Supplementary Material

for

Classification and energetics of the base-phosphate interactions in RNA

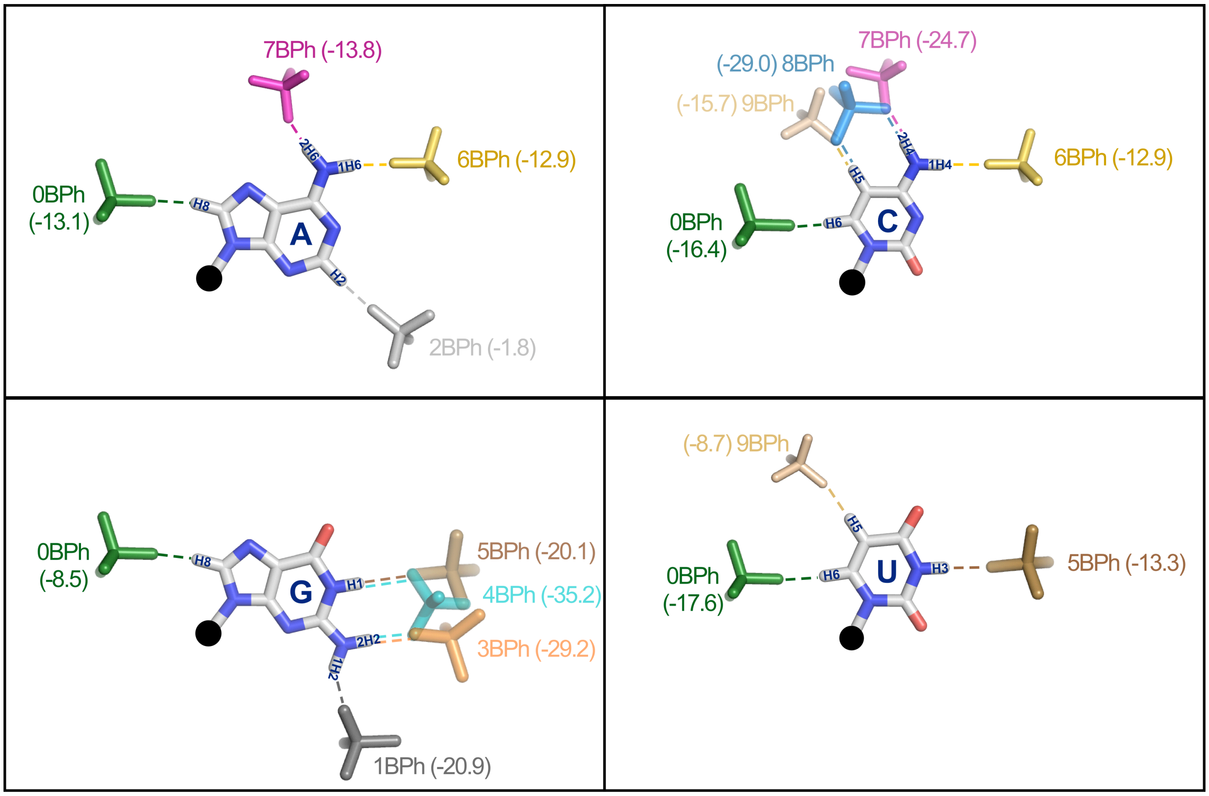
Craig L. Zirbel[1,4], Judit E. Šponer[5], Jiri Šponer[5], Jesse Stombaugh[2,4], and Neocles B. Leontis[2,4,*]

[1]Department of Mathematics and Statistics, [2]Department of Chemistry, and [3]Center for Biomolecular Sciences, Bowling Green State University, Bowling Green, OH 43403 USA, [4]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

*To whom correspondence should be addressed
Telephone: (419) 372-8663
Fax: (419) 372-9809
Email: leontis@bgsu.edu

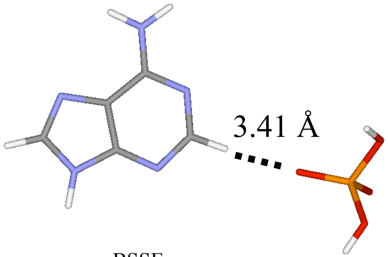
In addition to this pdf file, the supplementary material consists of one additional file:

