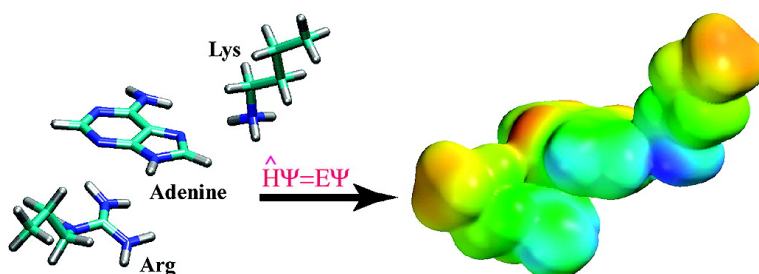Communication

# Multiple Intermolecular Interaction Modes of Positively Charged Residues with Adenine in ATP-Binding Proteins

Lisong Mao, Yanli Wang, Yuemin Liu, and Xiche Hu

## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 2 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

# Multiple Intermolecular Interaction Modes of Positively Charged Residues with Adenine in ATP-Binding Proteins

Lisong Mao, Yanli Wang, Yuemin Liu, and Xiche Hu*

*Department of Chemistry, University of Toledo, Toledo, Ohio 43606*

Received May 12, 2003; E-mail: xhu@utoledo.edu

Adenosine 5′-triphosphate (ATP) plays an essential role in all forms of life. It functions as a carrier of energy to fuel cellular machines via hydrolysis of the high energy phosphate bonds and controls the process of cell signaling via phosphorylation of proteins, etc. Molecular recognition of ATP in ATP-binding proteins is a subject of great importance for understanding enzymatic mechanisms and for drug design.[1] The ATP molecule is made of the adenine base linked to three phosphate groups via ribose. Shown in Figure 1a is the molecular structure of the adenine base. It has the capacity to form five hydrogen bonds, acting as two hydrogen bond donors at the N6 position and hydrogen bond acceptors at N1, N3, and N7 positions. Furthermore, the $\pi$ face of the adenine ring system within the geometric proximity of a cation can experience a strong attractive interaction, which is commonly referred to as cation−$\pi$ interactions.[2] Several structure-based analyses of molecular recognition of adenine in proteins exist,[3−8] which focus on individual modes of intermolecular interactions such as hydrogen bonding or cation−$\pi$ interactions. Here, we report the role of positively charged residues Arg and Lys in binding adenine, in particular, their involvement in multiple modes of nonbonded intermolecular interactions with adenine.

A large-scale data mining of the Protein Data Bank (PDB Release, April 14, 2003) was carried out through the RELIBASE+ program to identify proteins that contain bound ATP, ADP, AMP, and 5′-adenyl-imido-triphosphate (ANP); the latter is an inert ATP substitute commonly used in place of ATP to facilitate protein crystallization. A subset of the resulting protein complexes was selected for further analysis based on two criterions: sequence homology (less than 30% sequence identity) and resolution (2.5 Å or better). Overall, a total of 68 high-resolution nonredundant complex structures were obtained (see Table S1), representing proteins from 16 different kinds of folds according to the SCOP classification.[9]

The binding environments of the adenine base of ATPs were systematically analyzed. The 3-D structures of all 68 protein complexes were aligned by superimposition of the adenine bases. Patterns of intermolecular interactions were then examined to identify cation−$\pi$ interactions on the basis of the aligned 3-D complex structures. For each protein, positively charged residues (Lys and Arg) within 5.0 Å of the adenine base were displayed with the program VMD for visual identification. The resulting pattern of cation−$\pi$ interactions between side chains of positively charged residues and the adenine base is depicted in Figure 1b and analyzed in Table 1. In 40 out of 68 complexes (59%), cation−$\pi$ interactions do exist. Furthermore, a careful examination of residue distribution indicates that lysine residues tend to occupy the major groove N7 side of the adenine base, and the arginine residues situate preferentially above or below the adenine bases with the conjugated guanidinium groups positioned parallel to the $\pi$ rings of the adenine bases. As a result of such an arrangement, multiple modes of intermolecular interactions exist between positively charged residues

***Table 1.*** Number and Percentage Occurrence of Positively Charged Residues Involved in Cation−$\pi$ Interactions[a]

| number | percentage |
|---|---|
| $N_{K+R} = 40$, $N_T = 68$ | 59% |
| $N_{K,dual} = 8$, $N_K = 17$ | 47% |
| $N_{R,dual} = 14$, $N_R = 28$ | 50% |

[a] $N_T$, total sample size; $N_{K+R}$, number of proteins with either Lys and/or Arg; $N_K$, number of proteins with Lys; $N_{K,dual}$, number of proteins with dual-mode Lys; $N_R$, number of proteins with Arg; $N_{R,dual}$, number of proteins with dual-mode Arg.

