A Computational Study of Cation−π Interactions vs Salt Bridges in Aqueous Media: Implications for Protein Engineering

Justin P. Gallivan and Dennis A. Dougherty*

Contribution from the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Received May 27, 1999. Revised Manuscript Received December 1, 1999

Abstract: A direct comparison of the energetic significance of a representative salt bridge vs a representative cation−π interaction in aqueous media and in a range of organic solvents is presented using ab initio electronic structures and the SM5.42R/HF solvation model of Cramer and Truhlar. The cation−π interaction shows a well depth of 5.5 kcal/mol in water, significantly larger than the 2.2 kcal/mol seen for the salt bridge. Consistent with this idea, a survey of the Protein Data Bank reveals that energetically significant cation−π interactions are rarely completely buried within proteins, but prefer to be exposed to solvent. These results suggest that engineering surface-exposed cation−π interactions could be a novel way to enhance protein stability.

Introduction

Of the myriad noncovalent interactions that contribute to protein stability, few are both specific and strong when fully exposed to an aqueous medium. Hydrogen bond formation, for example, is opposed by competing interactions with water, and the hydrophobic effect is not specific for any particular pair of residues. Although salt bridges are specific and quite strong in nonpolar media, there is debate over the extent to which water-exposed salt bridges contribute to protein stability.[1] Many studies estimate that a water-exposed salt bridge contributes 1−2 kcal/mol to protein stability, while others suggest a salt bridge is worth very little energetically, and may even be destabilizing. There is, however, a relatively underappreciated noncovalent binding force that is potentially both specific and strong in an aqueous environment. Recent studies have shown that cation−π interactions,[2] the electrostatically favorable attraction between a cation and a π-system, are not only quite strong in aqueous media but also commonly found in protein structures.[3a] In addition, it appears that cation−π interactions are frequently found on the surfaces of proteins, exposed to aqueous solvation.[3] Here we present a computational study of the relative strengths of salt bridges and cation−π interactions both in aqueous media and in a broad range of organic solvents. While there have been other computational studies of salt bridges[4] and of cation−π interactions[5] in water, a wide range of theoretical models of differing complexities and reliabilities has been employed. Our goals in this study are 3-fold. First, we wish to apply a state-of-the-art solvation methodology, the SM5.42R/HF method of Cramer and Truhlar,[6] to these two types of interactions. Second, we present the first direct comparison of cation−π interactions and salt bridges at the same level of theory, so as to determine the potential significance of the less-studied cation−π interaction vs the well-appreciated salt bridge. Finally, we present an analysis of the solvent exposure of energetically significant cation−π interactions taken from the Protein Data Bank (PDB).[7] Our results suggest that in water, a cation−π interaction is more stabilizing than an analogous salt bridge.

Previous studies of cation−π interactions in water have relied primarily upon force-field based methods and statistical perturbation theory.[5a−c] Although traditional two-body force field-based methods are able to reproduce the trends in cation−π binding abilities, they are challenged to accurately reproduce the results of both experiment and high-level ab initio calculations. Kollman and co-workers have attempted to improve two-body descriptions of the cation−π interaction by adding "10−12" functions to the Amber force field.[5b] Although these

(1) See, for example: (a) Matouschek, A.; Kellis, J. T., Jr.; Serrano, L.; Bycroft, M.; Fersht, A. R. *Nature* **1990,** *346,* 440−445. (b) Brown, L. R.; De Marco, A.; Richarz, R.; Wagner, G.; Wüthrich, K. *J. Biochem.* **1978,** *88,* 87−95. (c) Dao-pin, S.; Sauer, U.; Nicholson, H.; Matthews, B. W. *Biochemistry* **1991,** *30,* 7142−7153. (d) Horovitz, A.; Serrano, L.; Avron, B.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1990,** *216,* 1031−1044. (e) Serrano, L.; Horovitz, A.; Avron, B.; Bycroft, M.; Fersht, A. R. *Biochemistry* **1990,** *29,* 9343−9352. (f) Lyu, P. C.; Gans, P. J.; Kallenbach, N. R. *J. Mol. Biol.* **1992,** *223,* 343−350. (g) Sali, D.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1991,** *220,* 779. (h) Tissot, A. C.; Vuilleumier, S.; Fersht, A. R. *Biochemistry* **1996,** *35,* 6786−6794. (i) Scholtz, J. M.; Qian, H.; Robbins, V. H.; Baldwin, R. L. *Biochemistry* **1993,** *32,* 9668−9676. (j) Waldburger, C. D.; Schildbach, J. F.; Sauer, R. T. *Struct. Biol.* **1995,** *2,* 122−128. (k) Blasie, C. A.; Berg, J. M. *Biochemistry* **1997,** *36,* 6218−6222. (l) Schneider, J. P.; Lear, J. D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1997,** *119,* 5742−5743. (m) Schneider, H.-J.; Schiestel, T.; Zimmerman, P. *J. Am. Chem. Soc.* **1992,** *114,* 7698−7703. (n) Merz, A.; Knöchel, T.; Jansonius, J. N.; Kirschner, K. *J. Mol. Biol.* **1999,** *288,* 753−763.

