Ab Initio Interaction Energies of Hydrogen-Bonded Amino Acid Side Chain−Nucleic Acid Base Interactions

Alan C. Cheng and Alan D. Frankel*,†

Department of Biochemistry and Biophysics and Graduate Group in Biophysics, University of California, San Francisco, California 94143-0448

Received July 14, 2003; E-mail: frankel@cgl.ucsf.edu

Hydrogen-bonding interactions often make substantial contributions to the specificity of protein−nucleic acid complexes. In a seminal study, Seeman et al.[1] postulated that two hydrogen bonds between an amino acid side chain and base could be used to uniquely distinguish among all possible base pairs in the DNA major groove. Indeed, their predicted Arg−G and Asn/Gln−A doubly hydrogen-bonded (bidentate) interactions (Figure 1; nos. 9 and 19) are the most frequently observed in protein−nucleic acid complexes.[2,3] Within RNA structures, bases also can be found in non-Watson−Crick pairs or in unpaired contexts,[2] providing new opportunities for specific amino acid−base hydrogen-bonding interactions. Using a geometric modeling approach, we previously identified 28 possible bidentate interactions to the four unpaired RNA bases.[2] Here we present interaction energies of these models, calculated by ab initio quantum chemical methods, and describe a correlation between the computed energies and observed frequencies of the interactions.[2,3]

Using a similar approach as in previous ab initio studies of amino acid interactions[4] and nucleic acid interactions,[5,6] we first optimized interaction geometries of the 28 modeled interactions at the HF/6-31G** level, resulting in five models moving away from their initial geometry and losing one of their hydrogen bonds[7] and two pairs of models each minimizing to identical arrangements.[8] The optimized geometries of the remaining 21 interactions are considered here and are shown in Figure 1. We include Lys−C (no. 5), which initially was modeled as a bidentate interaction, but we note that the geometry-optimized model has lost one hydrogen bond. Following geometry optimization, we next utilized a local MP2 method (LMP2)[9] to account for electron correlation in calculating interaction energies. A study of hydrogen-bonded formamides, formamidines, and DNA bases concluded that MP2/6-31G** energies underestimate the stabilization energy by 0.2−1.3 kcal/mol compared to aug-cc-pVDZ calculations.[6] Because our amino acid−base interaction energies range from 10 to 50 kcal/mol (see below), a medium LMP2/6-31G**//HF/6-31G** calculation was considered sufficient to rank-order our models with some confidence. We calculated in vacuo interaction energies[10] by computing $\Delta E = E_{\text{complex}} - E_{\text{base}} - E_{\text{aa}}$, using the counterpoise method to correct for basis set superposition error (BSSE) of the HF component.[11]

The calculated interaction energies of the 21 modeled interactions are presented in Table 1. Interactions involving charged amino acid side chains (Lys, Asp/Glu, and Arg) show the most favorable interaction energies, consistent with previous experiments indicating that hydrogen bonds from charged side chains may be energetically the most favorable, contributing as much as −5 kcal/mol in the context of a protein.[12] The most favorable calculated interaction is Lys−G (no. 18); it is ∼10 kcal/mol more favorable than the Arg−G
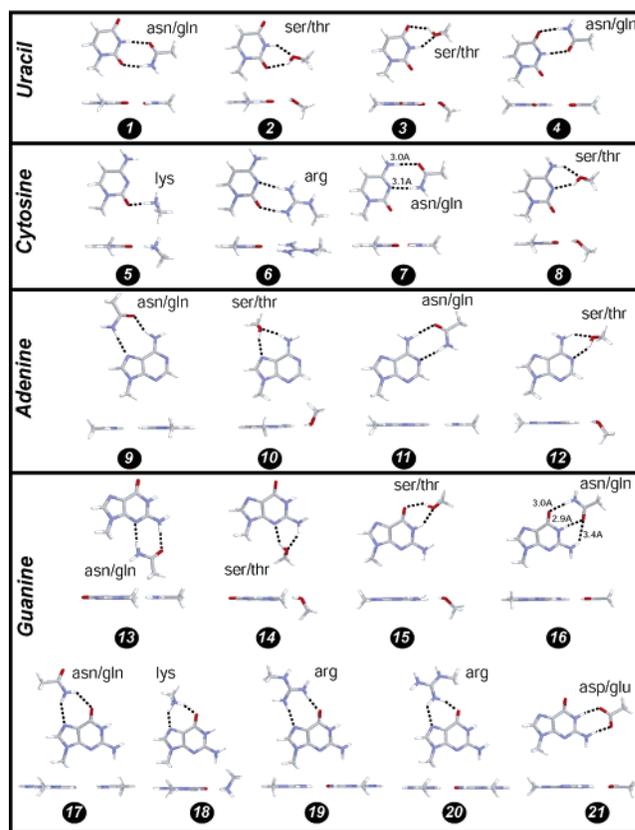
Figure 1. Minimized geometries of side chain−base interactions.

