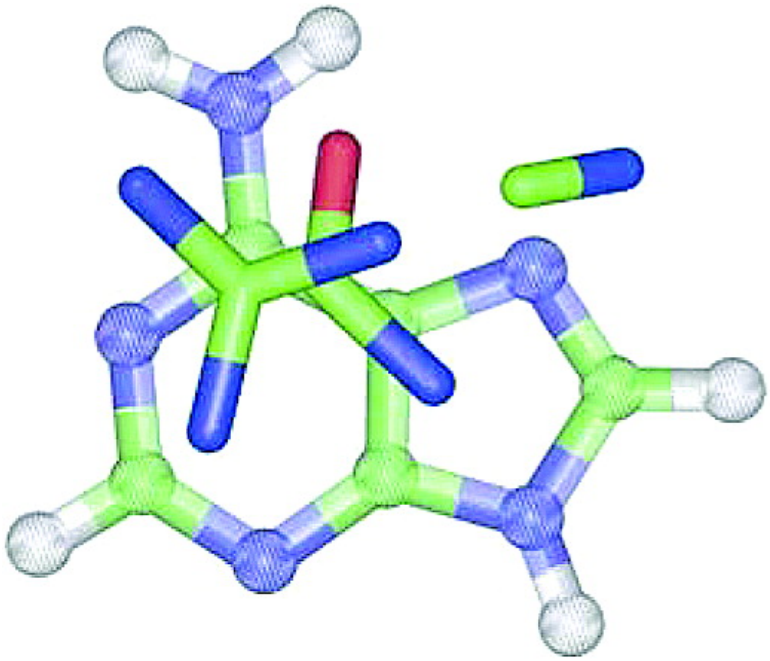
Article

Free-Energy Calculations of Protein–Ligand Cation–π and Amino–π Interactions: From Vacuum to Proteinlike Environments

Christophe Biot, Eric Buisine, and Marianne Rooman

J. Am. Chem. Soc., **2003**, 125 (46), 13988-13994• DOI: 10.1021/ja035223e • Publication Date (Web): 25 October 2003



J|A|C|S
ARTICLES
Published on Web 10/25/2003

Free-Energy Calculations of Protein−Ligand Cation−π and Amino−π Interactions: From Vacuum to Proteinlike Environments

Christophe Biot,*,† Eric Buisine,§ and Marianne Rooman†

Contribution from the Ingénierie Biomoléculaire, Université Libre de Bruxelles, CP 165/64, 50 Avenue Roosevelt, B-1050 Bruxelles, Belgium, and Laboratoire de Chimie Organique et Macromoléculaire, UPRESA-CNRS 8009, Ecole Nationale Supérieure de Chimie de Lille, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq Cedex, France

Received March 19, 2003; E-mail: cbiot@ulb.ac.be

Abstract: To probe the role of cation−π and amino−π interactions in the context of protein−ligand interactions, the stability of 55 X-ray cation/amino−π motifs involving the Ade moieties of cofactor molecules and Arg, Lys, Asn, or Gln side chains of their host protein was evaluated using quantum chemistry calculations. The conjunction of vacuum interaction energies, vibrational entropy, and solvation contributions led to identify Arg−Ade as the most favorable cation/amino−π complex in the solvents considered, followed by Asn/Gln−Ade and Lys−Ade: their minimum interaction free energies are approximately equal to −7, −4, and −2 kcal/mol, respectively, in the solvents of dielectric constant similar to that estimated for proteins (i.e., acetone, THF, and CCl_4). Remarkably, these free-energy values of cation/amino−π interactions correlate well with their frequency of occurrences in protein−ligand structures, which corroborates our approach in the absence of experimental data.

Introduction

Probing the noncovalent interactions that determine the three-dimensional structure of a protein and its interactions with other molecules—natural or synthetic ligands, DNA, or proteins—is of primary importance for a large series of applications ranging from protein design to drug discovery. These interactions are basically, but not fully, understood. In particular, their relative weight in protein environments and in molecular recognition remain far from settled. Moreover, the role of cation−π interactions between aromatic rings and positively charged groups has only recently started to be appreciated in the biomolecule context.[1−9]

Experimental and in silico studies have emphasized the frequent occurrence of cation−π interactions in proteins, where they are preferentially located near the surface[10] or across protein−protein interfaces or zip up domain-swapped oligomers.[11] Related interactions, sometimes termed amino−π or polar−π but classed in what follows with cation−π interactions, involve amino acids carrying a partial positive charge on their side-chain amino group (Asn and Gln).[12,13] Cation−π interactions have also been observed in several biomolecular association processes such as ligand-antibody binding and receptor−ligand interactions.[6,14,15] More recently, cation−π interactions involving the aromatic rings of nucleic acid bases and amino acids with a net positive charge (Arg or Lys) or a partially charged group (Asn or Gln) have been shown to be quite common at the interface between protein and DNA, where they have been suggested to play a role in the specificity of protein−DNA recognition and in the charge transport known to occur through double-stranded DNA.[16,17]

Moreover, cofactor molecules containing nucleic acid bases, such as ADP/ATP and GDP/GTP, the cell's most important energy source, or NAD and FAD, involved in electron transfer, frequently feature cation−π interactions with their host protein.[18] Ab initio quantum chemistry energy calculations in a vacuum,

† Université Libre de Bruxelles.
§ Université des Sciences et Technologies de Lille.