(2) (a) Dougherty, D. A. *Science* **1996,** *271,* 163−168. (b) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997,** *97,* 1303−1324.

(3) (a) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96,* 9459−9464. (b) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994,** *235,* 709−717.

(4) (a) No, K. T.; Nam, K.-Y.; Scheraga, H. A. *J. Am. Chem. Soc.* **1997,** *119,* 12917−12922. (b) Saigal, S.; Pranata, J. *Bioorg. Chem.* **1997,** *25,* 11−21. (c) Barril, X.; Alemán, C.; Orozco, M.; Luque, F. J. *Proteins* **1998,** *32,* 67−79. (d) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994,** *3,* 211−226. (e) Honig, B.; Nicholls, A. *Science* **1995,** *268,* 1144−1149. (f) Elcock, A. H. *J. Mol. Biol.* **1998,** *284,* 489−502. (g) Luo, R.; David, L.; Hung, H.; Devaney, J.; Gilson, M. K. *J. Phys. Chem. B* **1999,** *103,* 727−736.

(5) (a) Duffy, E. M.; Kowalczyk, P. J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1993,** *115,* 9271−9275. (b) Chipot, C.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996,** *118,* 2998−3005. (c) Kumpf, R. A.; Dougherty, D. A. *Science* **1993,** *261,* 1708−1710. (d) Gao, J.; Chou, L. W.; Auerbach, A. *Biophys. J.* **1993,** *65,* 43−47. (e) Gaberscek, M.; Mavri, J. *Chem. Phys. Lett.* **1999,** *308,* 421−427. (f) Eriksson, M. A. L.; Morgantini, P.-Y.; Kollman, P. A. *J. Phys. Chem. B* **1999,** *103,* 4474−4480.

(6) (a) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998,** *288,* 293−298. (b) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999,** *99,* 2161−2200. (c) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998,** *102,* 1820−1831.
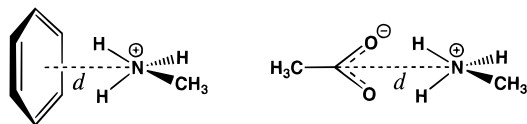
(7) http://www.rcsb.org/pdb/.

Figure 1. Geometries for methylammonium···benzene (left) and methylammonium···acetate (right). The distance, *d*, is reported in Figure 2.

functions allow the force field method to reproduce the results of high-level ab initio calculations in the gas phase, they nonetheless ignore the effects of both polarization and charge transfer, which are important in the accurate description of cation−π interactions.[8] Similarly, a previously described hybrid quantum mechanical/molecular mechanical study[5d] is compromised by the use of the AM1 Hamiltonian, which is similarly poor in treating cation−π interactions. Since ab initio calculations more accurately reproduce experimental results with regard to cation−π interactions,[2b] we anticipated that a modern implementation of aqueous solvation in a quantum mechanical calculation might be useful. The SM5.42R/HF method combines self-consistent reaction field calculations at the Hartree−Fock level with the generalized Born approximation plus surface tension terms to calculate free energies of solvation.[6] This method accurately predicts the free energies of solvation of hundreds of neutral molecules and ions, including the monomers used in this study. It has recently been used to explore solvent effects on bimolecular reactions in solution.[9] In addition, the calculations exhibit rapid convergence and proper dissociation behavior in supermolecule calculations, and they are very fast.[10] Most importantly, the method treats the intermolecular interactions at the Hartree−Fock level, thus accounting for the effects of polarization and charge transfer.[8]