interaction (no. 19) predicted by Seeman et al.,[1] which uses the same O6 and N7 acceptors on the base. Despite its apparent energetic advantage, the Lys−G interaction is observed only about one-third as frequently as the Arg−G interaction.[3b] One rationalization is that the rotational symmetry of the Lys amino group provides more ways to interact with the phosphate backbone, thus decreasing its utility in base-specific recognition. Nevertheless, Lys−G (no. 18) still is the third most commonly observed bidentate interaction.[2,3] Asp−G (no. 21) and Lys−C (no. 5) interactions, which involve the Watson−Crick faces of bases and thus cannot occur in double helices, also are calculated to be more favorable than the Arg−G (no. 19) interaction. The Asp/Glu−G interaction may be especially favorable for recognition of unpaired G's in RNAs, where repulsive interactions to the backbone are expected to be minimal. Indeed, there are 17 Asp/Glu-G interactions in the 11-subunit TRAP−RNA complex, and some have been shown to be essential for specific binding.[13] Of the interactions involving uncharged side chains, the Asn−G (no. 16) and Asn−C (no. 7) interactions, initially modeled with bifurcated hydrogen bonds,[8] are the most favorable but have not yet been observed.

Table 1. Calculated in Vacuo Energies (kcal/mol) of Interactions

rank	model	base face	PDB	ΔE (LMP2)	ΔE (HF)
1	Lys−G, no. 18	major	yes	−47.34	−49.47
2	Asp−G, no. 21	WC	yes	−43.41	−47.49
3	Lys−C, no. 5	WC	yes	−41.48	−46.97
4	**Arg−G, no. 19**	**major**	**yes**	**−36.54**	**−39.12**
5	Arg−C, no. 6	WC	yes	−35.29	−38.31
6	Arg−G, no. 20	major	yes	−33.16	−35.81
7	Asn−G, no. 16	WC	no	−20.48	−22.22
8	Asn−C, no. 7	WC	no	−18.95	−17.85
9	Asn−A, no. 11	WC	yes	−15.61	−13.59
10	Asn−G, no. 13	minor	yes	−15.43	−13.82
11	Ser−G, no. 15	WC	no	−15.33	−14.29
12	**Asn−A, no. 9**	**major**	**yes**	**−14.85**	**−13.02**
13	**Ser−A, no. 10**	**major**	**yes**	**−14.80**	**−12.26**
14	Ser−C, no. 8	WC	yes	−14.72	−13.01
15	Ser−G, no. 14	minor	yes	−13.41	−10.99
16	Ser−A, no. 12	WC	yes	−13.15	−11.16
17	Ser−U, no. 3	WC	yes	−12.46	−11.13
18	Asn−U, no. 1	WC	no	−12.30	−13.09
19	Asn−U, no. 4	WC	yes	−12.19	−14.07
20	Asn−G, no. 17	major	no	−10.21	−9.24
21	Ser−U, no. 2	WC	no	−9.85	−10.41

Table 2. Observed Ser/Thr/Tyr Intermolecular Hydrogen Bonds

PDB id	type	complex	interaction	structure
1RBJ	crystal	ribonuclease−DNA	Thr45/A201	mid-strand
5MSF	crystal	MS2 coat−RNA aptamer	Thr45/A11	mid-strand
1DZ5	NMR	U1A-PIE−RNA	Ser45/A25	strand end
1DZ5	NMR	U1A-PIE−RNA	Thr88/A44	strand end
1IJW	crystal	Hin recombinase−DNA	Ser174/A10	helix end
IG9Y	crystal	hom. endonuclease−DNA	Tyr33/A403	β turn
1KQQ	NMR	dead ringer−DNA	Thr92/A410	β turn
1BP7	crystal	CreI endonuclease−DNA	Tyr33/A13	β turn
1TN9	NMR	Tn916 integrase−DNA	Tyr40/A120	mid-strand

In Watson−Crick helices, all possible interactions in the major and minor grooves are observed, except for Asn−G (no. 17) which ranks as the second least favorable (Table 1). The most frequently observed interactions are Arg−G (no. 19) and Asn−A (no. 9), which constitute 43 and 26% of all bidentate arrangements, respectively (or 61 and 18% when "probable" Arg−G interactions are included).[3] It is interesting that these frequencies roughly parallel the calculated interaction energies of −36.54 kcal/mol for Arg−G and −14.85 kcal/mol for Asn−A. The correlation between inter-action energies and observed frequencies is provocative, but it should be recognized that our calculations have not taken into account other important energetic factors, such as solvation/desolvation and conformational entropy,[14] and that the database of interactions is still relatively small.