1. Zirbel_et_al_Sup_Mat_S2.xls



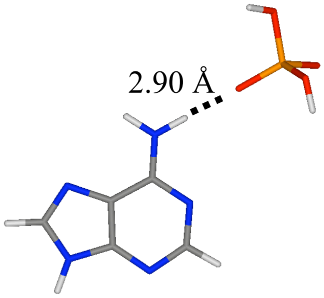
Supplemental Materials S1. Gas-phase interaction energy data (kcal/mol) for the base-phosphate interactions computed at RIMP2/aug-cc-pVDZ level of theory.

2BPh



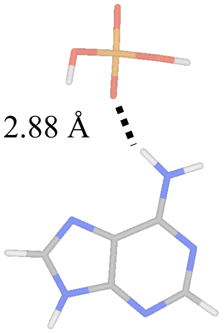
ΔE_{COSMO}^{BSSE} = -0.1 kcal/mol

6BPh



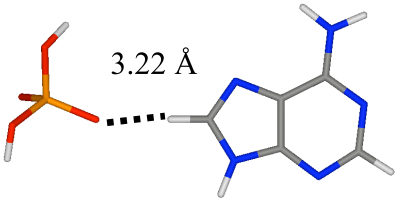
ΔE_{COSMO}^{BSSE} = -3.1 kcal/mol

7BPh



ΔE_{COSMO}^{BSSE} = -2.8 kcal/mol

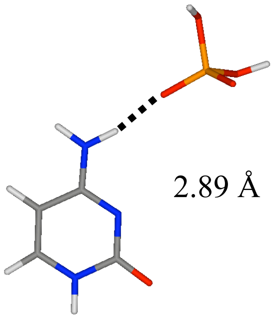
0BPh



ΔE_{COSMO}^{BSSE} = -1.1 kcal/mol

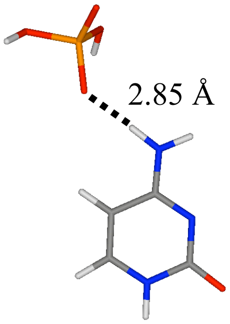
Supplemental Materials S3. Optimized geometries of the four base-phosphate binding patterns found for adenine. Geometry optimizations were carried out at the B3LYP level of theory using the 6-31G** basis set augmented with diffuse functions on phosphorus and the anionic oxygens of the phosphate. The COSMO dielectric continuum method (ε=78.4) was applied throughout the calculations. The dotted lines indicate H-bonding contacts between the base and the phosphate.

6BPh



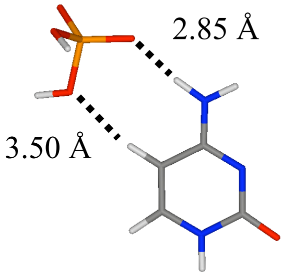
$\Delta E_{\mathrm{COSMO}}^{\mathrm{BSSE}} = -3.5$ kcal/mol

7BPh



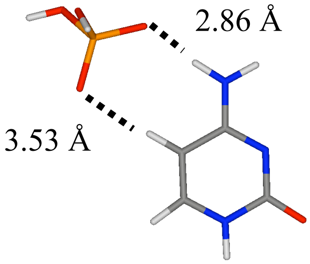
$\Delta E_{\mathrm{COSMO}}^{\mathrm{BSSE}} = -4.8$ kcal/mol

8BPh/1



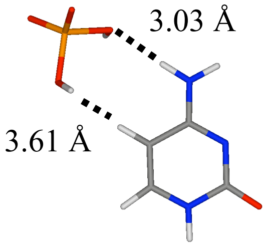
$\Delta E_{\mathrm{COSMO}}^{\mathrm{BSSE}} = -5.4$ kcal/mol

