J Mol Model (2012) 18:3805–3820
DOI 10.1007/s00894-012-1385-4

ORIGINAL PAPER

Role of wobble base pair geometry for codon degeneracy: purine-type bases at the anticodon wobble position

Gunajyoti Das · R. H. Duncan Lyngdoh

Received: 8 September 2011 / Accepted: 15 February 2012 / Published online: 8 March 2012

Abstract Codon degeneracy is a key feature of the genetic code, explained by Crick (J Mol Biol 19:548-555, 1966) in terms of imprecision of base pairing at the codon third position (the wobble position) of the codon-anticodon duplex. The Crick wobble rules define, but do not explain, which base pairs are allowed/disallowed at the wobble position of this duplex. This work examines whether the H-bonded configurations of solitary RNA base pairs can in themselves help decide which base pairs are allowed at the wobble position during codon-anticodon pairing. Taking the purine-type bases guanine, hypoxanthine, queuine and adenine as anticodon wobble bases, H-bonded pairing energies and optimized configurations of numerous RNA base pairs are calculated in gas and modeled aqueous phase at the B3LYP/6-31 G(d,p) level. Calculated descriptors of alignment of these solitary base pairs are able to screen between allowed and disallowed base pairs for all cases studied here, except two cases which invoke base-sugar interactions in the codon wobble nucleoside. The exclusion of adenine from the anticodon wobble position cannot be explained on the basis of pairing facility or base pair geometry. These DFT results thus account for the specificity and degeneracy of the genetic code for all cases involving guanine, hypoxanthine and queuine as anticodon wobble bases.

Keywords Anticodon wobble base · Codon degeneracy · Crick wobble hypothesis · Density functional theory · H-bonded pairing configuration · Wobble base pairs

Introduction

Protein synthesis involves the translation of the genetic information encoded in DNA into the amino acid sequence of proteins. This is dictated by the genetic code [2, 3], which refers to the unambiguous correspondence between the base sequence of the triplet DNA/mRNA codon and the amino acid incorporated at that point into the synthesized protein. Hydrogen bonding between the mRNA codon and the corresponding triplet anticodon of tRNA leads to the formation of cognate codon-anticodon duplexes. Only Watson-Crick type base pairs occur between the first base of the codon and the third base of the anticodon, and between the second base of the codon and the second base of the anticodon. Some pairs are allowed and others disallowed between the codon third base and the anticodon first base (called the *wobble position*). Note that the term "wobble pair" here refers to *any* hydrogen-bonded base pair considered at the wobble position, whether allowed or not.

Codon degeneracy means that most of the 20 amino acids of proteins are encoded for by more than one codon. Apart from Met and Trp, codon degeneracy may be two-, three-, four-, or six-fold. The patterns of codon degeneracy admit no simple generalizations and do not correspond to consistent arrangement of cognate codons into boxes and/or half-boxes. In the famous *wobble hypothesis* [1], Crick explained codon degeneracy by suggesting that more than one codon may pair with a single anticodon present in the cognate transfer RNA. Each anticodon may pair with one, two or three cognate codons arising from the same box.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-012-1385-4) contains supplementary material, which is available to authorized users.

G. Das · R. H. D. Lyngdoh (✉)
Department of Chemistry, North-Eastern Hill University, Shillong 793022, India
e-mail: harkuduncanl@yahoo.co.in

The Crick *wobble rules* [4–6] define, but do not explain, what pairs are allowed at the wobble position. Computational studies may predict whether a given RNA base pair occurs or not at the wobble position by calculating factors that screen between disallowed and allowed base pairs. Such factors may be energetic (base pair stability) or spatial (three-dimensional geometry). This work attempts to examine whether the configuration of a given wobble base pair, considered in complete *isolation*, may afford criteria which allow or disallow it at the wobble position during codon-anticodon pairing. This hypothesis starts by excluding nearest neighbor interactions, stacking interactions, and any factor apart from the isolated wobble base pair itself. This approach may at first seem unrealistic. However, we may cite the instance of the canonical DNA base pairs whose solitary configurations gauged computationally even in gas phase are to a large degree conserved in double-helical DNA, being of the greatest relevance for understanding the structure and function of DNA.

Pairing at the wobble position is less precise than at the other two positions of the codon-anticodon duplex. The *third* base of the codon (the *codon wobble base* or CWB) is always a major RNA base - adenine (Ade), guanine (Gua), uracil (Ura) or cytosine (Cyt). Each CWB pairs with the *first* base of the anticodon (the *anticodon wobble base* or AWB) which may be Gua, Ura, Cyt and various minor RNA bases. Ade usually never appears as an AWB (see later). The flexibility of wobble base pairing means a given AWB may pair with more than one CWB. This flexibility has its bounds, however, where some base pairs are allowed at the wobble position, and others are not.

Anticodon base sequences for transfer RNA's are derived from the full base sequences of known transfer RNA's for amino acids compiled by Sodd [7] and updated by Juhling et al. [8]. This data generates a list of all known anticodon wobble bases along with the bases they pair with at the wobble position. Note that any base pairing combination at the wobble position may be described *invariably* as either "allowed" or "forbidden", regardless of which amino acid is the cognate. The term "allowed" for a wobble pair means that the pair occurs at the wobble position during at least one case of cognate codon-anticodon pairing.

This complete list of allowed and disallowed wobble base pairs thus fully dictates which codon-anticodon pairs are allowed to occur as per the Crick wobble rules. The specificity and degeneracy of the genetic code may thus be entirely described in terms of allowed and forbidden wobble pairs. This study seeks to explain the wobble rules by modeling the pairing situation at the wobble position alone in clear molecular terms using a DFT method. Criteria are sought which may consistently label any candidate wobble base pair as allowed or disallowed *solely* on the basis of the *configuration* of the solitary base pair. The degree of deviation of a given base pair from the Watson-Crick type of alignment is assessed by comparing the values of well-defined descriptors of pairing configuration (see later) for the given pair with those for the canonical Watson-Crick base pairs.

Wobble pair stability and configuration

Stability of a wobble pair is estimated here as the energy change occurring when a given hydrogen-bonded base pair is formed from its constituent bases, which quantity is termed here as the *H-bonded pairing energy*. Base pair stability is a necessary but insufficient condition for a given base pair to occur at the wobble position. Since H-bonded pairing energies for disallowed RNA base pair mismatches do not differ much from those of Watson-Crick base pairs [9, 10], pairing facility in itself cannot decide whether a given wobble pair is allowed or not. It is not the objective of this study to accurately reproduce the actual value of the H-bonded pairing energy of any base pair. The pertinent factor here is that of a *suitable pairing configuration* for the wobble pair, ideally approaching the Watson-Crick alignment (assumed as optimal for the mini-helical situation of the codon-anticodon pair). This is linked to need for the codon to maintain the undeformed A-form of RNA during cognate codon-anticodon duplex formation.

Room is left for flexibility, however, and deviations from this alignment are allowed to a lesser or greater degree at the wobble position. Limits of configuration exist within which a given base pair may be accommodated at the wobble position. The DFT strategy used here is deemed as sufficiently reliable to describe the configurations of the H-bonded systems studied here, especially in the context of their use to furnish simple screening criteria which can differentiate between allowed and disallowed base pairs.

Computational studies on RNA base pairs

The few attempts by computational chemists to study the role of RNA base pairing for the genetic code include semi-empirical molecular orbital studies [11, 12] and a later DFT study [13] which reiterated the earlier studies using a more accurate methodology. Outside the context of codon-anticodon pairing, nucleobase pairs have been much investigated using computational chemistry, including the work of Hobza, Sponer and others [14]. H-bonded interactions between possible primordial nucleobase analogues was studied by high level theory and molecular dynamics simulations, though the results did not correlate with pKa's and melting points [15]. H-bonded and stacking interactions among nucleic acid bases have been studied computationally [16]. The pairing energies of H-bonded base pairs have been estimated to a high degree of accuracy [17]. Unusual

RNA base pairs having rare tautomers have been studied by MP2 theory [18]. H-bonded uracil dimers were studied by classical potential, quantum chemical and statistical approaches [19]. C-H…O carbon hydrogen bonds in RNA have been studied by *ab initio* theory [20]. The cytosine dimer has also been studied this way [21]. Non-canonical RNA base pairing interactions have been analyzed theoretically [22]. The structure, stability and dynamics of canonical and non-canonical base pairs were studied using quantum methods [23]. An analysis of the structure and stability of various RNA base pairs was carried out using quantum chemical theory [24].

The various RNA base pair families treated using *ab initio* and DFT methods include sugar edge/sugar edge RNA base pairs [25] and *cis* Watson-Crick/sugar edge RNA base pairs [26]. The *trans* Watson-Crick/sugar edge base pair families were studied using quantum chemical and molecular mechanics methods [27]. The *trans* Watson-Crick/ Watson-Crick RNA base pair family was also studied computationally [28]. Hoogsteen edge/sugar edge interactions in RNA base pairs were treated by topological and NBO analyses [29]. The role of *cis* Hoogsteen/sugar edge RNA base pairs in forming platforms and triplets was studied using quantum chemical methods [30]. The role of Hoogsteen-Hoogsteen interactions in RNA was also studied by *ab initio* methods [31].

The above survey shows that even simple base pair studies can shed light on the structure and function of RNA, although some phenomena like RNA folding cannot be explained simply on the basis of solitary base pair studies [32]. However, quantum chemical studies on nucleic acids in general may be justifiably regarded as a bridge to understanding their complex biology and bioinformatics [33]. The solitary RNA base pair studies of this work will prove, as seen later, to be of some relevance for explaining codon-anticodon pairing and the genetic code. We seek for suitable *physical* criteria which define factors that enable an anticodon to recognize its cognate codons, hoping to explain the Crick wobble rules by appealing to the key role of allowed RNA base pairing at the wobble position alone. We propose that the real physical factor involved here is the *geometry* or *configuration* of the H-bonded wobble base pair rather than thermodynamic facility of base pairing.