The Asn−A (no. 9) interaction, which involves hydrogen bonds to the N6 and N7 groups of A, is the next most frequent interaction in DNA complexes after Arg−G (no. 19) but, unexpectedly, is observed infrequently in RNA complexes.[2,3] In contrast, the Ser−A (no. 10) interaction, which also involves hydrogen bonds to the N6 and N7 groups of A, comprises 15% of all bidentate interactions to RNA but relatively few to DNA.[3] Our calculations indicate that the interaction energies of Asn−A (no. 9) and Ser−A (no. 10) are similar (−14.85 and −14.80 kcal/mol, respectively), suggesting that factors other than this isolated interaction energy account for the different frequencies observed. One possible explanation is that the deep, narrow major groove of an A-form RNA helix can more easily accommodate the hydroxyl group of Ser/Thr/Tyr side chains than the bulkier carboxamide of Asn/Gln. However, additional geometry-based model calculations suggest that both A-form and B-form helices can accommodate the two types of interactions, at least using isolated side chains [ACC and ADF, in preparation]. Upon examining the observed cases further, we noted that eight of the nine Ser/Thr/Tyr-A interactions (with both DNA and RNA) involve nonhelical regions of protein structure (Table 2), whereas 56% of observed Asn/Gln−A interactions are found in α-helices.[2,15] Previ-ous studies have shown that the three-dimensional protein context critically influences what side chain−DNA interactions are possible and, in general, precludes a master "recognition code".[16] From the limited data set so far, it seems that Ser/Thr/Tyr side chains may be well-suited to the wide variety of binding modes and RNA and protein tertiary structures found at RNA−protein interfaces.

The results presented here provide a measure of the intrinsic energies of discrete sequence-specific interactions that may be used to help deconvolute the various contributions to RNA- and DNA-binding specificity. Amino acid interactions that span base pairs[2] or steps of base pairs in a helix provide additional ways to achieve specificity, and it would be of interest to calculate their interactions energies as well. While effects other than isolated interaction energies clearly influence the selection of a particular hydrogen-bonding scheme in a sequence-specific complex, it is encouraging that the statistical distributions of interactions observed in protein structures, including hydrogen bonds, have been found to be accurately modeled by thermodynamic Boltzmann distributions, even when treated in isolation.[17]

References

(1) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci.* **1976**, *73*, 804−808.

(2) Cheng, A. C.; Chen W. W.; Fuhrmann C. N.; Frankel A. D. *J. Mol. Biol.* **2003**, *327*, 781−796.

(3) (a) Allers, J.; Shamoo, Y. *J. Mol. Biol.* **2001**, *311*, 75−86. (b) Luscombe, N. M.; Laskowski, R. A.; Thornton J. M. *Nucleic Acids Res.* **2001**, *29*, 2860−2874.

(4) Kim, K.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 12952−12961.

(5) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136−1150.

(6) Sponer J.; Hobza P. *J. Phys. Chem. A* **2000**, *104*, 4592−4597.

(7) These five models were observed only when allowing a very wide (38°) hydrogen bond donor angle (see ref 2). Additional energetic contributions may be needed to support these interactions.

(8) Both cases involve a bifurcated three-hydrogen bond interaction and a related nonbifurcated two-hydrogen bond interaction that minimized to identical structures. The Asn-G interaction (Figure 1, no. 16) retains a bifurcated interaction while the Asn-C interaction (Figure 1, no. 7) does not.

(9) (a) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622. (b) Sæbø, S.; Tong, W.; Pulay, P. *J. Chem. Phys.* **1993**, *98*, 2170−2175.

(10) *Jaguar Software*, version 4.1; Schrodinger Inc.: Portland, OR, 2001.

(11) (a) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566. (b) Schuetz, M.; Rauhut, G.; Werner, H.-J. *J. Phys. Chem. A* **1998**, *102*, 5997−6003.

(12) Fersht A. R.; Shi, J.-P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. *Nature* **1985**, *314*, 235−238.

(13) Antson A. A.; Dodson, E. J.; Dodson, G.; Greaves R. B.; Chen X.; Gollnick P. *Nature* **1999**, *401*, 235−242.

(14) Yoshida, T.; Nishimura, T.; Aida, M.; Pichierri, F.; Gromiha, M. M.; Sarai, A. *Biopolymers* **2002**, *61*, 84−95.

(15) Berman, H. M.; Westbrook J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242. (In cases where redundant structures exist, only one structure is shown in Table 2.)

(16) (a) Pabo, C.; Nekludova, L. *J. Mol. Biol.* **2000**, *301*, 597−624. (b) Suzuki, M. *Structure* **1994**, *2*, 317−326. (c) Kono, H.; Sarai A. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 114−131.

(17) Shortle, D. *Protein Sci.* **2003**, *12*, 1298−1302.

JA037264G