X-ray structures of proteins bound to ligand molecules containing a nucleic acid base were systematically searched for cation-π interactions between the base and a positively charged or partially charged side chain group located above it, using geometric criteria. Such interactions were found in 38% of the complexes and are thus even more frequent than π-π stacking interactions. They are moreover well conserved in families of related proteins. The overwhelming majority of cation-π contacts involve Ade bases, as these constitute by far the most frequent ligand building block; Arg-Ade is the most frequent cation-π pair. Ab initio energy calculations at MP2 level were performed on all recorded pairs. Though cation-π interactions involving the net positive charge carried by Arg or Lys side chains are the most favorable energetically, those involving the partial positive charge of Asn and Gln side chain amino groups (sometimes referred to as amino-π interactions) are favorable too, owing to the electron correlation energy contribution. Chains of cation-π interactions with a nucleobase bound simultaneously to two charged groups or a charged group sandwiched between two aromatic moieties are found in several complexes. The systematic association of these motifs with specific ligand molecules in unrelated protein sequences raises the question of their role in protein-ligand structure, stability, and recognition.

To be able to follow the protein folding process or the association of a protein with other (macro)molecules, by computer or real experiments, it is essential to have a fundamental knowledge of the energetic and entropic factors that drive these processes or to be able to precisely reproduce them by means of the correct descriptors. Indeed, even though the inter- and intramolecular non-covalent interactions, including electrostatic, hydrogen bond, London, Pauli, and electron charge transfer contributions, are known in principle, we still have a limited appreciation of their precise nature and of the way in which they compete or reinforce one another in complex systems. In particular, the importance of cation-π interactions between aromatic rings and positive charges has only recently started to be appreciated in the biomolecular context (1, 2), and the question regarding their true nature is far from settled.

Both experimental and theoretical studies have emphasized the presence of this favorable interaction in protein structure (3, 4), as well as in biomolecular association processes such as ligand-antibody binding and receptor-ligand interactions (5–9).

Cation-π interactions have also been shown to be quite common at the interface between protein and DNA (10, 11), where they involve positively charged Arg or Lys side chains and aromatic rings of nucleic acid bases. According to ab initio calculations the most favorable energy is obtained with Gua bases; hence, cation-π interactions contribute to the specificity of protein-DNA recognition.

Related interactions, termed amino-π or simply cation-π interactions, involve a partial instead of a full positive charge, carried by the polar amino acids Asn or Gln (10–15). These amino acids, albeit globally neutral, possess a partial positive charge \( \delta^+ \) on the amino group of their side chain amide group, which, when positioned above the aromatic ring, forms a cation-π interaction. Such interactions have been observed at the interface between protein and DNA (10, 11) and between receptor and ligand (15), but their favorable nature has not yet been demonstrated.

Biological processes such as enzymatic reactions generally need cofactor molecules to perform catalytic transformations of substrates. Most of these intervening molecules include nucleic acid bases in their molecular structure, and we shall focus on them. These nucleobases contribute to the protein-ligand binding but are not directly involved in the biological processes and thus not chemically altered during the catalytic reaction. One main group of cofactors embodies ATP/ADP, GTP/GDP, and CTP/CDP containing an Ade, Gua, and Cyt base, respectively. When bound to their target, they liberate a phosphate group by hydrolysis of the high energy phosphoanhydride bonds and constitute the most important energy source for the cell. Another important group of cofactors comprehends NAD and FAD, which are involved in the transfer of an electron pair from reduced NAD or FAD to the final electron acceptor in an elaborate electron transport cascade. The electrons pass through a series of discrete steps that permit the gradual harvesting of electron energy, whose dissipation is coupled to the synthesis of ATP molecules.
The present study is focused on probing the structural and biological role of amino acid-nucleobase cation-π interactions and lies within the scope of our recent research devoted to protein-DNA complexes (10, 11). Here, we consider proteins bound to ligand molecules and cation-π interactions linking the nucleobases contained in these molecules to amino acid side chain groups carrying a net or partial positive charge. In particular, we searched a non-redundant structure set of protein-ligand complexes for cation-π interactions and performed ab initio quantum mechanics energy calculations on all recorded cation-π pairs. The conservation of these interactions in families of related proteins was also investigated, so as to further evaluate their importance compared with other non-covalent interactions. The agreement and sometimes disagreement between the frequent occurrence of specific cation-π interactions in x-ray structures and their computed energies is debated. Finally, we discuss the predominance of Ade-containing ligands.

**MATERIALS AND METHODS**

**Structural Dataset of Protein-Ligand Complexes**—To generate a non-redundant set of protein-ligand complexes, we first extracted from the PDB database the list of protein structures complexesed with a ligand containing Ade, Gua, Cyt, or Thy, deposited in the protein data bank (PDB) (17) and having a resolution of 2.0 Å or better. This yielded 580 protein-ligand complexes. We then extracted from these complexes a subset that contained only protein chains displaying less than 25% sequence identity. Information on sequence identity was gathered from the FSSP fold classification data base (18). In this way, we obtained a final set of 188 protein-ligand structures. The subset of these structures containing protein-ligand cation-π interactions is listed under “Supplementary Material.”