(1) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303−1324.
(2) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459−9464.
(3) Paschall, C. M.; Hasserodt, J.; Jones, T.; Lerner, R. A.; Janda, K. D.; Christianson, D. W. *Angew. Chem., Int. Ed.* **1999**, *38*, 1743−1747.
(4) De Wall, S. L.; Meadows, S. L.; Barbour, L. J.; Gokel, G. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6271−6276.
(5) Pletneva, E. V.; Laederach, A. T.; Fulton, D. B.; Kostic, N. M. *J. Am. Chem. Soc.* **2001**, *123*, 6232−6245.
(6) Niedzwiecka, A.; Marcotrigiano, J.; Stepinski, J.; Jankowska-Anyszka, M.; Wyslouch-Cieszynska, A.; Dadlez, M.; Gingras, A. C.; Mak, P.; Darzynkiewicz, E.; Sonenberg, N.; Burley, S. K.; Stolarski, R. *J. Mol. Biol.* **2002**, *319*, 615−635.
(7) Chakravarty, S.; Varadarajan, R. *Biochemistry* **2002**, *41*, 8152−8161.
(8) Gromiha, M. M.; Thomas, S.; Santhosh C. *Prep. Biochem. Biotechnol.* **2002**, *32*, 355−362.
(9) Gromiha, M. M. *Biophys. Chem.* **2003**, *103*, 251−258.
(10) Gallivan, J. P.; Dougherty, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 870−874.
(11) Dehouck, Y.; Biot, C.; Gilis, D.; Kwasigroch, J. M.; Rooman, M. *J. Mol. Biol.* **2003**, *330*, 1215−1225.
(12) Burley, S. K.; Petsko, G. A. *FEBS Lett.* **1986**, *203*, 139−143.
(13) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315−331.
(14) Stauffer, D. A.; Karlin, A. *Biochemistry* **1994**, *3*, 6840−6849.
(15) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12088−12093.
(16) Wintjens, R.; Liévin, J.; Rooman, M.; Buisine, E. *J. Mol. Biol.* **2000**, *302*, 395−410.
(17) Rooman, M.; Lievin, J.; Buisine, E.; Wintjens, R. *J. Mol. Biol.* **2002**, *319*, 67−76.
(18) Biot, C.; Buisine, E.; Kwasigroch, J. M.; Wintjens, R.; Rooman, M. *J. Biol. Chem.* **2002**, *277*, 40816−40822.

10.1021/ja035223e CCC: $25.00

performed at the second order of Møller−Plesset (MP2) perturbational level of theory, indicated that a nucleic base and an amino acid possessing a fully or partially charged group form favorable cation−π complexes. The strength of these associations strongly depends on the type of nucleobase and the position of the amino acid above the aromatic cycle. These results, as well as the conservation of cation−π interactions in families of related proteins and the recurrent occurrence of specific cation−π patterns in unrelated protein sequences, concur to suggest that these interactions could play an important role in protein−ligand structure, stability, and molecular recognition.

Here we extend this analysis by investigating the contributions of solvation, zero-point energies (ZPE), and atomic vibrations to the interaction free energy of protein−ligand cation−π pairs. We focus on cation−π interactions between an Ade base, by far the most frequent ligand building block, and Arg, Lys, Asn, and Gln side chains. In view of probing the stability of cation−π interactions in protein environments, water and four organic solvents are considered: DMSO, acetone, THF, and CCl_4. Their contribution to the free energy of cation−π formation is evaluated using two different continuum solvent models, IEF−PCM and SM5.4/A.

Methods

Set of Protein−Ligand Cation−π Interactions. The ensemble of cation−π interactions between an Ade moiety included in a protein ligand and an amino acid side chain carrying a net positive charge (Arg, Lys) or a partial positive charge on its amino group (Asn, Gln) was inherited from previous work[18] and is listed in the Supporting Information. Note that the cation−π interactions with Asn or Gln are usually called amino−π or polar−π interactions, but are here for simplicity grouped with cation−π interactions, though they involve only a partial positive charge. In summary, a nonredundant set of 188 high-resolution X-ray structures of protein−ligand complexes was searched for cation−π interactions linking the ligand to the protein and yielded 57 Ade-involving cation−π interactions. Cation−π interactions were identified according to a distance and an angle criterion. The distance criterion required that at least one of the atoms of the aromatic ring be located no further than 4.5 Å from one of the atoms carrying the positive charge. The angle criterion demanded the latter atom to be situated above the plane defined by the aromatic ring, more precisely, inside a cylinder of height 4.5 Å, whose base included the ring and had a radius equal to the ring diameter.

The 57 cation−π pairs were simplified for computational study. Each ligand was reduced to its Ade base. Lys was represented as an ammonium, Arg as a guanidinium, and Gln and Asn as a formamide group. The H-atoms were added by construction, which was unique except for Lys. In this case, one of the H-atoms of the ammonium group was positioned along the N_ζ-C_ϵ axis, but there was an indeterminacy for the three others due to rotational symmetry. Accordingly, we considered two different geometries. In the first, one of the three remaining H-atoms was positioned as close as possible to the center of the aromatic ring, considering the constraint induced by the first H-atom. In the second, one of the three remaining H-atoms was positioned as far as possible from the center of the aromatic ring. The latter two H-atoms were then unambiguously fixed.

Ab Initio Quantum Chemistry Energy Calculations. All ab initio energy calculations were carried out using the Gaussian 98 suite of programs.[19] Since crystal structures sometimes display unrelaxed intramolecular geometries yielding distorted wave functions and wrong energies, we first replaced the crystal coordinates of the individual cation−π partners by coordinates optimized at the Hartree−Fock (HF) level and the 6-31G** basis set. The independently optimized cation−π partners were then superimposed onto the original crystal structure using the U3BEST algorithm.[20]

In a second stage, the gas-phase interaction energies of cation−π systems were calculated at the MP2 perturbation theory[21,22] as the sum of the HF interaction energy ΔE_{HF} and the electron correlation energy ΔE_{Cor}:

$$\Delta E_{MP2} = \Delta E_{HF} + \Delta E_{Cor}$$

The interaction energy ΔE is defined as the difference between the energy of the complex A−B and the energies of the isolated partners, i.e., $\Delta E = E(A-B) - E(A) - E(B)$.

