

Pairing interactions between nucleobases and ligands in aptamer:ligand complexes of riboswitches: crystal structure analysis, classification, optimal structures, and accurate interaction energies

PREETHI P. SEELAM,¹ ABHIJIT MITRA,¹ and PURSHOTAM SHARMA²

¹Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology (IIIT-H), Gachibowli, Hyderabad, Telangana 500032, India

²Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh 160014, India

ABSTRACT

In the present work, 67 crystal structures of the aptamer domains of RNA riboswitches are chosen for analysis of the structure and strength of hydrogen bonding (pairing) interactions between nucleobases constituting the aptamer binding pockets and the bound ligands. A total of 80 unique base:ligand hydrogen-bonded pairs containing at least two hydrogen bonds were identified through visual inspection. Classification of these contacts in terms of the interacting edge of the aptamer nucleobase revealed that interactions involving the Watson–Crick edge are the most common, followed by the sugar edge of purines and the Hoogsteen edge of uracil. Alternatively, classification in terms of the chemical constitution of the ligand yields five unique classes of base:ligand pairs: base:base, base:amino acid, base:sugar, base:phosphate, and base:other. Further, quantum mechanical (QM) geometry optimizations revealed that 67 out of 80 pairs exhibit stable geometries and optimal deviations from their macromolecular crystal occurrences. This indicates that these contacts are well-defined RNA aptamer:ligand interaction motifs. QM calculated interaction energies of base:ligand pairs reveal a rich hydrogen bonding landscape, ranging from weak interactions (base:other, -3 kcal/mol) to strong (base:phosphate, -48 kcal/mol) contacts. The analysis was further extended to study the biological importance of base:ligand interactions in the binding pocket of the tetrahydrofolate riboswitch and thiamine pyrophosphate riboswitch. Overall, our study helps in understanding the structural and energetic features of base:ligand pairs in riboswitches, which could aid in developing meaningful hypotheses in the context of RNA:ligand recognition. This can, in turn, contribute toward current efforts to develop antimicrobials that target RNAs.

INTRODUCTION

Riboswitches are important functional RNA molecules that are generally present in the 5'-untranslated mRNA regions. Riboswitches have the ability to monitor and effectively regulate the expression of fundamental genes associated with the biosynthesis and transport of small metabolites or ligands (Breaker 2008; McCown et al. 2017). Typically, riboswitches consist of two structural domains: the aptamer domain that recognizes and binds the ligand, and the expression platform that regulates the gene expression in response to ligand binding (Mandal and Breaker 2004). Due to their crucial role in regulation of a large number of metabolites, detailed mechanisms of

functioning of riboswitches, including cognate ligand recognition and binding; and alteration of transcription and translation events, constitute a major area of research (Winkler et al. 2003; Greenleaf et al. 2008; Sharma et al. 2009a; Perdrizet et al. 2012; Savinov et al. 2014; Aboul-ela et al. 2015; Wang et al. 2016).

X-ray crystal structures of riboswitch aptamer domains, bound to different ligands, are available at atomic resolution (McCown et al. 2017). These crystal structures reveal the significant roles of noncovalent interactions in the ligand-recognition process. For example, electrostatic interactions between the positively charged sulfonium

Corresponding authors: abi_chem@iiit.ac.in, psharma@pu.ac.in
Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.071530.119>.

© 2019 Seelam et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://majournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

group of S-adenosylmethionine (SAM) and two universally conserved base pairs within the aptamer domain of the SAM-I riboswitch help in discriminating the SAM ligand from the structurally similar ligand (i.e., S-adenosylhomocysteine [Montange et al. 2010]). In addition to a number of other examples (Ren et al. 2012), recent crystal structures of the guanidinium riboswitches underscore the importance of electrostatic and base:ligand hydrogen bonding interactions in ligand discrimination and binding (Huang et al. 2017). Further, ligand binding through hydrogen bonding between the ligand and the aptamer nucleobases of purine-sensing riboswitches is responsible for triggering corresponding gene expression (Batey 2012; Krepl et al. 2018). Furthermore, van der Waals stacking interactions between the bound ligand and the aptamer nucleobases also play an important role in the organization of ligand-binding pockets in a number of riboswitches (Roth and Breaker 2009).

The diversity in noncovalent interactions harnessed during the ligand-binding process as exemplified by riboswitch crystal structures underscores the importance of characterization of the physicochemical forces that regulate riboswitch structure and dynamics. Specifically, it is necessary to develop and validate hypotheses on how various RNA structural elements within different aptamers, and their respective structurally different cognate ligands, harness unique structural principles in different functional contexts. Since base:ligand hydrogen bonding is one of the most important noncovalent forces that help in ligand recognition by the aptamer domain of riboswitches, statistical analysis of this class of interactions in terms of the identity of the participating aptamer nucleobases and bound ligands constitute the first steps in this direction. Further, quantitative analysis of the structure and strength of these interactions, and careful assessment of their role in the ligand recognition may contribute to the understanding of the overall ligand-binding process by riboswitches. Structural bioinformatics analysis of crystal structures of RNA aptamer:ligand complexes, coupled with quantum mechanical (QM) calculations, constitute a comprehensive implementation strategy in this direction. QM calculations are especially useful, since they can provide an accurate description of individual energy components that influence the structural stability of noncovalently bonded chemical entities on one hand, and the intrinsic strengths of noncovalent interactions on the other.

Of course, given the differential effects of solvation on different energy components, the results of QM calculations in vacuo are not directly comparable to the binding free energies in solution. Furthermore, QM binding energies describe interactions on the molecular potential energy surface (PES), whereas binding strengths in experimental measurements (or sufficiently converged molecular dynamics simulations) are dictated by the free

energy surface (FES), which is primarily derived from relative populations of different molecular states (Šponer et al. 2013). As a result, the relation between the PES of pairing interaction and FES of the molecular system is generally complex. Specifically, as previously discussed in context of base stacking interactions (Šponer et al. 2013) and which is equally applicable to pairing interactions, while interaction energy derived from PES of a pairing interaction reflects the direct electronic structure interaction between the monomers, the pairing free energy is derived from a population of a given base-pair state relative to the population of the unpaired configurations in the context of a given system (bound vs. unbound). Although PES contributes to FES, the interrelation is complex and context-dependent. Nevertheless, QM calculations can accurately quantify the intrinsic strength and stability of hydrogen bonding interactions between aptamer nucleobases and the bound ligands, which may, in turn, help in assessing the contribution of individual pairwise interactions toward the overall strength of binding between the ligand and riboswitches.

A substantial number of crystal structure analysis and QM studies have focused on the analysis of hydrogen bonding interactions that constitute structural motifs present in RNA macromolecules (e.g., nucleobase pairs [Šponer et al. 2004, 2005a,b,c; Das et al. 2006; Oliva et al. 2006; Sarver et al. 2008; Sharma et al. 2008, 2010a, b; Seelam et al. 2017]), higher-order hydrogen bonding contacts including base triplets and quartets (Oliva et al. 2006; Abu Almakarem et al. 2012), and their posttranscriptionally modified counterparts (Chawla et al. 2015; Seelam et al. 2017) and base-phosphate contacts (Zirbel et al. 2009). However, in spite of the previous classification of base pairs and pseudo pairs in RNA:ligand complexes (Kondo and Westhof 2010), very few QM studies have analyzed the role of noncovalent interactions in the context of specific riboswitches (Sharma et al. 2009b; Chawla et al. 2014). Thus, existing literature does not provide a comprehensive understanding of the structural and energetic aspects of aptamer:ligand hydrogen-bonding interactions.

To address these aspects, we have carried out crystal structure database analysis of aptamer-ligand complexes, and have analyzed the nature of base:ligand hydrogen-bonding interactions as a function of nucleobase and ligand types, the interacting nucleobase edges, and the nature of H-bonds present within the pair. In addition, with the help of QM calculations, we have obtained optimal geometries of base:ligand pairs and have calculated their accurate binding energies. Overall, our combined analysis of crystal structure data sets and QM geometries and energies provide unique insights into the molecular features of ligand recognition by riboswitches.

RESULTS

Statistical overview of base:ligand interactions from crystal structural occurrences

In the data set of chosen crystal structures (Table 1) the size of the bound ligands varies considerably, ranging from small chemical entities (e.g., glycine and lysine) to complex

chemical structures (e.g., folate-related ligands, S-adenosyl methionine, and thiamine pyrophosphate derivatives, Supplemental Figs. S1–S4). The total 80 unique base:ligand interactions, thus identified, can be further classified based on the interacting edge (i.e., Watson–Crick edge [W edge], Hoogsteen edge [H edge] or sugar edge [S edge]) of the aptamer nucleobase (adenine [A], cytosine [C], guanine [G] or uracil [U]). It is observed that 78% of these

TABLE 1. PDB accession codes of the 67 riboswitch:ligand complexes studied in the present work

Class ^a	Riboswitch	Ligand code ^b	Ligand name	PDB code
Coenzymes and their derivatives	Thiamine	2QB	5-(azidomethyl)-2-methylpyrimidin-4-amine	4nya
		BFT	s-benzoylthiamine o-monophosphate	2hoo
		D2X	oxythiamine pyrophosphate	3d2x
		PYI	pyrithiamin pyrophosphate	3d2v
		TPP	thiamine diphosphate	2cky, 2gdi, 2hoj, 3k0j
	Tetrahydrofolate	TPS	thiamin phosphate	2hom
		6AP	9h-purine-2,6-diamine	4lvz
		7DG	7-deazaguanine	4lwv
		FFO	[6S]-5-formyl-tetrahydrofolate	3suh, 4lvw
		H4B	5,6,7,8-tetrahydrobiopterin	4lvx
		LYA	pemetrexed	4lvy
	S-adenosyl methionine	THF	5-hydroxymethylene-6-hydrofolic acid	3sux
		SAM	s-adenosylmethionine	2gis, 2qwy, 3e5c, 3gx5, 3iqr, 3v7e, 4b5r, 4kqy, 4oqu, 5fkh
		SAH	s-adenosyl-l-homocysteine	3e5e, 3gx3, 3nqn
		SFG	adenosyl-ornithine	3gx2
Flavin mono nucleotide	EEM	Se-Adenosylselenomethionine	3e5f	
	FMN	flavin mononucleotide	2yie, 3f2q	
	RBF	Riboflavin	3f4g	
	RS3	Roseoflavin	3f4h	
Nucleobases and their derivatives	Adenine	ADE	Adenine	1y26, 4lw0, 4txx
		29G	pyrimido[4,5-d]pyrimidine-2,4-diamine	4lx5
		29H	2-aminopyrimido[4,5-d]pyrimidin-4(3h)-one	4lx6
	Guanine	5AZ	6-amino-1,3,5-triazin-2(1h)-one	3la5
		2BP	9h-purin-2-amine	3g4m
		3AY	pyrimidine-2,4,6-triamine	2g9c
		6GO	6-o-methylguanine	3fo6
		6GU	6-chloroguanine	3gog
		A2F	2-fluoroadenine	3got
		N6M	n-methyl-9h-purin-6-amine	2xo1
	deoxy guanine	XAN	Xanthine	3gao
		5GP	guanosine-5'-monophosphate	3slq
		DGP	2'-deoxyguanosine-5'-monophosphate	3slm
		GMP	Guanosine	3skz
		PQ0	7-deaza-7-cyano-guanine	3gca
	preQ ₁	PRF	7-deaza-7-aminomethyl-guanine	3fu2, 3k1v, 3q50, 4jf2, 5d5l
		ZTP	aminoimidazole 4-carboxamide ribonucleotide	4xw7, 4znp, 5btq
Amino acids	c-di-GMP	C2E	Cyclic diguanosine monophosphate	3q3z, 4yb0, 3mxh
	Lysine	LYS	Lysine	3d0u,
	Glycine	GLY	Glycine	3owi, 4erl
Sugars	glms ribozyme	G6P	α -d-glucose-6-phosphate	2z74

^aRiboswitches are categorized according to Sherwood and Henkin (2016).

