pointing from its centroid toward the nitrogen atom of the lysine side chain. For arginine, the approach of the guanidinium moiety toward the π −electron cloud was characterized by two angles, χ_1 and χ_2. χ_1 is the angle formed by the normal to the ring and the unitary vector pointing from its centroid toward the C_ζ atom of arginine. χ_2 is the angle between the normal of the guanidinium moiety and the unitary vector pointing from its C_ζ atom toward the centroid of the aromatic ring.

To rationalize the results obtained from the examination of the Brookhaven PDB, we further investigated cation−π interactions in molecular systems of biological relevance quantum mechanically, using toluene, *p*-cresol, and methyl−indole, together with the ammonium and the guanidinium ions, as respective models of the phenylalanine, tyrosine, tryptophan, lysine, and arginine side chains (see Figure 1). A comprehensive exploration of the potential energy surface near the minimum was carried out at the ab initio MP2/6-31G** level of approximation for the six different cation−π dimers, using the Gaussian 94 suite of programs.[17] For all pairs, the approach of the ion toward the center of the aromatic ring proceeded by increments of 0.05 Å, for which the basis set superposition error (BSSE) -corrected interaction energy was evaluated. In the case of the ammonium ion, preliminary, full geometry optimizations at the MP2/6-31G** level of approximation confirmed that, regardless of the aromatic compound, the bidentate complex, in which two N−H bonds point toward the π −electron cloud, is preferred over both the mono− and the tridentate motifs. Consequently, for each distance separating the cation from the center of the aromatic ring, three possible values of the dihedral angle, θ, between the plane of the ring and that formed by the two N−H bonds pointing toward it, namely, $\theta = -60°$, 0°, and +60°, were considered. For the guanidinium ion, two possible approaches were investigated, namely a parallel one, whereby the aromatic ring and the planar cation are stacked, and a perpendicular, T-shaped one. Whereas it can be expected a priori that the perpendicular arrangement is the most favorable energetically, it has been ascertained that planar stacking motifs of arginine and aromatic side chains are common in proteins.[7,8] In the case of the parallel interaction, four possible values of the angle, φ, between the vector borne by one of the three C−N bonds of the cation and the unitary vector, $\hat{\mathbf{u}}$, pointing from the center of the aromatic ring toward one of its carbon atom, namely $\varphi = 0°$, +30°, +60°, and +90°, were considered. For the perpendicular approach, only two values of the dihedral angle, ω, formed by the plane of the ring and that of the guanidinium ion, namely, $\omega = 0°$ and +90°, were explored. At the minimum of the potential energy surface of each dimer, for the optimal orientation, a single point energy was computed at the MP2/6-311++G** level of approximation, with full counterpoise correction.[18]

(16) Hobohm, U.; Sander, C. Pdb_Select: Representative list of PDB chain identifiers, 1992 (see http://www.sander.embl−heidelberg.de/pdbsel/).

(17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al−Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head−Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*; Gaussian Inc., Pittsburgh, PA, 1995.

Considering the magnitude of the binding energy in cation−π complexes, the zero-point energy (ZPE) was not estimated.[19,20]

BSSE−corrected quantum chemical calculations are CPU intensive, which explains why the potential energy surface of the cation−π complexes was only investigated at the MP2/6-31G** level of theory. In the case of the tryptophan−arginine assembly, employing this basis set (namely, 285 functions), one point of the potential energy surface required 8.42 CPU hours on a Silicon Graphics R10000 processor (180 MHz), with 520 Mb of memory. At the MP2/6-311++G** level, for the same system, 67.47 CPU hours were necessary for the completion of one single point with counterpoise correction.