The 6-31G**(0.2) basis set was used for computing the interaction energies. It corresponds to the standard 6-31G** basis set, augmented by a Gaussian α_d-exponent equal to 0.2 on the heavy atoms C, N, and O. It has indeed been shown that this extended description of the d-polarization functions allows a more accurate description of cation−π interaction energies, comparable to that obtained with more extended basis sets.[23,24] The standard counterpoise (CP) method was applied to correct interaction energies for the basis set superposition error (BSSE).[25,26]

Normal modes of vibration of cation−π pairs were determined at the HF/6-31G**(0.2) level of theory to evaluate the ZPE, the thermal corrections to the energy E_{th}, and the entropy S_{gas}. The gas-phase interaction free energy of the complexes, ΔG_{gas}, is given by:

$$\Delta G_{gas} = \Delta E_{MP2} + \Delta ZPE + \Delta E_{th} - T\Delta S_{gas}$$

The temperature is taken to be $T = 298.15$ K. We assume that the gas-phase entropy and thermal energy corrections can be restricted to vibrational motions: $\Delta E_{th} \approx \Delta E_{vib}$ and $\Delta S_{gas} \approx \Delta S_{vib}$. Indeed, the change in energy and entropy due to rotational and translational degrees of freedom upon formation of the cation−π complex is small in a folded protein environment.[27,28] It is, moreover, reasonable to assume that the amount of rotational and translational entropy lost upon cation−π formation within a protein, after all other neighboring interactions have been formed, is roughly identical for all complexes.

Note that the normal mode computations, in principle, require the systems to be at an energy minimum or saddle point. Here, the separate cation−π entities are optimized but the complexes are not; the reason for this choice is discussed at the end of this section. Using Taylor expansions of energy, we can verify straightforwardly that the error on the normal-mode frequencies is, in first approximation and supposing $\partial^2_x E_{HF}$ positive, proportional to $\Delta x\ \partial^3_x E_{HF}/(\partial^2_x E_{HF})^{1/2}$, where Δx is the difference between the coordinates of the optimized and nonoptimized conformations and the derivatives are taken at the nonoptimized point. Hence, considering non-fully optimized geometries comes to assume

(19) 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. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.
(20) Kabsch W. *Acta Crystallogr.* **1978**, *A34*, 827−828.
(21) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622.
(22) Head-Gordon, M.; Pople, J. A.; Frisch M. J. *Chem. Phys. Lett.* **1988**, *153*, 503−506.
(23) Wintjens, R.; Biot, C.; Rooman, M.; Liévin, J. *J. Phys. Chem. A* **2003**, *107*, 6249−6258.
(24) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247−3276.
(25) Jansen, H. B.; Ros, P. *Chem. Phys. Lett.* **1969**, *3*, 140−143.
(26) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.
(27) Jusuf, S.; Loll, P. J.; Axelsen, P. H. *J. Am. Chem. Soc.* **2002**, *124*, 3490−3491.
(28) Jusuf, S.; Loll, P. J.; Axelsen, P. H. *J. Am. Chem. Soc.* **2003**, *125*, 3988−3994.

that this error is small in comparison with the actual frequencies. We would, moreover, like to emphasize that the vibrational entropy S_{vib} and especially the vibrational energy E_{vib} are relatively insensitive to errors on the normal-mode frequencies.[29]

Finally, the interaction free energy of the cation−π pairs in the presence of a solvent, noted $\Delta\mathbf{G}$, was evaluated as:

$$\Delta\mathbf{G} = \Delta G_{gas} + \Delta\Delta G_{solv}$$

where the free-energy difference $\Delta\Delta G_{solv}$ is defined as the difference between the solvation energy of the complex A−B and that of the separate partners: $\Delta\Delta G_{solv} = \Delta G_{solv}(A-B) - \Delta G_{solv}(A) - \Delta G_{solv}(B)$. Note that the $\Delta\Delta G_{solv}$ term contains both energetic and entropic contributions.

Two different formalisms were used to evaluate the solvation free energies. The former is the integral equation formalism version (IEF) of the polarized continuum model (PCM) implemented in the Gaussian 98 program.[30,31] It is a continuum solvation model in a quantum mechanical framework, where the solvent is mimicked by a polarizable continuum surrounding a cavity having the shape and dimension of the solute molecule. The cavity is described by interlocking spheres centered on solute atoms; we used the default values for the atomic radii (UATM), multiplied by a default factor (1.2 or 1.4 according to the solvent) that accounts for the fact that the distance between the solvent and solute atoms is normally somewhat larger than the van der Waals radii.

The IEF−PCM calculations were performed at the HF/6-31G**(0.2) level. In principle, they should be performed up to the MP2 level[32] similar to the gas-phase calculations. More precisely, the HF wave function perturbed by the presence of the solvent should be used to estimate the electronic correlation contributions ΔE_{Cor}. To check the validity of restricting the solvation effects to the HF level, we calculated the ΔE_{Cor} interaction energies in gas phase and water for all Arg−Ade complexes and found that they differ by only 0.1 kcal/mol on the average. We thus chose the hybrid approach consisting of performing MP2-level calculations in gas phase and HF-level calculations in solution. This allows a gain of a factor of 5 of computational time and to maintain the BSSE corrections on the gas-phase energy contributions.

The second formalism we used to evaluate the solvation free energies was the SM5.4/A model[33,34] implemented in the Linux version of the Spartan 02 program.[35] Note that this version only allows water as solvent. It also considers the solvent as a continuum and divides the free energy of solvation into two contributions. The first includes the change in the solute's internal free energy upon insertion in the solvent and the solute−solvent electrostatic interactions. It is evaluated using a self-consistent reaction field model and charges derived from AM1 wave functions.[36] The second contribution is semiempirical and accounts for first solvation shell effects.