**Geometric Definition of Cation-π Interactions**—We define cation-π interactions geometrically by a distance and an angle criterion (10). The distance criterion requires that at least one of the atoms of the aromatic ring is located no further than 4.5 Å from one of the atoms carrying the positive charge (i.e. Ngu, Ngu, C=O, or N for Arg and N, or C (for Lys) or the partial positive charge (i.e. Ngu, C=O, or C for Glu and Ngu, or C for Asn). Note that nitrogen atoms are considered instead of hydrogen atoms because the latter are lacking in x-ray structures. Atoms carrying a partial positive charge designate atoms where the intramolecular electron distribution presents a positive peak due to a lack of electrons. The angle criterion requires that the (partially) positively charged atom is situated above the plane defined by the aromatic ring, more precisely, inside a cylinder of height 4.5 Å, whose base includes the ring and has a radius equal to the ring diameter.

**Simplified Representation of Cation-π Partners**—Each pair was reduced to a simple system that could be studied computationally. The ligands were reduced to the nucleic acid base they contain. Lys and Arg were represented as ammonium and guanidinium groups, respectively, and Gln and Asn as formamide groups. In the case of Lys, two different orientations of the ammonium moiety toward the π-electron cloud were considered, with one of the N=O bonds pointing either toward the aromatic ring or in the opposite direction.

**Ab Initio Quantum Mechanics Energy Calculations**—All ab initio energy calculations were carried out using the Gaussian 98 suite of programs (19). In a first step, the nucleic acid bases as well as the ammonium, guanidinium, and formamide groups, which represent the cation-π partners, were optimized separately, using the Hartree-Fock (HF) and the 6-31G** basis set. The optimized coordinates were then superimposed onto the crystal coordinates using the U3BEST algorithm (20).

In a second step, the energies of the cation-π systems were calculated at second order level of Möller-Plesset perturbation theory (MP2) (21). The basis set used is 6-31G**(0.2), which corresponds to the standard 6-31G** basis set where the Gaussian απ-exponent of the d-polarization functions on the heavy atoms C, N, and O is equal to 0.8, with an additional απ-exponent equal to 0.2. It has indeed been shown that this extended description of the d-polarization functions allows a more accurate description of cation-π interactions (21, 22).

The interaction energy \( \Delta E_{\text{interp}} \) is calculated as the sum of the Hartree-Fock energy \( \Delta E_{\text{HF}} \) and a correlation energy contribution \( \Delta E_{\text{COR}} \), evaluated here in Equation 1 as the second order term of the Möller-Plesset perturbation expansion.

\[
\Delta E_{\text{interp}} = \Delta E_{\text{HF}} + \Delta E_{\text{COR}}
\]  

For a given complex A-B, \( \Delta E_{\text{HF}} \) and \( \Delta E_{\text{COR}} \) are evaluated as the differences at HF and MP2 level, respectively, between the energy of the complex E(A-B) and the energies of isolated partners of the system, E(A) and E(B) as shown in Equations 2 and 3.

\[
\Delta E_{\text{HF}} = E_{\text{HF}}(A-B) - E_{\text{HF}}(A) - E_{\text{HF}}(B)
\]

\[
\Delta E_{\text{COR}} = E_{\text{COR}}(A-B) - E_{\text{COR}}(A) - E_{\text{COR}}(B)
\]

\( \Delta E_{\text{HF}} \) contains electrostatic, polarization, electron transfer, and exchange energy contributions, while \( \Delta E_{\text{COR}} \) contains electron correlation (dispersion) contributions. Together, they are expected to represent the major contributions to the Van der Waals energies.

All calculations were corrected for the basis set superposition error (BSSE) by using the standard function counterpoise (CP) method (24, 25). Details on the ab initio calculation methods can be found in Ref. 10.

**Conservation of Cation-π Interactions in Protein Families**—To investigate the conservation of the cation-π motifs in related proteins, we scanned the PIR annotation and similarity data base. The PIR protein entry displaying the highest degree of sequence identity was considered PDB sequence (found to vary between 98 and 100%) was used as probe. The sequences similar to each probe and belonging to the same family as classified in the PIR data base, were collected with the FASTA sequence alignment program (27) using the threshold e-value < 0.0001. Within each family, we discarded protein sequences shorter than 75% of the original PDB sequence. The conservation of the amino acids forming the cation-π motif was evaluated on the basis of multiple sequence alignments performed with the ClustalW program (28).