^bChemical structures of all the ligands are provided in Supplemental Figures S1–S4.

The names and three-letter codes of 40 different bound ligands are provided.

interactions involve the nucleobase W edge (7% A, 9% G, 36% U, and 26% C), followed by the S edge of purines (21% total [i.e., 5% A and 16% G]) and the H edge of U (1%, Table 2; Fig. 1C). Thus, pyrimidines in the binding

pocket of aptamers interact with the ligands, almost exclusively, through their W edge. On the other hand, the purines mainly use the W edge or S edge, where the highest binding occurrence is shown by U (37%), followed by

TABLE 2. Distribution of base:ligand interactions in the riboswitch aptamers with respect to base-pairing families^a

Interaction	Total occurrences	Base:ligand interaction (interaction code; PDB code)			
		Coenzymes	Nucleobases	Amino acids	Sugars
B:B	19 (cVWW)	A:FMN (99:200; 3f2q), A:RBF (99:200; 3f4g)	U:29G (74:501; 4lx5), U:3AY (51:91; 2g9c), U:3AY (74:91; 2g9c), U:A2F (74:91; 3got), U:ADE (74:90; 1y26), U: N6M (74:1081; 2x01), C:29H (74:601; 4lx6), C:2BP (74:91; 3g4m), C:5AZ (74:84; 3la5), C:5GP (80:120; 3slq), C:6GO (74:91; 3fo6), C:6GU (74:91; 3gog), C:C2E (92:1; 3mxh), C:DGP (80:120; 3slm), C:PQ0 (15:34; 3gca), 1C:PRF (17:101; 3fu2), C:XAN (74:90; 3gao)		
	7 (tVWW)	U:7DG (42:101; 4lvw), U:H4B (42:102; 4lvx), C:7DG (53:102; 4lvw), C:FFO (53:102; 4lvv), C:H4B (53:101; 4lvx), C:THF (58:103; 3sux)	C:PRF (30:101; 4jf2)		
	16 (tVWS)	U:FFO (7:101, 4lvv), U:FFO (25:102; 4lvv), U:7DG (7:101; 4lvw),	A:PQ0 (29:34; 3gca), A:PRF (30:101; 3fu2), U:2BP (51:91; 3g4m), U:6GO (51:91; 3fo6), U:6GU (51:91; 3gog), U: ADE (51:90; 1y26), U:N6M (51:1081; 2x01), U:XAN (51:90; 3gao), C:29G (51:501; 4lx5), C:29H (51:601; 4lx6), C:5AZ (51:84; 3la5), C:5GP (58:120; 3slq), C:GMP (58:120; 3skz)		
	4 (cVWH)	U:SAH (57:301; 3gx3), U:SAM (78:201; 4kqy), U:SAM (44:100; 4l81), U:SFG (57:301; 3gx2)			
	3 (tVWH)	U:SAM (44:100; 2qwy)	G:C2E (20:1; 3mxh), U:6AP (7:101; 4lvz), U:PQ0 (6:34; 3gca)		
	1 (cHS)		U:PRF (41:101; 4jf2),		
	4 (cVWS)	U:H4B (25:101; 4lvx), U:LYA (25:101; 4lvy), U:THF (28:103; 3sux)			
	3 (cVSW)	A:SAH (45:301; 3gx3), A:SAM (46:201; 4kqy), A:SFG (45:301; 3gx2)			
	7 (tVSW)	G:2QB (40:110; 4nya), G:BFT (40:95; 2hoo), G:EEM (7:216; 3e5f), G:SAM (8:101; 4l81), G:TPP (28:1084; 2cky), G:TPS (40:97; 2hom), G:D2X (28:85; 3d2x)			
B:A	9	A:SAM (46:201; 4kqy), G:SAM (S edge, 79:201; 4kqy), G:SAH (11:301; 3gx3), G:SFG (W edge, 11:301; 3gx2), G:SAM (W edge, 48:101; 4l81), 3G:SAM (S edge, 48:101; 4l81), G:SFG (S edge, 58:301; 3gx2)		G:LYS (111:205; 3d0u), U: GLY (69:89; 3oww)	
B:Ph	3	G:D2X (66:85; 3d2x), G:PYI (66:82; 3d2v)			G:G6P (65:1; 2z74)
B:S	2	G:RS3 (W edge, 11:200; 3f4g), G:RS3 (S edge, 84:200; 3f4h)			
B:O	2		A:PRF (70:101; 4jf2), U:AMZ (66:101; 4znp)		

^aOnly the hydrogen atoms were optimized for base:ligand pairs that are represented in bold font.

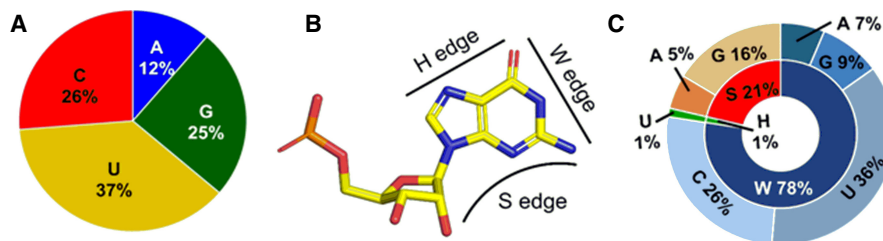


FIGURE 1. (A) Statistical distribution of the aptamer nucleobases that interact with the ligand in riboswitches in the crystal structures. (B) Interacting edges (Watson–Crick edge [W edge]; Hoogsteen edge [H edge] and Sugar edge [S edge]) of the aptamer nucleobases. (C) Distribution of base:ligand pairs as a function of the interacting edge (*inner circle*) and the aptamer base (*outer circle*).

C (26%). A close third is G (25%) using its S edge (16%) and W edge (9%) in the ratio of around 5:3, whereas the least frequent is A (12%) using its S edge (5%) and W edge (7%) in an ~1:1 ratio. In contrast, to multiple examples of W-edge and S-edge interactions, we could find only a single occurrence of H-edge interaction, where U interacts

with PQ0 in the aptamer domain of the preQ₁ riboswitch (PDB code: 3gca).

Classification of base:ligand interactions

Classification of base:ligand pairs on the basis of the chemical composition of the ligand, as well as on the nature of the ligand-aptamer interactions at the molecular level, reveals five basic interaction types: base:base (B:B), base:amino acid (B:A), base:sugar (B:S), base:phosphate (B:Ph), and base:other (B:O, Table 2; Fig. 2). Although B:B (80%) and B:A (11%) interactions have significant occurrences, instances of the other three types of interactions (B:S [two instances], B:Ph [three instances], and B:O [two instances]) constitute only the remaining 8% of occurrences. Interestingly, the aptamer nucleobase component of all three B:Ph interactions involves the W edge of the aptamer base, whereas B:S and B:O interactions involve either W edge or S edge of the base.

The most abundant (B:B) contacts involve ligands possessing partial or complete structural similarity with the canonical RNA nucleobases. Consequently, the pairing geometries of aptamer-ligand B:B interactions are comparable with the base-pairing geometries observed in folded RNA structures (Leontis and Westhof 2001). Notably, about 30% of these interactions adopt cWW base-pairing, followed by tWS (25%), tWW (11%), cWS (6%), cWH (6%), and tWH (6%, Table 2) pairings (Leontis and Westhof 2001). All these interactions involve the W edge of the aptamer nucleobase and constitute 84% of the B:B interactions. In contrast, only two pairing geometries involve the S edge of the aptamer nucleobase (cSW, 5% and tSW, 11%). The single interaction involving the H edge of the aptamer nucleobase forms cHS pairing. Overall, the observed B:B interactions involve nine unique base:ligand edge interaction patterns (c/tWW, c/tWH, c/tWS, c/tSW, and cHS).

In contrast, B:A interactions (Fig. 2B) span nine instances of base:amino acid contacts, and can be further classified into two interaction types: two interactions involving a free amino acid and seven contacts where the interacting amino acid moiety is a part of the more complex chemical

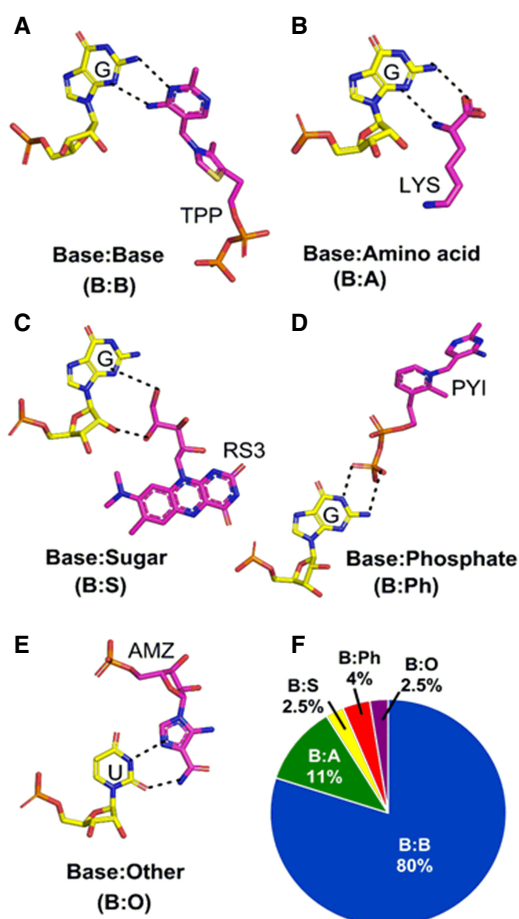


FIGURE 2. (A–E) Examples of each base:ligand interaction type (base:base [B:B], base:amino acid [B:A], base:sugar [B:S], base:phosphate [B:Ph], and base:other [B:O]) observed in the crystal structures of ligand-bound aptamer domains of riboswitches. (F) Statistical distribution of different base:ligand interactions in the crystal structures.