We chose not to optimize the (free) energy of the cation−π complexes—only that of the separate molecular groups—for several reasons. Full geometry optimization in a vacuum has no biological sense because it neglects the protein and solvent environment and leads to completely unrealistic structures. In addition, the BSSE correction is not well defined in this approach. In principle, the cation−π systems may be optimized in the presence of solvent with the IEF−PCM method, but this procedure is quite computer time-consuming and must be repeated for each solvent. Furthermore, it gives rise to energy convergence problems and sometimes yields unexpected planarity distortions. Such optimizations are, moreover, only possible at the HF level of theory; the MP2 contributions are indeed calculated from the solvent-corrected HF wave functions. The (free) energy of the so-optimized structures have optimal electrostatic components but usually poor dispersion contributions, which again is unrealistic.

Results

We focused on the 57 cation−π interactions between an Arg, Lys, Asn, or Gln side chain and an Ade moiety included in a protein ligand, identified in X-ray structures of protein−ligand complexes (see Methods section). Among these, 38 involve an Arg, 7 a Lys, 6 an Asn, and 6 a Gln residue. Two of the Asn−Ade complexes (between ATP and N175 in 1A82 and between NAD and N211 in 1DXY) are at the limit of our cation−π definition criteria and form an H-bond in addition to a cation−π interaction. As we are only interested in cation−π interactions, these two complexes are dropped, reducing the number of cation−π interactions to 55.

In the large majority of the Arg− and Asn/Gln−Ade cation−π pairs, the guanidinium or formamide planes are approximately parallel to the Ade plane. Indeed, the angles between these planes exceed 45° in 7 of the 38 Arg−Ade pairs and in 2 of the 10 Asn/Gln−Ade only. This is in agreement with previous findings showing that the parallel orientation of guanidinium groups with aromatic moieties is preferred to the T-shape both in the protein structure context[13,37] and in aqueous solution.[38] In contrast, the perpendicular or T-shaped configuration has been shown to be the most stable in gas phase by HF- and MP2-level calculations using medium-size basis sets.[2,39] It can, however, been argued that these descriptions are not sufficient to account for stacking interactions.[23,40] Alternatively, it can be argued that in proteins and water the higher energy of the parallel conformation is compensated by a better interaction network with the surroundings.

The gas-phase interaction free energies ΔG_{gas} of the 55 X-ray cation−π complexes, including ZPE and vibrational energy and entropy contributions, were estimated by means of ab initio calculations and summarized in Table 1. Table 2 contains the interaction free energies $\Delta\mathbf{G}$ of the cation−π pairs immersed in various solvents, i.e., water, DMSO, acetone, THF, and CCl_4. Detailed (free) energy values are available in the Supporting Information.

Gas-Phase Interaction Free Energies. Interaction energy calculations at MP2/6-31G**(0.2) level of theory show the favorable nature of the considered cation−π interactions in a vacuum, with average ΔE_{MP2} energies ranging from −1.4 to −5.6 kcal/mol (Table 1). The Lys−Ade interactions are essentially due to electrostatic forces ($\langle\Delta E_{HF}\rangle < 0$; $\langle\Delta E_{Cor}\rangle \approx 0$), the Asn/Gln−Ade interactions are of dispersive nature ($\langle\Delta E_{HF}\rangle > 0$; $\langle\Delta E_{Cor}\rangle < 0$), and the Arg−Ade pairs are stabilized to the same extent by electrostatic and dispersive effects ($\langle\Delta E_{HF}\rangle \approx \langle\Delta E_{Cor}\rangle < 0$). The importance of the electron correlation contributions in Arg− and Asn/Gln−Ade complexes results from the stacking of the guanidinium and formamide planes with the Ade plane and, hence, from the overlap of their π-orbitals.

(29) Scott, P. A.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502−16513.
(30) Tomasi, J.; Mennucci, B.; Cancès, E. *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 211−226.
(31) Mennucci, B. *J. Am. Chem. Soc.* **2002**, *124*, 1506−1515.
(32) Nielsen, P. A.; Norrby, P.-O.; Liljefors, T.; Rega, N.; Barone, V. *J. Am. Chem. Soc.* **2000**, *122*, 3151−3155.
(33) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385−16398.
(34) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(35) *Spartan 02 for Linux*; Wavefunction, Inc.: Irvine, CA, 2002.
(36) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(37) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709−717.
(38) Duffy, E. M.; Kowalczyk, P. J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1993**, *115*, 9271−9275.
(39) Minoux, H.; Chipot, C. *J. Am. Chem. Soc.* **1999**, *121*, 10366−10372.
(40) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802−11808.

Table 1. Interaction Energy and Vibrational Entropy Contributions in Gas Phase for the 55 X-ray Cation−π Pairs[a]

X–Ade	*n*	ΔE_{HF}	ΔE_{Cor}	ΔZPE	ΔE_{vib}	$-T\Delta S_{vib}$	ΔG_{gas}
Lys	7	−4.2 ± 2.4 (−7.5)	−0.4 ± 0.3 (−0.8)	0.1 ± 0.1 (−0.0)	0.7 ± 0.3 (0.5)	−2.5 ± 1.1 (−4.3)	−6.2 ± 2.6 (−10.5)
Arg	38	−2.6 ± 2.0 (−5.2)	−3.0 ±1.6 (−6.8)	0.0 ± 0.1 (−0.2)	1.6 ± 0.5 (0.6)	−5.0 ± 1.6 (−8.8)	−9.0 ± 3.6 (−16.4)
Asn/Gln	10	1.9 ± 1.4 (−0.9)	−3.3 ± 1.8 (−5.8)	0.3 ± 0.1 (0.0)	1.8 ± 0.5 (1.1)	−5.1 ± 1.3 (−7.8)	−4.5 ± 1.8 (−7.3)

[a] All values are in kcal/mol. ΔE_{HF}, ΔZPE, ΔE_{vib}, and ΔS_{vib} were computed at HF/6-31G**(0.2) level of theory and ΔE_{Cor} at MP2/6-31G**(0.2). ΔE_{HF} and ΔE_{Cor} were corrected for the BSSE. The correction on ΔE_{HF} is equal to +0.5, +1.7, and +0.7 kcal/mol for Lys, Arg, and Asn/Gln, on the average, while that on ΔE_{Cor} is equal to +0.2, +2.2, and +1.9 kcal/mol. The mean values plus or minus the standard deviations are given, with the minimum values in parentheses. For Lys, two different orientations of the H-atoms with respect to the Ade moiety are considered (see Methods section), and that yielding the lowest energy value is retained.