**RESULTS**

**Cation-π Interactions in X-ray Protein-Ligand Complexes**—A non-redundant dataset of 188 high-resolution x-ray structures of protein-ligand complexes, defined as described under “Materials and Methods,” was searched for cation-π interactions linking protein to ligand. All the ligand molecules considered contain a nucleic acid base, which is however not directly involved in the enzymatic reaction. The overwhelming majority of the ligands (81%) contain an Ade; Gua, Thy, and Cyt only occur in 12, 4, and 3% of the ligands, respectively. These numbers could be biased, as our structure set contains few membrane proteins that are known to frequently bind Gua-containing ligands; however, a survey of sequence data bases confirms the abundance of Ade in ligands. The most frequent ligands are adenosine mono-, di-, or tri-phosphates (AMP, ADP, ATP) and ligands containing a nicotinamide adenine dinucleotide moiety (NAD, NAP, NAQ, NDP); they occur 37 and 38 times in the structure set, respectively. We focused on cation-π interactions between an aromatic ring of the ligand’s nucleic acid base and an Arg or Lys side chain carrying a net positive charge or an Asn or Gln side chain carrying a partial positive charge on its amino group. We did not consider cation-π interactions involving other aromatic moieties of the ligand, nor positively charged ligand groups interacting with aromatic amino acid side chains. We also overlooked His residues, due to their ubiquitous nature. They indeed play the role of cations under protonated form and behave like simple aromatic systems under unprotonated form. All these interactions are equally interesting to study, but fall outside the scope of the present work.

In the set of 188 complexes, 77 protein-ligand cation-π interactions (see “Supplementary Material”) were identified using the geometrical criterion described under “Materials and Methods,” distributed over 71 structures (listed under “Supplementary Material”). More than half of these interactions (40 occur-
rences) involve Arg, followed by Lys (19), Asn (12), and Gln (6). The frequency differences are even more marked for nucleic acid bases: by far the most frequent base is Ade (54 occurrences), followed by Gua (19), Thy (2), and Cyt (2). In the case of Gua, the cation-π partner is always positioned above the 5-member cycle, whereas for Ade, this happens in 72% of the cases. With its 36 occurrences, the Arg-Ade is the preferred cation-π pair. Note that in 60% of the observed cation-π interactions, the charged group is positioned right above the aromatic cycle, and in the remaining interactions it is situated above the extracyclic atoms.

The preferential occurrence of Arg residues in cation-π interactions is in agreement with previous observations in proteins (3, 12, 29, 30) and in protein-DNA complexes (10, 31). The lower frequency of Asn and Gln residues compared with Arg and Lys also argues with previous observations (13) and is related to the fact that they carry only a partial positive charge on their side chain amino group. The observation that Ade is almost three times more frequently involved in cation-π interactions than Gua seems a priori in contradistinction with what happens in protein-DNA complexes where Gua is more frequent (10). But it is not. Indeed, we find that 58% of the Gua-involving ligands form a cation-π interaction, against 40% for Ade-involving ligands.

To probe the orientation of the Arg, Asn, and Gln side chains relative to the nucleic acid base, we computed the angle between the aromatic plane and that formed by the guanidinium group of Arg or the amide group of Asn or Gln side chains. This angle is, on average, equal to 32, 35, and 41° for Arg, Gln, and Asn, respectively. The planes are thus usually somewhat more parallel than perpendicular, which has already been noted for guanidinium groups in aqueous solution (32) and in proteins (13, 33) although the perpendicular or T-shaped conformation seems more stable in vacuum (3, 30). Moreover, the deviation from the stacked conformation is largest for Gln and Asn. This allows the partial positive charge carried by the amino group to be closer to the aromatic ring than the partial negative charge.

**Table I**

| Cation-π chain motif |
|---------------------|
| π  | +  | π  |
| 1B8S Trp (A67) | Arg (A110) | Ade (FAD 510) |
| 1BA1 Tyr (15) | Arg (272) | Ade (ADP 486) |
| 1BG0 Trp (221) | Arg (124) | Ade (ADP 400) |
| 1H8E Tyr (A433) | Gln (A432) | Ade (ADP A600) |
| 1LVK Phe (129) | Asn (127) | Ade (MNT 999) |
| 1CIY Asn (A116) | Gua (GTP 170) | Lys (A117) |
| 1EXM Asn (A136) | Gua (GNP 406) | Lys (A157) |
| 1HUQ Asn (A134) | Gua (GNP 200) | Lys (A135) |
| 1HUR Asn (A126) | Gua (ADP A1) | Lys (A127) |
| 1QF5 Lys (A18) | Gua (GDP 1) | Lys (A331) |

**Fig. 1. Three-dimensional representation of cation-π chain motifs.** The ligand is colored using the following code: carbon (green), nitrogen (blue), oxygen (red), and phosphate (purple). The images were generated using Insight II (Accelrys Inc.). a, +:π:+, motif in 1QF5, involving the GDP ligand, Lys (A18) and Lys (A331); b, +:π:+, motif in 1CIY involving the GTP ligand, Asn (A118) and Lys (A117); c, π:+:π, motif in 1LVK, involving the MNT ligand, Asn (127) and Phe (129); d, π:+:π, motif in 1H8E, involving the ADP ligand, Gln (A432) and Tyr (A433).
carried by the oxygen of the carbonyl group, thereby improving the electrostatic contribution (13). Note that the deviation from the stacked conformation of Arg, Gln, and Asn is larger here than in the protein-DNA context (11), where the simultaneous formation of a double H-bond between the guanidinium or amide group and a subsequent nucleic acid base along the DNA stack imposes geometrical constraints favoring the parallel conformation.