Scope of this study

This study starts from the anticodon wobble base (AWB) as the point of reference rather than the codon wobble base (CWB). This is because there are only four CWB's (Ade, Gua, Ura and Cyt), while the total number of AWB's in nature is not fully known. The total number of base pair combinations to be considered for any single AWB in order to make the picture complete is *four* and no more. For each CWB, though, an as yet unspecified number of AWB's would have to be considered to make the picture complete.

This study focuses on *purines* acting as anticodon wobble bases, including guanine (Gua) and hypoxanthine or inosine aglycoside (Ino). Queuosine aglycoside or queuine (Que) is also considered, though not strictly a purine, since its N7-atom is replaced by carbon. The topology of this 7-deazaguanine resembles the purines, and queuosine is a well-recognized wobble nucleoside in tRNA [34–36]. We study adenine as a candidate purine anticodon wobble base representing another situation (see below). Table 1 lists all known anticodons having these bases at the wobble position [7], along with their cognate amino acids and codons, giving a list of allowed and disallowed wobble pairs involving these bases.

The near total absence in nature of adenine (Ade) at the anticodon wobble position is attributed to evolutionary exclusion [37]. Any adenosine residues at the anticodon wobble position are usually converted to inosine by anticodon adenine deaminase (highly specific for this site). Adenine is found, however, at the anticodon wobble position in threonine tRNA for *Mycoplasma capricolum* [38]. This almost universal absence of Ade poses the question as to whether this might be due to an inability of the AWB Ade to pair satisfactorily with the third codon base. We study the base pairing properties of Ade (functioning as an AWB) as it H-bonds with the four major RNA bases.

Base pairs are described by the scheme of Leontis and Westhof [39, 40] which includes definition of the interacting base edges as well as description of a pair as *cis* or *trans* depending on the orientation of the two glycosidic N-C1' bonds. Only *cis* pairs are relevant here since the *trans* geometries are not expected to fit into the wobble position during codon-anticodon pairing. Base pairs are also classified with respect to the *edges* that interact with each other during H-bonding, which include the Watson-Crick (WC) edge, the Hoogsteen (H) edge and the sugar edge (SE). Base pairs with H-bonds involving carbon are considered only if no other *cis* pairing option exists.

The four AWB's Gua, Ino, Que and Ade are each paired with the four CWB's Ade, Gua, Ura and Cyt in turn, each yielding four pairing combinations in the *cis* orientation. For some combinations, more than one configuration is considered. For each combination, note is taken of the status (allowed or disallowed) of each combination in the context of codon-anticodon pairing, the interacting edges, values of the H-bonded pairing energies and of the configuration descriptors for the wobble pairs. Both gas and simulated aqueous phases are used to estimate pairing energies and configuration descriptors, since most previous studies on RNA base pairs have employed only the gas phase. For a given wobble pair **A…B**, the base **A** on the left in the figures is defined as the AWB and the base **B** on the right as the

Table 1 Allowed and disallowed wobble pairs, derived from tRNA sequences [7], involving Gua, Ino and Que as the anticodon wobble base (AWB). Wobble pairs are reckoned starting from the anticodon side

AWB	Amino acid	Anticodons	Codons	Allowed pairs	Disallowed pairs
Gua	Cys	GCA	UGU; UGC	Gua:Ura, Gua:Cyt	Gua:Ade, Gua:Gua
	Gly	GCC	GGU; GGC		
	Ile	GAU	UUU; UUC		
	Leu	GAG	CUU; CUC		
	Phe	GAA	UUU; UUC		
	Ser	GCU	AGU; AGC		
	Thr	GGU	ACU; ACC		
	Val	GAC	GUU; GUC		
Ino	Ala	IGC	GCA; GCU; GCC	Ino:Ade, Ino:Ura,	Ino:Gua
	Arg	ICG	CGA; CGU; CGC	Ino:Cyt	
	Ser	IGA	UCA; UCU; UCC		
	Val	IAC	GUA; GUU; GUC		
Que	Asp	QUC	GAU; GAC	Que:Ura, Que:Cyt	Que:Ade,
	His	QUG	CAU; CAC		Que:Gua
	Tyr	QUA	UAU; UAC		

CWB. The sugar moiety is not explicitly considered here, being modeled here by a methyl group attached to the N9-atom of purines and the N1-atom of pyrimidines. N-methylated nucleobases were used in an earlier computational study on DNA bases [41].

Computational methods

The B3LYP DFT model [42, 43] with a 6-31 G(d,p) basis set gave optimized geometries for all systems in gas and aqueous phase using a polarizable continuum model [44]. Frequency analyses showed all structures to be true minima. Zero point energy (ZPE) corrections were applied to the total energies of all the gas phase and solvated systems, using a scaling factor of 0.9613 [45]. All calculations were carried out using the Gaussian 2003 program [46].

Base pairing facility is estimated here as the energy change occurring upon formation of the H-bonded base pair from its constituent free bases. This energy change due to hydrogen bonded base pair formation is termed here as the *pairing energy*. This quantity is expressed *(a)* the gas phase pairing energy ΔE_p obtained from B3LYP/6-31 G(d,p) calculations, *(b)* the pairing energy ΔE_p(aq) in simulated aqueous phase, *(c)* the B3LYP/6-31 G(d,p) gas phase free energy change ΔG_p at 298 K accompanying base pair formation (with entropy terms ΔS incorporated explicitly), and *(d)* the corresponding free energy change ΔG_p(aq) in the simulated aqueous phase. All these energy changes are defined as the energy difference between the H-bonded base pair and the free bases infinitely apart. For instance, the pairing energy ΔE_p for a base pair **A…B** is obtained from the total energies of the given pair and of its constituent bases **A** and **B** as shown below in Eqn. (1):

$$\Delta E_p = E_t(\mathbf{A}\ldots\mathbf{B}) - [E_t(\mathbf{A}) + E_t(\mathbf{B})], \quad (1)$$

where E_t(**A…B**), E_t(**A**) and E_t(**B**) are the total ZPE-corrected electronic energies of the **A…B**, **A** and **B** systems respectively. Quantum theoretical work on nucleobase pairs has so far by and large made use only of total energy changes, and not free energy differences. Most computations on nucleobase pairs have also been carried out only in gas phase.

Base pairing configuration is estimated using geometry determinants used earlier [47, 48], schematically shown in Fig. 1 for the Gua:Cyt base pair as an example. The

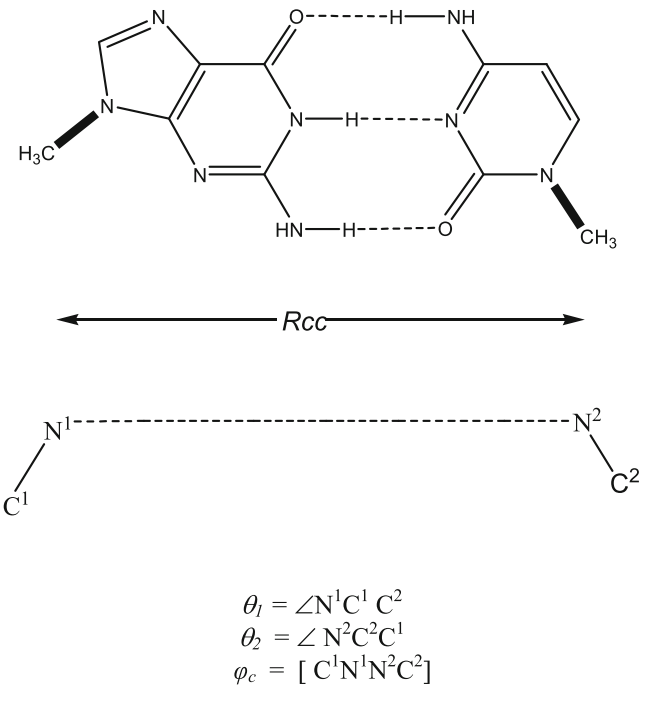


Fig. 1 Configuration markers for the Gua:Cyt base pair

configuration descriptors include *(a)* the distance R_{cc} between the C1' carbon termini of the two would-be glycoside N-C1' bonds of the two bases in a pair, *(b)* the angles θ_1 and θ_2 between the N-C1' bonds and the C1'-C1' vector, and *(c)* the dihedral angle φ_c spanning the two N-C1' bonds connected by the N-N vector. R_{cc} describes base pair width, an important factor affecting accommodation of a base pair at the wobble position. Similar θ_1 and θ_2 values mean that a pair **A…B** has the same configuration as the reversed **B…A** pair, as is true for the canonical DNA base pairs which are interchangeable between strands. Dihedral φ_c is related to the propeller twist [49] (defined by the IUPAC/IUB convention) which points to how co-planar the two bases within a pair are, where greater co-planarity stabilizes codon-anticodon pairs through promoting stacking interactions between two adjacent base pairs.

Optimized geometries in gas and solvent phase yield values for configuration descriptors as well as the H-bond geometry determinants. More emphasis is laid here on the simulated solvent phase data than on the gas phase data, since the natural environment in which RNA base pairs occur during codon-anticodon pairing is an aqueous phase situation. The configuration of each base pair is compared with the B3LYP/6-31 G(d,p) alignments of the canonical pairs Gua:Cyt and Ade:Ura calculated here. The aqueous phase R_{cc} values are 10.75 and 10.64 Å for Gua:Cyt and Ade:Ura respectively, θ_1 and θ_2 values ranging from 53.4 to 55.2°, while the dihedral φ_c approaches 0º. In gas phase, R_{cc} values are 10.76 and 10.61 Å for Gua:Cyt and Ade:Ura respectively, θ_1 and θ_2 values between 53.5 and 55.2°, while φ_c approaches 0º. The Watson-Crick alignment for a wobble pair is deemed optimal in the context of the mini-helical environment around the codon-anticodon pairing region, approaching the A-form of RNA (the structural invariant for the codon strand).

Geometries of select RNA base pairs obtained from crystallographic data stored in the Protein Data Bank [50] are compared with our aqueous phase B3LYP/6-31 G(d,p) data. Relevant RNA base pair geometries were extracted using the Swiss-PDBViewer software [51]. Of these pairs, only two may relate to codon-anticodon pairing.