Despite the continuous decrease of the price/performance ratio of computational resources, molecular simulations of solvated protein structures still remain handled using additive potential energy functions for obvious cost-effectiveness reasons. Whereas a full treatment of multibody effects is acceptable for small systems, it becomes rapidly untractable for large molecular assemblies, like a protein in its aqueous environment, for which the exploration of sufficiently long time scales is a key issue. To probe the adequacy of additive potential energy functions to deal with cation−π interactions, we compared the above benchmark, MP2/6-311++G**//MP2/6-31G** ab initio computations to molecular mechanical energy minimizations employing three alternative, commercial force fields, namely, Amber,[21] Cvff,[22] and Cff97.[23,24]

Results and Discussion

Over the 1718 representative protein structures selected, 802 phenylalanine−lysine, 1254 tyrosine−lysine, and 415 tryptophan−lysine pairs were found at distances, *d*, separating the onium group from the centroid of the six-membered aromatic ring less than 5.0 Å. As a basis of comparison, by using the same number of protein structures, we observed 3795 phenylalanine−alanine, 1696 phenylalanine−serine, and 1100 phenylalanine−cysteine at distances between the centroid of the ring and the β-carbon, the oxygen, or the sulfur atom, respectively, less than 5.0 Å. Not too surprisingly, Figure 2 reveals that most of the pairs formed by the lysine and the aromatic side chains correspond to separations greater than 4.0 Å. Broadly speaking, one can consider that a cation−π interaction is formed when *d* is less than 3.7 Å and the angle of approach, χ, does not exceed 45°.[25,36]This reduces the number of interactions to 93, 74, and 63, for lysine bound to phenylalanine, tyrosine, and tryptophan, respectively. A more stringent restriction in the approach of the onium group toward the center of the π-electron cloud, for which χ is less than 15°, gives a population of 31, 18, and 11 cation−π interactions, respectively. The differences in the number of phenylalanine−lysine, tyrosine−lysine, and tryptophan−lysine pairs encountered in the Brookhaven PDB analysis may be ascribed to different factors. First, in the presence of a hydroxyl moiety, tyrosine is more likely to associate with lysine than phenylalanine is, hence, the larger number of pairs. It further explains why phenylalanine and lysine form more cation−π complexes than tyrosine and lysine, the onium group of the latter

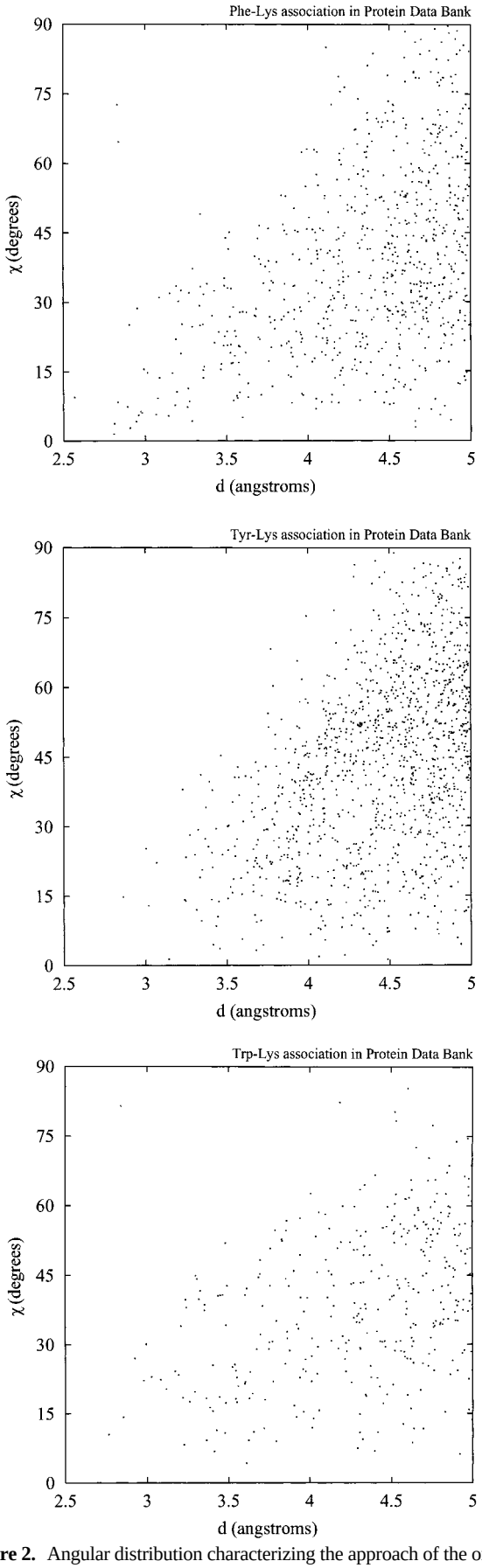

Figure 2. Angular distribution characterizing the approach of the onium group of lysine toward phenylalanine (a), tyrosine (b), and tryptophan (c), as a function of their separation, *d*. χ is the angle formed by the normal to the aromatic ring and the unitary vector pointing from its centroid toward the nitrogen atom of the lysine side chain.