Methods

All calculations were performed using Gaussian 94, Rev. D3,[11] enhanced with the Minnesota Gaussian Solvation Module Version 98.2.5 (GSM).[6a] For the methylammonium···benzene complex, the nitrogen was constrained to the C_6 axis of benzene, the complex had C_s symmetry, and the monomer geometries were fixed at their gas-phase minima. For the methylammonium···acetate complex, the distance between the N and the carboxylate carbon was varied, keeping all other parameters fixed. The solution phase minimum energy geometry was a "doubly hydrogen bonded" structure that placed the carboxylate oxygens nearly coplanar with the ammonium hydrogens (Figure 1). Consideration of other geometries gave comparable results. For solution phase calculations, supermolecule calculations with the SM5.42R/HF/6-31+G* method produced an "interaction energy", defined as

$$\text{interaction energy} = E_{ab(gas)} - E_{a(gas)} - E_{b(gas)} + \Delta G_{ab(sol)} - \Delta G_{a(sol)} - \Delta G_{b(sol)} \quad (1)$$

where $E_{ab(gas)}$ is the gas-phase energy of the complex, $E_{a(gas)}$ and $E_{b(gas)}$

are the gas-phase energies of monomers *a* and *b*, $\Delta G_{ab(sol)}$ is the free energy of solvation of the complex, and $\Delta G_{a(sol)}$ and $\Delta G_{b(sol)}$ are the free energies of solvation of the monomers *a* and *b*. To validate the performance of the method, we compared the calculated $\Delta G_{(sol)}$ for the monomers methylammonium, benzene, and acetate to the experimentally determined values in water. In all cases, these agree well with experiment. For acetate, $\Delta G_{(sol)}$calcd = −77 kcal/mol, $\Delta G_{(sol)}$expt = −77 kcal/mol; methylammonium $\Delta G_{(sol)}$calcd = −76 kcal/mol, $\Delta G_{(sol)}$expt = −73 kcal/mol; benzene $\Delta G_{(sol)}$calcd = −0.6 kcal/mol, $\Delta G_{(sol)}$expt = −0.9 kcal/mol.[6a,12] Anilinium, a cation comparable in size to the cation−π complex of Figure 1, is also well treated by the method ($\Delta G_{(sol)}$calcd = −68 kcal/mol, $\Delta G_{(sol)}$expt = −68 kcal/mol). Since the method treats all the individual molecules properly, as well as a larger molecule that contains the relevant functional groups, it is reasonable to expect good performance for the supermolecules of Figure 1. The largest discrepancy between experiment and theory involves methylammonium ion, although it should be noted that the error in *experimental* values for ions is often several kilocalories per mole. In this study, methylammonium is used in calculations for both the cation−π interaction and the salt bridge. Thus, to the extent that this value may differ from experiment, it will affect both calculations in the same way.

Calculations on protein structures were performed using the CaPTURE program (Cation−π Trends Using Realistic Electrostatics)[3a,13] modified to use the GEPOL algorithm[14] to determine the solvent accessible surface area of individual amino acids and pairs. A solvent radius of 1.4 Å was used for the probe and van der Waals radii were taken from the work of Rose et al.[15]

Results and Discussion

As a model for the cation−π interaction, we chose methylammonium ion binding to benzene (Figure 1), which mimics the interaction between lysine and phenylalanine in protein structures. For the salt bridge, we chose methylammonium ion binding to acetate, which represents the interaction between lysine and glutamate or aspartate (Figure 1). To calculate the interaction energies, we first optimized the geometries of the individual molecules in the gas phase using HF/6-31+G* calculations. Then, using fixed monomer geometries, we optimized the complexes in the gas phase at the HF/6-31+G* level. We find that gas-phase binding energies for methylammonium to benzene and acetate are −12.5 and −125.5 kcal/mol, respectively. Not surprisingly, the ion-pair is much stronger than the cation−π interaction in the gas phase.[16]

To perform calculations in solution, we used the gas-phase monomer geometries, altered the distances between the monomers (*d* in Figure 1), and performed SM5.42R/HF/6-31+G* supermolecule calculations to determine an "interaction energy" at each point. The results of these calculations with water as the solvent are shown in Figure 2.