Table 3. Root mean square deviations (rmsd, Å) with respect to the source (i.e., PDB) structures and gas phase interaction energies (kcal/mol) for B:B type of base:ligand pairs belonging to the cWW base-pairing family

Structure #	Interaction code; PDB code	Base:ligand interaction	rmsd	Interaction energy
1	99:200; 3f2q	A:FMN	2.386	-15.1
2	99:200; 3f4g	A:RBF	1.799	-14.8
3	74:501; 4lx5	U:29G	0.273	-18.5
4	51:91; 2g9c	U:3AY	0.203	-18.2
5	74:91; 2g9c	U:3AY	0.197	-18.2
6	74:91; 3got	U:A2F	0.186	-13.2
7	74:90; 1y26	U:ADE	0.266	-15.3
8	74:1081; 2x01	U:N6M	0.260	-15.3
9	74:601; 4lx6	C:29H	0.289	-29.2
10	74:91; 3g4m	C:2BP	0.378	-15.2
11	74:84; 3la5	C:5AZ	0.208	-31.6
12	80:120; 3slq	C:5GP ^a	-	-30.0
13	74:91; 3fo6	C:6GO	0.359	-11.1
14	74:91; 3gog	C:6GU	0.347	-16.5
15	92:1; 3mxh	C:C2E ^a	-	-32.3
16	80:120; 3slm	C:DGP	1.430	-29.9
17	15:34; 3gca	C:PQ0	0.247	-30.2
18	17:101; 3fu2	C:PRF	0.279	-28.6
19	74:90; 3gao	C:XAN	0.336	-17.9

^aOnly the hydrogen atoms were optimized for these structures, and heavy-atom coordinates were derived from crystal structures (see Materials and Methods). As a result, no RMSD information is provided for these pairs.

Table 4. Root mean square deviations (rmsds, Å) with respect to the source (i.e., PDB) structures and gas phase interaction energies (kcal/mol) for B:B type of base:ligand pairs belonging to the tWW and tWS base-pairing families

Structure #	Pairing geometry	Interaction code; PDB code	Base:ligand interaction	rmsd	Interaction energy
20	tWW	42:101; 4lww	U:7DG	0.339	-18.6
21		42:102; 4lvx	U:H4B	0.541	-18.4
22		53:102; 4lww	C:7DG	0.984	-29.4
23		53:102; 4lww	C:FFO ^a	-	-17.8
24		53:101; 4lvx	C:H4B	0.772	-12.7
25		30:101; 4jf2	C:PRF	0.714	-13.5
26		58:103; 3sux	C:THF	0.780	-12.6
27	tWS	29:34; 3gca	A:PQ0	0.172	-12.8
28		30:101; 3fu2	A:PRF	0.290	-12.6
29		51:91; 3g4m	U:2BP	0.204	-19.8
30		51:91; 3fo6	U:6GO	0.370	-12.3
31		51:91; 3gog	U:6GU	0.195	-19.9
32		7:101; 4lww	U:7DG	0.327	-13.7
33		51:90; 1y26	U:ADE	0.165	-17.3
34		7:101; 4lvv	U:FFO	0.468	-19.6
35		25:102; 4lww	U:FFO ^a	-	-16.2
36		51:1081; 2x01	U:N6M	0.217	-17.6
37		51:90; 3gao	U:XAN	0.825	-17.9
38		51:501; 4lx5	C:29G	0.762	-19.9
39		51:601; 4lx6	C:29H	0.751	-22.7
40		51:84; 3la5	C:5AZ	0.118	-19.8
41		58:120; 3slq	C:5GP ^a	-	-19.6
42		58:120; 3skz	C:GMP	0.525	-19.3

^aOnly the hydrogen atoms were optimized for these structures, and heavy-atom coordinates were derived from crystal structures (see Materials and Methods). As a result, no RMSD information is provided for these pairs.

Table 5. Root mean square deviations (rmsds, Å) with respect to the source (i.e., PDB) structures and gas phase interaction energies (kcal/mol) for B:B type of base:ligand pairs belonging to the cWH, tWH, cHS, cWS, cSW and tSW base-pairing families

Structure #	Pairing geometry	Interaction code; PDB code	Base:ligand interaction	rmsd	Interaction energy
43	cWH	78:201; 4kqy	U:SAM	0.664	-18.5
44		47:100; 4l81	U:SAM	0.630	-20.7
45		57:301; 3gx2	U:SFG	1.744	-22.9
46		57:301; 3gx3	U:SAH	0.634	-28.6
47	tWH	20:1; 3mxh	G:C2E ^a	-	-24.7
48		7:101; 4lvz	U:6AP	0.196	-16.5
49		44:100; 2qwy	U:SAM	2.039	-14.1
50	cHS	6:34; 3gca	U:PQO	1.491	-10.6
51	cWS	25:101; 4lvx	U:H4B	0.769	-18.3
52		25:101; 4lvy	U:LYA	0.479	-13.7
53		41:101; 4jf2	U:PRF	0.277	-20.1
54		28:103; 3sux	U:THF	0.246	-18.5
55	cSW	45:301; 3gx2	A:SFG	0.830	-17.5
56		45:301; 3gx3	A:SAH ^a	-	-15.5
57		46:201; 4kqy	A:SAM	1.013	-18.0
58	tSW	40:110; 4nya	G:2QB	1.070	-15.9
59		40:95; 2hoo	G:BFT	1.443	-12.6
60		28:85; 3d2x	G:D2X ^a	-	-13.5
61		7:216; 3e5f	G:EEM	0.715	-12.2
62		8:101; 4l81	G:SAM	1.014	-12.4
63		28:1084; 2cky	G:TPP	2.175	-10.9
64		40:97; 2hom	G:TPS	0.462	-10.8

^aOnly the hydrogen atoms were optimized for these structures, and heavy-atom coordinates were derived from crystal structures (see Materials and Methods). As a result, no RMSD information is provided for these pairs.

Table 6. Root mean square deviations (rmsds, Å) with respect to the source (i.e., PDB) structures and gas phase interaction energies (kcal/mol) for B:A, B:Ph, B:S and B:O type of base:ligand pairs

Structure #	Interaction type	Interaction code; PDB code	Base:ligand interactions	rmsd	Interaction energy
65	B:A	4kqy; 46:201	A:SAM	0.953	-18.8
66		3gx3; 11:301	G:SAH	1.933	-18.2
67		4l81; 8:101	G:SAM ^a	-	-4.1
68		3d0u; 111:205	G:LYS	0.932	-20.8
69		4kqy; 79:201	G:SAM	1.727	-24.9
70		4l81; 48:101	G:SAM	1.526	-39.8
71		3gx2; 11:301	G:SFG (S edge)	1.168	-22.5
72		3gx2; 58:301	G:SFG (W edge) ^a	-	-7.0
73		3oww; 69:89	U:GLY	1.420	-8.0
74	B:Ph	3d2x; 66:85	G:D2X ^a	-	-3.7
75		2z74; 65:1	G:G6P	0.726	-31.1
76		3d2v; 66:82	G:PYI	2.716	-48.4
77	B:S	3f4g; 11:200	G:RS3 (W edge)	0.698	-19.9
78		3f4h; 84:200	G:RS3 (S edge) ^a	-	-12.8
79	B:O	4jf2; 70:101	A:PRF ^a	-	-2.6
80		4znp; 66:101	U:AMZ	0.646	-21.1

^aOnly the hydrogen atoms were optimized for these structures, and heavy-atom coordinates were derived from crystal structures (see Materials and Methods). As a result, no RMSD information is provided for these pairs.

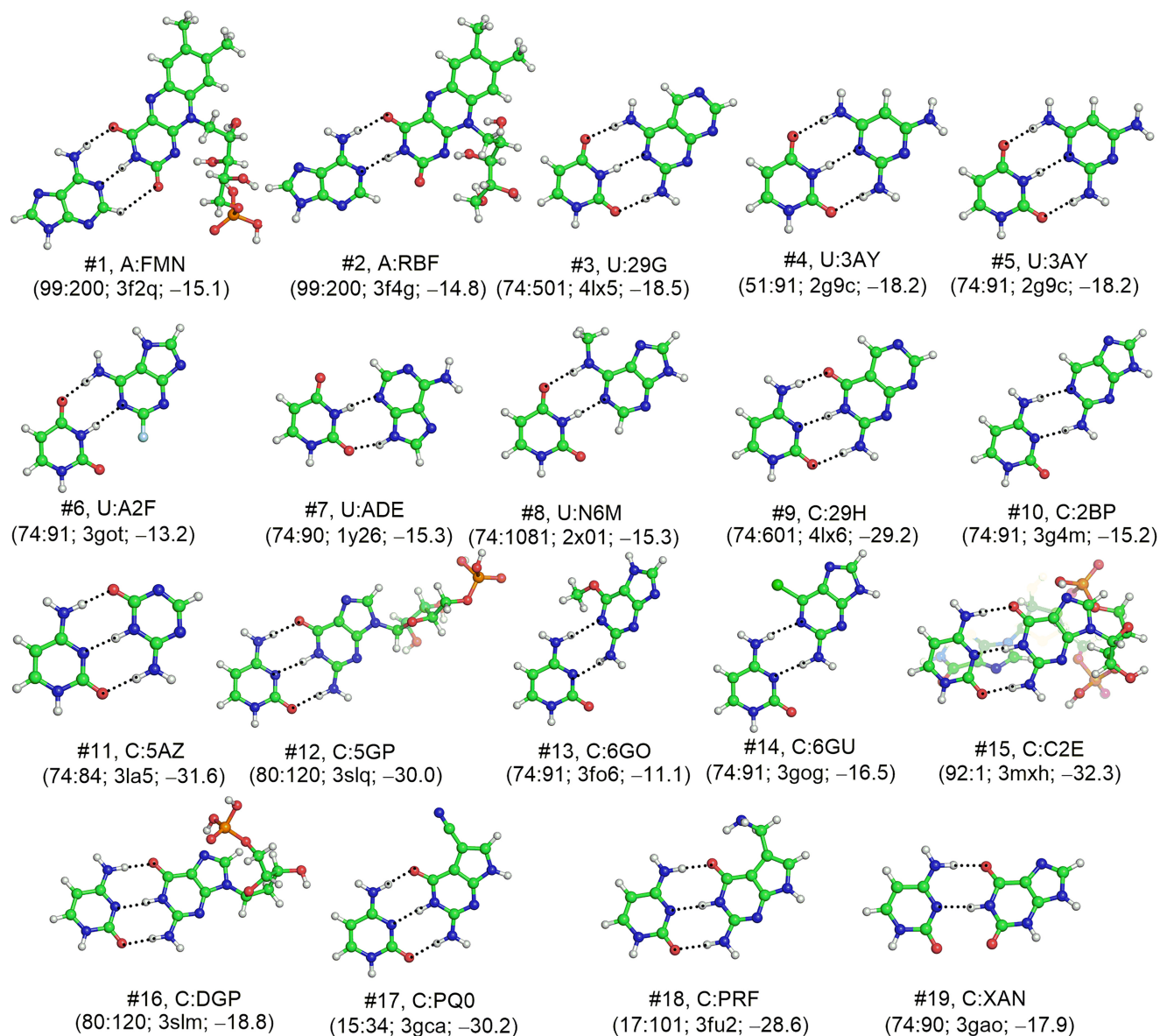


FIGURE 3. B3LYP/6-31G(d,p) optimized geometries of base:base (B:B) type of base:ligand interactions that resemble the *cis* Watson Crick–Watson Crick (cWW) RNA base pairs. Structure number (#) and base:ligand interaction identity are provided against each structure for identification from tables. Base-pairing interaction code and PDB code of the crystal structure used for optimization, as well as the MP2/aug-cc-pVDZ interaction energy (kcal/mol) are provided for each structure in parentheses. Hydrogen bonding donor–acceptor (D–A) distance (Å), acceptor–hydrogen (A–H) distance (Å), and D–H–A (deg) are provided for each hydrogen bond (dotted lines) in Supplemental Table S1. For structures #12 and #15, only the positions of hydrogen atoms were optimized, whereas full optimization was carried out for all other structures. Structures #4 and #5 relaxed to identical geometry after optimization. None of the structures carry a charged moiety.