(18) Boys, S. J.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.
(19) Kim, K. S.; Lee, J. Y.; Lee, S. J.; Ha, T. K.; Kim, D. H. *J. Am. Chem. Soc.* **1994**, *116*, 7399−7400.
(20) Pullman, A.; Berthier, G.; Savinelli, R. *J. Comput. Chem.* **1997**, *18*, 2012−2022.
(21) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. C.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(22) Dauber−Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 31−47.
(23) Dinur, U.; Hagler, A. T. New approaches to empirical force fields. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D. B., Eds.; VCH: New York, 1991; Vol. 2, Chapter 4, pp 99−164.
(24) Maple, J. R.; Hwang, M. L.; Jalkanen, K. J.; Stockfisch, T. P.; Hagler, A. T. *J. Comput. Chem.* **1998**, *19*, 430−458.
(25) Levitt, M.; Perutz, M. F. *J. Mol. Biol.* **1988**, *201*, 751−754.

interacting preferentially with the oxygen atom of the former. Second, the presence of the five-membered ring in tryptophan partially hinders the approach of the side chain of lysine, thereby reducing the possibility of favorable interactions.[1] In addition, there are significantly less tryptophan residues in proteins than in phenylalanine residues. In the analysis of protein structures reported here, 22096 phenylalanine, 21530 tyrosine, and 8596 tryptophan amino acids were detected. It should be noted, however, that, considering interactions at $d \leq 5.0$ Å, the ratio of tryptophan−lysine to phenylalanine−lysine pairs is about 33% larger than the ratio of tryptophan to phenylalanine residues, implying a bias in favor of tryptophan.

Similarily to the interactions of π −systems with lysine, the largest number of pairs involving arginine occurs for tyrosine. From the 1718 nonredundant protein structures, 2331 phenylalanine−arginine, 2778 tyrosine−arginine, and 1243 tryptophan−arginine pairs were found, for which the distance, d, separating the C_ζ atom of arginine from the centroid of the six-membered aromatic ring is less than 6.0 Å. It should be noted that, for all complexes, no separation smaller than 3.0 Å was observed. In Figure 3, large dots characterize small separations. In particular, it can be noted that the largest dots, for which $3.0 \leq d \leq 3.5$ Å, lie mainly in the four corners of the graphs, indicating that, at short separations, cation−π interactions involving arginine mostly correspond to stacked arrangements.[7,8] In this case, the amino units of the guanidinium moiety are free to engage in hydrogen bonding interactions with surrounding functional groups, while the side chain of arginine is conformationally restrained.[7] At increased values of d, the distributions of (χ_1, χ_2) become more uniform, with a non-negligible population of perpendicular, T-shaped motifs, namely, $\chi_1 \simeq 0°$ or 180° and $\chi_2 \simeq 90°$. Interestingly enough, the opposite T-shaped motif, for which $\chi_1 \simeq 90°$ and $\chi_2 \simeq 0°$ or 180°, appears to be more populated. Complexes in which the aromatic ring and the guanidinium moiety are coplanar are scarce and only occur at large separations.

The results of the MP2/6-311++G**//MP2/6-31G** exploration of the potential energy surfaces of the complexes formed by ammonium and guanidinium with toluene, *p*-cresol, and methyl−indole, near the minimum, are summarized in Table 1. From the onset, it can be seen that the effect of the counterpoise correction[18] on the relative binding energies, ΔE^{QM}, is substantial. For instance, in the case of methyl−indole, the BSSE lowers the interaction energy by ∼2.5 kcal/mol for ammonium, and almost 4 kcal/mol for guanidinium. Whereas counterpoise-corrected energies are generally accepted to be more reliable than uncorrected ones, there is a strong body of evidence that the former might be overestimated.[26] Another noteworthy finding concerns the magnitude of the electron correlation contribution to the binding energy. Compared to neutral molecular dimers, the presence of a positive charge magnifies the induction energy and the charge transfer in the π lowest unoccupied molecular orbital (LUMO), resulting in larger correlation effects. In the example of methyl−indole, when the same basis set is used and the BSSE is corrected, the difference between the MP2 and the Hartree−Fock (not provided in Table 1) binding energies not too surprisingly reaches −4.90, −7.75, and −6.21 kcal/mol for the interaction with ammonium and guanidinium, stacked and perpendicular, respectively. This contribution arises mainly from dispersion forces, which are expected to be larger for stacked arrangements than for T-shaped ones. Dispersion is further anticipated to be smaller for the six-