Figure 2 clearly shows that both the salt bridge and the cation−π interaction have well-defined minima in water at this level of theory. The salt bridge experiences a broad, shallow minimum of −2.2 kcal/mol at a distance of 3.5 Å. This result is compatible with other studies, and the overall interaction surface is smooth and well-behaved, further supporting the reliability of the SM5.42R method as applied here.[17] In contrast,

(8) Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4177−4178. Cubero, E.; Luque, F. J.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5976−5980.

(9) Chuang, Y.-Y.; Radhakrishnan, M. L.; Fast, P. L.; Cramer, J.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 4893−4909.

(10) Once the gas-phase minima had been determined, all of the data in Figure 2 could be obtained in under 10 h of CPU time on a 195 MHz Silicon Graphics R10000 processor.

(11) *Gaussian 94 (Revision D.3)*; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; 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. A.; 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., Ed.; Gaussian, Inc.: Pittsburgh, PA, 1995.
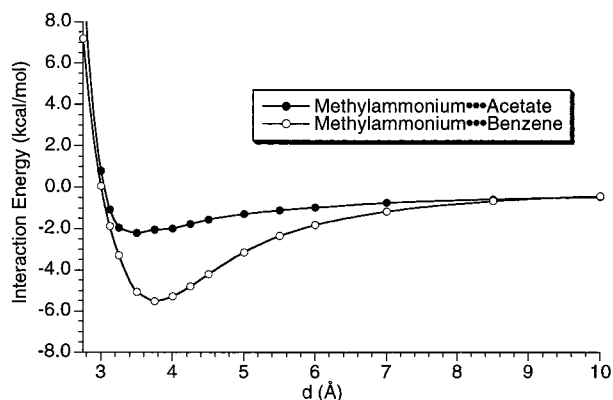
(12) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J. *Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(13) http://www.its.caltech.edu/~dadgrp/.

(14) Silla, E.; Tuñón, I.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 1077−1088.

(15) Rose, G. D.; Geselowitz, A. R.; Lesser, G. J.; Lee, R. H.; Zehfus, M. H. *Science* **1985**, *229*, 834−838.

(16) We have restricted the gas-phase geometries to exist as the ion pair. Allowing the proton to transfer to form a neutral hydrogen bonded geometry would lower the energy further. See: Zheng, Y.-J.; Ornstein, R. L. *J. Am. Chem. Soc.* **1996**, *118*, 11237−11243.

Figure 2. Interaction energies at the SM5.42R/HF/6-31+G* level for methylammonium···acetate and methylammonium···benzene in water.

Table 1. Interaction Energies (kcal/mol) at the SM5.42R/HF/6-31+G* Level for Methylammonium···Acetate and Methylammonium···Benzene in a Range of Solvents

solvent	dielectric	cation−π interaction energy	salt bridge interaction energy
none	1.0	−12.5	−125.5
CCl$_4$	2.23	−7.8	−53.4
CH$_3$CO$_2$Et	5.99	−6.2	−19.7
CH$_3$CH$_2$OH	24.85	−5.6	−5.2
CH$_3$CN	37.5	−5.6	−3.8
H$_2$O	78.0	−5.5	−2.2

the cation−π interaction experiences a substantially deeper minimum, with −5.5 kcal/mol of interaction energy at a distance of 3.75 Å. We note that HF theory underestimates the magnitude of the methylammonium···benzene interaction energy in the gas phase. In particular, the level of theory used here (HF/6-31+G*) gives a value of 12.5 kcal/mol, while the experimental ΔH is 18.8 kcal/mol.[18] As such, it may well be that the approach used here underestimates the cation−π well depth in water. While higher levels of theory perform better on the gas phase binding energy (e.g., the MP2/6-31+G* value is 17.3 kcal/mol), such methods are not presently compatible with the SM5.42R method. Since any change to a higher level of theory is expected to increase the magnitude of the cation−π interaction in water, our basic conclusion that the cation−π interaction is stronger than a typical salt bridge in water remains.