Table 2. Solvation and Interaction Free Energies in a Range of Solvents Characterized by Their Dielectric Constant ε for the 55 X-ray Cation−π Pairs[a]

solvent/model/ε	X–Ade	$\Delta\Delta G_{solv}$	ΔG
water	Lys	6.7 ± 3.9 (1.9)	0.5 ± 2.6 (−2.5)
SM5.4/A	Arg	8.6 ± 4.1 (−2.5)	−0.4 ± 1.8 (−4.9)
78.4	Asn/Gln	3.6 ± 1.6 (0.7)	−1.0 ± 0.9 (−2.4)
water	Lys	6.9 ± 2.7 (3.1)	0.8 ± 1.0 (−1.4)
IEF−PCM	Arg	8.2 ± 2.9 (−0.7)	−0.8 ± 1.7 (−4.8)
78.4	Asn/Gln	3.3 ± 1.4 (0.7)	−1.3 ± 1.0 (−3.0)
DMSO	Lys	7.6 ± 2.2 (4.4)	1.5 ± 1.2 (−0.6)
IEF−PCM	Arg	8.5 ± 2.9 (0.3)	−0.5 ± 2.2 (−5.2)
46.7	Asn/Gln	3.5 ± 1.3 (1.3)	−1.0 ± 1.0 (−3.0)
Acetone	Lys	7.2 ± 2.3 (4.0)	1.1 ± 1.3 (−1.3)
IEF−PCM	Arg	7.6 ± 2.8 (0.1)	−1.4 ± 2.2 (−6.9)
20.7	Asn/Gln	2.9 ± 1.2 (0.9)	−1.6 ± 0.9 (−3.8)
THF	Lys	6.9 ± 2.3 (4.2)	0.8 ± 1.3 (−1.1)
IEF−PCM	Arg	0.8 ± 1.3 (−1.1)	−0.9 ± 1.9 (−5.8)
7.6	Asn/Gln	3.3 ± 1.1 (1.4)	−1.2 ± 1.0 (−3.3)
CCl_4	Lys	6.2 ± 1.4 (3.9)	0.2 ± 1.7 (−2.7)
IEF−PCM	Arg	6.8 ± 2.0 (1.0)	−2.1 ± 2.5 (−8.4)
2.2	Asn/Gln	2.9 ± 0.7 (1.7)	−1.6 ± 1.3 (−3.7)
gas phase	Lys		−6.2 ± 2.6 (−10.5)
MP2	Arg		−9.0 ± 3.6 (−16.4)
1.0	Asn/Gln		−4.5 ± 1.8 (−7.3)

[a] All values are in kcal/mol. The mean values plus or minus the standard deviations are given, with the minimum values in parentheses. For Lys, two different orientations of the H-atoms with respect to the Ade moiety are considered (see Methods section), and that yielding the lowest energy value is retained. The last row displays the gas phase ΔG_{gas} values taken from Table 1.

Comparing the MP2 energies of these cation−π interactions with those calculated in the context of protein−DNA complexes[23] reveals that Arg−Ade pairs display similar energies on the average, whereas Lys−Ade are more favorable and Asn/Gln−Ade less favorable in protein−DNA. This result can be explained by the fact that, at the protein−DNA interface, the charged or partially charged group is in general located above the endocyclic nitrogen atoms rather than above the cycle center, due to the poor accessibility of the nucleobases within DNA. This improves the electrostatic contribution for Lys−Ade complexes and is energetically less favorable for Asn/Gln−Ade pairs.

The ZPE contributions to the interaction free energies are totally negligible at the level of accuracy of the calculations. In contrast, the contributions due to internal vibrations are not. The interaction free-energy contributions due to vibrational energy are unfavorable, whereas those due to the vibrational entropy are favorable for all cation−π pairs. The sum of these contributions, $\Delta E_{vib} - T\Delta S_{vib}$, is more negative for Asn/Gln and Arg (about −3.4 kcal/mol on the average) than for Lys (−1.8 kcal/mol). Note that this difference is not related to a different number of atoms (10 for Arg, against 6 for Asn/Gln and 5 for Lys). Rather, it can be attributed to the difference in shape and complementarity of the complexes, where the planarity of the guanidinium and formamide groups could favor concomitant vibration with the Ade plane. This interpretation is supported by the fact that the Arg−Ade complex with the most unfavorable $-T\Delta S_{vib}$ (−2.0 kcal/mol in 1FPX) has its guanidinium group in T-shaped conformation above the Ade moiety, whereas the pair with most favorable $-T\Delta S_{vib}$ (−8.8 kcal/mol in 5TMP) is almost perfectly stacked.

The analysis of vibrational modes shows that the frequency values calculated for the cation−π pairs are markedly lower than those obtained for the two isolated partners and thus contribute more to the vibrational entropy. For the cation−π pair in 1AYL, the smallest vibrational frequencies are 190, 238, 302, and 328 cm^{-1} for Ade, 123, 256, 265, and 405 cm^{-1} for Arg, and only 21, 27, 62, and 72 cm^{-1} for the cation−π complex. This indicates the presence of a broad energy well linking the cation−π partners. The $-T\Delta S_{vib}$ contribution corresponding to these frequencies is quite favorable, i.e., −2.1 kcal/mol. When considering all vibrational frequencies, this contribution becomes even more negative, reaching −5.5 kcal/mol. Note that the lowest frequency of 21 cm^{-1} corresponds to a relative rotational motion of the Ade and Arg moieties.