A hydrogen bond in a pair is written as **X…H-Y** or **X-H…Y**, where **X** and **Y** are electro-negative atoms belonging to two different base moieties. The dots "…" signify the actual H-bond, while the dash "–" refers to the covalent bond. H-bond geometry is defined by *(a)* the length R_{hb} of the actual H-bond **X…H** or **H…Y**, *(b)* the length R_{xy} between atoms **X** and **Y**, *(c)* the H-bond angle θ_{hb} of the moiety **X…H-Y** or **X-H…Y**, and *(d)* the dihedral φ_{hb} spanning the non-H atoms involved within the zone encompassing two adjacent H-bonds, where pairs with two H-bonds have one φ_{hb} value, while pairs with three H-bonds have two φ_{hb} values for the two adjacent H-bonds. Sadlej-Sosnowska [13] had used such a feature to assess local planarity within the H-bonding zone for numerous RNA wobble base pairs.

Results and discussion

Tables 2, 3, 4 and 5 list gas and aqueous phase data for pairing facility and pairing configuration. Values of the pairing energies ΔE_p and ΔE_p(aq) are negative for all base pairs, with ΔE_p ranging from -6.68 to -28.69 kcal mol^{-1}. The pairing energies ΔE_p(aq) in aqueous phase are markedly smaller in magnitude in each case, ranging from -1.72 to -9.68 kcal mol^{-1}. These small values for pairing energies in solvent phase have been noted by Sharma et al. [31] and attributed to dampening of electrostatic interactions through solvation by the polar aqueous medium. These trends do indicate, however, that H-bonded formation of the base pair in each case is accompanied by a lowering of total electronic energy.

A noticeable contrast in trend is furnished by the free energy changes ΔG_p and ΔG_p(aq) in gas and simulated aqueous phases. The gas phase free energy change ΔG_p, though negative for most cases, is not always so, ranging from -15.18 to 18.69 kcal mol^{-1}. It is the simulated aqueous phase free energy change ΔG_p(aq) that departs radically

Table 2 Pairing facility and configuration[a] of wobble pairs arising from Gua, Ino and Que as the anticodon wobble bases with Ade as the codon wobble base

Pair	Status	Edges	ΔE_p	ΔG_p	ΔE_p(aq)	ΔG_p(aq)	R_{cc}	θ_1	θ_2	φ_c
Gua:Ade	Disallwd	WC/WC	-16.98	-3.15	-3.87	9.26	12.94 (12.97)	45.0 (45.6)	45.3 (44.1)	17.1 (16.4)
		WC/H	-16.23	-2.28	-3.96	9.60	10.95 (10.97)	50.9 (52.2)	40.2 (38.8)	21.3 (20.4)
Ino:Ade	Allowed	WC/WC	-16.41	-3.98	-5.00	7.83	12.97 (12.95)	45.5 (46.0)	45.0 (45.0)	0.1 (0.1)
		WC/H	-15.51	-3.77	-4.89	7.92	10.94 (10.98)	52.2 (52.8)	41.2 (39.7)	0.0 (0.0)
Que:Ade	Disallwd	WC/WC	-16.75	-3.95	-3.44	11.01	12.92 (12.91)	47.4 (48.2)	45.4 (45.0)	15.1 (13.9)
		WC/H	-16.03	-3.23	-3.66	10.60	10.93 (10.93)	53.1 (54.6)	40.6 (39.0)	20.0 (19.2)

[a] Pairing energies in kcal mol^{-1} ; distances in angstrom; angles in degrees

Table 3 Pairing facility and configuration[a] of wobble pairs arising from Gua, Ino and Que as anticodon wobble bases with Gua as the codon wobble base

Pair	Status	Edges	ΔE_p	ΔG_p	ΔE_p(aq)	ΔG_p(aq)	R_{cc}	θ_1	θ_2	φ_c
Gua:Gua	Disallwd	H/WC	-14.12	-1.42	-4.97	7.68	11.42 (11.20)	29.3 (33.0)	61.3 (60.8)	43.0 (60.9)
		WC/SE	-16.62	18.69	-3.06	11.01	7.93 (8.03)	3.6 (53.9)	127.3 (123.2)	58.3 (64.6)
Ino:Gua	Disallwd	H/WC	-13.05	-0.69	-4.24	8.63	11.46 (11.00)	29.3 (35.5)	61.2 (59.3)	42.7 (64.8)
		WC/H	-8.76	-11.96	-3.53	7.86	10.64 (10.65)	72.9 (76.4)	34.3 (34.0)	2.5 (2.6)
		WC/SE	-14.22	-0.50	-2.73	10.86	7.92 (7.82)	54.0 (56.3)	128.3 (130.7)	55.3 (48.3)
Que:Gua	Disallwd	WC/H	-14.94	-3.01	-3.28	10.16	11.38 (11.10)	63.7 (63.7)	29.1 (33.2)	46.5 (58.6)
		WC/SE	-15.34	-1.11	-2.27	13.96	7.81 (7.92)	57.3 (56.6)	129.3 (124.9)	54.7 (62.0)

[a] Pairing energies in kcal mol^{-1} ; distances in angstrom; angles in degrees

from the trends described above. All ΔG_p(aq) values are positive (4.29 to 13.96 kcal mol^{-1}). Such a situation is clearly unrealistic, since one cannot imagine that none of these pairs would be thermodynamically stable in aqueous phase. Such a result arises since water molecules in the solvation model do not have their degrees of freedom taken into account explicitly. The entropy lowering ΔS is thus underestimated for the base pair system, which actually has a larger number of degrees of freedom than the two component bases infinitely apart. We infer that free energy changes in simulated aqueous phase here do not represent a real situation, and so no more appeal will be made in this paper to such. However, it is noted that small positive ΔG_p(aq) values are usually associated with large negative ΔE_p and ΔE_p(aq) values and vice versa.

The negative values of ΔE_p and ΔE_p(aq) for all cases suggest that thermodynamic facility of H-bonded pairing by itself cannot differentiate between allowed and disallowed wobble pairs, though necessary to ensure a stable pair. Canonical and mismatched RNA base pairs may both be easily adopted into RNA mini-helices when the configuration of the pair is not of great importance, unlike during codon-anticodon pairing. A wrong codon-anticodon duplex may thus be stable enough to be observed in vitro outside the biological context of the translation process, as found to occur within a crystalline yeast tRNA minihelix [52]. This suggests that the *configuration* (not the stability) of the base pair is the real criterion for deciding whether a given RNA base pair may be accommodated or not at the wobble position during codon-anticodon pairing in the context of protein synthesis.

The B3LYP base pair geometries usually present only small variations between gas and aqueous phases. This contrasts with the large divergences in value for the pairing energies between gas and aqueous phase. Configurations of the pairs formed from Gua, Ino and Que as anticodon wobble bases predict that these three bases have somewhat similar pairing properties. A wide range of configurations is predicted, where aqueous phase values of the φ_c marker range from 0.0° to 58.3°, while R_{cc} distances range between 7.81 and 12.97 Å.

Values for descriptors of H-bond geometry (see Supplementary information) do not show much divergence between gas and aqueous phases. Lengths of the C = O…H-N type of H-bond are somewhat longer in aqueous phase than in gas phase. Differences in H-bond lengths between aqueous and gas phases are less marked for the less polar N…H-N type of H-bond. Solvation by a polar species like water affects the more polar C = O…H-N type of bond more than the less polar N…H-N type. Values of the H-bond angles θ_{hb} closely approach 180° for the vast majority of cases, indicating H-bond linearity in general.

Table 4 Pairing facility and configuration[a] of wobble pairs arising from Gua, Ino and Que as the anticodon wobble bases with Ura and Cyt as codon wobble bases

Pair	Status	Edges	ΔE_p	ΔG_p	ΔE_p(aq)	ΔG_p(aq)	R_{cc}	θ_1	θ_2	φ_c
Gua:Ura	Allowed	WC/WC	-16.02	-3.03	-3.31	9.30	10.57 (10.46)	41.0 (42.7)	67.2 (68.9)	0.3 (1.0)
Ino:Ura	Allowed	WC/WC	-14.67	-1.66	-3.53	9.20	10.56 (10.57)	41.0 (41.5)	67.7 (66.6)	0.1 (0.1)
Que:Ura	Allowed	WC/WC	-16.35	-3.20	-2.59	11.52	10.54 (10.31)	43.4 (45.4)	67.8 (65.4)	4.8 (23.9)
Gua:Cyt	Allowed	WC/WC	-28.69	-15.18	-9.68	4.29	10.75 (10.76)	53.4 (53.5)	54.9 (54.4)	0.4 (0.4)
Ino:Cyt	Allowed	WC/WC	-22.18	-9.32	-6.79	6.67	10.44 (10.48)	55.3 (55.3)	58.0 (57.5)	0.1 (0.1)
Que:Cyt	Allowed	WC/WC	-28.15	-14.71	-8.91	5.98	10.72 (10.70)	55.6 (55.8)	55.2 (55.1)	2.7 (0.2)

[a] Pairing energies in kcal mol^{-1} ; distances in angstrom; angles in degrees

Table 5 Pairing facility and configuration[a] of wobble pairs arising from Ade as the anticodon wobble base with the four major RNA bases as the codon wobble base

Pair	Status	Edges	ΔE_p	ΔG_p	ΔE_p(aq)	ΔG_p(aq)	R_{cc}	θ_1	θ_2	φ_c
Ade:Ade	Disallwd	WC/WC	-6.68	4.52	-2.18	7.44	12.81 (12.59)	39.9 (42.1)	63.8 (66.5)	2.7 (3.5)
Ade:Gua	Disallwd	WC/WC	-16.98	-3.15	-3.87	9.26	12.94 (12.97)	45.3 (41.1)	45.0 (45.6)	17.1 (16.4)
Ade:Ura	Disallwd	WC/WC	-15.60	-3.43	-5.45	7.30	10.64 (10.61)	54.5 (55.2)	55.2 (53.9)	0.1 (0.0)
Ade:Cyt	Disallwd	WC/WC	-7.06	4.32	-1.72	8.83	10.65 (10.41)	73.4 (75.1)	49.4 (51.9)	5.6 (3.9)

[a] Pairing energies in kcal mol^{-1} ; distances in angstrom; angles in degrees

Base pairs with Ade as CWB

Table 2 gives data for base pairs arising from Gua, Ino and Que as AWB's with Ade as the CWB. The Ino:Ade combination is allowed at the wobble position for the amino acids Ala, Arg, Ser and Val, while the Gua:Ade and Que:Ade combinations are disallowed. Values of the gas phase pairing energy ΔE_p (-15.51 to -16.98 kcal mol^{-1}) show appreciable H-bonded stabilization in gas phase for all base pairs. Solvent phase pairing energy ΔE_p(aq) values are smaller (-3.44 to -5.00 kcal mol^{-1}), showing a marked effect of the simulated aqueous phase. The gas phase free energy change ΔG_p (-2.28 to -3.98 kcal mol^{-1}) is negative for all the pairs, and predicts thermodynamic stability for all cases.