What is perhaps most striking is that in aqueous solution, the magnitude of the salt bridge interaction is reduced by over 50-fold relative to the gas phase, whereas the cation−π interaction is reduced by less than 3-fold. This suggests that cation−π interactions will remain strong across a wide range of dielectric constants. To test this notion, we performed calculations analogous to those shown in Figure 2 in differing solvents. The results of calculations using carbon tetrachloride, ethyl acetate, ethanol, and acetonitrile as solvents are shown in Table 1.

As might be anticipated, both the cation−π interaction and the salt bridge show slightly deeper minima in acetonitrile than in water. In the nonpolar solvents carbon tetrachloride and ethyl acetate, the results tend to approach the gas phase numbers,

(17) An earlier ab initio study using the PCM method[4a] concluded that this ion pair was repulsive at *all* distances, showing a destabilizing interaction of 30 kcal/mol even at essentially infinite separation. We have also seen such physically unreasonable results, as well as severe discontinuities along the path, when using the PCM method as implemented in Gaussian 94.

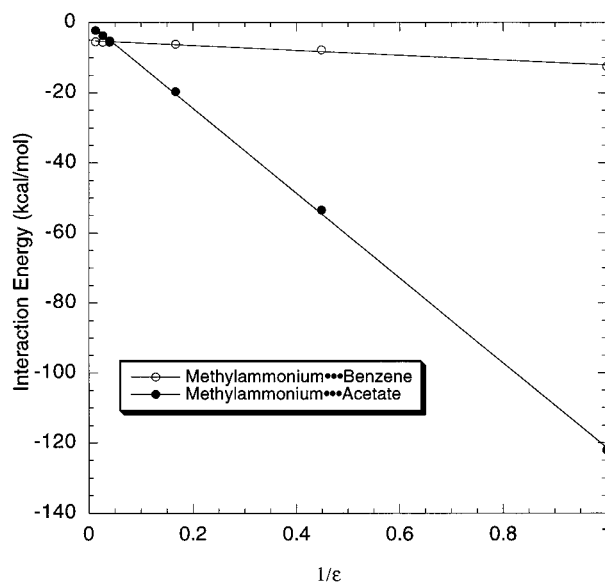(18) Deakyne, C. A.; Meot-Ner (Mautner), M. *J. Am. Chem. Soc.* **1985**, *107*, 474−479.

Figure 3. Interaction energies at the SM5.42R/HF/6-31+G* level for methylammonium···acetate and methylammonium···benzene in water, acetonitrile, ethanol, ethyl acetate, and carbon tetrachloride along with a calculation in vacuo, plotted against the inverse of the dielectric constant of the medium. *R* is 0.999 for methylammonium···acetate and 0.998 for methylammonium···benzene. The intersection of the two lines is at $\epsilon = 22.7$.

and the energy minimum of the salt bridge is substantially deeper than that of the cation−π interaction. Interestingly, the minima for both the cation−π interaction and the salt bridge are very similar in ethanol. A plot of the dielectric dependence of the binding energies is shown in Figure 3. Clearly variations in the dielectric constant of the medium are the overwhelming contributor to variations in the strength of the interactions. This is to be expected because of the dependence of the Born solvation model as implemented here on the dielectric constant of the medium. Nonetheless, Figure 3 predicts that the strengths of the cation−π interaction and the salt bridge are approximately equal in a solvent with a dielectric constant of 22.7, slightly less polar than ethanol and slightly more polar than acetone (ϵ = 20.7).

Given these results, it is perhaps tempting to suggest that the minima in Figure 2 and Table 1 may represent ΔG for the association of the two partners.[19] This would suggest that these interactions have association constants in the range observable by methods such as NMR. However, to obtain a better estimate of the association constant of the binding process, one should use eq 2 to integrate the potential of mean force (PMF) curves of Figure 2.[5a]

$$K_a = 4\pi \int_0^{r_{cut}} r^2 e^{-w(r)/kT}\, dr \qquad (2)$$

This approach has shortcomings, because the curves above were derived using only a very small number of configurations, rather than by sampling the large number of configurations typically examined in a statistical perturbation study. Because of the limited number of configurations sampled, any estimate of the association constant will be too favorable, and will thus represent an upper limit to the magnitude of the interactions. The calculated association constant for the cation−π interaction between methylammonium and benzene in water is 401 M^{-1},