Combining these different contributions, we find that the average vacuum interaction free energy $\langle\Delta G_{gas}\rangle$ is the most favorable for Arg−Ade cation−π pairs, followed by Lys− and Asn/Gln−Ade pairs. We would like to stress that the favorable nature of cation−π pairs strongly depends on the level of approximation. When restricting to the HF energy level, we find indeed:

$$\langle\Delta E_{HF}(\text{Lys}-\text{Ade})\rangle < \langle\Delta E_{HF}(\text{Arg}-\text{Ade})\rangle < 0 < \langle\Delta E_{HF}(\text{Asn/Gln}-\text{Ade})\rangle \quad (1)$$

In contrast, by adding electron correlation and vibrational contributions, we obtain the reversed tendencies:

$$\langle\Delta G_{gas}(\text{Arg}-\text{Ade})\rangle < \langle\Delta G_{gas}(\text{Lys}-\text{Ade})\rangle < \langle\Delta G_{gas}(\text{Asn/Gln}-\text{Ade})\rangle < 0 \quad (2)$$

This result clearly shows that the HF level of theory is not sufficient for estimating the stability of cation−π complexes and, more generally, of all types of stacking-involving interactions. These require MP2-level calculations at least and the use of extended basis sets with diffuse d-polarization functions.[40,23]

The geometries of the cation−π pairs were not optimized; only the geometries of the separate molecular groups were. We chose this procedure because the geometries of the complexes optimized in gas phase differ from the optimal geometries in protein environments probably more than the X-ray geometries

themselves (see discussion at the end of the Methods section). Hence, the calculated energy and free-energy values of the X-ray structures must be considered as upper bounds of the actual ones, i.e., of those that would be optimized within the host proteins. Along this line of thought, we will usually consider the minimum interaction (free) energies of cation−π pairs instead of the average ones, expecting that some of the X-ray complexes have geometries close to optimal. The number of complexes is of course too low for this to be true, but this approximation can be considered as reasonable.

As a matter of fact, all average and minimal (free) energy values exhibit similar tendencies, with the latter being obviously more negative than the former. In particular, the mean $\langle \Delta G_{gas} \rangle$ values vary from −4.5, −6.2 to −9.0 kcal/mol for Asn/Gln−, Lys−, and Arg−Ade pairs, respectively, and the minimal ΔG_{gas} values vary from −7.3, −10.5 to −16.4 kcal/mol. Equation 2 can thus be rewritten in terms of minimal interaction free energies as:

$$\mathrm{Min}[\Delta G_{gas}(\mathrm{Arg-Ade})] < \mathrm{Min}[\Delta G_{gas}(\mathrm{Lys-Ade})] < \mathrm{Min}[\Delta G_{gas}(\mathrm{Asn/Gln-Ade})] < 0 \quad (3)$$

Interaction Free Energies in Different Solvents. Solvation free-energy differences $\Delta\Delta G_{solv}$ occurring upon cation−π complex formation were estimated using the IEF−PCM and SM5.4/A continuum models (Table 2). To assess the reliability of these models, we compared, when possible, the calculated and measured ΔG_{solv} values in water. For the ammonium group, we found ΔG_{solv} equal to −80 kcal/mol with IEF−PCM and −88 kcal/mol with SM5.4/A, whereas the experimental value varies between −79 and −84 kcal/mol.[41−43] For Ade, we obtained −15 kcal/mol with IEF−PCM and −20 kcal/mol with SM5.4/A; no experimental values were available for Ade, but a value of −13.6 kcal/mol was reported for 9-methyladenine.[44] The two solvent models considered, and especially IEF−PCM, can thus be considered as appropriate for estimating the free energies of our systems with reasonable accuracy.

Let us concentrate on the interaction free energies $\Delta \boldsymbol{G}$ obtained with the IEF−PCM method. Overall, they become less favorable when the dielectric constant increases; some departures from this trend can be attributed to the size or other characteristics of the solvent molecules. The change in interaction energy is especially marked for Lys− and Arg−Ade complexes: the average interaction free energy increases by about 7−8 kcal/mol from vacuum to water, whereas the increase is only 3 kcal/mol for the Asn/Gln−Ade complexes. The inclusion of solvent effects therefore modifies the trends observed in a vacuum, which are summarized in eq 3. Indeed, we find that in all five solvents:

$$\mathrm{Min}[\Delta \mathbf{G}(\mathrm{Arg-Ade})] < \mathrm{Min}[\Delta \mathbf{G}(\mathrm{Asn/Gln-Ade})] < \mathrm{Min}[\Delta \mathbf{G}(\mathrm{Lys-Ade})] < 0 \quad (4)$$

Note that this inequality does not hold for average $\Delta \mathbf{G}$ values. We find indeed that $\langle \Delta \mathbf{G}(\mathrm{Arg-Ade}) \rangle \geq \langle \Delta \mathbf{G}(\mathrm{Asn/Gln-Ade}) \rangle$. This discrepancy is related to the larger standard deviations on the $\Delta \mathbf{G}$ values for Arg−Ade than for Asn/Gln−Ade and thus to an apparent greater sensitivity of the Arg−Ade pairs on the optimality of the geometries. As discussed above, in absence of a reliable optimization procedure within protein environments, the geometries with lowest $\Delta \mathbf{G}$ can be expected to be close to optimal. The validity of this hypothesis is supported by the similarity of the Arg− and Asn/Gln−Ade geometries with minimal interaction free energies (given that guanidinium and formamide groups are chemically analogous), which are depicted in Figure 1. We are thus led to focus on minimal rather than on average interaction free energies.