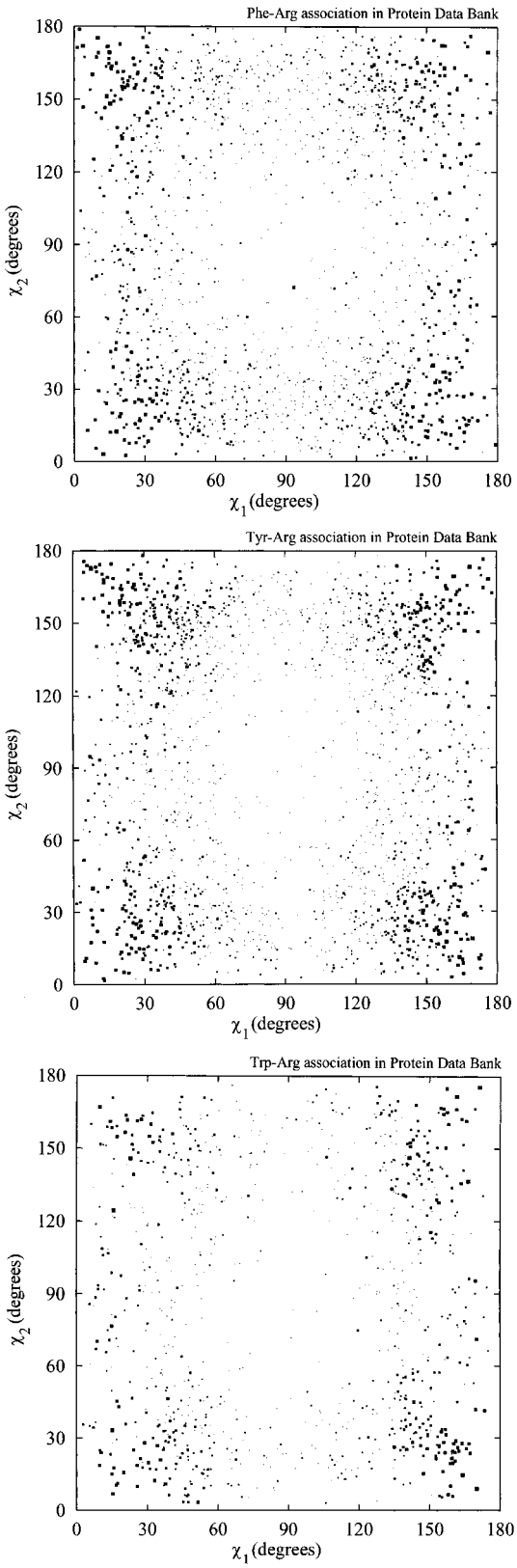

Figure 3. Correlation plot of the angles χ_1 and χ_2 at various distances, d, separating the centroid of the six-membered aromatic ring of phenylalanine (a), tyrosine (b), and tryptophan (c) from the C_ζ atom of arginine. The largest dots correspond to the smallest separations, that is, $3.0 \leq d \leq 3.5$ Å, whereas the smallest dots correspond to the largest separations, that is, $5.5 \leq d \leq 6.0$ Å. χ_1 is the angle between the normal to the ring and the unitary vector pointing from its centroid toward the C_ζ. χ_2 is the angle formed by the normal of the guanidinium moiety and the unitary vector pointing from its C_ζ atom toward the centroid of the π-electron cloud.

(26) Davidson, E. R.; Chakravorty, S. J. *Chem. Phys. Lett.* **1994**, *217*, 48−54.