(19) Previous workers have suggested that this interaction energy approximates ΔG_{bind}. See: Lamb, M. L.; Jorgensen, W. L. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449−457.

while that for the salt bridge between methylammonium and acetate is 6 M^{-1} using in both instances a value of 10 Å for r_{cut}. This suggests that the upper limit for ΔG_{bind} for the cation−π interaction is approximately −3.6 kcal/mol, while that of the salt bridge is approximately −1.0 kcal/mol. We again stress that these values must be considered with caution because of the limited number of configurations and the symmetry constraints imposed.[20] We do note that these numbers are consistent with earlier theoretical and computational studies that employed similar constraints. In particular, the value of ΔG_{bind} for the cation−π interaction (−3.6 kcal/mol) agrees remarkably well with the value of −3.2 kcal/mol determined by Kollman and co-workers for a similarly constrained simulation.[5b] In addition, the value for ΔG_{bind} for the salt bridge is in good agreement with experimental values ranging from −0.2 to −1.25 kcal/mol for surface exposed salt bridges determined by Fersht and co-workers.[1d] This agreement with both experiment and theory further demonstrates the applicability of the SM5.42R/ HF method for calculating weak interactions in solution.

We propose the following simple explanation for why the formation of a cation−π interaction may be favored over the formation of a salt bridge in water. To form an ion pair in water, two well-solvated ions must become desolvated to a considerable extent. The binding energy of the complex may be very strong, but the desolvation penalty may be of comparable magnitude. In addition, although the ion pair complex is strongly polar, to some extent there is a neutralization of charge in bringing two opposite charges in close contact. This will diminish the long range, Born solvation of the ion pair relative to the isolated ions. To form a cation−π interaction, a cation must pay a similar penalty to partially desolvate. However, the aromatic partner is poorly solvated by water, and so the desolvation associated with cation binding is most likely beneficial. In addition, no "charge neutralization" occurs, and so the cationic complex can still reap the substantial benefits of long-range solvation by water.