and adenine. The positively charged amino group of lysine interacts with the adenine base through both cation−$\pi$ interaction and a hydrogen bond with the N7 atom of adenine, and the charged guanidinium group of arginine interacts with the conjugated adenine rings by both cation−$\pi$ and $\pi$−$\pi$ stacking interactions. Out of a total of 17 complexes containing Lys···Adenine interactions, as many as 8 complexes (i.e., tryptophanyl-tRNA synthetase, kinesin motor Ncd, yeast glutathione synthase, TRP Ca-channel kinase domain, N5-carboxyaminoimidazole ribonucleotide synthetase, D-Ala-D-Ala ligase, glycinamide ribonucleotide transformylase, biotin carboxylase subunit of acetyl-CoA carboxylase) contain dual-mode Lys residues. The 8 proteins represent 4 different folds (i.e., ATP-grasp, adenine nucleotide alpha hydrolase, P-loop containing nucleotide triphosphate hydrolases, protein kinase). In the 28 complexes that contain Arg···Adenine interactions, dual-mode Arg residues exist in 14 complexes (i.e., NH3-dependent NAD+ − synthetase, heat shock protein, asparagine synthetase, aspartyl-tRNA synthetase, lysyl-tRNA synthetase, glutaminyl-tRNA synthetase, arginine kinase, creatine kinase, cell division regulator MinD, kinesin motor Ncd, adenylate kinase, hypothetical protein HI0065, thymidylate kinase, adenine PRTase); the 14 complexes represent 6 unique folds (i.e., ribonuclease H-like motif, Class II aaRS and biotin synthetase, glutamine synthase/guanido kinase, PRTase, adenine nucleotide alpha hydrolase, P-loop containing nucleotide triphosphate hydrolase). On the basis of the fact that the observed dual-mode interactions occur with a high frequency (~50%), and the proteins in which they occur come from many different folds, it can be concluded that the dual-mode interactions represent a significant pattern of adenine−protein interactions.

To quantify the strength of the observed dual-mode intermolecular interactions, a high level quantum chemical analysis was performed for the two representative Lys···Adenine and Arg···Adenine cases as shown in Figure 2. The intermolecular interaction energies between adenines and their surrounding residues were calculated by ab initio electronic structure calculations at the MP2 level using the supermolecular approach as in ref 12. The energy of interaction between molecules A and B is defined as

$$\Delta E^g = E_{AB} - E_A - E_B \qquad (1)$$

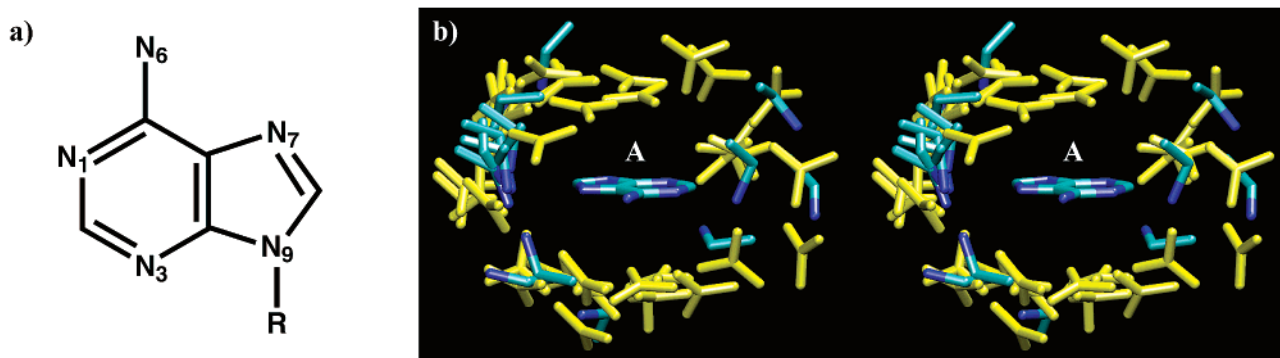The MP2 calculations were carried out using the Gaussian 98

**Figure 1.** Pattern of cation−π interactions between charged residues and adenine. (a) Molecular structure of adenine. (b) A stereo diagram of adenine surrounded by positively charged residues in 68 adenine-binding proteins. All 68 protein complexes are aligned by superimposition of the adenine bases. Truncated side chains of lysine (in cyan and blue) and arginine (in yellow) within 5.0 Å of the adenine base are shown.
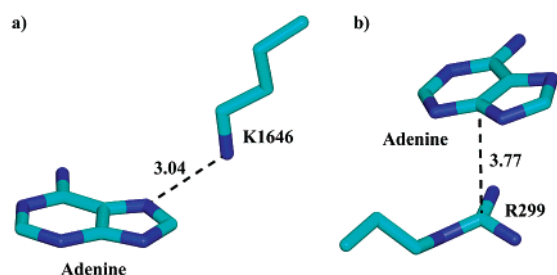


**Figure 2.** Representative intermolecular interactions between adenine and positively charged residues. (a) Lys•••Adenine pair in the TRP Ca-channel kinase domain[10] (PDB ID: 1IA9); (b) Arg•••Adenine pair in asparagine synthetase[11] (PDB ID: 12AS). The atom−atom distance is indicated by a dashed line and is given in angstroms.