Table 1. Gas Phase, Relative ab Initio Binding Energies, ΔE^{QM}, for the Different Cation−π Complexes of Toluene, *p*-Cresol and Methyl−Indole with Ammonium and Guanidinium, at the MP2/6-311++G**//MP2/6-31G** Level of Approximation

			ΔE^{QM} (kcal/mol)			
cation	aromatic compound	cation−π complex	no BSSE	BSSE correction	*d* (Å)	θ (deg)
ammonium	toluene	(a)	−19.72	−17.58	2.90	−60
	p-cresol	(d)	−19.67	−17.44	2.90	+60
	methyl−indole	(g)	−25.91	−23.41	2.85	+60
			ΔE^{QM} (kcal/mol)			
cation	aromatic compound	cation−π complex	no BSSE	BSSE correction	d (Å)	φ (deg)
guanidinium (stacked)	toluene	(b)	−9.81	−6.83	3.60	+90
	p-cresol	(e)	−10.48	−7.13	3.55	+90
	methyl−indole	(h)	−14.73	−10.77	3.45	+30
			ΔE^{QM} (kcal/mol)			
cation	aromatic compound	cation−π complex	no BSSE	BSSE correction	*d* (Å)	ω (deg)
guanidinium (T-shaped)	toluene	(c)	−16.94	−13.67	4.00	+90
	p-cresol	(f)	−17.06	−13.53	4.00	+90
	methyl−indole	(i)	−22.18	−18.23	3.95	+90

membered ring of toluene than for methyl−indole,[27] which accounts for the witnessed differences in affinity, regardless of the cation. On the basis of the statistical analysis of protein structures presented above, this result appears to be in line with the observed bias in favor of tryptophan, with respect to phenylalanine, for forming cation−π interactions with lysine, albeit a protein environment is clearly distinct from a low-pressure gaseous state characteristic of the quantum chemical calculations reported here.

An interesting feature highlighted by these quantum chemical computations is the preferred perpendicular arrangement of guanidinium−π complexes. Such a position of the cation with respect to the aromatic ring leads to an optimal overlap of the π and σ^*_{N-H} orbitals. This result is in line with the data of Duffy et al.,[28] who note that, whereas the gas phase energy minimum is characteristic of a T-shaped motif, the stacked geometry is energetically favored in an aqueous solution. Accordingly, it can be inferred that all stacked cation−π complexes found in the Brookhaven PDB probably correspond to participating amino acid residues either accessible to the solvent or located in a polar environment.

The preferential relative orientations of the cations with respect to the aromatic rings can be rationalized by simple electrostatic considerations. In the case of the three perpendicular arrangements (c), (f), and (i), the angle ω is equal to +90° to avoid unfavorable repulsion (see Figure 1). The difference in the BSSE-corrected relative binding energies between $\omega = 0°$ and +90° amounts to ∼0.4 kcal/mol for these complexes. A similar situation is observed with the stacked motifs (b), (e), and (h), for which φ is either +90° or +30°, which corresponds to a staggered position of the nitrogen atoms of guanidinium with respect to the carbon atoms of the π-systems. Differences in the relative binding energies between eclipsed and staggered arrangements are not representative, amounting to ∼0.1 kcal/mol. For the three cation−π complexes with ammonium, that is, (a), (d), and (g), the investigated values of θ always correspond to eclipsed hydrogen atoms of the ion with the carbon atoms of the aromatic rings. However, positions at +60° or −60° systematically minimize the electrostatic repulsion characteristic of $\theta = 0°$. This is particularly true for the complex involving *p*-cresol, for which the difference in ΔE^{QM} when $\theta = 0°$ and +60° amounts to ∼0.7 kcal/mol.