The possibility that cation−π interactions are strong and specific in aqueous media suggests that exposed cation−π interactions may also contribute to protein structure and stability. We recently reported an extensive study demonstrating that cation−π interactions are quite common within proteins.[3a] In that study it was suggested, though not proven, that cation−π interactions tend to occur on the surfaces of proteins, rather than buried in the cores. To more rigorously explore this hypothesis in the context of the present paper, we have determined the extent to which cation−π interactions are exposed to water in protein structures. To do this, we implemented within the CaPTURE (Cation−π Trends Using Realistic Electrostatics) program[3a,13] a variant of the GEPOL method[14] to calculate the solvent accessible surface area (SASA) of individual interactions. Traditionally, calculation of relative surface accessibility of amino acids has relied on comparing the SASA of an individual amino acid within the protein to the SASA of the same amino acid in a GXG peptide (where G is glycine and X is the amino acid). However, in a cation−π interaction, since the interacting partners are in contact with one another, their surface exposed area is by definition less than that of the amino acid in a GXG peptide, even though the interacting pair *as a unit* may be highly exposed. To overcome this difficulty, we have chosen to calculate the accessibility of a cation−π
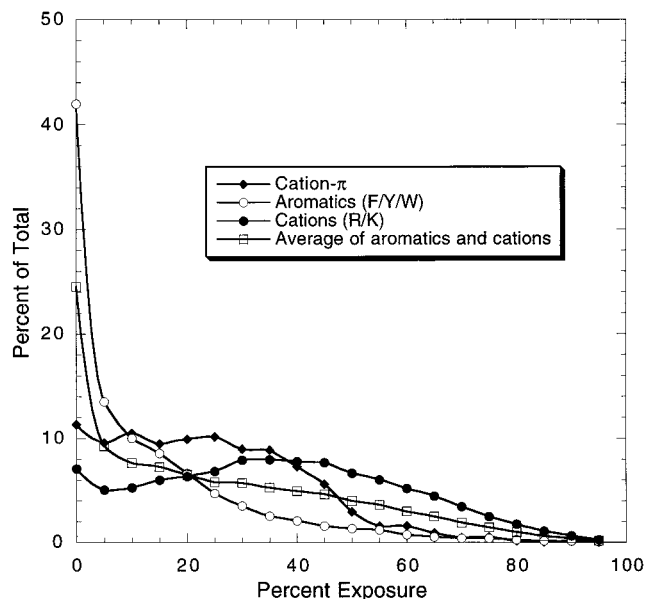
(20) For example, in a study of a similar system using molecular mechanics methods, Kollman found that relaxing the constraints of the sort involved here diminished the binding constant by a factor of 30.[5b] The key point is that all these changes should have comparable effects on the cation−π interaction and the salt bridge, and so our essential conclusion that the former is stronger remains.
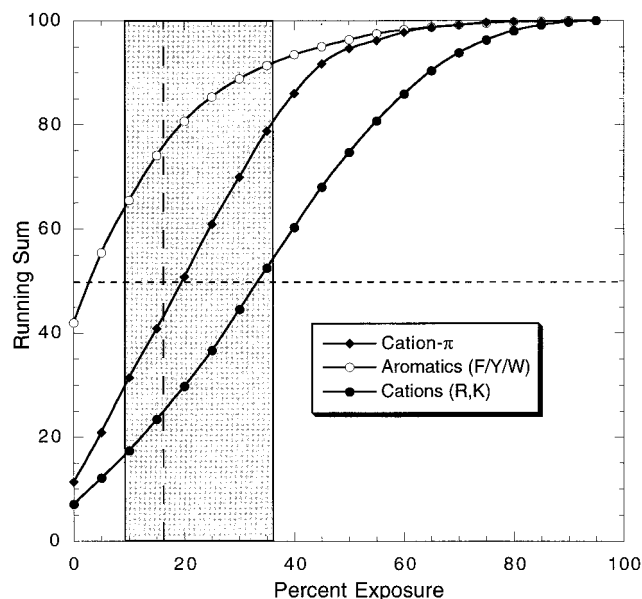

Figure 4. The percentage of the total number of residues (interactions) that have a given relative exposure. Abbreviations: F = phenylalanine, Y = tyrosine, W = tryptophan, R = arginine, K = lysine.

interaction considered as a unit within the protein, relative to the accessibility of the amino acids in their identical orientation, but not in the context of the protein. To do this, the solvent accessible surface area for each partner in the cation−π interaction is calculated. The total exposed area of the two partners is then compared to the maximum exposed area of the two partners. This is determined by "excising" the interaction from the rest of the protein. The coordinates of the amino acids involved in the cation−π interaction, as well as the backbone coordinates of the (*i−1*) and (*i+1*) residues of both the cation and the aromatic are retained, and the rest of the atoms in the protein are deleted. The solvent accessible surface area of the cation−π interaction in this minimal model is calculated. This represents the maximum possible surface area of the cation−π interaction taken *as a unit*. The ratio of the exposed area of the cation−π interaction within the protein to the maximum area of the cation−π interaction considered by itself can be expressed as a "percent exposed". Using this method, we can also determine the "percent exposed" value for each amino acid within the dataset.

As a dataset of proteins, we chose the 593 nonhomologous proteins evaluated in our earlier study.[3a] The exposure of each amino acid, as well as 2878 energetically significant cation−π interactions, was calculated using the method outlined. The results are shown in Figure 4.

The results in Figure 4 are largely as expected for both the cationic and aromatic amino acids. Phe, Tyr, and Trp tend to be buried within the proteins with over 40% of the aromatics completely buried. In addition, the aromatic amino acids are rarely highly exposed, only 20% have at least 20% of their surfaces exposed to water. Not surprisingly, the cationic amino acids Lys and Arg tend to be exposed, 70% of all cationic amino acids expose more than 20% of their surfaces to water, and much less than 10% of these residues are completely buried. It might be anticipated that cation−π interactions might resemble the average of the cationic and aromatic amino acid curves in Figure 4. However, this is not the case. Instead, cation−π interactions tend to behave as a hybrid between the cationic and aromatic amino acids. On one hand, cation−π interactions are rarely completely buried within proteins. Only 11% of cation−π