**Cation-π Chains in Protein-Ligand Complexes**—In some instances, two or more cation-π interactions occur concomitantly, hence forming cation-π chain motifs (11). Two types of chains motifs can be distinguished: either a side chain carrying a net or partial positive charge interacts with two aromatic rings (type π+:+π chain, where π denotes an aromatic cycle; + a net or partial positive charge; and ... a cation-π interaction) or an aromatic ring interacts with two side chains carrying a net or partial positive charge (type +:π+:+ chain). A total of 10 cation-π chains are identified in the set of protein-ligand complexes, 5 of each type (Table I).

In the 5 occurrences of type π+:+π cation-π chains, a guanidinium or amide group of Arg, Asn, or Gln is sandwiched between an Ade base and an aromatic Trp, Tyr, or Phe side chain. The protein-ligand cation-π interaction is thus prolonged with a protein-protein cation-π interaction (Fig. 1, c and d).

The π+:+π cation-π chains can be subdivided in two subsets, one in which the two side chains are at the same side of the aromatic ring (noted +:π+:+π) and one in which they are at opposite sides (+:π+:π:+). There is only one occurrence of +:π+:π:+ chain, with the 5-member cycle of a Gua base sandwiched between two Lys side chains (Fig. 1a), whereas there are four occurrences of +:π+:π:+ chains, where a Gua base interacts simultaneously with two consecutive residues along the polypeptide chain, an Asn and a Lys, both situated above the 5-member cycle (Fig. 1b). Note the simultaneous presence of one partial and one full positive charge above the aromatic cycle; the presence of two net positive charges would probably be unfavorable.

Though the number of cation-π chains is too limited to allow firm conclusions, it is striking that π+:+π cation-π chains always involve Ade bases included in ADP-like ligands and that the +:π:+ chains always involve Gua bases contained in GDP-like ligands. Furthermore, the five proteins presenting the π+:+π motif present neither sequence nor structure similarity. In contradiction, the four proteins having the π+:+π motif have unrelated sequences but are structurally similar; the structure superposition program SoFiSt (34) superposes about 130 residues with a root mean square deviation of heavy backbone atoms in the 1.0–2.7 Å range. Strikingly, in the resulting sequence alignments, the Asn-Lys pair forming the cation-π chain with the ligand is aligned in all four structures. Among the other residues that interact with the ligand, in particular through H-bonds, the pattern Gly-Lys-Thr-Thr, or its variant Gly-Lys-Ser-(Ser/Ala), is also well conserved and aligned, as well as an Asp residue at position A119 in 1CIY numbering. There is thus a residue conservation in the immediate neighborhood of the ligand but not in the rest of the protein.

In general, it can be argued that the presence of a specific cation-π chain motif could be due to a local sequence conservation (a global sequence similarity is excluded because of the requirement of less than 25% protein sequence identity, see “Materials and Methods”) and/or have a structural role. In particular, the conservation of the +:π+:π motif could possibly reflect some sequence similarity left after (highly) divergent evolution or acquired during divergent evolution. Alternatively, this motif, as well as the π+:+π motif, could play a particular role in the recognition process, especially as the former is exclusively found in GDP/protein complexes and the other in ADP/protein complexes. However, it must be noted that not all ADP- or GDP-containing proteins exhibit these motifs. At this point, the exact role of these chain motifs is thus far from settled.

**Comparison with Other Protein-Ligand Interactions**—The ligands are situated in protein pockets that are sometimes close to the protein surface and sometimes completely buried. They make thus extensive contacts with protein residues. The most frequent interactions linking protein and nucleobases are H-bonds; the large majority of the bases form 1 up to 4 H-bonds, with an average of 1.4, as detected by the HBPLUS program (35). Moreover, the bases are often involved in hydrophobic contacts or π-π stacking interactions. In the dataset of 188 protein-ligand complexes, we found with the help of the LigPlot program (36) 122 contacts between the nucleobases and the hydrophobic Val, Ile, Leu, and Met side chains, and 86 between