8BPh/2



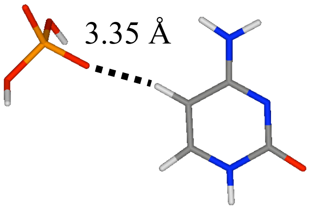
$\Delta E_{\mathrm{COSMO}}^{\mathrm{BSSE}} = -5.6$ kcal/mol

8BPh/3



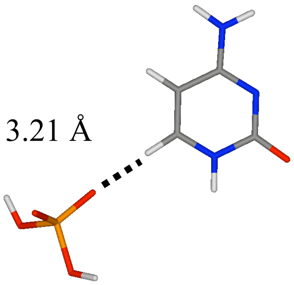
$\Delta E_{\mathrm{COSMO}}^{\mathrm{BSSE}} = -2.5$ kcal/mol

9BPh



$\Delta E_{\mathrm{COSMO}}^{\mathrm{BSSE}} = -0.6$ kcal/mol

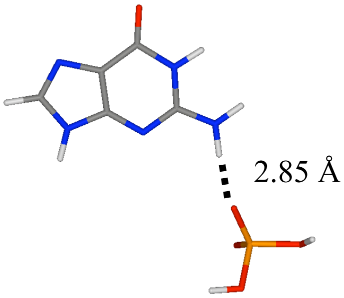
0BPh



$\Delta E_{\mathrm{COSMO}}^{\mathrm{BSSE}} = -1.0$ kcal/mol

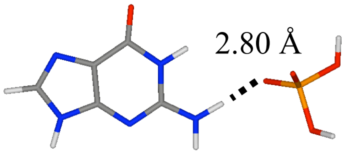
Supplementary Materials S4. Optimized geometries of the base-phosphate binding patterns obtained for cytosine. Geometry optimizations were carried out at the B3LYP level of theory using the 6-31G** basis set augmented with diffuse functions on phosphorus and the anionic oxygens of the phosphate. The COSMO dielectric continuum method (ε=78.4) was applied throughout the calculations. The dotted lines indicate H-bonding contacts.

1BPh



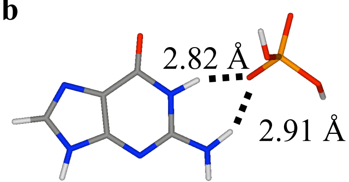
$\Delta E_{COSMO}{}^{BSSE}$= -4.3 kcal/mol

3BPh



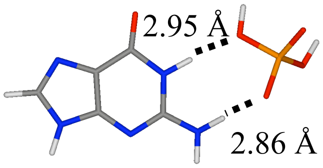
$\Delta E_{COSMO}{}^{BSSE}$ = -5.4 kcal/mol

4BPb



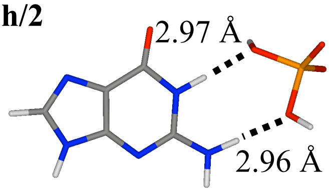
$\Delta E_{COSMO}{}^{BSSE}$ = -8.8 kcal/mol

4BPh/1



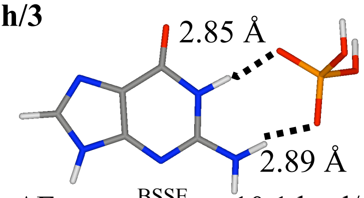
$\Delta E_{COSMO}{}^{BSSE}$ = -7.8 kcal/mol

4BPh/2



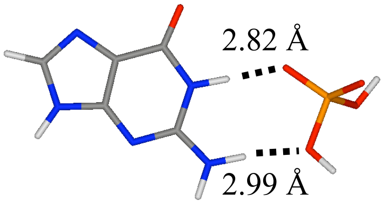
$\Delta E_{COSMO}{}^{BSSE}$ = -5.3 kcal/mol

4BPh/3



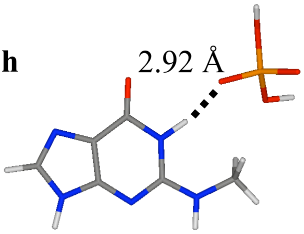
$\Delta E_{COSMO}{}^{BSSE}$ = -10.1 kcal/mol