Assuming that (i) the repulsion and dispersion components of ΔE^{QM} sum up to zero, namely, based on molecular mechanical calculations, these components contribute to ∼0.3, 0.1, and −0.4 kcal/mol, for ammonium bound to toluene, *p*-cresol, and methyl−indole, respectively, and that (ii) both the polarization of the cation by the aromatic ring and the charge transfer between the two entities are negligible, it is possible to provide a rough estimate of the induction energy, U_{ind}, for all three ammonium−π complexes, using

$$U_{ind} \simeq \Delta E^{QM} - \sum_{l,\kappa} \tilde{Q}_l^{cation}\, T_{00,l\kappa}^{cation-\pi}\, \tilde{Q}_\kappa^{\pi} \qquad (1)$$

where $T_{00,l\kappa}^{cation-\pi}$ is a matrix element of the electrostatic tensor of order l and rank κ,[29] and $\tilde{Q}_l^{cation}$ and $\tilde{Q}_\kappa^{\pi}$ are the charge distributions of ammonium and the aromatic ring, respectively. The product $T_{00,l\kappa}^{cation-\pi}\ \tilde{Q}_\kappa^{\pi}$ represents the electrostatic potential created by the π-system, which is computed at the MP2/6-311++G** level of theory, at the position of atoms *k* pertaining to the cation. $\tilde{Q}_l^{cation}$ is approximated to the distribution of net atomic charges derived from the electrostatic potential due to the isolated ion, namely, −0.8392 for nitrogen and +0.4598 for hydrogen. It ensues that U_{ind} amounts to −6.73, −7.19, and −8.77 kcal/mol for the interaction of ammonium with toluene, *p*-cresol, and methyl−indole, respectively. Interestingly enough, Kim et al. estimated that contribution to be equal to −6.88 kcal/mol in the ammonium−benzene complex,[19] a result very close to that obtained here for the ammonium−toluene dimer.

The relative binding energies, ΔE^{MM}, reproduced by the three different force fields, for complexes (a), (c), (d), (f), (g), and (i), are reported in Table 2. It is interesting to note that, in each case, compared to the MP2/6-311++G**//MP2/6-31G** data, the Amber potential energy function of Cornell et al.[21] is the most successful in describing cation−π interactions. As expected, however, complexes involving ammonium lead to the

(27) Pullman, A.; Berthier, G.; Savinelli, R. *J. Am. Chem. Soc.* **1998**, *120*, 8553−8554.

(28) Duffy, E. M.; Kowalczyk, P. J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1993**, *115*, 9271−9275.

(29) Stone, A. J. Classical electrostatics in molecular interactions. in *Theoretical models of chemical bonding*; Maksić, H., Ed.; Springer−Verlag: Berlin, Germany, 1991; Vol. 4, pp. 103−131.

Table 2. Gas Phase, Relative Molecular Mechanical Binding Energies, ΔE^{MM}, for the Different Cation−π Complexes of Toluene, *p*-Cresol, and Methyl−Indole with Ammonium and Guanidinium, Using Alternative Potential Energy Functions[a]

cation	aromatic compound	cation−π complex	ΔE^{MM} (kcal/mol) Amber	Cvff	Cff97	ΔE^{QM} (kcal/mol)
ammonium	toluene	(a)	−14.22 (2.86)	−8.95 (3.27)	−12.03 (2.84)	−17.58 (2.90)
	p−cresol	(d)	−13.93 (2.87)	−9.97 (3.27)	−11.41 (2.90)	−17.44 (2.90)
	methyl−indole	(g)	−18.21 (2.83)	−10.15 (3.28)	−16.85 (2.81)	−23.41 (2.85)
guanidinium (T-shaped)	toluene	(c)	−13.85 (3.71)	−10.19 (4.04)	−10.00 (3.89)	−13.67 (4.00)
	p-cresol	(f)	−14.18 (3.69)	−11.70 (4.03)	−10.55 (3.92)	−13.53 (4.00)
	methyl−indole	(i)	−16.24 (3.75)	−12.46 (4.05)	−14.96 (3.85)	−18.23 (3.95)

[a] Distances separating the cation from the centroid of the aromatic ring are given in parentheses.