---

**Table II**

Conservation of the cation-π chain motifs (listed in Table I) in related sequences from the PIR protein sequence data base

| PDB code | Number of sequences | Average sequence identity | Conservation of the motifs | Cation-π motifs |
|----------|---------------------|---------------------------|---------------------------|-----------------|
| 1BBS     | 2                   | 57%                       | 2/2                       | Trp-Ade-Ade     |
| 1BA1     | 266                 | 64%                       | 129/266                   | Tyr-Ade-Ade     |
|          |                     |                           | 145/266                   | Arg-Ade         |
|          |                     |                           | 130/266                   | Tyr(G or Lys)-Ade |
| 1BG0     | 10                  | 65%                       | 10/10                     | Trp-Ade-Ade     |
| 1HE5     | 102                 | 64%                       | 51/102                    | Tyr-Gln-Ade     |
|          |                     |                           | 80/102                    | Gln-Ade         |
|          |                     |                           | 81/102 (Gln or Arg)-Ade   |
|          |                     |                           | 84/102 (Gln or Asn or Arg)-Ade |
| 1LVK     | 1                   | —                         | 1/1                       | Phe-Asn-Ade     |
| 1CY1     | 83                  | 58%                       | 83/83                     | Asn-Gua-Lys     |
| 1EXM     | 113                 | 70%                       | 111/113                   | Asn-Gua-Lys     |
| 1HU5     | 22                  | 66%                       | 22/22                     | Asn-Gua-Lys     |
| 1HUB     | 53                  | 77%                       | 53/53                     | Asn-Gua-Lys     |
| 1QF5     | 32                  | 60%                       | 22/32                     | Lys-Gua         |
|          |                     |                           | 30/32                     | Lys-Gua         |
|          |                     |                           | 24/32                     | Lys-Gua         |

* The PIR entries corresponding to the PDB sequences are (within parentheses, the PDB code, the chain limits, and the percentage of sequence identity between the PIR and PDB sequences): A32260 (1BBS, A43–A546, 99.8%); S11455 (1BA1, 1–386, 99.7%); S22098 (1BG0, 2–357, 99.2%); PWBOA (1HE5, 44–543, 99.8%); A26655 (1LVK, 1–759, 98.4%); A33283 (1CY1, A1–A167, 100.0%); S17146 (1EXM, 1–106, 100.0%); S65933 (1HU5, A19–A182, 100%); A35283 (1HUB, A1–A181, 100.0%); A3ECDS (1QF5, A1–A402, 100.0%).

---
the nucleobases and the aromatic residues Tyr, Trp, and Phe, among which 49 correspond to \( \pi-\pi \) interactions. Hence, the nucleobase-amino acid cation-\( \pi \) interactions, with their 78 occurrences, appear to be more frequent than \( \pi-\pi \) stacking interactions in protein-ligand complexes, and its importance should thus not be overlooked.

Furthermore, most of the amino acid side chains involved in cation-\( \pi \) interactions simultaneously make additional interactions with their local environment. These interactions vary among the complexes, and involve solvent molecules, other amino acid side chains or other ligand moieties adjacent to the purine or pyrimidine ring.

This observation demonstrates an important difference between protein-ligand and protein-DNA cation-\( \pi \) interactions. Indeed, the latter occur almost systematically with an H-bond, generally double, that links the guanidinium, amide, or ammonium group with the next base along the DNA stack, hence forming a recurrent motif with typical stair shape (11). This concomitant occurrence has sometimes been used to shed doubt on the reality on cation-\( \pi \) interactions, suspected to result only from the geometric constraint imposed by the H-bond. In contrast, in 9\% of the protein-ligand cation-\( \pi \) interactions no concomitant H-bond is observed, and in the 91\% remaining cases, the H-bond forms with different partners positioned differently with respect to the cation-\( \pi \) partners, as a consequence of a much more flexible environment. Thus, the cation-\( \pi \) interaction is here certainly not due to a geometric constraint.

**Conservation of Cation-\( \pi \) Interactions in Protein Families**—The conservation of the amino acid residues involved in cation-\( \pi \) interactions in families of related proteins was analyzed by scanning the PIR sequence data base, as described under “Materials and Methods.” For the sake of simplicity, we focused on the cation-\( \pi \) chain motifs given in Table I. The results are summarized in Table II. The two residues carrying a partial or net positive charge, involved in the \(+.:.+o\) cation-\( \pi \) chain motifs, are almost perfectly conserved, in the 4 protein families in which they occur. In the only family only presenting a \(+.:.+o\) motif, at least one of the Lys residues, and thus at least one of the two cation-\( \pi \) interactions, is conserved. Finally, in the 5 families displaying a \(\pi.:.+.\) chain, the conservation of the charged residue, which forms the cation-\( \pi \) interaction with the ligand is quite fair; the conservation of the aromatic residue, which extends the cation-\( \pi \) motif inside the protein is somewhat lower. Note that the observed substitutions often maintain the aromatic or (partially) charged nature of the amino acid side chains. In summary, the rate of residue conservation of the different motifs is somewhat variable, but overall quite good, which supports the importance of cation-\( \pi \) interactions in the formation and/or stabilization of protein-ligand complexes.

**Energy Calculations on Cation-\( \pi \) Partners**—With the purpose of explaining the preferred occurrence of specific cation-\( \pi \) partners, we performed *ab initio* quantum mechanics energy calculations at MP2/6-31G**+(0.2) level on the 77 cation-\( \pi \) pairs extracted from their environment, as described under “Materials and Methods.” The energy values are given under “Supplementary Material” and their averages in Table III, the cation-\( \pi \) interactions involving Ade bases and guanidinium groups are depicted and colored according to their energy in Fig. 2. For Lys residues, two different orientations of the ammonium group relative to the nucleic acid base were considered; because both orientations gave almost the same energy value (see Table III), only the most favorable energy was retained.