4BPh/4



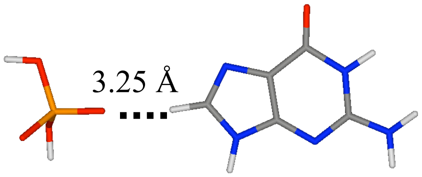
$\Delta E_{COSMO}{}^{BSSE}$= -8.3 kcal/mol

5BPh



$\Delta E_{COSMO}{}^{BSSE}$= -4.0 kcal/mol

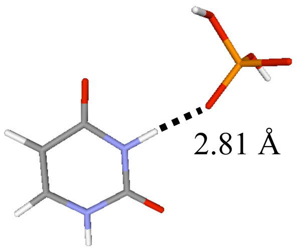
0BPh



$\Delta E_{COSMO}{}^{BSSE}$ = -1.1 kcal/mol

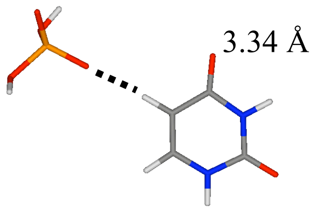
Supplemental Materials S5. Optimized geometries of the base-phosphate binding patterns found for guanine. Geometry optimizations were carried out at the B3LYP level of theory using the 6-31G** basis set augmented with diffuse functions on phosphorus and the anionic oxygens of the phosphate. The COSMO dielectric continuum method (ε=78.4) was applied throughout the calculations. The dotted lines indicate H-bonding contacts between the base and the phosphate.