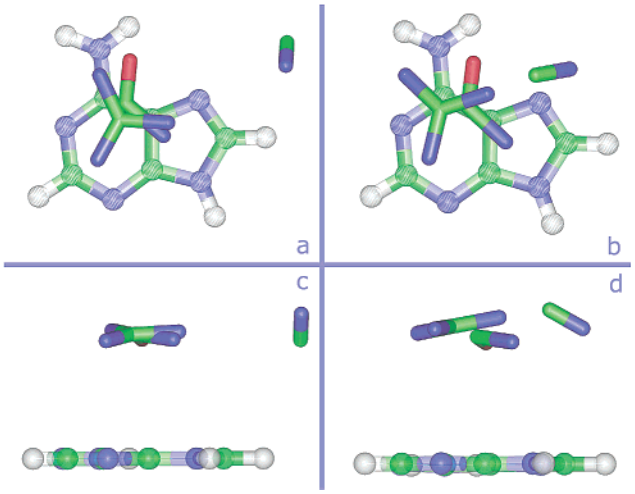


Figure 1. Representation of the X-ray structures of the Arg−, Asn/Gln−, and Lys−Ade cation−π complexes which display minimal interaction free energies $\Delta \mathbf{G}$ in water (a, top view and c, side view) or acetone (b, top view and d, side view). All Ade bases are superimposed. C, N, O, and H atoms are colored in green, blue, red, and white, respectively. The images were generated using InsightII (Accelrys, Inc.). The Arg−Ade pair with minimal $\Delta \mathbf{G}_{water}$ occurs in 1QFL ($\Delta E_{MP2} = -10.8$ kcal/mol, $\Delta G_{gas} = -16.3$ kcal/mol, $\Delta \mathbf{G}_{acetone} = -6.0$ kcal/mol, and $\Delta \mathbf{G}_{water} = -4.6$ kcal/mol), and that with minimal $\Delta \mathbf{G}_{acetone}$ occurs in 1ZIN ($\Delta E_{MP2} = -9.5$ kcal/mol, $\Delta G_{gas} = -13.1$ kcal/mol, $\Delta \mathbf{G}_{acetone} = -6.8$ kcal/mol, and $\Delta \mathbf{G}_{water} = -1.7$ kcal/mol). The Asn/Gln−Ade system which displays both minimal $\Delta \mathbf{G}_{water}$ and $\Delta \mathbf{G}_{acetone}$ occurs in 1C14 ($\Delta E_{MP2} = -2.5$ kcal/mol, $\Delta G_{gas} = -6.9$ kcal/mol, $\Delta \mathbf{G}_{acetone} = -3.4$ kcal/mol, and $\Delta \mathbf{G}_{water} = -2.6$ kcal/mol). The Lys−Ade complex with minimal $\Delta \mathbf{G}_{water}$ occurs in 1E19 ($\Delta E_{MP2} = -1.4$ kcal/mol, $\Delta G_{gas} = -2.62$ kcal/mol, $\Delta \mathbf{G}_{acetone} = 1.6$ kcal/mol, and $\Delta \mathbf{G}_{water} = -1.4$ kcal/mol), and that with minimal $\Delta \mathbf{G}_{acetone}$ occurs in 1EQ2 ($\Delta E_{MP2} = -8.1$ kcal/mol, $\Delta G_{gas} = -10.3$ kcal/mol, $\Delta \mathbf{G}_{acetone} = -1.0$ kcal/mol, and $\Delta \mathbf{G}_{water} = -0.1$ kcal/mol).

To further justify this choice and check the statistical significance of the minimal $\Delta \mathbf{G}$ values, Min[$\Delta \mathbf{G}$], in particular for seldom cation−π pairs, we selected all possible sets of N pairs among the 38 Arg−Ade, 10 Asn/Gln−Ade, and 7 Lys−Ade geometries. We computed the Min[$\Delta \mathbf{G}$]'s on each set and noted the average of these $\langle$Min[$\Delta \mathbf{G}$]$\rangle$. For $N = 7$, corresponding to the number of Lys−Ade complexes, we found that the $\langle$Min[$\Delta \mathbf{G}$]$\rangle$'s are only slightly higher than the Min[$\Delta \mathbf{G}$]'s on the whole set. In water, for example, these Min[$\Delta \mathbf{G}$] values are equal to −4.8 and −3.0 kcal/mol for Arg−Ade and Asn/Gln−Ade pairs, respectively (see Table 2), whereas the $\langle$Min[$\Delta \mathbf{G}$]$\rangle$'s on the restricted sets are equal to −4.3 and −2.8 kcal/mol. This result supports the validity of the Min[$\Delta \mathbf{G}$] values and the choice of considering them instead of average interaction free energies.

The differences between the stability of the different cation−π pairs observed in a vacuum and in the solvents, summarized in eqs 3 and 4, can be explained by the screening of the electrostatic interactions in solution. As a consequence, the

(41) Klots, C. E. *J. Phys. Chem.* **1981**, *85*, 3585−3588.
(42) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109−6114.
(43) Ford, G. P.; Wang, B. *J. Am. Chem. Soc.* **1992**, *114*, 10563−10572.
(44) Ferguson, D. M.; Pearlman, D. A.; Swope, W. C.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 362.

interaction free energies of Lys− and Arg−Ade pairs, both of which carry a net positive charge and whose electrostatic contribution is important, are drastically less favorable in the solvents than in gas phase.

The SM5.4/A model yields interaction free energies in water which are quite similar to those computed with IEF−PCM, whereas these models differ profoundly: IEF was originally dedicated to extend PCM-like methods to ionic solutions or anisotropic solvents and is now the default method for standard isotropic solvents, whereas the SM5.4/A model is based on a low computer cost, semiempirical approach. The fact that both give similar results in water—we do not dispose of an SM5 module that allows using other solvents—adds confidence to our results.

Protein environments do not correspond to water. Rather, they can be considered as media with dielectric constants ranging roughly from 2 to about 25,[45] depending on the distance from the protein surface and the specific environment. Protein−ligand cation−π interactions are usually rather buried, though not far from the protein surface. Acetone, THF, and CCl_4, which have dielectric constants of 20.7, 7.6, and 2.2, can thus be viewed as mimicking typical protein environments. In these solvents, the minimal $\Delta\mathbf{G}$ values are in the range −2.7 to −1.1 kcal/mol for Lys−Ade, −3.8 to −3.3 kcal/mol for Asn/Gln−Ade, and −8.4 to −5.8 kcal/mol for Arg−Ade.