The first remarkable feature, clearly visible in Fig. 2, is that roughly all positions in a slice situated about 4 Å above the aromatic cycles are favorable energetically to groups carrying a net or partial positive charge. The positions with the lowest energies are just above the cycles or above some of the extra-cyclic atoms; the energies become less favorable when moving...
further away from the cycles.

Another observation is that the most favorable energies are reached for cation-π interactions involving the positively charged Lys or Arg side chains and the purine bases Ade or Gua. The Arg-Gua, Arg-Ade, Lys-Ade, and Lys-Gua pairs display an average interaction energy of −10.4, −5.6, −4.5, and −3.9 kcal/mol, respectively. Note that the significance of the average energy of Arg-Gua pairs is questionable as it is computed on the basis of only two occurrences. The interaction energies of cation-π pairs between purine bases and Asn/Gln residues involving only a partial positive charge are somewhat less favorable but usually still negative: −3.7, −2.4, and −1.3 kcal/mol on average for Asn-Ade, Asn-Gua, and Gln-Ade. Finally, cation-π interactions involving pyrimidine bases Cyt and Thy are very rare, there is at most one occurrence of each pair. Their energy is equal to −1.6 or −6.1 kcal/mol when involving Arg, and to −0.5 to 1.6 kcal/mol when involving Asn/Gln. Cation-π interactions involving pyrimidine bases seem thus slightly less favorable than those involving purine bases, but there are too few examples to be sure of this. Note that these conclusions do not depend on whether the charged group is located above the aromatic ring itself or above the extracyclic atoms. Indeed, the average energies remain essentially unchanged.

A feature that distinguishes Lys as cation-π partner is that the ΔE_{HF} and ΔE_{MP2} interaction energies are almost equal, whereas for Arg, Asn, and Gln, ΔE_{MP2} is of the order of 3 kcal/mol lower than ΔE_{HF}. This means that the cation-π interaction energy is essentially of electrostatic nature when Lys residues are involved, whereas the electron correlation contribution is non-negligible for Arg, Asn, and Gln residues, owing to the stacking between the aromatic and the guanidinium or amide planes when these are parallel and to the delocalization of the charge. This difference could perhaps explain that Arg is more frequently involved in cation-π interactions than Lys. Note that it is the electron correlation contribution, which renders in general the cation-π interactions involving Asn or Gln favorable. Indeed, ΔE_{HF} is usually positive or zero and ΔE_{MP2} is usually negative, both for the parallel and perpendicular (or T-shaped) conformations.

The lowest cation-π interaction energies are found for specific geometries. The Lys-Gua pair in the 1QF5 complex, with the lowest of all ΔE_{MP2} energies computed (−16.4 kcal/mol), has one hydrogen atom of the ammonium group close to the extracyclic oxygen atom O6 yielding a strongly negative electrostatic term. Interaction energies of −10 to −11 kcal/mol are reached for Arg-Ade pairs in the 1E2F, 1QFL, and 1IZN complexes where the guanidinium ion is stacked above the 6-member aromatic ring in close proximity of the exocyclic nitrogen; this particularly favorable conformation has already been noted before (30, 37). The two Asn-Ade cation-π pairs identified in 1AS2 and 1DXY structures are also quite particular; the same partners form both a cation-π and an H-bond interaction, which explains the low energy values obtained (−8.2 and −7.4 kcal/mol). More precisely, the H-bond links the O3 of Asn to the H-atom on N6 of Ade, whereas the Nε1 atom of Asn is in cation-π interaction above the N7 atom of Ade; these cation-π interactions are however at the limit of our detection criteria, as they are situated near the border of the cylinder defining them (see “Materials and Methods”).

In protein-DNA complexes, the Arg-Gua pairs were shown to constitute the most stable and frequent cation-π interactions (10). Here, in the protein-ligand context, only two such associations have been identified, and their energy is so different (−5.6 and −15.2 kcal/mol) that it is difficult to determine if they are really more stable than the other cation-π pairs. As a matter of fact, there is no clear correlation between the energy values and the frequency of occurrences, in contrast to the protein-DNA cation-π interactions.

An important difference between the protein-DNA and protein-ligand contexts is that in the former the amino acid side chains can approach only some sides of the nucleic acid bases and remain in general far from the center; because of steric constraints. Moreover, since the exocyclic heteroatoms accessible to amino acid side chains are different in the four nucleic acid bases, the cation-π energy values are very different. This renders the protein-DNA cation-π interactions quite specific. This view is in agreement with the recent observations that the strength of the cation-π interactions is very sensitive to the electron density on the face of the aromatic ring (38).

In protein-ligand complexes, there are much fewer constraints, and amino side chains can come right above the aromatic cycles of the ligands, with the consequence that the cation-π energy values are more similar for the four nucleic acid bases and that these interactions lose specificity.