Two *cis* base pairing configurations are tried out for each pairing combination. The first is the c*is* WC/WC type, listed in Fig. 2 as Gua:Ade(I), Ino:Ade(I) and Que:Ade(I). These pairs are interchangeable between strands, where θ_1 and θ_2 values range from 45.0 to 47.4° in aqueous phase. The dihedral φ_c points to essential co-planarity of the Ino:Ade(I) pair while the other pairs are somewhat less planar.

All these *cis* WC/WC pairs are characterized in aqueous and gas phase by long R_{cc} values of nearly 13 Å, much longer than the values of about 10.7 Å obtained for the canonical pairs. This markedly wide base pair width begs the question of whether any of these *cis* WC/WC pairs would be able to fit in at the wobble position at all. Moreover, this type of configuration cannot distinguish between the allowed Ino:Ade wobble pair and the disallowed Gua:Ade and Que:Ade wobble pairs since their configurations are all quite similar (except that φ_c values in the Gua:Ade (I) and Que:Ade (I) pairs deviate much from zero). We thus exclude the *cis* WC/WC type of configuration for these wobble pairs.

A second configuration – the *cis* WC/H type (Fig. 2) for the Gua:Ade(II), Ino:Ade(II) and Que:Ade(II) pairs – involves the N7 atom of the CWB Ade in H-bonding. The adenine nucleoside on the codon side (not explicitly considered here) adopts a conformation other than the usual *anti* one. These WC/H pairs are not reversible and θ_1 and θ_2 differ by up to 12.5°, which, however, is probably still admissible in the wobble pairing context. The Ino:Ade(II) pair shows base co-planarity while the other two pairs deviate by about 20° from base co-planarity. These pairs all have R_{cc} values (10.93 to 10.95 Å) that approach those for the Watson-Crick base pairs. We infer that the Gua:Ade, Ino:Ade and Que:Ade pairs in the *cis* WC/H configuration should be all admissible at the wobble position.

However, the Gua:Ade and Que:Ade pairing combinations do not occur at the wobble position during codon-anticodon pairing, while the Ino:Ade pair is allowed. To explain why the Gua:Ade (II) and Que:Ade (II) pairs are not allowed, we propose that the sugar moiety of the CWB nucleoside adenosine blocks formation of a wobble pair for these cases due to steric interaction with the 2-amino group present in both Gua and Que, but not in Ino. Such interactions could arise if the ribose moiety protrudes to some extent into the region between the two bases of the pair, which could happen if the adenine nucleoside has a conformation intermediate between *syn* and *anti*. This explanation thus invokes factors external to the solitary base itself, calling for consideration of the adenine nucleoside with its sugar moiety instead of just the free base. This also calls for involvement of a non-*anti* conformation for the codon nucleoside. All this thus goes beyond our original hypothesis which invokes just the solitary base.

The above rationale stands in contradiction to the experimental findings of Murphy and Ramakrishnan [53] who used X-ray crystallography to study the Ino:Ade wobble pair at the ribosome decoding center. They found it to be of the *anti-anti* WC/WC configuration, not the *anti-syn* WC/H configuration proposed by us. The WC/WC Ino:Ade(I) wobble pair has a markedly long R_{cc} distance (12.97 Å in our study and 12.30 Å in the experimental study). Murphy and Ramakrishnan explained the accommodation of the *anti-anti* Ino:Ade pair at the wobble position by appealing to compensations in the distances between the backbone phosphate ester moieties. These crystallographic findings present the difficulty that the width of a wobble pair is dispensed with as a factor dictating whether or not it may be admissible in the codon-anticodon pairing context. As shown later, base pair width does indeed emerge as a convincing factor to label wobble pairs as allowed or disallowed.

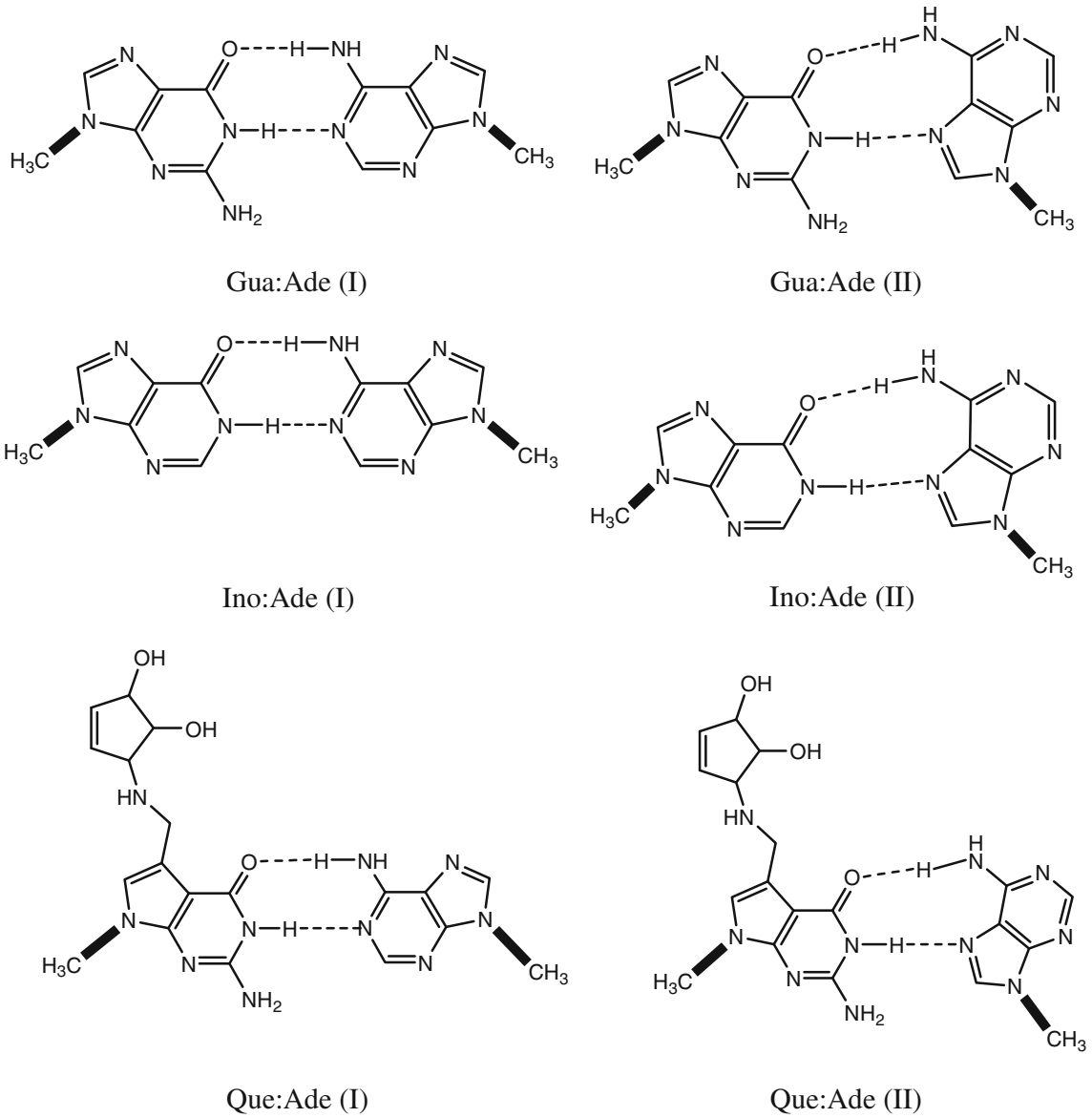


Fig. 2 Wobble pairs arising from Gua, Ino and Que as anticodon wobble bases with Ade as codon wobble base

In this case, large base pair width rules out the *cis* WC/WC configuration as inadmissible for the Gua:Ade, Ino:Ade and Que:Ade pairs. Occurrence of the Ino:Ade pair is possible if one proposes an alternative *cis* WC/H alignment with shorter base pair width as discussed above. We suggest that the system studied by Murphy and Ramakrishnan may not really represent an allowed codon-anticodon pair at the moment of cognate pairing during protein synthesis. A non-cognate codon-anticodon pair outside the situation of protein synthesis has, in fact, been characterized experimentally as mentioned earlier [52].

H-bond data for the Gua:Ade, Ino:Ade and Que:Ade pairs (both types) suggest that H-bond geometry does not vary much between solvent and gas phase. The H-bonds are quite linear (θ_{hb} values from 166.1 to 179.5° in aqueous phase). The first type of configuration (WC/WC) leads to a more linear O6…H-N6 hydrogen bond than the second type (WC/H). This is because the first base pair type involves a hexagonal ring of 8 atoms around the H-bonding region (Fig. 2) like the Watson-Crick pairs. The second type has a ring of 9 atoms, leading to diminished linearity for the O6…H-N6 H-bond. This may explain the smaller pairing energies in gas phase for the second type of pair in each case (Table 2). Solvation tends to equalize these differences in H-bond length between the two configuration types.