structure of the ligand (Supplemental Figs. S1, S4). Both the interaction types involve the W edge or S edge of aptamer nucleobases. Further, the B:A interactions in aptamer complexes differ from those occurring in the context of ribonucleotide:protein (Kondo and Westhof 2011) and RNA:protein complexes (Cheng et al. 2003) in terms of the source of amino acid moiety (free ligand in the case of aptamer complexes and as part of a protein chain in the context of ribonucleotide:protein and RNA:protein

complexes). On the other hand, B:S, B:Ph or B:O interactions with very few available examples, involve the interaction of aptamer nucleobase with the sugar, phosphate or side chain moieties of the ligands.

QM calculations

QM geometry optimization of base:ligand pairs, isolated from their respective crystal environments, is expected

to reveal their intrinsic structural features. This may, in turn, provide insights into the contribution of base:ligand pairing to the overall stabilization of ligand within the aptamer binding pocket. Accordingly, geometry optimizations, followed by interaction energy calculations were attempted on all base:ligand pairs (Tables 3–6). Although the hydrogen bonding patterns of 8 B:B, 2 B:A, and 1 each of B:Ph, B:S, and B:O pairs, either could not be optimized or deviated significantly from their crystal structures on full optimization (Table 2), all the remaining pairs retained their respective hydrogen bonding patterns on optimization.

B:B pairs

Among the B:B interactions, all 53 pairs that interact through the W edge of the aptamer nucleobase show linear N–H...N/O or O–H...N H-bonds on optimization, indicating strong base-pairing characteristics (Figs. 3–6; Supplemental Tables S1–S7); 19 of these interactions possess a minimum of three strong (N–H...N/O or O–H...N type) hydrogen bonds (Supplemental Tables S1–S7). Except for complex ligands with flexible functional groups (e.g., SAM and its derivatives, TPP, FMN, and THF), all the W-edge B:B pairs showed relatively small deviations from their respective crystal structure geometries on optimization (rmsd ranging from $0.1 \text{ \AA} \leq 1 \text{ \AA}$), which points toward their significance as well-defined base:ligand interaction motifs. However, in the case of complex ligands containing flexible functional groups (e.g., thiazole and pyrophosphate groups of TPP derivatives, sugar moiety of FMN derivatives, p-amino benzoate and glutamate of THF derivatives), the structural deviation increased upon optimization, resulting in higher rmsd (1.2 \AA – 2.4 \AA), albeit without affecting the base:ligand hydrogen bonding. Regardless, in all cases, interaction energy calculations reveal substantial (-14 to -32 kcal/mol) strength of W-edge B:B pairing (Tables 3–5).

With the exception of three pairs (G:TPS, G:EEM, and A:SFG), the structural deviation on optimization of all the S-edge B:B interactions (rmsd of 1.0 \AA – 2.2 \AA) is high (Table 5). This occurs partly due to the inherent torsional flexibility associated with ribose, which decreases to the

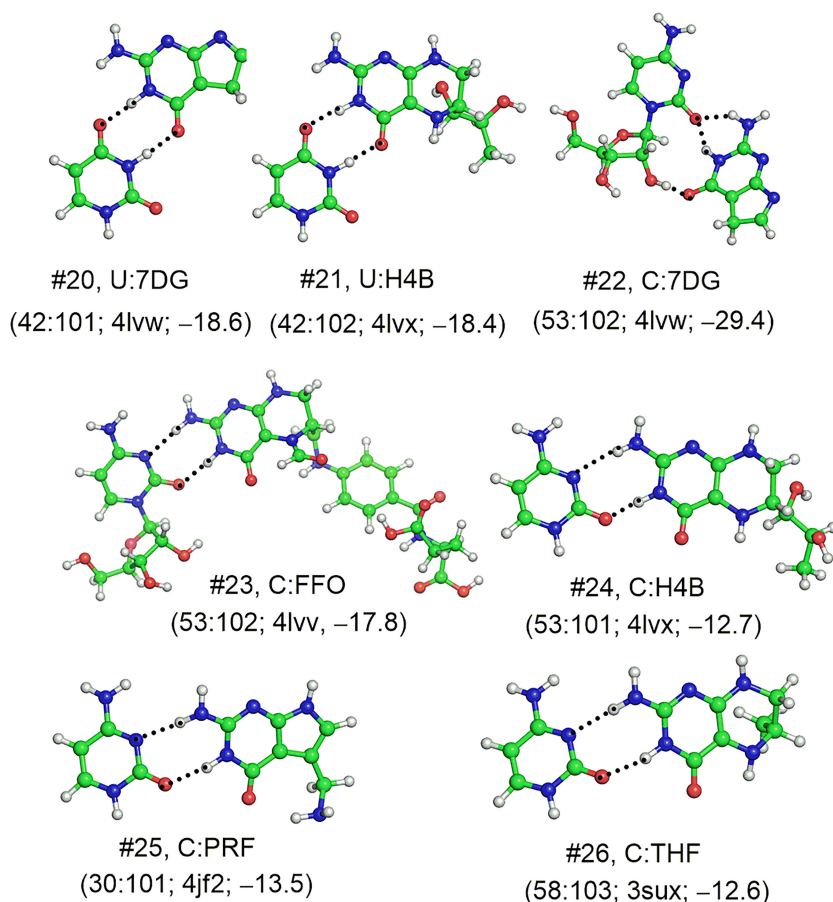


FIGURE 4. B3LYP/6-31G(d,p) optimized geometries of base:base (B:B) type of base:ligand interactions that resemble the *trans* Watson Crick–Watson Crick (tWW) RNA base pairs. Structure number (#) and base:ligand interaction identity is provided against each structure for identification from tables. Base-pairing interaction code and PDB code of the crystal structure used for optimization, as well as the MP2/aug-cc-pVDZ interaction energy (kcal/mol) are provided for each structure in parentheses. Hydrogen bonding donor–acceptor (D–A) distance (Å), acceptor–hydrogen (A–H) distance (Å), and D–H–A (deg) is provided for each hydrogen bond (dotted lines) in Supplemental Table S2. For structure #23, only the positions of hydrogen atoms were optimized, whereas full optimization was carried out for all other structures. None of the structures carry a charged moiety.

nearest local minima on optimization. Further, except for three interactions mediated by at least three H-bonds (G:2QB, G:EEM, and 2G:SAM), all other S-edge pairs possess only two hydrogen bonds on optimization (Figs. 6, 7; Supplemental Tables S8, S9). Albeit smaller than for W-edge pairs, the strengths of S-edge B:B base pairs are evidenced by the calculated interaction energies (-11 to -18 kcal/mol , Table 5).

B:A pairs

For the B:A interactions, optimizations reveal three distinct hydrogen bonding patterns (Fig. 8; Supplemental Table S10). The first pattern includes the interaction of both amino ($-\text{NH}_2$) and carboxylic groups ($-\text{CO}_2\text{H}$) of the amino acid moiety of the ligand. In this type, the amino group