**DISCUSSION**

Protein-ligand cation-π interactions between charged residues and nucleic acid bases are observed in 38% of the complexes of the dataset and are thus rather common. They are even more common than π-π interactions between aromatic residues and these bases. The more specific cation-π chain motifs that simultaneously involve two cation-π interactions appear in 5% of the complexes. These observations concur with the generally favorable energies of cation-π pairs estimated using *ab initio* quantum mechanics calculations at MP2 level. These energies are, on average, in the −4 to −10 kcal/mol range when involving the net charge of Arg or Lys, and in the −1 to −4 kcal/mol range when involving the partial positive charge of Asn or Gln. The frequent occurrence of protein-ligand cation-π pairs in protein structures and their favorable energy lead to the conclusion that they contribute to the stability of the complex.

We would like to stress that the energy of the Asn/Gln involving cation-π interactions, also termed amino-π interactions, is on average slightly positive at HF calculation level and becomes negative at MP2 level (see Table III and “Supplementary Material”). These interactions are thus only computed as being favorable when electron correlation contributions are taken into account.

An *a priori* surprising result of the present analysis is the poor correlation between the frequency of occurrence of specific cation-π pairs and their computed energy. Indeed Arg-Ade is more frequent and more favorable energetically than Lys-Ade, on average, but Arg-Gua is less frequent though presenting a more negative average energy than Lys-Gua. By far the most often occurring cation-π pairs is Arg-Ade, which constitutes 47% of all observed pairs, whereas its energy is not more favorable than that of Arg-Gua pairs, which forms less than 3% of the cation-π pairs. There is thus only a very limited correlation between the frequency of cation-π pairs and their energies, in contrast with what happens in the protein-DNA context, where an almost perfect correlation has been noted (10).

There are different explanations to this limited correlation. First, the *ab initio* energy computations were performed in vacuum and may not be directly transposed to more realistic environments consisting of water and/or protein residues. *Ab initio* energies of H-bonds and especially salt bridges are largely overestimated in vacuum compared with water (4, 39), whereas stacking interactions are less overestimated (39). In other words, electrostatic contributions, estimated at HF calculation level, are more overestimated than electron correla-
tion contributions introduced at MP2 level. Since both the electrostatic and electron correlation contributions are important in cation-π interactions (Table III), albeit at different degrees according to the type of cation-π partners and their relative positioning, the transposition of vacuum energy values to energy values in solvent can entail modifications in the energy ranking of the different cation-π pairs. Note that we chose not to optimize the geometry of the interactions. Indeed, optimization is likely to induce non-realistic geometries, because the relative weight of the different interactions is not reliable, and also because extracting two partners from their environment overlooks the natural structural constraints (40).

Another reason of the preferential occurrence of certain cation-π pairs could be their possible role in protein-ligand recognition and/or in functional features. This view is supported by the conservation in unrelated sequences of the +π:+π+, cation-π chain, with a Guanine from a GDP-type ligand in simultaneous cation-π interaction with an Asn and a Lys, and of the π:+:π chain motif, where a (partially) charged side chain group is sandwiched between an aromatic amino acid and an Ade base from an ADP-type ligand. However, considering that the proteins in which these chain motifs occur do not seem to have particular functional features compared with proteins that are bound to the same ligands but that present no or different cation-π interaction patterns, this view appears rather dubious. The comparison of protein subsets with and without cation-π interactions (not necessarily cation-π chains) leads to a similar conclusion. We indeed found that each type of ligand is linked to its target protein, sometimes through cation-π interactions and sometimes not. In particular, ATP is involved in cation-π interactions in some transferases and kinases and not in others. It remains nevertheless conceivable that cation-π (chain) interactions are associated with a more precise, not yet identified, recognition or functional feature. Alternatively, there could several ways to achieve a given goal, either by means of cation-π (chain) motifs or through other types of interactions.

Finally, the overwhelming majority of protein-ligand cation-π interactions involving Ade bases, though their energy is not more favorable than that of those containing Guan, can simply be explained by the abundance of Ade as natural ligand building blocks. A rapid calculation shows indeed that there is a slightly larger fraction of Guan-containing ligands that forms cation-π interactions than of Ade-containing ligands. This result shifts the question to why is Ade the most frequent ligand skeleton. A possible answer is that Ade was the first purine to be incorporated in living matter (41), but this prebiotic hypothesis is still a subject of debate (26). It is in agreement with the fact that purine bases are easier to synthesize than pyrimidine bases, thereby explaining the very low number of Cyt- and Thy-containing ligands. However, the de novo purine nucleotide biosynthesis requires, at least today and in most eukaryotes, an equivalent number of steps. The predominance of Ade could also be a random initial choice of nature that would be kept through evolution, or be related to the fewer number of possible H-bonds of Ade compared with Guan and thus to its lesser specificity. It could also be related to the property of Guan to be easily oxidizable (23), which could be problematic for certain protein-ligand interactions. Clearly, this issue is far from being understood.