The local planarity dihedral φ_{hb} deviates from 180° in gas and solvent phase for the Gua:Ade and Que:Ade pairs, but approaches 180° for the Ino:Ade pairs. This is due to non-planarity of the 2-amino group of the AWB's Gua and Que as it participates in H-bonding. Amino group

pyramidalization, more appreciable in the free bases, is still partly retained in the base pairs Gua:Ade and Que:Ade. The Ino base, however, has no such amino group.

Base pairs with Gua as CWB

Pairs involving the AWB's Gua, Ino and Que with the CWB Gua are all disallowed. Table 3 presents data for pairs arising out of the Gua:Gua, Ino:Gua and Que:Gua combinations. Gas phase ΔE_p values range from -8.76 to -16.62 kcal mol^{-1}. Introduction of aqueous phase greatly diminishes the pairing energy, with ΔE_p(aq) values ranging from -2.27 to -4.97 kcal mol^{-1}. The gas phase free energy change ΔG_p has either positive values or small negative values, indicating that these base pairs are unstable or have reduced stability. Use of aqueous phase also reduces the dihedral φ_c, leading to somewhat more planar pairs. The R_{cc} values differ between gas and solvent phase by up to 0.46 Å.

Several arrangements are tried out for the pairing combinations having Gua as the CWB. The first motif is the *cis* H/WC type. For the Gua:Gua and Ino:Gua pairing combinations, this is labeled (Fig. 3) as Gua:Gua(I) and Ino:Gua(I), where the AWB nucleoside guanosine (not considered explicitly here) is not in the usual *anti* conformation. Such an arrangement is not possible for the Que:Gua pair due to the exocyclic group on the C7 atom of Que. For this pair and the Ino:Gua combination, a second type – the *cis* WC/H type of motif – is also adopted, labeled as Que:Gua(II) and Ino:Gua (II) in Fig. 3. These involve the nucleoside guanosine (not explicitly considered) in a conformation other than the usual *anti* one. In a third series of pairs, the *cis* WC/SE pairing motif is used, the pairs being labeled as Gua:Gua(III), Ino:Gua(III) and Que:Gua(III), all involving the CWB Gua sugar edge.

For the H/WC pairs Gua:Gua(I) and Ino:Gua(I), the R_{cc} values (11.42 and 11.46 Å in aqueous phase) are somewhat longer than those for normal Watson-Crick pairs. These pairs are not reversible, with θ_1 and θ_2 diverging by about 32° in aqueous phase, which may not itself be the prime disallowing factor. Greater deviation from Watson-Crick alignment occurs in the φ_c dihedral (from 43.0 to 42.7°), indicating significant departure from base co-planarity. Together with the large R_{cc} distances, this is perhaps the real factor which disallows the occurrence of these candidate wobble pairs during codon-anticodon pairing.

The WC/H type Ino:Gua(II) pair involves participation of the C2-proton of Ino in H-bonding, giving a carbon hydrogen bond, which is not strong or well-defined enough to impart a fixed and stable configuration to the Ino:Gua(III) pair. The ΔE_p and ΔE_p(aq) values are -8.76 and -3.53 kcal mol^{-1} respectively, while the free energy change ΔG_p has the small positive value of 0.04 kcal mol^{-1}. This motif is quite co-planar and has an acceptable R_{cc} value of 10.64 Å in solvent phase. Its non-occurrence at the wobble position is probably due more to its lack of fixity rather than its low pairing energy, or else due to difficulty of the codon nucleoside guanosine to assume a conformation other than the *anti* one. The WC/H type Que:Gua(I) pair yields an R_{cc} value of 11.38 Å, a φ_c value of 46.5° and a large difference in values (about 34.6°) between the θ_1 and θ_2 angles in solvent phase. All of these factors may serve to screen out the Que:Gua(I) pair as disallowed at the wobble position.

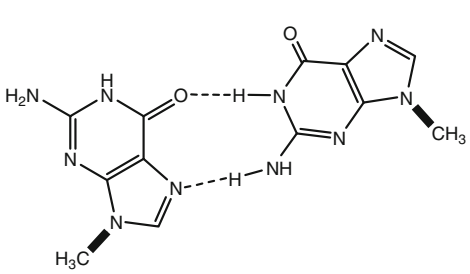
The WC/SE pairs Gua:Gua(III), Ino:Gua(III) and Que:Gua(III) all show R_{cc} values (7.81 to 7.93 Å in aqueous phase) that are much shorter than the standard Watson-Crick values. These pairs also deviate much from base co-planarity, with large φ_c dihedral values (54.7 to 58.3°). The θ_1 and θ_2 angle values also show that these pairs are far from being reversible between strands. All these features emerge as factors that disallow these WC/SE pairs at the wobble position. Another factor may be the difficulty of the nucleoside guanosine to adopt a conformation that allows for pairing through its sugar edge. Note also that the free energy change ΔG_p for these WC/SE pairs (-1.11 to 18.69 kcal mol^{-1}) does not point to any appreciable stability in gas phase, even though the ΔE_p and ΔE_p(aq) values are all negative. Some or all of these factors may serve to screen out these WC/SE wobble pairs.

H-bond geometry data for these pairs is given in the Supplementary Information. Though all H-bonds are quite linear, the local planarity dihedral φ_{hb} for these pairs (except the WC/H Ino:Gua pair) deviates much from 180°, thus explaining the non-planarity predicted by the φ_c dihedral for these base pairs. The H-bond lengths R_{xh} and R_{xy} for the C=O…H-N type H-bonds are shorter by about 0.1 Å those for the N…H-N type H-bonds. The weak C_2-H : N_7 hydrogen bond in the WC/H Ino:Gua(II) pair is noticeably longer (R_{xh}=2.34 Å in aqueous phase) than the other H-bonds and is also quite non-linear.

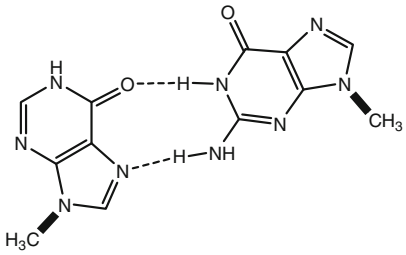
Base pairs with Cyt as CWB

Wobble pairs where the AWB's Gua, Ino and Que pair with Cyt as the CWB are all allowed at the wobble position. The Gua:Cyt pair occurs in codon-anticodon pairing for the amino acids Cys, Gly, Ile, Leu, Phe, Ser, Thr and Val. The Ino:Cyt pair occurs during codon-anticodon pairing for Ala, Arg, Ser and Val. The Que:Cyt pair occurs for the amino acids Asp, His and Tyr. The only pairing motif required to be adopted here for all these pairing combinations is the canonical *cis* WC/WC alignment (Fig. 4), which adequately accounts for the occurrence of these pairing combinations at the wobble position.

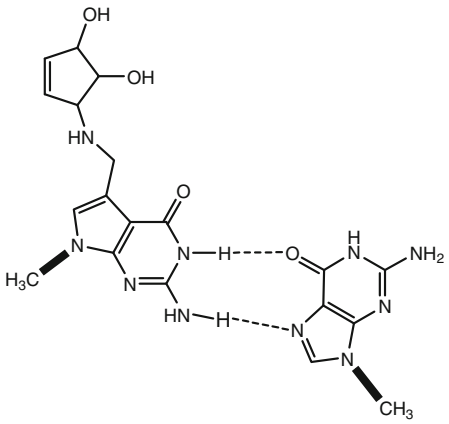
Fig. 3 Wobble pairs arising from Gua, Ino and Que as anticodon wobble bases with Gua as codon wobble base



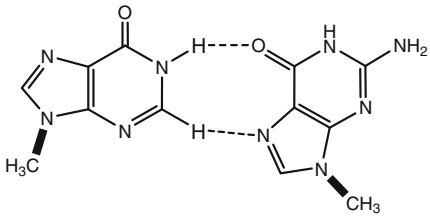
Gua:Gua (I)



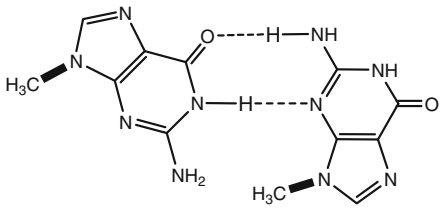
Ino:Gua (I)



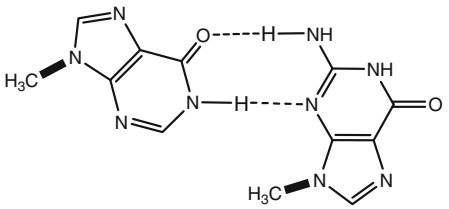
Que:Gua (II)



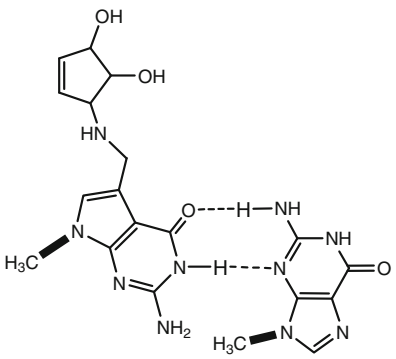
Ino:Gua (II)



Gua:Gua (III)



Ino:Gua (III)



Que:Gua (III)

Table 4 gives our data for these pairs having Cyt as the CWB. The Gua:Cyt and Que:Cyt pairs with three H-bonds show larger ΔE_p values (-28.69 and -28.15 kcal mol^{-1} respectively) than the Ino:Cyt pair (-22.18 kcal mol^{-1}) with two H-bonds. This trend is followed by the aqueous phase pairing energy ΔE_p(aq) values, and by the free energy change ΔG_p. The free energy change ΔG_p for these Watson-Crick pairs is appreciable (-9.32 to -15.18 kcal mol^{-1}).