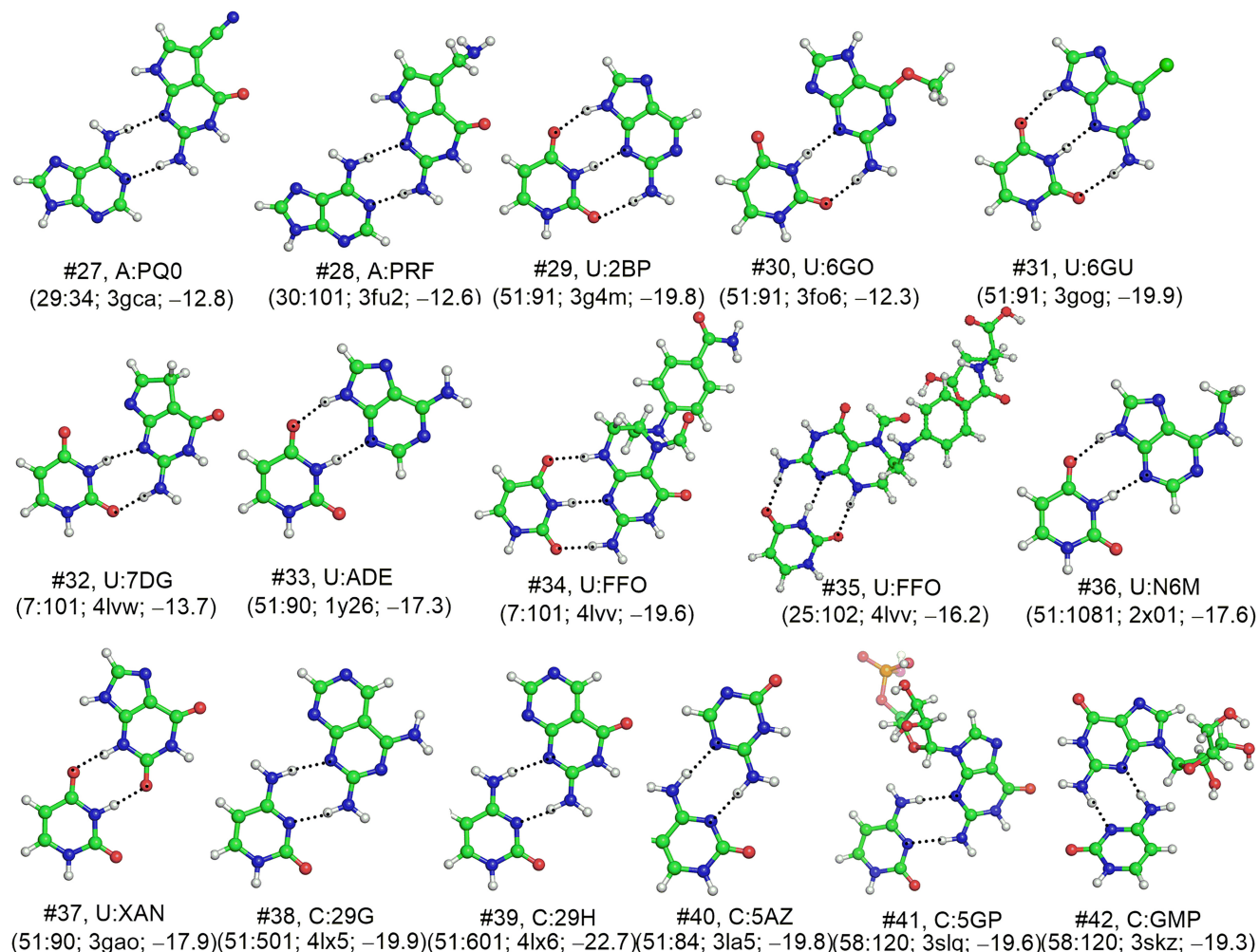


FIGURE 5. B3LYP/6-31G(d,p) optimized geometries of base:base (B:B) type of base:ligand interactions that resemble the *trans* Watson Crick-Sugar (tWS) RNA base pairs. Structure number (#) and base:ligand interaction identity is provided against each structure for identification from tables. Base-pairing interaction code and PDB code of the crystal structure used for optimization, as well as the MP2/aug-cc-pVDZ interaction energy (kcal/mol) are provided for each structure in parentheses. Hydrogen bonding donor-acceptor (D-A) distance (Å), acceptor-hydrogen (A-H) distance (Å), and D-H-A (deg) are provided for each hydrogen bond (dotted lines) in Supplemental Table S3. For structures #34 and #41, only the positions of hydrogen atoms were optimized, whereas full optimization was carried out for all other structures. None of the structures carry a charged moiety.

acts a donor, whereas the carboxylic oxygen atoms either act as a donor or acceptor, thus stabilizing each pair with a minimum of two strong H-bonds. Optimized structures of five of the total nine B:A pairs exhibit this interaction pattern, which spans a wide range (-8 to -40 kcal/mol) of interaction strength. The second pattern observed only in one (G:SFG) pair, simultaneously harnesses both the donor as well as acceptor properties of the amino nitrogen of the amino acid and possesses significant (-23 kcal/mol) interaction energy. Specifically, the -NH₂ group of the ornithine moiety of SFG acts as a H-bond donor [N-H(SFG)⋯N3(G)] as well as acceptor [N2-H(G)⋯N(SFG)] in the optimized G:SFG pair, and. In contrast, the third interaction pattern observed in three (G:SAH, G:SAM, and G:SFG) B:A pairs involves the acceptor interaction of the -

CO₂H group of the amino acid moiety with the W edge of G. However, relatively small magnitude of interaction strength (-8 to -18 kcal/mol) suggest the inherent weakness of this hydrogen-bonding pattern compared with the other two patterns (Table 6; Fig. 8). Regardless, the B:A pairs span a relatively smaller range of rmsd (0.9 Å-1.9 Å) compared with B:B pairs (0.1 Å-2.4 Å, Table 6).

B:Ph pairs

Among the B:Ph interactions, except for G:D2X pair which could not be fully-optimized, the other two pairs have more than two hydrogen bonds. For example, G:G6P is stabilized by three hydrogen bonds, whereas four acceptor-bifurcated hydrogen bonds are observed for the G:

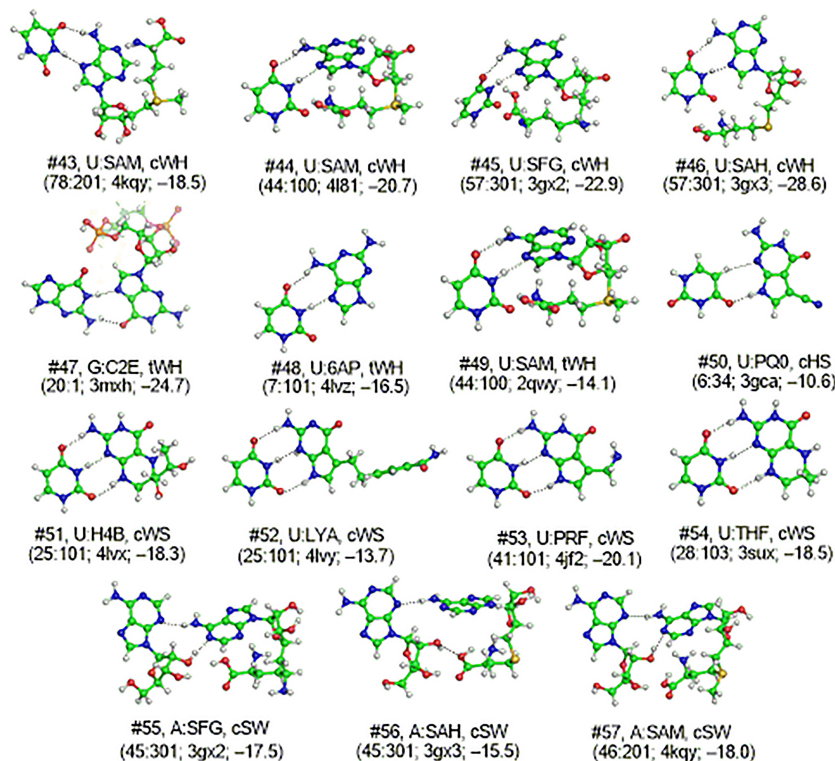


FIGURE 6. B3LYP/6-31G(d,p) optimized geometries of base:base (B:B) type of base:ligand interactions that resemble the *cis* Watson Crick–Hoogsteen (cWH), *trans* Watson Crick–Hoogsteen (tWH), *cis* Hoogsteen–Sugar (cHS), *cis* Watson Crick–Sugar (cWS) or *cis* Sugar–Watson Crick (cSW) RNA base pairs. Structure number (#), base:ligand interaction identity, and base pair type (cWH, tWH, cHS, cWS or cSW) are provided against each structure for identification from tables. Base-pairing interaction code and PDB code of the crystal structure used for optimization, as well as the MP2/aug-cc-pVDZ interaction energy (kcal/mol) are provided for each structure in parentheses. Hydrogen bonding donor–acceptor (D–A) distance (Å), acceptor–hydrogen (A–H) distance (Å), and D–H–A (deg) are provided for each hydrogen bond (dotted lines) in Supplemental Tables S4–S8. For structures #47 and #56, only the positions of hydrogen atoms were optimized, whereas full optimization was carried out for all other structures. Structures #43, #44, #49, and #57 contain positively charged ligand moieties.

PYI pair (Supplemental Table S10). Regardless, interaction energies (–31 kcal/mol and –48 kcal/mol) of the fully optimized pairs reveal that B:Ph pairs are stronger than all other categories of base:ligand pairs (Table 6; Fig. 9).

B:S pairs

In the context of B:S interactions, in spite of the presence of sugar moiety in the chemical structure of all flavin-related ligands, only RS3 shows significant (i.e., involving at least two hydrogen bonds) interaction, where the sugar group of RS3 interacts with W edge or S edge of G. While the optimized geometry of the W-edge pair is similar to the crystal geometry (rmsd of 0.7 Å) and is stabilized by a substantial (–20 kcal/mol) binding energy, the optimized S-edge pair shows relatively weaker (–13 kcal/mol) binding. Further, interaction energies of B:S pairs are similar to the S-edge interactions of B:B pairs (Fig. 9; Supplemental Table S10).

B:O pairs

Among the two B:O pairs, the A:PRF pair involves a weak (–3 kcal/mol) interaction of the side chain of PRF with the S-edge of adenine, which is mediated by two C–H···N/O H-bonds. Consequently, this structure shows a high rmsd (2.1 Å) on optimization. However, the second (U:AMZ) pair involves the strong interaction between imadazole group of AMZ ligand and W edge of U, and is mediated by three (two N–H···N/O and one O–H···O) hydrogen bonds. Thus, this pair exhibits small rmsd (0.7 Å) and high (–21 kcal/mol) interaction energy on optimization (Fig. 9; Supplemental Table S10).

Biological relevance of base:ligand pairs: case studies and insights

Alternate conformation of thiamine pyrophosphate (TPP) aptamer induced by 2QB ligand

The chemical structure of the TPP ligand is composed of three distinct chemical moieties: pyrimidine, thiazole, and pyrophosphate. The pyrimidine and pyrophosphate moieties are recognized by the pyrimidine domain and the pyrophosphate sensor domain of the aptamer, respectively (Fig. 10A; Thore et al. 2008).

However, since thiazole moiety is not specifically recognized, ligands similar to TPP, but lacking the thiazole moiety (e.g., PYI) may also bind with the aptamer, and consequently trigger riboswitch action (Thore et al. 2008). Among 11 crystal structures of TPP aptamer analyzed in the present study, only six (2QB, BFT, D2X, PYI, TPP, and TPS, Supplemental Fig. S5) interact with the G in the pyrimidine sensor domain to form B:B pairs mediated by two H-bonds with the following order of strength: G:2QB (–16 kcal/mol) > G:D2X (–14 kcal/mol) > G:BFT (–13 kcal/mol) > G:TPP (–11 kcal/mol) = G:TPS (–11 kcal/mol). However, in spite of strongest B:B (i.e., G:2QB) pairing, 2QB induces a different conformation of the aptamer, compared with binding with other ligands (Warner et al. 2014). Analysis of B:Ph interactions between the ligands and the pyrophosphate domain reveal that due to the absence of the pyrophosphate group in its chemical structure, 2QB misses out on a very significant B:Ph interaction, whereas other ligands form strong B:Ph contacts. For example, D2X forms B:B contacts with G28 and