Acknowledgment—We thank J. Lièvren for help with the ab initio energy calculations and for useful discussions.

REFERENCES

1. Dougherty, D. A. (1996) Science 271, 163-166
2. Ma, J. C., and Dougherty, D. A. (1997) Chem. Rev. 97, 1303-1324
3. Gallivan, J. P., and Dougherty, D. A. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 9459-9464
4. Gallivan, J. P., and Dougherty, D. A. (2000) J. Am. Chem. Soc. 122, 870-874
5. Dougherty, D. A., and Stauffer, D. A. (1990) Science 250, 1558-1560
6. Dougherty, D. A., and Kumpf, R. A. (1993) Science 261, 1708-1710
7. Stauffer, D. A., and Karlin, A. (1984) Biochemistry 3, 6640-6649
8. Mezzetti, W., West, A. P. Jr., and Dougherty, D. A. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 10566-10571
9. Zhong, W., Gallivan, J. P., Zhang, Y., Li, L., Lester, H. A., and Dougherty, D. A. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 12088-12093
10. Wintjens, R., Lièvren, J., Roisman, M., and Buisine, E. (2000) J. Mol. Biol. 302, 395-410
11. Roisman, M., Lièvren, J., Buisine, E., and Wintjens, R. (2002) J. Mol. Biol. 319, 67-76
12. Barley, S. K. and Petsko, G. A. (1986) FEBS Lett. 203, 139-143
13. Mitchell, B. E. O., Nandi, C. L., McDonald, I. K., Thornton, J. M., and Price, S. L. (1994) J. Mol. Biol. 239, 315-331
14. Scrutton, N. S., and Raine, A. R. C. (1996) Biochem. J. 319, 1-8
15. Miura, S., Feng, Y., Hussain, A., and Kurnik, S. S. (1999) J. Biol. Chem. 274, 1003-1110
16. Hendlich M. (1998) Acta Crystallogr. B Biol. Crystallogr. 54, 1178-1182
17. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, N. T., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000) Nuclear Acids Res. 28, 255-262
18. Holm, L., and Sander, C. (1996) Science 273, 595-602
19. Frisch, M. J., Trucks, G. W., Schelgel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A., Stratmann, R. E., Farkas, O., Raghavachari, K., Lim, J. J., Galick, J. R., Rendell, A. P., Cioslowski, J., Peng, C. Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M., Willems, B., Mompiou, F., Solar, P., and Gomperts, R. (1998) Gaussian, Revision A. 7. Gaussian, Inc., Pittsburgh, PA
20. Kabsch W. (1978) Acta Crystallogr. A34, 827-828
21. Sponer, J., Leszczynski, J., and Hobza, P. (1996) J. Phys. Chem. 100, 5590-5596
22. Hasegawa, P., and Sponer, J. (1999) Chem. Rev. 99, 3247-3276
23. Steenken, S., and Jovanovic, S. V. (1997) J. Am. Chem. Soc. 119, 617-618
24. Jansen, B. H., and Bos, P. (1969) Chem. Phys. Lett. 3, 140-143
25. Boys, S. F., and Bernardi, F. (1970) Mol. Phys. 19, 593-596
26. Shapiro, R. (1995) Orig. Life Evol. Biosph. 1-3, 83-98
27. Pearson, W. R., and Lipman, D. J. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 2444-2448
28. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nuclear Acids Res. 22, 4673-4680
29. Singh, J., and Thornton, J. M. (1990) J. Mol. Biol. 211, 595-615
30. Mignard, H., and Chipot, C. (1999) J. Am. Chem. Soc. 121, 10366-10372
31. Pabo, C. O., and Sauer, R. T. (1992) Annu. Rev. Biochem. 61, 1053-1095
32. Duffy, E. M., J., and Jorgensen, W. L. (1993) J. Am. Chem. Soc. 115, 9271-9275
33. Plocco, M. M., and Mowbray, S. L. (1994) J. Mol. Biol. 235, 709-717
34. Boutonnet, N., Roisman, M., Ochagavia M.-E., Richelle, J., and Wodak S. (1995) Protein Eng. 8, 647-662
35. McDonald, I. K., and Thornton, J. M. (1994) J. Mol. Biol. 238, 777-793
36. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) Protein Eng. 8, 127-1341
37. Luisi, B., Orozco, M., Sponer, J., Luque, F. J., and Shakked, Z. (1998) J. Mol. Biol. 279, 1123-1136
38. Hunter, C. A., Low, C. M. R., Rotger, C., Vazter, J. G., and Zonta, C. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 4873-4876
39. Florian, J., Sponer, J., and Warschel, A. (1999) J. Phys. Chem. B 103, 884-892
40. Shi, Z., Olson, C. A., and Kallenbach, N. R. (2002) J. Am. Chem. Soc. 124, 3284-3291
41. Schwartz, A. W., and Bakker, C. G. (1989) Science 245, 1102-1104