Gas and solvent phase values of the markers R_{cc}, θ_1, θ_2 and φ_c (Table 4) establish these pairs as being Watson-Crick type and therefore allowed at the wobble position. Solvent phase R_{cc} values are 10.75, 10.44 and

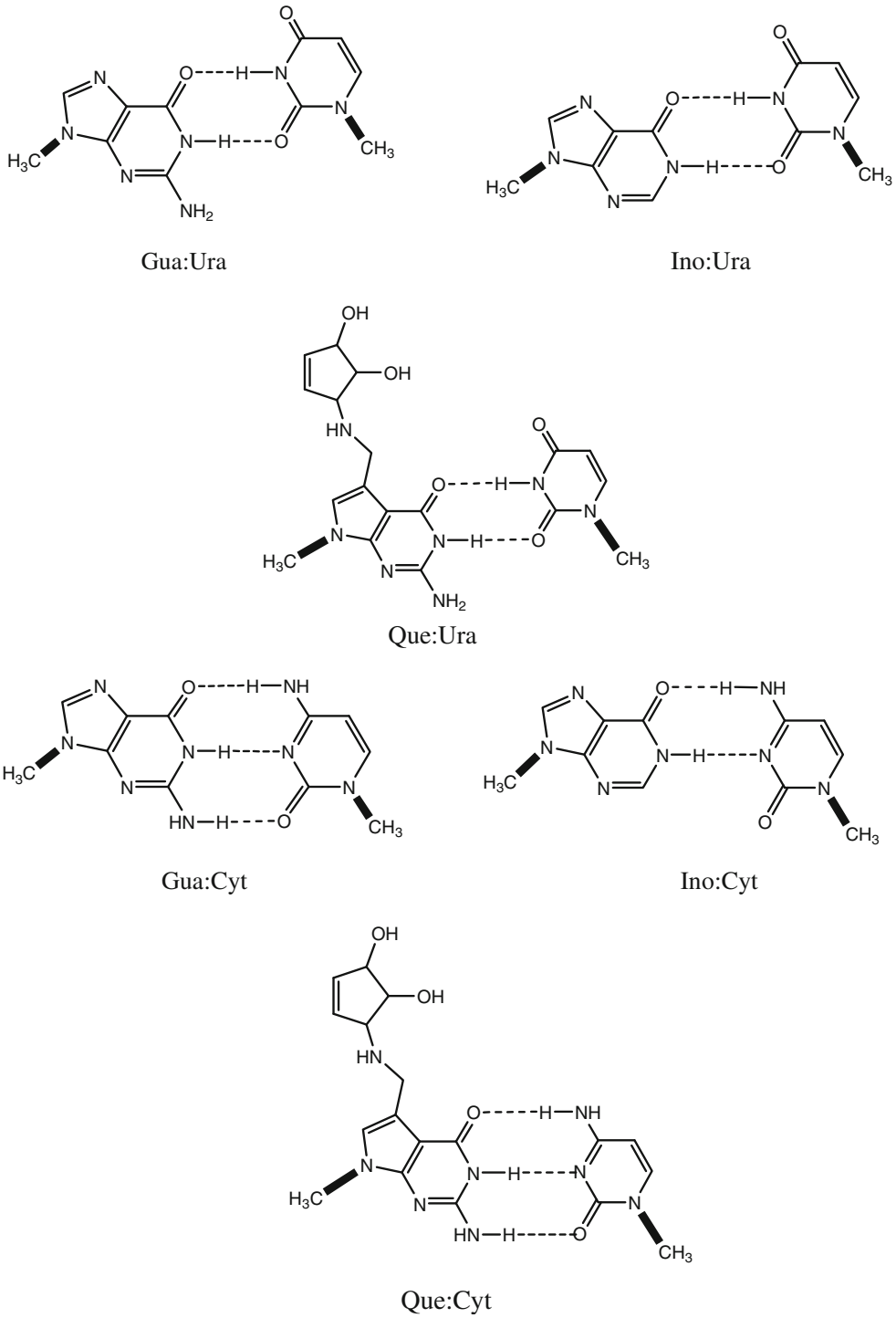


Fig. 4 Wobble pairs arising from Gua, Ino and Que as anticodon wobble bases with Ura and Cyt as codon wobble bases

10.72 Å respectively for the Gua:Cyt, Ino:Cyt and Que:Cyt pairs. These pairs are planar (φ_c values close to 0°) and reversible between strands (with small differences between θ_1 and θ_2). Their H-bonds are all linear, contained within the hexagonal 8-atom framework characteristic of Watson-Crick pairs. Values of the local planarity dihedral φ_{hb} also indicate planarity around the H-bonding zone for all these pairs.

Base pairs with Ura as CWB

Pairs involving the AWB's Gua, Ino and Que with the CWB Ura are all allowed at the wobble position. The Gua:Ura pair, an oft-quoted example of RNA base pair mismatch [54], occurs for the amino acids Cys, Gly, Ile, Leu, Phe, Ser, Thr and Val. The Ino:Ura pair occurs for Ala, Arg, Ser and Val, while the Que:Ura pair occurs for Asp, His and Tyr.

The motifs adopted for studying these pairing combinations are all of the *cis* WC/WC type (Fig. 4). Other configurations are not considered since this type of configuration itself explains adequately why these pairing combinations are allowed. The gas phase ΔE_p values (Table 4) (-16.02, -14.67 and -16.35 kcal mol^{-1} for Gua:Ura, Ino:Ura and Que:Ura respectively), show facile H-bonding in gas phase for these pairs. Corresponding solvent phase E_p(aq) values are smaller but all negative (-3.31, -3.53 and -2.59 kcal mol^{-1}), as also are the gas phase free energy changes (-3.03, -1.66 and -3.20 kcal mol^{-1}). Note that these non-Watson-Crick pairs are all less stable than the canonical type pairs formed with Cyt, and the values of ΔE_p, ΔE_p(aq) and ΔG_p are all smaller than those for the Watson-Crick pairs.

Table 4 shows that the solvent phase R_{cc} values (10.57, 10.56 and 10.54 Å for Gua:Ura, Ino:Ura and Que:Ura respectively) approach those for the canonical pairs. These pairs are also more or less co-planar (with φ_c equal to 0.3, 0.1 and 4.8° respectively). The differences between θ_1 and θ_2 are 26.2, 26.7 and 24.4° respectively, so these pairs are not exchangeable between strands. However, since these pairs are all allowed, we conclude that differences to this extent do not disqualify them from occurring at the wobble position. These *cis* WC/WC pairing motifs are thus most likely to be the ones that actually occur in nature for these wobble pairs during codon-anticodon pairing.

H-bond data for the Gua:Ura, Ino:Ura and Que:Ura pairs (see Supplementary information) indicate that the H-bonds are linear in aqueous phase and relatively short, thus contributing to H-bonded pairing facility for these pairs. The H-bond geometries of these pairs are similar in both gas and aqueous phases except for the Que:Ura pair.

Ade as candidate AWB

Study of the pairing properties of Ade as a candidate AWB is carried out in the context of the almost total absence of Ade at the anticodon wobble position in nature, notably in eukaryotes. However, adenosine may pair with any major RNA base at the wobble position if it does indeed occur there, as in some prokaryotes [5]. Adenosine as an AWB allows for translation of synonymous family boxes of codons by *Mycoplasma capricolum* tRNA's [6].

We use *cis* WC/WC alignments to represent the pairing combinations between the AWB Ade and the CWB's Ade, Gua, Ura and Cyt (Fig. 5). Table 5 presents pairing energy values and configuration data for these pairs. While pairing energy values are appreciable for the Ade:Gua and Ade:Ura pairs, they are noticeably smaller for the Ade:Ade and Ade:Cyt pairs owing to the presence of carbon H-bonds in these latter two pairs. These carbon H-bonds arise from our insistence upon a *cis* configuration for all the pairs of this study.

From the configuration data of Table 5, the Ade:Ade and Ade:Gua pairs may be regarded as disallowed owing to their large R_{cc} values of 12.81 and 12.94 Å respectively in aqueous phase. Moreover, the Ade:Ade pair with its weak N_1:H-C_2 carbon hydrogen bond has only one proper H-bond and unfavorable pairing energy values, where ΔE_p, ΔE_p(aq) and ΔG_p are respectively -6.68, -2.18 and 4.52 kcal mol^{-1}. The WC/WC alignment for the Ade:Ade pair can thus guarantee neither fixity nor stability, and may be excluded.

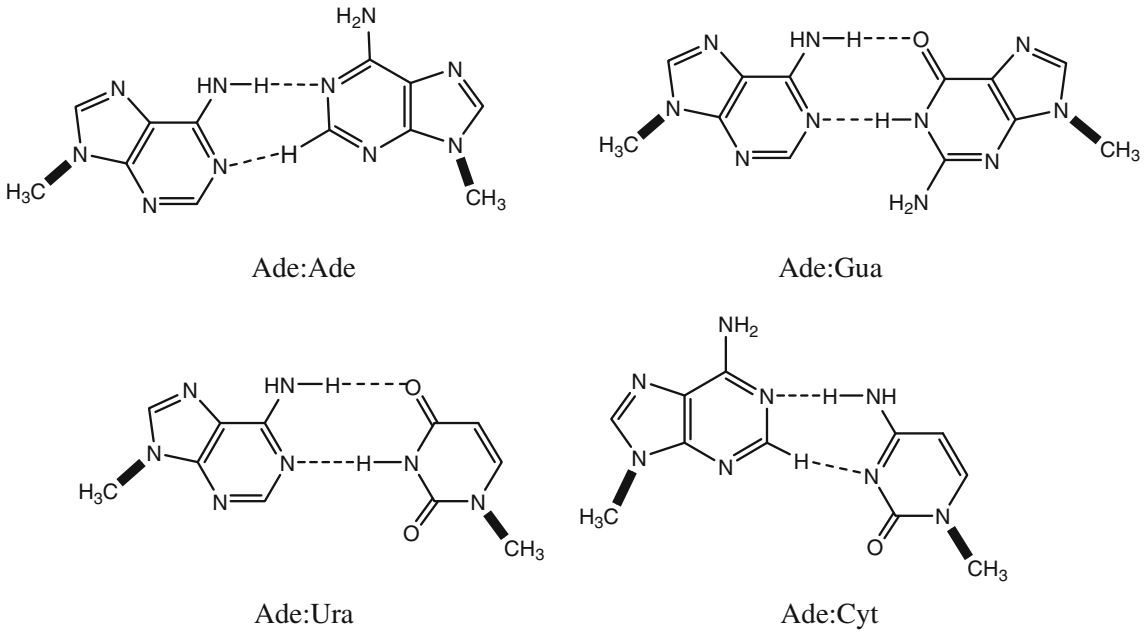


Fig. 5 Wobble pairs arising from Ade as the anticodon wobble base with the four major RNA bases as the codon wobble base

However, the two WC/WC pairs Ade:Ura and Ade:Cyt yield acceptable values for the R_{cc} marker, the θ_1 and θ_2 angles and the dihedral φ_c. The Ade:Ura pair is, of course, Watson-Crick and should be able to be accommodated at the wobble position. Its absence here is thus not due to its inability to assume a configuration that fits into the wobble position. For the WC/WC Ade:Cyt wobble pair, we may infer that its absence could arises from its lack of fixedness and its small pairing energy, where E_p, E_p(aq) and ΔG_p values are respectively -7.06, -1.72 and 4.32 kcal mol^{-1} due to presence of the weak C_2-H:N_3 carbon hydrogen bond. So the only pair allowed by the norms of this study is the canonical Ade:Ura WC/WC pair.