5BPh



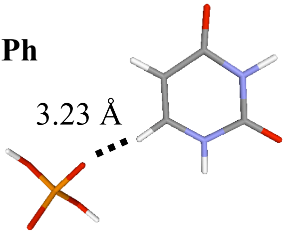
ΔE_{COSMO}^{BSSE} = -4.2 kcal/mol

9BPh



ΔE_{COSMO}^{BSSE} = -0.6 kcal/mol

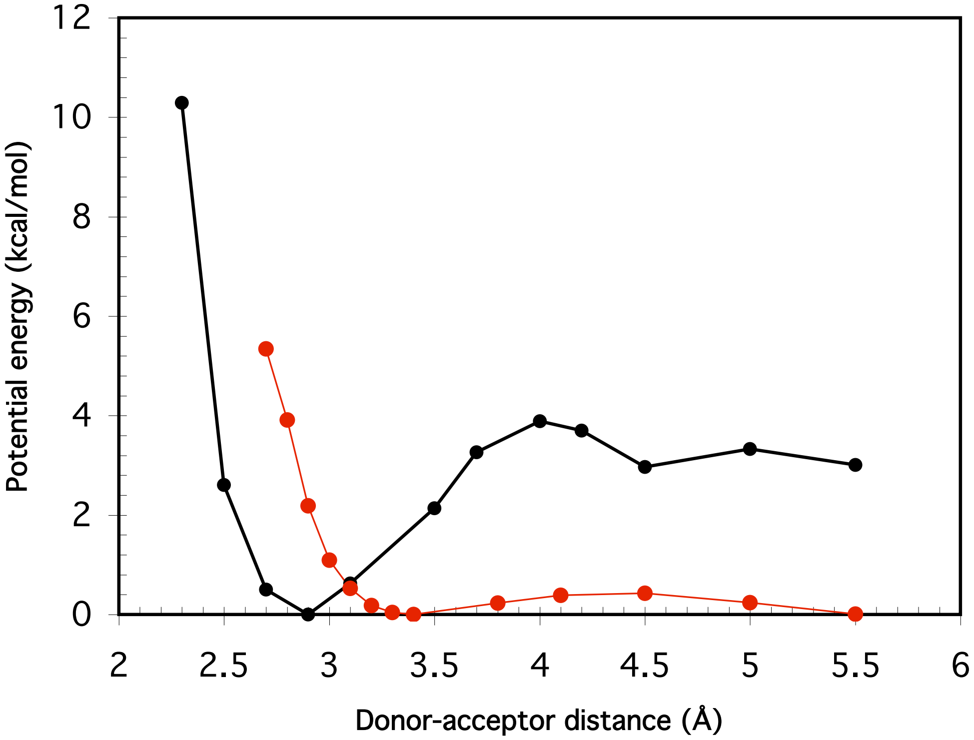
0BPh



ΔE_{COSMO}^{BSSE} = -1.1 kcal/mol

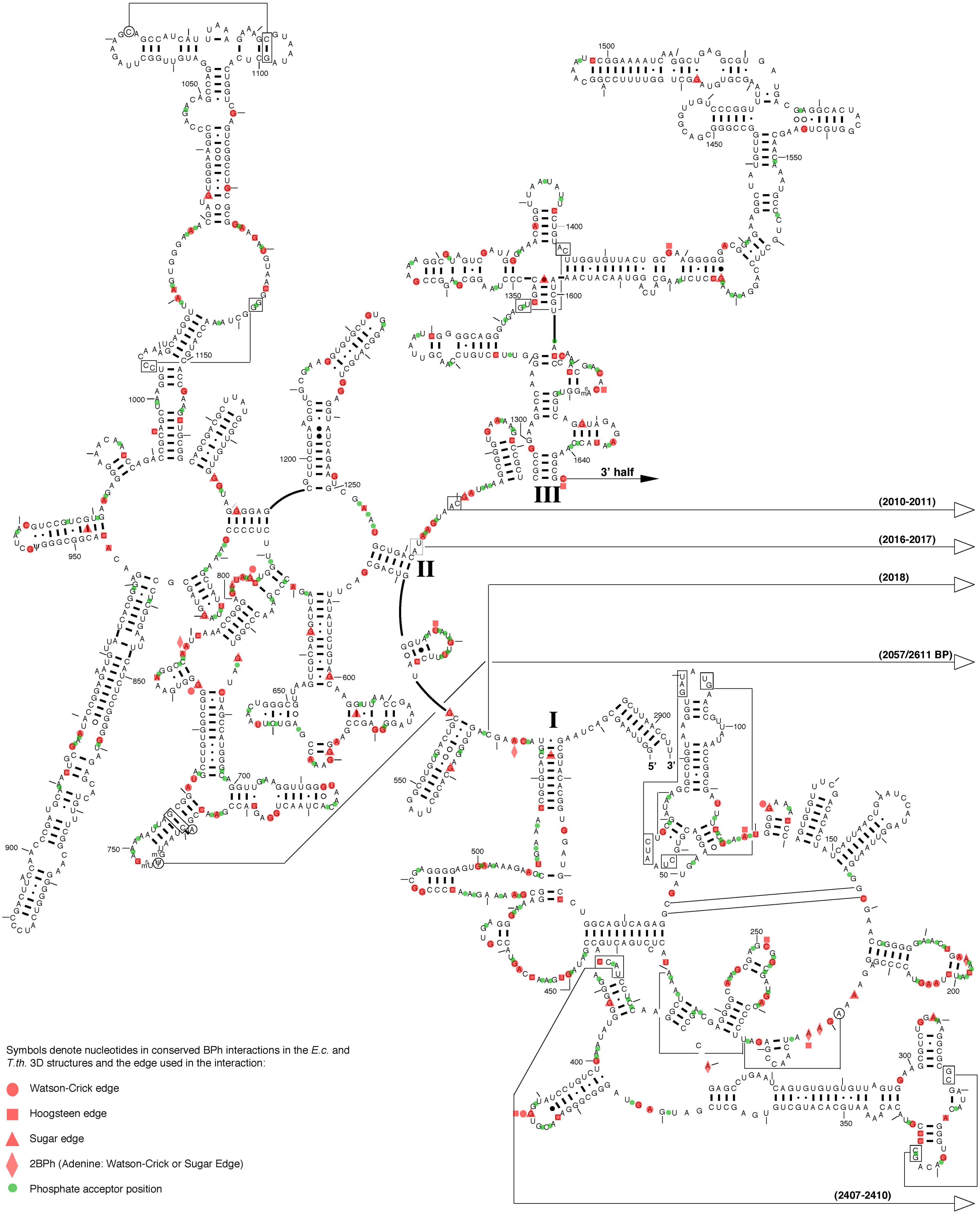
Supplemental Materials S6. Optimized geometries of the three base-phosphate binding patterns found for uracil. Geometry optimizations were carried out at the B3LYP level of theory using the 6-31G** basis set augmented with diffuse functions on phosphorus and the anionic oxygens of the phosphate. The COSMO dielectric continuum method (ε=78.4) was applied throughout the calculations. The dotted lines indicate H-bonding contacts between the base and the phosphate.