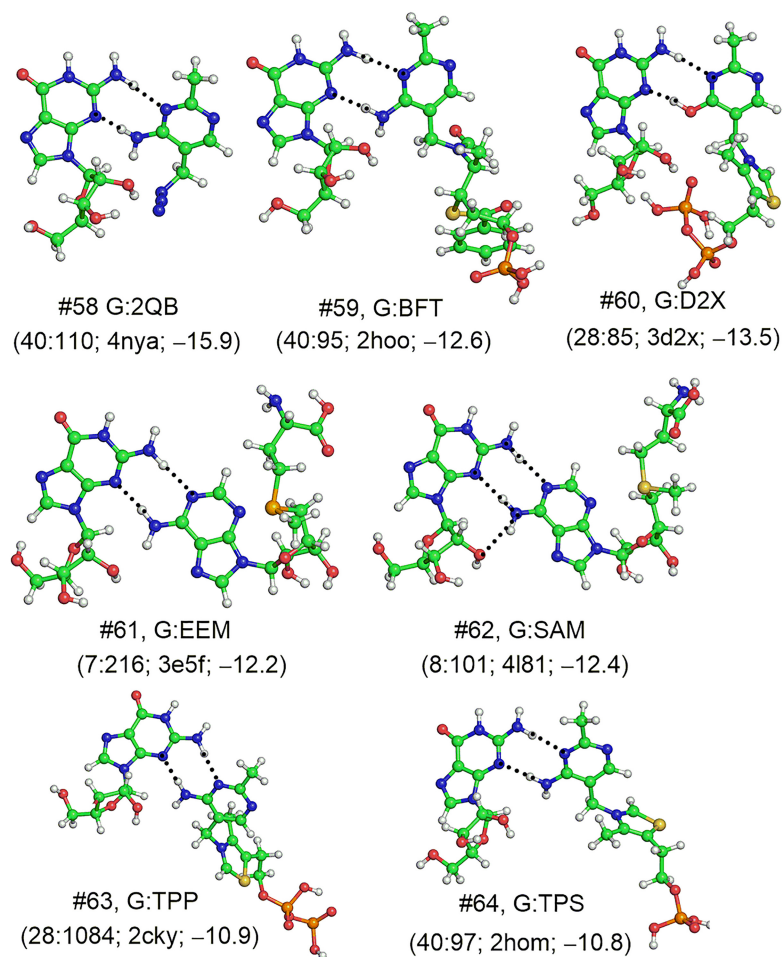


FIGURE 7. B3LYP/6-31G(d,p) optimized geometries of base:base (B:B) type of base:ligand interactions that resemble the *trans* Sugar–Watson Crick (cSW) RNA base pairs. Structure number (#) and base:ligand interaction identity are provided against each structure for identification from tables. Base-pairing interaction code and PDB code of the crystal structure used for optimization, as well as the MP2/aug-cc-pVDZ interaction energy (kcal/mol) are provided for each structure in parentheses. Hydrogen bonding donor–acceptor (D–A) distance (Å), acceptor–hydrogen (A–H) distance (Å), and D–H–A (deg) are provided for each hydrogen bond (dotted lines) in Supplemental Table S9. For structure #60, only the positions of hydrogen atoms were optimized, whereas full optimization was carried out for all other structures. Structures #60 to #64 contain positively charged ligand moieties.

G11 of the pyrimidine sensor element and additional B:Ph contacts with G66 of the pyrophosphate sensor element (PDB code: 3d2x). Thus, the absence of crucial B:Ph interactions explains the overall weaker binding of the 2QB. Overall, this example illustrates the importance of different interacting moieties in the chemical structure of the ligand in the optimal base:ligand binding.

Importance of ligand chemical structure and allosteric binding pockets in ligand binding with the aptamer domain of THF riboswitch

The aptamer domain of THF riboswitch binds folate-related ligands that are composed of three structural elements:

pterin ring, p-aminobenzoate (pABA) group, and glutamate group (Fig. 10B). Although the pterin ring plays the most important role in molecular recognition by forming B:B interactions with the aptamer, the pABA region also interacts with the aptamer bases either through hydrogen bonding or stacking interactions. However, apart from facilitating proper ligand orientation, the glutamate group does not directly participate in noncovalent interactions with the aptamer domain.

Within the studied crystal structures, the aptamer domain of THF riboswitch is bound to six different ligands (6AP, 7DG, FFO, H4B, LYA, and THF, Supplemental Fig. S6). These ligands either contain all three structural moieties of THF (FFO, LYA or THF) or contain tailored purine bases similar to the pterin ring of THF (6AP, 7DG or H4B). Our QM calculations reveal that the observed order of stability of B:B pairs involving the pterin ring of the ligand is 7DG (−43 kcal/mol) > FFO (−34 kcal/mol) > THF (−31 kcal/mol) = H4B (−31 kcal/mol). However, although the B:B pair involving 7DG possesses the highest interaction energy, the aptamer possesses overall low affinity toward this pterin analog (Trausch and Batey 2014), mainly due to the absence of pABA moiety, which stabilizes the ligand through stacking interactions with the aptamer domain. This further reiterates the synergy in interactions involving different chemical moieties and the aptamer in determining the overall ligand:aptamer affinities.

The aptamer domain of THF riboswitch can simultaneously bind two ligands through two different binding pockets—the pseudoknot (PK) site and the three-way junction (3WJ) site (PDB code: 4lvx). The aptamer bases U7 and U42 of the PK-site interact with the ligand through at least two hydrogen bonds. Similar interactions were observed between ligand present in 3WJ-site and aptamer bases U25 and C53. However, our calculations reveal that the sum of binding energies of nucleobase:ligand pairs at the PK site (−37 kcal/mol) is greater than the 3WJ site (−31 kcal/mol). Thus, the inability of ligand at 3WJ site to reach the interaction energy threshold required for triggering gene expression explains the experimentally observed generalization that only the PK site ligand triggers the

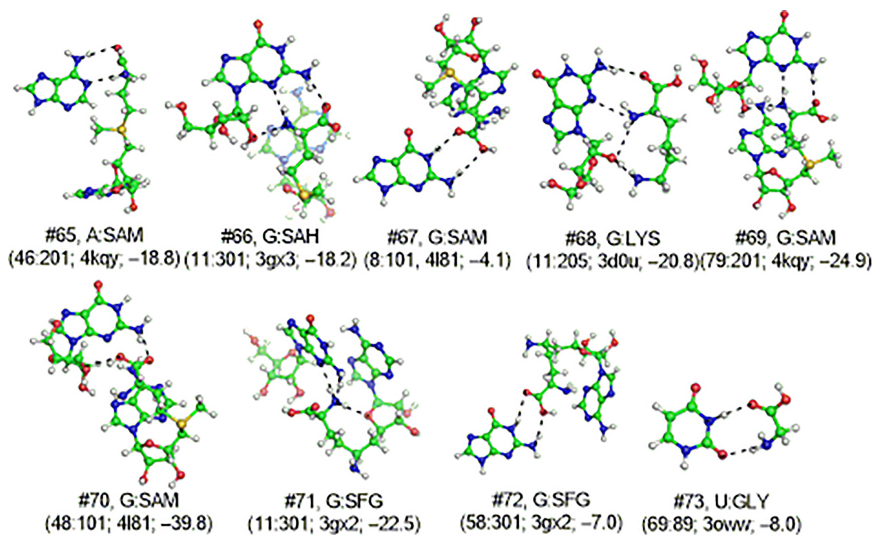


FIGURE 8. B3LYP/6-31G(d,p) optimized geometries of base:amino acid (B:A) type of base:ligand pairs. Structure number (#) and base:ligand identity are provided against each structure for identification from tables. Base-pairing interaction code and PDB code of the crystal structure used for optimization, as well as the MP2/aug-cc-pVDZ interaction energy (kcal/mol) are provided for each structure in parentheses. Hydrogen bonding donor-acceptor (D-A) distance (Å), acceptor-hydrogen (A-H) distance (Å), and D-H-A (deg) are provided for each hydrogen bond (dotted lines) in Supplemental Table S10. For structures #70 and #71, only the positions of hydrogen atoms were optimized, whereas full optimization was carried out for all other structures. Structures #65 and #67 to #70 contain positively charged ligand moieties.

gene expression in THF riboswitches (Tausch and Batey 2014).

DISCUSSION

In the present work, we combined the techniques of structural bioinformatics and quantum chemical calculations to analyze the nature of hydrogen-bonded contacts between ligands and RNA in aptamer:ligand complexes. Altogether, 80 nonredundant base:ligand pairs were identified that contain at least two hydrogen bonds. The bound ligands differ in size and complexity, where the specific functional groups act as key instruments of molecular recognition. In synchrony with a previous study (McCown et al. 2017), we found that coenzymes constitute the highest proportion of the bound ligands, which is followed by nucleobases. Further, the aptamer nucleobases most commonly interact with the ligand through the W edge, and the uracil is the most commonly interacting aptamer nucleobase.

On the basis of the chemical structure of the bound ligand, we classified the base:ligand pairs into five basic types: B:B, B:A, B:Ph, B:S, and B:O. Among these, B:B interactions were further classified into two types: those involving nucleobases as ligands (e.g., 2U:ADE, PDB code: 1y26), and those involving nucleobase moiety as a part of the chemical structure of the ligands (e.g., G:TPP, PDB code: 2cky). Such contacts span nine (cWW, tWW, cWH,

tWH, cWS, cSW, tWS, tSW, and cHS) distinct base:ligand edge interactions and largely possess structural characteristics similar to base pairs found in available RNA crystal structures (Šponer et al. 2005c; Sharma et al. 2008). Our crystal structure analysis further revealed two types of B:A contacts involving either the free amino acid as a ligand (e.g., G:LYS pair [PDB code: 3d0u]) or amino acid moiety as a part of the chemical structure of the ligand (e.g., MET, HEY and ORN are a part of SAM [PDB code: 4l81], SAH [PDB code: 3gx3], and SFG [PDB code: 3gx2], respectively). The B:Ph interactions, on the other hand, involve a total of three contacts and possess interaction patterns similar to one of the base-backbone phosphate interaction types (4BPh) observed previously in RNA structures (Zirbel et al. 2009). However, B:S and B:O contacts involve only a few specific examples.

QM calculations reveal that more than one-third of the pairs belonging to the highly populated interaction type, the W-edge B:B pairing, possess three strong H-bonds, and significant (-14 to -32 kcal/mol) strengths, which are comparable to the strength of canonical (cWW) A:U and G:C RNA base pairs (-15 and -29 kcal/mol, respectively [Bhattacharya et al. 2015]). However, the strength of S-edge B:B base pairs (-11 to -18 kcal/mol) is relatively weaker than the W-edge pairs. In contrast, the B:A interactions exhibit three distinct geometries involving (i) both -NH₂ (donor) and -CO₂H (donor or acceptor) of the amino acid with the aptamer nucleobase and a relatively wider (-8 to -40 kcal/mol) range of stabilization energies, (ii) -NH₂ as both donor and acceptor, without the involvement of -CO₂H, and a significant (-23 kcal/mol) interaction strength, and (iii) Both oxygen atoms of -CO₂H as acceptors, without the involvement of -NH₂ and a relatively weaker (by up to -21 kcal/mol) interaction strength. On the other hand, B:Ph interactions, involve at least three hydrogen bonds. Further, significant binding strengths (-31 to -48 kcal/mol) indicate that B:Ph interactions are the strongest among all base:ligand pairs, and are even stronger than canonical base-pairing interactions in RNA. It is noteworthy that B:Ph interactions are stronger than the usual base-phosphate interactions (-1 to -10 kcal/mol) in functional RNA (Zirbel et al. 2009), mainly because of positive charge on the ligand, which possibly results in enhanced electrostatic interactions. Although B:S interactions have moderate (-13 kcal/mol and -20 kcal/mol) binding