Our data does not explain the almost universal absence of Ade at the anticodon wobble position, since it should be able to form at least the Ade:Ura pair if present. This absence of Ade as an AWB certainly does not arise from its inability to pair suitably with any CWB. Nor does this data explain the indiscriminate pairing of the AWB Ade in some organelles and lower organisms. One rationale for this latter observation is that ribosomal recognition of the different codon-anticodon pairs may be markedly less stringent for such cases in the context of the comparatively primitive decoding systems involved.

The H-bond geometry data for these wobble pairs involving Ade as AWB points to linear and well-formed H-bonds for most cases. The N1…H-C2 and C2-H…N3 carbon H-bonds of the Ade:Ade and Ade:Cyt pairs have anomalously large H-bond lengths (R_{xh}=2.61 and 2.55 Å respectively in aqueous phase). The local planarity dihedral φ_{hb} indicates essential planarity of the H-bonding zone for all pairs except the purine-purine Ade:Gua pair.

Comparison with crystal structure and other computational data

Table 6 lists nine select RNA base pairs located at certain H-bonded pairing regions of experimentally characterized RNA molecules, of which only two may pertain to codon-anticodon pairing. Crystal structure data for these RNA molecules is derived from the Protein Data Bank [50], where the select RNA base pair units are not necessarily canonical or double-helical in configuration. Listed in Table 6 for each base pair (AWB:CWB) are the interacting edges, the computational and experimental values of the configuration markers R_{cc}, θ_1, θ_2 and φ_c, as well as the distance R_{nn} between the two glycosidic nitrogens, along with the source of the experimental data (the PDB ID number and the sequence-based location of the concerned base pair).

The 1XNQ and 1XNR databases [53] include the Ino:Ade and Ino:Cyt pairs as found at the decoding center of the ribosome, which are the only two cases here which represent actual codon-anticodon pairing. The 1DRZ database [55] includes the WC/WC Gua:Ade and Gua:Cyt pairs in the context of the structure of a hepatitis delta viral ribozyme. The 1FFK database [56] includes the WC/H Gua:Ade pair within the large ribosomal subunit. The 1DUQ database [57] includes an H/WC Gua:Gua pair located within the Rev binding element of HIV-1. The 1ASY database [58] includes the WC/WC Ade:Ura and Gua:Ura pairs in the context of yeast aspartyl-tRNA synthetase complexated with tRNA(Asp). The 409D database [59] gives the WC/WC Ino:Ura pair present in a synthetic octamer duplex.

Two cases (pairs 6 and 8) are from the decoding center of the ribosome. Five other cases (pairs 1, 3, 4, 5 and 9) occur in an environment rather comparable with that occurring in codon-anticodon pairing, all with a sequence of two canonical pairs on one side of the concerned pair, but no base pair on the other side. Two cases (pairs 5 and 9) are the canonical pairs, for which experimental and computational data are plentiful in the literature in the context of various environments and hence are not discussed further here, apart from the general comment that these pairs conform closely to the standard Watson-Crick alignment.

Apart from the WC/H Gua:Ade pair No. 2, the H/WC Gua:Gua pair No. 3 and the WC/WC Ino:Ade pair No. 6, the computational and experimental data sets match fairly closely. Differences in the R_{cc} values are less than 0.3 Å, those in the θ_1 and θ_2 values less than 13° and in the φ_c dihedral less than 19°. This suggests that a macromolecular environment (as for all these experimental results) does not greatly affect the pairing motifs represented. Our computational data describes solitary base pairs in simulated aqueous phase as contrasted with the experimental data recorded in a macromolecular environment, which involves a variety of adjunct base sequences and presence of the sugar phosphate moieties. Note that for pairs 4, 5, 6, 7, 8 and 9, the dihedral φ_c indicates greater planarity of the base pair in our computed results than in the experimental databases. This is due to the macromolecular environment associated with the pairs in crystal phase as contrasted with the solitary base pair in our calculations, where the adjacent nucleotides and pairs present in the former may lead to non-planar conformations for the concerned base pair.

Differences between computational and experimental datasets are more noteworthy for the pairs 2, 3 and 6 (all non-Watson-Crick purine-purine pairs). Between the two data sets, the R_{cc} values for pairs 2 and 6 differ by up to 0.67 Å, where the experimentally observed values are smaller than our computational values. For pair 3, an appreciable difference of 34.8° in the dihedral φ_c occurs between experiment and computation. This could point to the effect of the environment in introducing planarity to this pair. Our computations in a solitary base pair environment allow these pairs to adopt conformations less restricted than expected in

Table 6 Values of configuration descriptors[a] for RNA base pairs as obtained from X-ray crystal structures of various RNA systems [53, 55–59] and compared with the present B3LYP/6-31 G(d,p) data as obtained in solvent phase

Base-pair	Edges	R_{cc}	R_{nn}	θ_1	θ_2	φ_c	Sources
1. Gua:Ade	WC/WC	12.79	10.90	50.5	48.5	14.2	1DRZ G162(B)-A143(B)
	WC/WC	12.94	10.89	45.0	45.3	17.1	Results of this work
2. Gua:Ade	WC/H	10.37	8.66	65.3	41.1	13.7	1FFK G25(9)-A3(9)
	WC/H	10.95	8.93	50.9	40.2	21.3	Results of this work
3. Gua:Gua	H/WC	11.24	9.43	36.3	63.6	8.5	1DUQ G106(A)-G124(B)
	H/WC	11.43	9.49	29.3	61.3	43.0	Results of this work
4. Gua:Ura	WC/WC	10.31	8.75	42.5	70.9	2.9	1ASY G610(R)-U625(R)
	WC/WC	10.57	8.91	41.0	67.2	0.3	Results of this work
5. Gua:Cyt	WC/WC	10.56	8.84	52.5	55.1	5.4	1DRZ G158(B)-C147(B)
	WC/WC	10.75	9.04	53.4	54.9	0.4	Results of this work
6. Ino:Ade	WC/WC	12.30	10.46	44.5	53.5	5.6	1XNQ I34(X)-A3(W)
	WC/WC	12.97	10.92	45.5	45.0	0.1	Results of this work
7. Ino:Ura	WC/WC	10.57	8.77	46.2	56.9	7.6	409D I4(A)-U13(B)
	WC/WC	10.56	8.91	41.0	67.7	0.1	Results of this work
8. Ino:Cyt	WC/WC	10.52	8.75	47.7	57.6	19.0	1XNR I34(X)-C3(W)
	WC/WC	10.44	8.84	55.3	58.0	0.1	Results of this work
9. Ade:Ura	WC/WC	10.49	8.83	54.8	56.5	16.3	1ASY A607(R)-U666(R)
	WC/WC	10.64	8.96	54.5	55.2	0.1	Results of this work

[a] Distances in angstrom; angles in degrees

the macromolecular environment of these systems experimentally observed in crystal phase. Use of high salt concentrations [60] and cryoprotective agents [61, 62] in crystal studies could also lead to environments which appreciably modify biomolecular structures from their natural state.

We also compare our B3LYP/6-31 G(d,p) results with previous computational results [63] choosing three pairing motifs (WC/WC Gua:Cyt, WC/WC Gua:Ade and WC/H Gua:Ade) common to the two sets of results. While we represent the sugar moieties of individual bases by methyl groups, the previous results concern only the unmethylated bases. Table 7 presents the gas phase pairing energies of the three pairs obtained at the B3LYP/6-31 g(d,p) [64], HF/6-31 G(d,p) [65] and HF/MIN1 [66] levels, which are compared with the B3LYP pairing energies E_p and E_p(aq) of this study in gas and simulated aqueous phase. Previous DFT values of the pairing energies are somewhat smaller (by 2.2 to 2.6 kcal mol^{-1}) than the corresponding values of this study (which uses N-methylated bases). The HF/6-31 G(d,p) and HF/MIN1 values exclude correlation corrections, being both smaller than our B3LYP/6-31 G(d,p) values for each case (by up to 4.9 kcal mol^{-1}). However, all four sets of gas phase values lead to the common order Gua:Cyt (WC/WC) > Gua:Ade (WC/WC) > Gua:Ade (WC/H) for magnitude of pairing energy with respect to the pair. Only the aqueous phase values do not exactly follow this order, being also smaller in magnitude.