The interaction free-energy values of specific structure motifs are expected to be related to their frequencies of occurrences in protein structures, the most favorable interactions being those that occur the most often. To check this for cation−π motifs, let us remember that there are 7 Lys−, 10 Asn/Gln−, and 38 Arg−Ade complexes in the protein−ligand dataset. These frequencies must be normalized by the abundance of Asn/Gln, Lys, and Arg in protein structures, which is about 7.7%, 6.4%, and 4.2%, respectively. The normalized frequencies are thus equal to 1.1 for Lys−, 1.3 for Asn/Gln−, and 9.0 for Arg−Ade pairs. These values follow quite well the minimal $\Delta\mathbf{G}$ values in proteinlike solvents, which are approximately equal to −2, −4, and −7 kcal/mol.

The cation−π pairs presenting the lowest interaction free-energy $\Delta\mathbf{G}$ in acetone and water are depicted in Figure 1. The guanidinium and formamide moieties are stacked against the Ade plane above the C6 cycle, near the C_6 atom, both in water and acetone. The ammonium group is situated above the N_7 atom of the C5 cycle; in water the lowest-energy complex is at the limit of the cation−π detection criteria and has a favorable $\Delta\mathbf{G}$ value only because of a low $\Delta\Delta G_{solv}$ contribution. Note that except in the latter case, the conformation with minimal $\Delta\mathbf{G}$ is also the conformation with minimal, or almost minimal, interaction energy ΔE_{MP2} and interaction free-energy ΔG_{gas} in gas phase.

Discussion

A striking result of this work is that the London dispersion and vibrational contributions to the interaction free energies of cation−π complexes may not be neglected and are favorable, especially when stacking interactions are involved such as in the Arg− and Asn/Gln−Ade complexes. As a consequence, Lys−Ade appears as the most stable pair when limiting the calculation to the HF level in gas phase, but no more when using more accurate descriptions.[23,24] Another noticeable result is that when immersing cation−π complexes in a solvent, the electrostatic energy contributions to the interaction free energy are drastically screened, whereas the electron correlation contributions are not. Because of the synergetic effects of electronic correlation, solvation,[46] and vibrational entropy, we find Arg−Ade to be the most favorable cation−π pairs in solution, followed by Asn/Gln−Ade and Lys−Ade.

The interaction free energies $\Delta\mathbf{G}$ of cation−π systems in proteinlike solvents CCl_4, THF, and acetone were found to follow quite well their normalized frequencies of occurrences in protein−ligand structures. Indeed, Arg−Ade, Asn/Gln−Ade, and Lys−Ade display minimal interaction free-energy values of about −7, −4, and −2 kcal/mol, respectively, and normalized frequencies of 9, 1.3, and 1.1, respectively. In absence of experimental free-energy values for Ade−amino acid cation−π systems, this good correlation corroborates our approach; in particular, it validates a posteriori the approximations of neglecting translational and rotational degrees of freedom and of considering complexes with minimal interaction free energies as mimicking optimal geometries. The key ingredient that allows us to obtain this correlation is to consider BSSE-corrected electron correlation contributions calculated at the MP2 level of theory, energetic and entropic contributions due to atomic vibrations, and solvation free energies.

A corollary of these findings is that all cation−π complexes are not equally favorable: their level of stability crucially depends on the type of partners and on the environment. It is thus not surprising that Ade−amino acid cation−π pairs are less stable than methylammonium−benzene in water and organic solvents.[10] Indeed, the presence of endo- and exocyclic nitrogen atoms in the Ade ring renders the solvation energy (ΔG_{solv}) of Ade more favorable than that of benzene and, moreover, affects the interaction energies (ΔE_{MP2}) by modifying the π-density, which depends on the balance between the electro-donating and withdrawing effects of the substituents.[47]

We may conclude that the recurrent occurrence of cation−π interactions in protein−ligand complexes may be explained by their stabilizing nature. This result may be expected to hold for cation−π interactions within proteins and at protein−DNA interfaces. The next step will be to unravel the possible functional role of these interactions in biological systems. This role is suggested, for example, by the experimentally measured effect of His-involving cation−π interactions in proteins on the pK_a,[48,49] which could, for example, allow aromatic residues to act as pH-dependent gates in ion channels.[49] It has also been hypothesized in protein−DNA complexes, where the partial intercalation of a positive charge between successive nucleobases along the DNA stack might affect the charge migration occurring along double-stranded DNA.[50,51]

Acknowledgment. We are grateful to M. Cossi, J. Tomasi, J. M. Kwasigroch, J. Liévin, and R. Wintjens for helpful

(45) Gillès de Pélichy, L.; Eidsness, M. K.; Kurtz, D. M., Jr.; Scott, R. A.; Smith, E. T. *Curr. Sep.* **1998**, *17*, 79−82.
(46) Florian, J.; Sponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884−892.
(47) Hunter, C. A.; Low, C. M. R.; Rotger, C.; Vinter, J. G.; Zonta, C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4873−4876.
(48) Fernandez-Recio, J.; Romero, A.; Sancho, J. *J. Mol. Biol.* **1999**, *290*, 319−330.
(49) Okada, A.; Miura, T.; Takeuchi, H. *Biochemistry* **2001**, *40*, 6053−6060.
(50) Boon, E. M.; Barton, J. K. *Curr. Opin. Struct. Biol.* **2002**, *12*, 320−329.
(51) Meggers, E.; Michel-Beyerle, M. E.; Giese, B. *J. Am. Chem. Soc.* **1998**, *120*, 12950−12955.

discussions. C.B. is supported by a BioVal research program of the Walloon Region, and M.R. is Research Director at the Belgian National Fund for Scientific Research (F.N.R.S.). C.B. and M.R. also acknowledge support from the Communauté Française de Belgique through an Action de Recherche Concertée.

Supporting Information Available: Table of interaction (free) energies of the 57 X-ray protein−ligand cation−π pairs in gas phase and in water and THF (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA035223E