Figure 5. Integration of the data in Figure 4 for cationic amino acids, aromatic amino acids, and cation$-\pi$ interactions. The vertical dotted line represents the 16% criterion in the two-state model. The shaded box represents the "intermediate" state of the three-state model (9% < exposure < 36%).

interactions are completely buried, compared to 7% for cationic amino acids and 42% for the aromatics. In this regard, they tend to behave more like cations than like aromatics. In contrast, while cation$-\pi$ interactions are more often highly exposed than are aromatics, they are certainly less often exposed than are the cationic amino acids. This is perhaps not surprising as a cation$-\pi$ interaction is by definition more hydrophobic than a cationic amino acid.

Nevertheless, the question of whether cation$-\pi$ interactions are "exposed" still remains. The choice of what degree of exposure constitutes "exposed" is somewhat arbitrary. Previous workers have offered varying criteria for exposure. In particular, Rost and Sander have suggested several models.[21] The first is a two-state model in which amino acids whose exposed area is greater than 16% are considered exposed, while those less than 16% are considered buried. The second is a three-state model in which amino acids whose exposed area is <9% are considered buried, those between 9 and 36% exposure are intermediate, and those >36% are considered exposed. Figure 5 integrates the data for cationic amino acids, aromatic amino acids, and cation$-\pi$ interactions shown in Figure 4. The vertical dotted line represents the 16% criterion in the two-state model: points to the right of this line are exposed, points to the left are buried. The shaded box represents the "intermediate" state of the three-state model (9% < exposure < 36%). Points to the left of the box are considered buried and those to the right are considered exposed in the three-state model. To determine the nature of a particular interaction, we define the PE_{50} as the percent exposure at which 50% of the amino acids are more exposed than this

point and 50% are less exposed. For a particular curve, PE_{50} is the intersection between it and the horizontal dotted line in Figure 5. PE_{50} is 3% for the aromatic amino acids, 20% for the cation$-\pi$ interactions, and 34% for the cationic amino acids. Using the two-state model suggests that the average aromatic amino acid is buried, and that the average cation$-\pi$ interaction or cationic amino acid is exposed. Using the three-state model, aromatic amino acids are still considered buried, whereas both cation$-\pi$ interactions and cationic amino acids are considered "intermediate", although the cationic amino acids have a higher PE_{50} than the cation$-\pi$ interactions. Thus, using two different sets of criteria for determining exposure, the cation$-\pi$ interaction falls into the same broad classification as the cationic amino acids. Regardless of the particular (arbitrary) cutoffs, the trend is clear, cation$-\pi$ interactions tend not to be buried, reflecting their cationic character, but they are not as exposed as cationic residues by themselves, reflecting their aromatic component.

Conclusion

These results represent the state-of-the-art in continuum calculations and compare, at a consistent level of theory, the strengths of salt bridges and cation$-\pi$ interactions in water and in a range of organic solvents. The SM5.42R/HF method is well behaved for this problem, and while future efforts may refine the present results, we believe the basic trend will survive. In particular, we suggest that in water, a cation$-\pi$ interaction is more stabilizing than an analogous salt bridge. Not surprisingly, the strengths of the interactions are strongly dependent on the dielectric constant of the medium, as would be expected for interactions with large electrostatic components. Interestingly, the cation$-\pi$ interaction maintains its strength across a range of solvents. Whereas the strength of the salt bridge is reduced over 50-fold on moving from the gas phase to water, the strength of a cation$-\pi$ interaction is weakened by less than a factor of 3. Thus, cation$-\pi$ interactions can provide strong, specific interactions on the surfaces of proteins. Consistent with this prediction, we find that cation$-\pi$ interactions are rarely buried within proteins in the Protein Data Bank, and more often than not, are exposed to water.

These results suggest a novel mechanism to increase the stability of proteins. Most efforts at protein design have emphasized optimization of packing in the core of the protein.[22] Engineering solvent exposed cation$-\pi$ interactions, especially at the interfaces between domains or subunits, may provide a complementary strategy.

(21) Rost, B.; Sander, C. *Proteins* **1994**, *20*, 216−226.

(22) See for example: Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87.