Conclusions

1. A candidate wobble pair should have at least two H-bonds of appreciable strength in order to maintain fixity of configuration. Base pairs having carbon hydrogen

Table 7 Interaction energies[a] of selected H-bonded base pairs as obtained in this study and compared with results obtained at the DFT/6-31 g(d,p) [64], HF/6-31 G(d,p) [65] and HF/MIN1 [66] levels. Geometries are fully optimized at all the respective levels

Base-pair	Type	HF/6-31 G(d,p)	HF/MIN1	DFT/6-31 g(d,p)	ΔE_p	ΔE_p(aq)
Gua:Cyt	WC/WC	-23.8	-25.6	-26.5	-28.69	-9.68
Gua:Ade	WC/WC	-14.2	-15.0	-14.5	-16.98	-3.87
Gua:Ade	WC/H	-13.7	-13.3	-13.7	-16.23	-3.96

[a] Pairing energies in kcal mol^{-1}

bonds are associated with reduced fixity and unfavorable pairing energies, which may be sufficient reason to exclude all such pairs from occurring during codon-anticodon pairing.

2. Base pair width, as estimated by the R_{cc} marker, is an important factor for screening a given wobble pair. Within the context of this study, a permissible range of about 10.4 to 11.0 Å for the R_{cc} marker may be suggested, gauged from the limits represented here by the allowed WC/WC Ino:Cyt and the WC/H Gua:Ade(III) pairs respectively.
3. The WC/H alignment emerges as a possible arrangement for the Ino:Ade pairing combination allowing for its occurrence at the wobble position. The disallowed Gua:Ade and Que:Ade pairing combinations are excluded on the basis of their WC/WC alignments (with long R_{cc} distances) and also by appealing to the role of the 2-amino group of the Gua and Que bases in these WC/H pairs (through interaction with the codon adenosine sugar).
4. Values of the φ_c dihedral should be small if a candidate pair is to be accommodated at the wobble position. This study suggests φ_c values up to 4.8° may be permissible in the context of our calculations. Large propeller twists (with φ_c values over 42.7° as gauged by this study) may screen out a candidate base pair from occurring at the wobble position.
5. Unequal θ_1 and θ_2 values do not in themselves prohibit a candidate base pair from occurring at the wobble position, where differences up to 26.2° (as in the Gua:Ura pair) in aqueous phase may still allow for accommodation at the wobble position.
6. The observed near total absence of Ade at the wobble position cannot be explained by any inability of Ade to form acceptable H-bonded pairs, since the Ade:Ura pair at least is acceptable as per the criteria proposed here for the other cases.
7. Apart from two exceptions, the H-bonded pairing configuration of the wobble pair considered in isolation is in itself sufficient to explain the specificity and degeneracy of the genetic code for all cases covered here. The only exceptions are the disallowed Gua:Ade and Que:Ade pairing combinations, where the role of the sugar moiety of the codon adenosine nucleoside has to be invoked in order to screen out these pairs.

Acknowledgments G. D. is grateful to the University Grants Commission, Government of India, New Delhi, for financial assistance through award of the Research Fellowship in Science for Meritorious Students. Financial assistance from the Special Assistance Program of the University Grants Commission to this Department is also gratefully acknowledged.

References

1. Crick FHC (1966) J Mol Biol 19:548–555
2. Woese CR (1967) The genetic code: The molecular basis for genetic expression. Harper and Row, New York
3. Ycas M (1969) The biological code (Frontiers of biology, volume 12). North-Holland, Amsterdam
4. Osawa S, Jukes TH, Watanabe K, Muto A (1992) Microbiol Mol Biol Rev 56:229–264
5. Boren T, Elias P, Samuelsson T, Claesson C, Barciszewska M, Gehrke CW, Kuo KC, Lustig F (1993) J Mol Biol 230:739–749
6. Inagaki Y, Kojima A, Bessho Y, Hori H, Ohama T, Osawa S (1995) J Mol Biol 251:486–492
7. Sodd MA (1969) Nucleic Acids. In: Fasman GD (ed) Handbook of biochemistry and molecular biology, vol II. CRC, Cleveland, p 423
8. Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, Putz J (2009) Nucleic Acids Res 37:D159–D162
9. Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH (1998) Biochemistry 37:14719–14735
10. Mathews DH, Sabina J, Juker M, Turner DH (1999) J Mol Biol 288:911–940
11. Mangang SU, Lyngdoh RHD (2001) Indian J Biochem Biophys 38:115–119
12. Das G, Lyngdoh RHD (2008) J Mol Struct (THEOCHEM) 851:319–334
13. Sadlej-Sosnowska N (2009) J Mol Struct (THEOCHEM) 913:270–276
14. Sponer J, Lankas F (2009) Computational Studies of RNA and DNA. Challenges and advances in computational chemistry and physics. 2. Springer, Heidelberg
15. Sponer JE, Vazquez-Mayagoitia A, Sumpter BJ, Leszczynski J, Sponer J, Otyepka M, Banas P, Feuntes-Cabrera M (2010) Chem Eur J 16:3057–3065
16. Sponer J, Leszczynski J, Hobza P (2001) Biopolymers 61:3–31
17. Sponer J, Jurecka P, Hobza P (2004) J Am Chem Soc 126:10142–10151
18. Zhanpeisov NU, Sponer J, Leszczynski J (1998) J Phys Chem A 102:10374–10379
19. Kratochvil M, Engkvist O, Sponer J, Jungwirth P, Hobza P (1998) J Phys Chem A 102:6921–6926
20. Hobza P, Sponer J, Cubero E, Orozco M, Luque FJ (2000) J Phys Chem B 104:6286–6292
21. Jurecka P, Sponer J, Hobza P (2004) J Phys Chem B 108:5466–5471
22. Bhattacharyya D, Koripella SC, Mitra A, Rajendran VB, Sinha B (2007) J Biosci 32:809–825
23. Roy A, Panigrahi S, Bhatatcharyya M, Bhattacharyya D (2008) J Phys Chem B 112:3786–3796
24. Sharma P, Mitra A, Sharma S, Singh H (2007) J Chem Sci 119:525–531
25. Sponer JE, Leszczynski J, Sychrovsky V, Sponer J (2005) J Phys Chem B 109:18680–18689
26. Sponer JE, Spackova N, Kulhanek P, Leszczynski J, Sponer J (2005) J Phys Chem B 109:2292–2301
27. Sponer JE, Spackova N, Leszczynski J, Sponer J (2005) J Phys Chem B 109:11399–11410
28. Sharma P, Mitra A, Sharma S, Singh H, Bhattacharyya D (2008) J Biomol Struct Dyn 25:709–732
29. Sharma P, Singh H, Mitra A (2008) Lect Notes Comput Sci 5102:379–386
30. Sharma P, Sponer JE, Sponer J, Sharma S, Bhattacharyya D, Mitra A (2010) J Phys Chem B 114:3307–3320
31. Sharma P, Chawla M, Sharma S, Mitra A (2010) RNA 16:942–957
32. Yildirim I, Turner DH (2005) Biochemistry 44:13225–13234

33. Sponer J, Sponer JE, Petrov AI, Leontis NB (2010) J Phys Chem B 114:15723–15741
34. Harada F, Nishimura S (1972) Biochemistry 11:301–308
35. Farkas WR (1983) Nucleosides Nucleotides 2:1–20
36. Nishimura S (1983) Prog Nucleic Acid Res Mol Biol 28:49–80
37. Balasubramanian R, Seetharamulu PA (1981) J Mol Evol 17:27–30
38. Andachi Y, Yamao F, Iwami M, Muto A, Osawa S (1987) Proc Natl Acad Sci USA 84:7398–7402
39. Leontis NB, Stombaugh J, Westhof E (2002) Nucleic Acids Res 30:3497–531
40. Leontis NB, Westhof E (2001) RNA 7:499–512
41. Bakalarski G, Grochowski P, Kwiatkowski JS, Lesyng B, Leszczynski J (1996) Chem Phys 204:301–311
42. Becke AD (1993) J Chem Phys 98:5648–5652
43. Lee C, Yang W, Parr RG (1993) Phys Rev B 37:785–789
44. Miertus S, Scrocco E, Tomasi J (1981) Chem Phys 55:117–129
45. Scott AP, Radom L (1996) J Phys Chem 100:16502–16513
46. Frisch MJ et al. (2009) Gaussian 09, Revision A.1. Gaussian Inc, Wallingford
47. Venkateswarlu D, Lyngdoh RHD (1995) J Chem Soc Perkin Trans 2:839–846
48. Venkateswarlu D, Lyngdoh RHD, Bansal M (1997) J Chem Soc Perkin Trans 2:621–625
49. El Hassan MA, Calladine CR (1996) J Mol Biol 259:95–103
50. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
51. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
52. Moras D, Comarmond MB, Fischer J, Weiss R, Thierry JC, Ebel JP, Giege R (1980) Nature 288:669–674
53. Murphy FV IV, Ramakrishnan V (2004) Nat Struct Mol Biol 11:1251–1252
54. Varani G, McClain WH (2000) EMBO Rep 1:18–23
55. Ferre-D Amare AR, Zhou K, Doudna JA (1998) Nature 395:567–574
56. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) Science 289:905–920
57. Hung LW, Holbrook EL, Holbrook SR (2000) Proc Natl Acad Sci USA 97:5107–5112
58. Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, Rees B, Thierry JC, Moras D (1991) Science 252:1682–1689
59. Pan B, Mitra SN, Sun L, Hart D, Sundaralingam M (1998) Nucleic Acids Res 26:5699–5706
60. Antao VP, Tinoco I Jr (1992) Nucleic Acids Res 20:819–824
61. Adamiak DA, Rypniewski WR, Milecki J, Adamiak RW (2001) Nucleic Acids Res 29:4144–4153
62. Rypniewski WR, Vallazza M, Perbandt M, Klussmann S, Delucas LJ, Betzel C, Erdmann VA (2006) Acta Crystallogr D: Biol Crystallogr 62:659–664
63. Hobza P, Sponer J (1999) Chem Rev 99:3247–3276
64. Sponer J, Florian J, Leszczynski J, Hobza P (1996) J Biomol Struct Dyn 13:827–833
65. Sponer J, Leszczynski J, Hobza P (1996) J Phys Chem 100:1965–1974
66. Florian J, Hrouda V, Hobza P (1994) J Am Chem Soc 116:1457–1460