poorest agreement, because the localized charge in the cation implies a stronger polarization of the π-systems. In contrast, the charge in guanidinium being delocalized over the whole cation reduces induction phenomena significantly. For such systems, the Amber force field conspicuously underestimates the distance separating the carbon atom of guanidinium from the centroid of the π-electron cloud, although the accord between ΔE^{MM} and ΔE^{QM} remains reasonably good, except, perhaps, for methyl−indole. Strikingly, in the case of ammonium−π dimers, the Cvff force field[22] always overestimates the distances separating the nitrogen atom of the cation from the centroid of the aromatic ring and underestimates the interaction energies. For methyl−indole, the difference between the molecular and the quantum mechanically calculated relative binding energies reaches ∼13.3 kcal/mol, which makes this force field inappropriate for estimating protein−ligand affinities involving such cation−π interactions. The improvement of Cff97[23,24] over Cvff is glaring, with distances separating the cation from the aromatic ring close to the ab initio values. However, the difference between ΔE^{MM} and ΔE^{QM} up to ∼6 kcal/mol for ammonium−π complexes and of about 3 kcal/mol for guanidinium−π complexes suggests that this potential energy function should be recalibrated to handle cation−π interactions more accurately. Last, we note in passing that a common trait of the three force fields is their tendency to prefer monodentate complexes for all ammonium−π dimers, when quantum chemical calculations predict bidentate motifs to be energetically more favorable. This is likely to result from a slight imbalance in the Coulomb and the dispersion contributions to the potential energy.

Ideally, the accurate description of cation−π interactions in large-scale statistical simulations should include an explicit treatment of induction phenomena.[30−32] A rigorous approach for taking into account multibody effects would be to introduce in the classical potential energy function distributed polarizabilities, namely, typically charge flows and dipolar polarizabilities. Such models, however, remain complex and not very tractable for large, solvated molecular assemblies. Moreover, the additional effort involved in the determination, self-consistently, by matrix inversion or as part of an extended Lagrangian approach, of the converged induced dipole moments borne by the participating polarizable sites can increase the overall computational investment substantially. Because of their short-range nature, it is tempting to classify cation−π interactions to an unconventional type of hydrogen bonding,[33] since, in principle, such a bond can form between a donor and a π-electron density, which, in the absence of lone pairs, acts as an acceptor.[34] A simplistic analytical description similar to that employed to model hydrogen bonding[35] and consisting of a 10−12 potential, based on quantum chemical calculations,[36] has, thus, been proposed to tackle induction effects in an average fashion. Such a representation is, however, arguable. Just like for regular van der Waals interactions, an $\mathbf{r}^{-12}$ term is only ad hoc to describe the repulsion between nuclei and should be replaced by the more accurate, exponential form introduced by Buckingham. However, the most critical aspect of the 10−12 potential lies in the range of the $\mathbf{r}^{-10}$ component, way too short for characterizing accurately the polarization-driven attraction part of cation−π interactions. This contribution results from two distinct factors.[29] First, the electric field of the cation polarizes the π-electron cloud of the aromatic system, which corresponds to an $\mathbf{r}^{-2}$ interaction. In turn, the induced dipole moment of the ring interacts with the polarizing charge via its electrostatic potential, which also contributes to $\mathbf{r}^{-2}$. Formally, the attractive part of the short-range potential accounting for induction effects should be described analytically using an $\mathbf{r}^{-4}$ term. It can be shown that, in the case of the three cation−π complexes involving ammonium, the difference between the quantum mechanical and the molecular mechanical relative binding energies, that is, $\Delta E^{QM} - \Delta E^{MM}$, which roughly corresponds to the part of the induction contribution that additive force fields do not take into account, is fitted more accurately by a 4−12 correcting potential than by a 10−12 one. This procedure involved additional single point MP2/6-311++G** calculations for various separations, namely, 2.5, 4.5, 6.0, and 8.0 Å. When the Amber potential energy function is employed, the correcting term between the nitrogen atom, namely, N3, and the carbon atoms pertaining to the six-membered aromatic rings, namely, CA for toluene, CA and C for *p*-cresol, and CA, CB, and CN for methyl−indole, is the following:

$$V_{4-12}(\mathbf{r}) = 964511 r^{-12} - 144.355\, r^{-4} \qquad (2)$$

where r is the distance between the CA, C, CB, or CN atoms,

(30) Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4177−4178.
(31) Meng, E.; Cieplak, P.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 12061−12062.
(32) Cubero, E.; Luque, F. J.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5976−5980.
(33) Deakyne, C. A.; Meot−Ner (Mautner), M. *J. Am. Chem. Soc.* **1985**, *107*, 474−479.
(34) Legon, A. C.; Millen, D. J. *Acc. Chem. Res.* **1987**, *20*, 39−46.
(35) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230−252.
(36) Chipot, C.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 2998−3005.

depending on the nature of the aromatic ring, and the N3 atom. With this correction, the equilibrium distances and binding energies become 2.90, 2.88, and 2.95 Å, and −17.52, −17.64, and −21.61 kcal/mol, for the complexes of ammonium with toluene, *p*-cresol, and methyl−indole, respectively. Despite underestimated interaction distances, the good accord between the quantum mechanical and the molecular mechanical energies for the complexes formed by the three aromatic compounds and guanidinium, using the Amber force field, obviates the inclusion of a correcting potential.