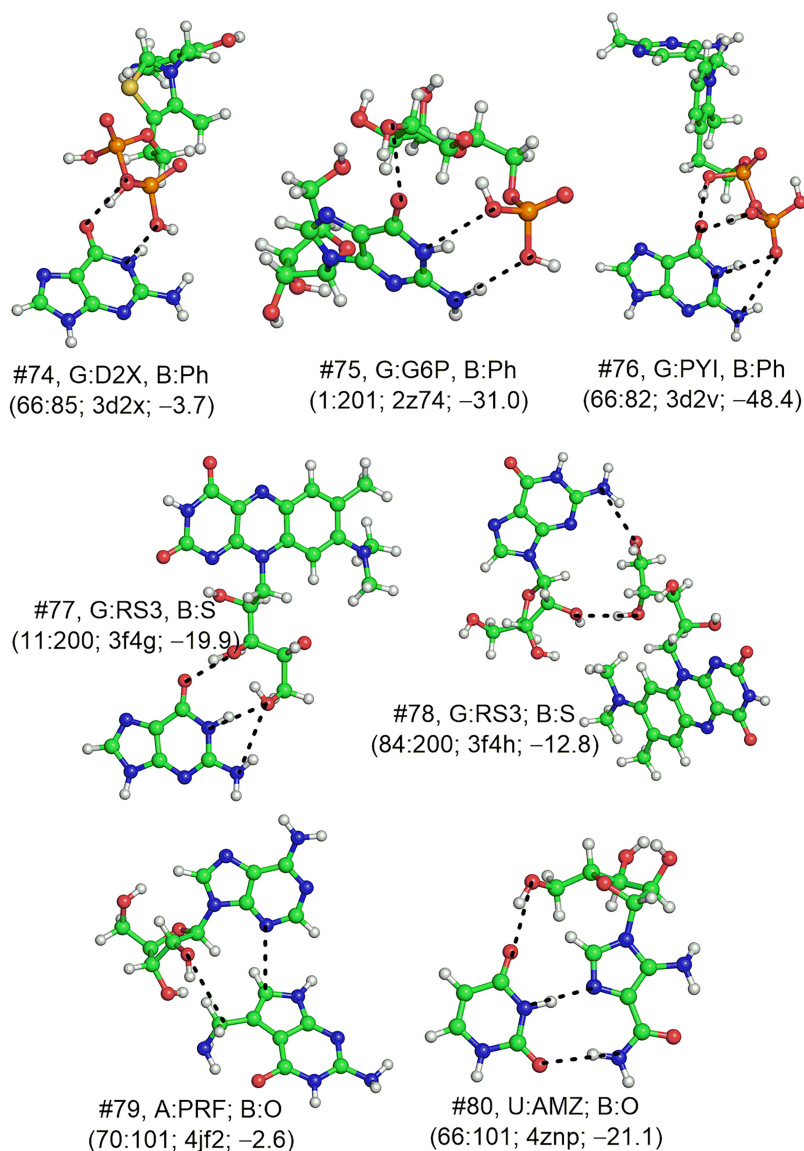


FIGURE 9. B3LYP/6-31G(d,p) optimized geometries of base:phosphate (B:Ph), base:sugar (B:S), and base:other (B:O) type of base:ligand interactions. Structure number (#), base:ligand interaction identity and interaction type (B:Ph, B:S or B:O) are provided against each structure for identification from tables. Base-pairing interaction code and PDB code of the crystal structure used for optimization, as well as the MP2/aug-cc-pVDZ interaction energy (kcal/mol) are provided for each structure in parentheses. Hydrogen bonding donor-acceptor (D-A) distance (Å), acceptor-hydrogen (A-H) distance (Å), and D-H-A (deg) is provided for each hydrogen bond (dotted lines) in Supplemental Table S10. For structures #74, #78, and #79, only the positions of hydrogen atoms were optimized, whereas full optimization was carried out for all other structures. Structures #74 and #76 contain positively charged ligand moieties.

strength, the strength of B:O pairs (-3 kcal/mol and -21 kcal/mol), however, depends on the interacting (W or S) edge of the aptamer base and the type of hydrogen bonds.

Conclusions

In summary, our calculations yield important insights into the abundance and strength of base:ligand interactions.

Due to close similarities with strong and structurally important canonical and noncanonical base pairs observed in RNA macromolecular structures, our results reiterate that base:ligand interactions provide significant stability to the aptamer:ligand complexes and thereby play an important role in ligand recognition. Indeed, examples of aptamer base:ligand pairs in the higher-order motifs in the ligand-binding pocket are crucial for the transfer of chemical information from binding pocket to expression platform. Thus, the present analysis of the base:ligand interactions may help trigger advances in RNA based drug design and development.

Although our QM calculations suggest that base:ligand pairs are intrinsically stable interaction motifs, a single ligand is generally recognized by (i.e., hydrogen-bonded with) multiple RNA bases in the ligand-binding pockets of the aptamer domains of riboswitches (Supplemental Figs. S7, S8). Thus, the ligand is geometrically arranged in such a way that a higher-order structure involving nucleobases and ligand (e.g., in purine riboswitches) is formed (Supplemental Fig. S8; Sharma et al. 2009b). In such contexts, our analysis of base:ligand pairs provide a limited structural description of the overall ligand-binding pocket. Thus, future studies must explore these aspects in larger assemblies involving simultaneous interaction of the single ligand with multiple bases of the aptamer binding pocket.

MATERIALS AND METHODS

A data set of 67 crystal structures (Table 1) spanning the first four of the five recently classified riboswitch categories (Sherwood and Henkin 2016) (namely coenzymes and related compounds, nucleotides and their derivatives, amino acids and sugars, Supplemental Figs. S1-S4), containing base:ligand pairs with at least two hydrogen bonds, was selected from protein data bank (PDB), on the basis of visual analysis, for preliminary analysis. This data set excludes crystal structures in which ions act as ligands, mainly because the associated aptamer:ligand noncovalent contacts are physicochemically different from the (hydrogen bonding) interactions considered in the present work. The data set was also curated to include only

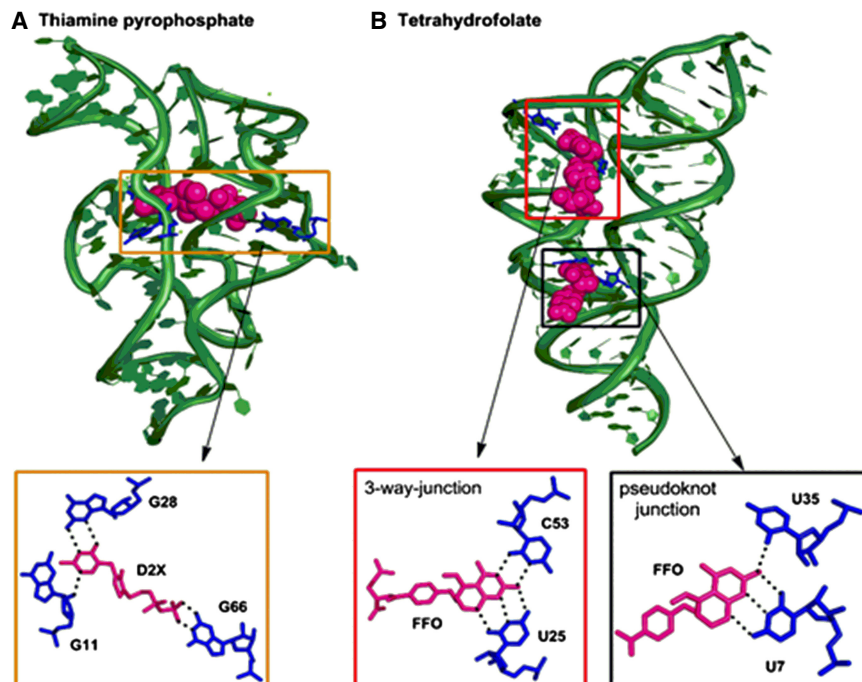


FIGURE 10. (A) D2X ligand bound to the aptamer domain of thiamine pyrophosphate riboswitch (PDB code: 3d2x). (B) FFO ligand bound to the aptamer domain of tetrahydrofolate riboswitch (PDB code: 4lv).

structures having at least 3.2 Å resolution, and by retaining only wild-type aptamer crystal structures. The final comprehensive data set contained structures of 38 aptamers bound to natural (biological) ligands and 29 aptamers bound to ligand derivatives.

A total of 80 base:ligand pairs connected through two hydrogen bonds were extracted from these crystal structures, statistically analyzed and used as initial models for QM analysis. For QM geometry optimizations, the missing hydrogen atoms were added to the heavy-atom coordinates of the ligands, extracted from the respective crystal structures, to complete the necessary covalency requirements in the initial models. Subsequently, the interacting aptamer nucleotides were truncated at their glycosidic bonds, and the C1' atoms were replaced with hydrogen atoms, wherever the ribose sugars of these nucleotides do not interact with bound ligands (Sharma et al. 2010b). In cases where the ribose sugar bound to the aptamer nucleobase interacts with ligands, the respective sugars were retained after replacing the 5'-phosphate group with a hydroxyl group, in order to retain a tractable, but relevant model size. Further, in the case of ligands with amino acid functionality, the $-NH_2$ and CO_2H were taken as neutral and for ligands with phosphate group interactions (D2X, PYI, and G6P), the phosphate group was considered as a charged moiety. Geometry optimization of the base:ligand pairs was carried out at the B3LYP/6-31G(d,p) level using Gaussian 09 (Frisch et al. 2009), which was selected in synchrony with previous studies on RNA base pairs (Šponer et al. 2004, 2005a,b,c; Sharma et al. 2008, 2010a).

To understand the difference in base:ligand geometry, in its crystal occurrence and in the corresponding fully optimized form, the root mean square deviation (rmsd) was calculated using

VMD v1.9 software (Humphrey et al. 1996). The rmsd values provide quantitative information on the conformational changes undergone by the base:ligand pair on optimization in isolated form (i.e., free from the constraints of the aptamer macromolecular environment) and are expected to be relatively higher in the case of flexible ligands. Further, the strength of each base:ligand pair was measured in terms of interaction energy, which is a measure of the stabilization acquired by the interacting system through hydrogen bonding.

Basis set superposition error (BSSE) corrected interaction energies were calculated using the counterpoise method (Boys and Bernardi 1970) at the RIMP2/aug-cc-pVDZ or MP2/aug-cc-pVDZ level (Weigend and Häser 1997; Ahlrichs et al. 1998) using Turbomole v6.2 (<http://www.turbomole.com>) or Gaussian 09 (Frisch et al. 2009), which was selected in analogy with previous studies on RNA base pairs (Šponer et al. 2005a,b,c; Mládek et al. 2009; Sharma et al. 2010b; Seelam et al. 2017). Further, in analogy with previous studies on pairing interactions in RNA involving flexible sugar moieties (Šponer et al. 2005a,b,c; Mládek et al.

2009; Sharma et al. 2010b), the monomer deformation energies were not included in final interaction energies. This is mainly due to the possibility of substantial structural alteration on the optimization of the flexible ligands in their isolated form. As a result, the inclusion of monomer deformation energies into the interaction energies can bias the relative energy trends. Although geometries of 67 out of total 80 base:ligand pairs could be fully optimized, the remaining 13 pairs either showed highly deviated optimized structures compared with the initial structures or could not be optimized (Table 2). In these cases, as in previous studies on RNA base pairs (Sharma et al. 2008), constrained optimizations, restricted to the optimization of the positions of the added hydrogen atoms, were performed by freezing all the heavy-atom coordinates. Interaction energies were calculated on these constrained-optimized systems at the RIMP2/aug-cc-pVDZ or MP2/aug-cc-pVDZ level.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was funded by the University Grants Commission (UGC FRP Scheme F.4-5 [176-FRP/2015 {BSR}]) and the Department of Science and Technology, Ministry of Science and Technology award no. IFA14-CH162.