**Table 2.** Basis Set Dependence of Intermolecular Interaction Energies for the Two Representatives in Figure 2

| basis set | Lys•••Adenine | | Arg••• Adenine | |
|---|---|---|---|---|
| | $\Delta E_{HF}{}^a$ | $\Delta E_{MP2}{}^a$ | $\Delta E_{HF}{}^a$ | $\Delta E_{MP2}{}^a$ |
| 6-31G | −10.93 | −13.43 | −4.81 | −7.00 |
| | (1.01) | (2.21) | (1.50) | (2.87) |
| 6-31+G* | −8.97 | −11.65 | −4.95 | −8.76 |
| | (0.64) | (2.11) | (0.89) | (3.78) |
| 6-311+G** | −9.08 | −12.26 | −4.95 | −9.10 |
| | (0.45) | (1.44) | (0.83) | (3.16) |
| 6-311++G** | −9.09 | −12.28 | −4.96 | −9.13 |
| | (0.47) | (1.58) | (0.86) | (3.30) |

$^a$ Gas-phase intermolecular interaction energies (in kcal/mol) at the MP2 level ($\Delta E_{MP2}$) and HF level ($\Delta E_{HF}$) after BSSE correction. The BSSE value for each complex is shown in parentheses.

program. The basis set superposition error (BSSE) was corrected by the Boys and Bernardi Counter Poise method.[13] Table 2 lists the BSSE corrected intermolecular interaction energies at both the Hartree−Fock (HF) and the MP2 levels using medium to large basis sets. At the MP2/6-311++G** level, an intermolecular interaction energy, $\Delta E^g$, of −12.28 and −9.13 kcal/mol is produced for the Lys•••Adenine and Arg•••Adenine pairs, respectively. Subsequently, free energy of solvation $G^{sol}$ was estimated by applying the SM5.42R model of Cramer and Truhlar[14] at the HF/6-31+G* level in the GAMESOL-v3.1 package. SCF scheme II was used to achieve better convergence, and the dielectric constant of water solvent was set to 78. Taking into account of free energy of solvation, we obtained the intermolecular interaction energy by

$$\Delta E^S = \Delta E^g + G^{sol}_{AB} - G^{sol}_A - G^{sol}_B \qquad (2)$$

Application of eq 2 to the two representative cases resulted in a

solution phase interaction energy, $\Delta E^S$, of −3.99 and −1.84 kcal/mol for the Lys•••Adenine and the Arg•••Adenine pairs, respectively. These are interaction energies of significant magnitude, which suggests that positively charged residues play an important role in molecular recognition of the adenine moiety of ATP. Sequence analysis done by us and by Biot et al.[7] indicated that these Arg and Lys residues are well conserved in their respective protein families. Comparison of MP2 and HF interaction energies in Table 2 shows a substantial correlation correction, which indicates that in addition to electrostatic (charge−quadrupole) interactions, dispersion energies (correlation effect) contribute substantially to the cation−π interaction energy.

In conclusion, a data mining analysis resulted in the discovery of multiple modes of intermolecular interactions between positively charged residues (Lys and Arg) and the adenine moiety of the ATP molecule in proteins. Contribution of the latter interactions to protein−adenine recognition was analyzed by means of the supermolecular approach at the MP2 level with solvation free energy correction, which gave rise to significant interaction strengths. In addition to contributing to the understanding of the molecular recognition of ATP, the finding reported here provides guidance for protein engineering and molecular design of ATP binding pocket targeted enzyme inhibitors.[1]

**Supporting Information Available:** A list of PDB IDs for the 68 protein complexes, coordinates of the two representative pairs, a list of computer programs used (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

**References**

(1) Chene, P. *Nat. Rev. Drug Discovery* **2002**, *1*, 665−673.
(2) Dougherty, D. A. *Science* **1996**, *271*, 163−168.
(3) Moodie, S. L.; Mitchell, J. B. O.; Thornton, J. M. *J. Mol. Biol.* **1996**, *263*, 486−500.
(4) Cappello, V.; Tramontano, A.; Koch, U. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 106−115.
(5) Denessiouk, K. A.; Johnson, M. S. *Proteins: Struct., Funct., Genet.* **2000**, *38*, 310−326.
(6) Denessiouk, K. A.; Rantanen, V. V.; Johnson, M. S. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 282−291.
(7) Biot, C.; Buisine, E.; Kwasigroch, J. M.; Wintjens, R.; Rooman, M. *J. Biol. Chem.* **2002**, *277*, 40816−40822.
(8) Kuttner, Y. Y.; Sobolev, V.; Raskind, A.; Edelman, M. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 400−411.
(9) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1995**, *247*, 536−540.
(10) Yamaguchi, H.; Matsushita, M.; Nairn, A. C.; Kuriyan, J. *Mol. Cell* **2001**, *7*, 1047−1057.
(11) Nakatsu, T.; Kato, H.; Oda, J. *Nat. Struct. Biol.* **1998**, *5*, 15−19.
(12) Wang, Y. L.; Hu, X. C. *J. Am. Chem. Soc.* **2002**, *124*, 8445−8451.
(13) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.
(14) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; A., L. D.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9−63.

JA036096P