It should be emphasized that short-range, correcting potentials are not intended to replace nonadditive force fields in molecular simulations. They constitute an alternative, cost-effective solution, but their validity for modeling large molecular assemblies remains problematic. In particular, one of their major drawbacks lies in the incorrect description of induction phenomena when more than one cation binds the aromatic ring. If, for instance, two positively charged ions approach the π-system, from each side, it is anticipated that the resulting induction contribution will zero out. Such is, however, not the case with a correcting potential, the two cations binding the aromatic ring similarily.

Conclusion

By using information from both experimentally determined three-dimensional structures of proteins and theoretical calculations at different levels of complexity, the present study offers new insights into the nature of cation$-\pi$ interactions in biological systems.

The structural analysis of the Brookhaven protein data bank (PDB)[14,15] reveals a considerable number of pairs formed by cationic and aromatic side chains. The number of true cation$-\pi$ interactions is, however, reduced significantly when stringent criteria characterizing the approach of the positively charged side chain toward the aromatic one are enforced. Examination of the Brookhaven PDB further indicates that cation$-\pi$ complexes involving arginine preferentially adopt stacked geometries. This result is at variance with gas phase MP2/6-311++G**//MP2/6-31G** ab initio calculations, for which the energy minima always correspond to perpendicular, T-shaped motifs. This is consistent with the fact that solvation stabilizes markedly stacked arrangements.[28]

Regardless of the cation, the most stable complexes encountered using quantum mechanical calculations recurrently involve methyl−indole, for which the MP2 correlation energy, and, hence, the dispersion contribution, is the largest.[27] The strong affinity of methyl−indole for either ammonium or guanidinium is in line with the inspection of the Brookhaven PDB, which highlights a bias in favor of tryptophan, although cation$-\pi$ interactions of the lysine or the arginine side chain with that of tryptophan are clearly less frequent in proteins than those involving phenylalanine or tyrosine. In addition to the smaller number of tryptophan residues in protein structures, compared to both phenylalanine and tyrosine, this result seems to stem from sterical reasons, the five-membered ring of tryptophan hindering partially the approach of the cationic side chain.[1]

Last, the ability of three commercial force fields to describe accurately cation$-\pi$ interactions is investigated. Even if two-body, additive potential energy functions are generally unable to reproduce faithfully this type of interaction, to this date, they probably constitute the most cost-effective approach for studying large biological systems. The Amber force field is shown to yield the best accord with the benchmark MP2/6-311++G**//MP2/6-31G** ab initio results. To enhance the accuracy of this force field, we introduced correcting parameters by means of a short-range 4−12 potential, derived from quantum mechanical computations and targeted at the reproduction of induction phenomena in an average sense. Despite its obvious limitations and shortcomings, this type of potential represents a sensible and economical solution for describing cation$-\pi$ interactions in large molecular assemblies of biological interest. It is further recommended that, as an alternative to costly, nonadditive models, such potentials be employed for studying ligand−receptor binding that involves pivotal cation$-\pi$ interactions, as it is the case in those receptors that bind acetylcholine.[4] Inaccurate calibration of force field parameters can result in severe misreproductions of the energetics associated to cation$-\pi$ interactions and lead to poor interpretations of the observed phenomena.

Acknowledgment. The authors wish to thank the Centre Charles Hermite, Vandœuvre-lès-Nancy, France, for generous allocation of computer time. H.M. acknowledges a doctoral fellowship from the Laboratoire de Chimie Théorique. C.C. gratefully thanks Drs. János G. Ángyán and Claude Millot for fruitful discussions.

JA990914P