Received April 12, 2019; accepted July 12, 2019.

REFERENCES

- Aboul-ela F, Huang W, Abd Elrahman M, Boyapati V, Li P. 2015. Linking aptamer-ligand binding and expression platform folding in riboswitches: prospects for mechanistic modeling and design. *Wiley Interdiscip Rev RNA* **6**: 631–650. doi:10.1002/wrna.1300
- Abu Almakarem AS A, Petrov AI, Stombaugh J, Zirbel CL, Leontis NB. 2012. Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res* **40**: 1407–1423. doi:10.1093/nar/gkr810
- Ahlich R, Elliott SD, Huniar U. 1998. Quantum chemistry: large molecules—small computers. *Ber Bunsenges Phys Chem* **102**: 795–804. doi:10.1002/bbpc.19981020602
- Batey RT. 2012. Structure and mechanism of purine-binding riboswitches. *Q Rev Biophys* **45**: 345–381. doi:10.1017/S0033583512000078
- Bhattacharya S, Mittal S, Panigrahi S, Sharma P, Preethi SP, Paul R, Halder S, Halder A, Bhattacharyya D, Mitra A. 2015. RNABP COGEST: a resource for investigating functional RNAs. *Database* **2015**: bav011. doi:10.1093/database/bav011
- Boys SF, Bernardi F. 1970. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol Phys* **19**: 553–566. doi:10.1080/00268977000101561
- Breaker RR. 2008. Complex riboswitches. *Science* **319**: 1795–1797. doi:10.1126/science.1152621
- Chawla M, Abdel-Azeim S, Oliva R, Cavallo L. 2014. Higher order structural effects stabilizing the reverse Watson–Crick guanine–cytosine base pair in functional RNAs. *Nucleic Acids Res* **42**: 714–726. doi:10.1093/nar/gkt800
- Chawla M, Oliva R, Bujnicki JM, Cavallo L. 2015. An atlas of RNA base pairs involving modified nucleobases with optimal geometries and accurate energies. *Nucleic Acids Res* **43**: 6714–6729. doi:10.1093/nar/gkv606
- Cheng AC, Chen WW, Fuhrmann CN, Frankel AD. 2003. Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J Mol Biol* **327**: 781–796. doi:10.1016/S0022-2836(03)00091-3
- Das J, Mukherjee S, Mitra A, Bhattacharyya D. 2006. Non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J Biomol Struct Dyn* **24**: 149–161. doi:10.1080/07391102.2006.10507108
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, et al. 2009. *Gaussian 09*. Gaussian Inc., Wallingford, CT.
- Greenleaf WJ, Frieda KL, Foster DA, Woodside MT, Block SM. 2008. Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **319**: 630–633. doi:10.1126/science.1151298
- Huang L, Wang J, Wilson TJ, Lilley DM. 2017. Structure of the guanine III riboswitch. *Cell Chem Biol* **24**: 1407–1415.e1402. doi:10.1016/j.chembiol.2017.08.021
- Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph* **14**: 33–38. doi:10.1016/0263-7855(96)00018-5
- Kondo J, Westhof E. 2010. Base pairs and pseudo pairs observed in RNA-ligand complexes. *J Mol Recognit* **23**: 241–252. doi:10.1002/jmr.978
- Kondo J, Westhof E. 2011. Classification of pseudo pairs between nucleotide bases and amino acids by analysis of nucleotide–protein complexes. *Nucleic Acids Res* **39**: 8628–8637. doi:10.1093/nar/gkr452
- Krepl M, Vögele J, Kruse H, Duchardt-Ferner E, Wöhnert J, Šponer J. 2018. An intricate balance of hydrogen bonding, ion atmosphere and dynamics facilitates a seamless uracil to cytosine substitution in the U-turn of the neomycin-sensing riboswitch. *Nucleic Acids Res* **46**: 6528–6543. doi:10.1093/nar/gky490
- Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**: 499–512. doi:10.1017/S1355838201002515
- Mandal M, Breaker RR. 2004. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**: 451–463. doi:10.1038/nrm1403
- McCown PJ, Corbino KA, Stav S, Sherlock ME, Breaker RR. 2017. Riboswitch diversity and distribution. *RNA* **23**: 995–1011. doi:10.1261/rna.061234.117
- Mládek A, Sharma P, Mitra A, Bhattacharyya D, Šponer J, Šponer JE. 2009. Trans Hoogsteen/sugar edge base pairing in RNA. Structures, energies, and stabilities from quantum chemical calculations. *J Phys Chem B* **113**: 1743–1755. doi:10.1021/jp808357m
- Montange RK, Mondragón E, van Tyne D, Garst AD, Ceres P, Batey RT. 2010. Discrimination between closely related cellular metabolites by the SAM-I riboswitch. *J Mol Biol* **396**: 761–772. doi:10.1016/j.jmb.2009.12.007
- Oliva R, Cavallo L, Tramontano A. 2006. Accurate energies of hydrogen bonded nucleic acid base pairs and triplets in tRNA tertiary interactions. *Nucleic Acids Res* **34**: 865–879. doi:10.1093/nar/gkj491
- Perdrizet GA, Artsimovitch I, Furman R, Sosnick TR, Pan T. 2012. Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc Natl Acad Sci* **109**: 3323–3328. doi:10.1073/pnas.1113086109
- Ren A, Rajashankar KR, Patel DJ. 2012. Fluoride ion encapsulation by Mg²⁺ ions and phosphates in a fluoride riboswitch. *Nature* **486**: 85–89. doi:10.1038/nature11152
- Roth A, Breaker RR. 2009. The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* **78**: 305–334. doi:10.1146/annurev.biochem.78.070507.135656
- Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. 2008. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol* **56**: 215–252. doi:10.1007/s00285-007-0110-x
- Savinov A, Perez CF, Block SM. 2014. Single-molecule studies of riboswitch folding. *Biochim Biophys Acta Gene Regul Mech* **1839**: 1030–1045. doi:10.1016/j.bbaggm.2014.04.005
- Seelam PP, Sharma P, Mitra A. 2017. Structural landscape of base pairs containing post-transcriptional modifications in RNA. *RNA* **23**: 847–859. doi:10.1261/rna.060749.117
- Sharma P, Mitra A, Sharma S, Singh H, Bhattacharyya D. 2008. Quantum chemical studies of structures and binding in noncanonical RNA base pairs: the trans Watson–Crick/Watson–Crick family. *J Biomol Struct Dyn* **25**: 709–732. doi:10.1080/07391102.2008.10507216
- Sharma M, Bulusu G, Mitra A. 2009a. MD simulations of ligand-bound and ligand-free aptamer: molecular level insights into the binding and switching mechanism of the add A-riboswitch. *RNA* **15**: 1673–1692. doi:10.1261/rna.1675809
- Sharma P, Sharma S, Chawla M, Mitra A. 2009b. Modeling the noncovalent interactions at the metabolite binding site in purine riboswitches. *J Mol Model* **15**: 633–649. doi:10.1007/s00894-008-0384-y
- Sharma P, Chawla M, Sharma S, Mitra A. 2010a. On the role of Hoogsteen:Hoogsteen interactions in RNA: ab initio investigations of structures and energies. *RNA* **16**: 942–957. doi:10.1261/rna.1919010
- Sharma P, Šponer JE, Šponer J, Sharma S, Bhattacharyya D, Mitra A. 2010b. On the role of cis Hoogsteen:sugar edge family of base pairs in platforms and triplets—quantum chemical insights into RNA structural biology. *J Phys Chem B* **114**: 10234–10234. doi:10.1021/jp106172b

- Sherwood AV, Henkin TM. 2016. Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. *Annu Rev Microbiol* **70**: 361–374. doi:10.1146/annurev-micro-091014-104306
- Šponer J, Jurečka P, Hobza P. 2004. Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J Am Chem Soc* **126**: 10142–10151. doi:10.1021/ja048436s
- Šponer JE, Leszczynski J, Sychrovský V, Šponer J. 2005a. Sugar edge/sugar edge base pairs in RNA: stabilities and structures from quantum chemical calculations. *J Phys Chem B* **109**: 18680–18689. doi:10.1021/jp053379q
- Šponer JE, Špačková NA, Kulhánek P, Leszczynski J, Šponer J. 2005b. Non-Watson–Crick base pairing in RNA. Quantum chemical analysis of the cis Watson–Crick/sugar edge base pair family. *J Phys Chem A* **109**: 2292–2301. doi:10.1021/jp050132k
- Šponer JE, Špačková NA, Leszczynski J, Šponer J. 2005c. Principles of RNA base pairing: structures and energies of the trans Watson–Crick/sugar edge base pairs. *J Phys Chem B* **109**: 11399–11410. doi:10.1021/jp051126r
- Šponer J, Šponer JE, Mládel A, Jurečka P, Banáš P, Otyepka M. 2013. Nature and magnitude of aromatic base stacking in DNA and RNA: quantum chemistry, molecular mechanics, and experiment. *Biopolymers* **99**: 978–988.
- Thore S, Frick C, Ban N. 2008. Structural basis of thiamine pyrophosphate analogues binding to the eukaryotic riboswitch. *J Am Chem Soc* **130**: 8116–8117. doi:10.1021/ja801708e
- Trausch JJ, Batey RT. 2014. A disconnect between high-affinity binding and efficient regulation by antifolates and purines in the tetrahydrofolate riboswitch. *Chem Biol* **21**: 205–216. doi:10.1016/j.chembiol.2013.11.012
- Wang W, Jiang C, Zhang J, Ye W, Luo R, Chen H-F. 2016. Dynamics correlation network for allosteric switching of PreQ1 riboswitch. *Sci Rep* **6**. doi:10.1038/srep31005
- Warner KD, Homan P, Weeks KM, Smith AG, Abell C, Ferré-D'Amaré AR. 2014. Validating fragment-based drug discovery for biological RNAs: lead fragments bind and remodel the TPP riboswitch specifically. *Chem Biol* **21**: 591–595. doi:10.1016/j.chembiol.2014.03.007
- Weigend F, Häser M. 1997. RI-MP2: first derivatives and global consistency. *Theor Chem Acc* **97**: 331–340. doi:10.1007/s002140050269
- Winkler WC, Nahvi A, Sudarsan N, Barrick JE, Breaker RR. 2003. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Mol Biol* **10**: 701. doi:10.1038/nsb967
- Zirbel CL, Šponer JE, Šponer J, Stombaugh J, Leontis NB. 2009. Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res* **37**: 4898–4918. doi:10.1093/nar/gkp468