Supplement to Section 2: Mapping the potential energy hypersurface at the optimized minima. The optimized geometry of a model corresponds to a local energy minimum on the potential energy hypersurface of the studied system. However, for non-covalent interactions, including H-bonds, other arrangements belonging to the low-energy potential energy surface can be attained with minor energy penalties besides the geometry of the exact local minimum. To examine the curvature of the potential energy hypersurface near the minimum energy geometries, we carried out one-dimensional potential energy searches along the coordinate representing the distance between the H-bond donor and acceptor atoms for each of two binding modes: the 6BPh interaction of cytosine, to represent nitrogen H-bond donors, and the 2BPh interaction of adenine, to represent carbon H-bond donors. **Supplemental Materials S7** reveals that decreasing the H-bond distances causes the total energy to rapidly increase due to the obvious short-range exchange repulsion of the interacting atoms, independently of the type of H-bond donor atom. Shortening the donor-acceptor distance by just 0.5 Å results in about 2 kcal/mol destabilization for both H-bonds. The N-H…O interaction is associated with a visibly shorter optimal donor-acceptor distance, as expected. However, upon elongation of the donor-acceptor distance, the N-H...O(Phosphate) and C-H...O(Phosphate) H-bonds behave differently. The N-H...O(Phosphate) H-bond is more rigid, as an 0.5 Å elongation of this H-bond requires about 2 kcal/mol energy investment. On the contrary, the C-H...O(Phosphate) H-bond is very flexible and the C-O(Phosphate) distance can be elongated with basically no energy penalty. Note that the energy gradient associated with elongation of the donor-acceptor distance can be fairly well assessed in **Table 1**, since the interaction energy basically determines the depth of the curve in **Supplemental Materials S7**. When the donor-acceptor separation increases by ~1Å compared to the minimum, the direct attractive force basically vanishes, as short-range attractive contributions fall out and the electrostatic attraction is eliminated by the solvent screening. This is a common behavior of ionic interactions upon aqueous solvation (1). The results mean that when compressed geometries of BPh interactions are observed in x-ray crystal structures, they can be attributed to uncertainty of the experimental data, given the large energy gradient associated with such a donor-acceptor distance compression. Similar considerations apply to anomalously extended N…O distances in BPh interactions. In fact, molecular dynamics simulation studies indicate that the N…O BPh interactions, once formed, are quite stiff, as expected for strong H-bonds (2). Thus, we expect that N…O type of BPh interactions can attenuate thermal fluctuations of respective RNA structures or motifs that comprise such interactions, so that formation of strong H-bonds between bases and phosphates may immobilize otherwise flexible local backbone segments. Such structuring may be relevant to molecular recognition of such RNA building blocks by proteins or other ligands. However, we also need to consider the competition between direct and water-mediated interactions. This was not investigated in this study, since the actual balance of such interactions may depend on the context of the interaction. On the other hand, the weakest C-H…O BPh interactions can be very flexible, although a more detailed characterization would require carrying out molecular dynamic simulations.



Supplemental Materials S7. One dimensional scan of the potential energy surface for the 6BPh binding mode of cytosine along the N4-O(Phosphate) distance coordinate (black curve) and for the 2BPh binding mode of adenine along the C2-O(Phosphate) distance coordinate (red curve). The energies of the optimal geometries were shifted to 0 kcal/mol to highlight the gradients of energies associated with compression and extension of the interaction.

23S *Escherichia coli* - 5' end

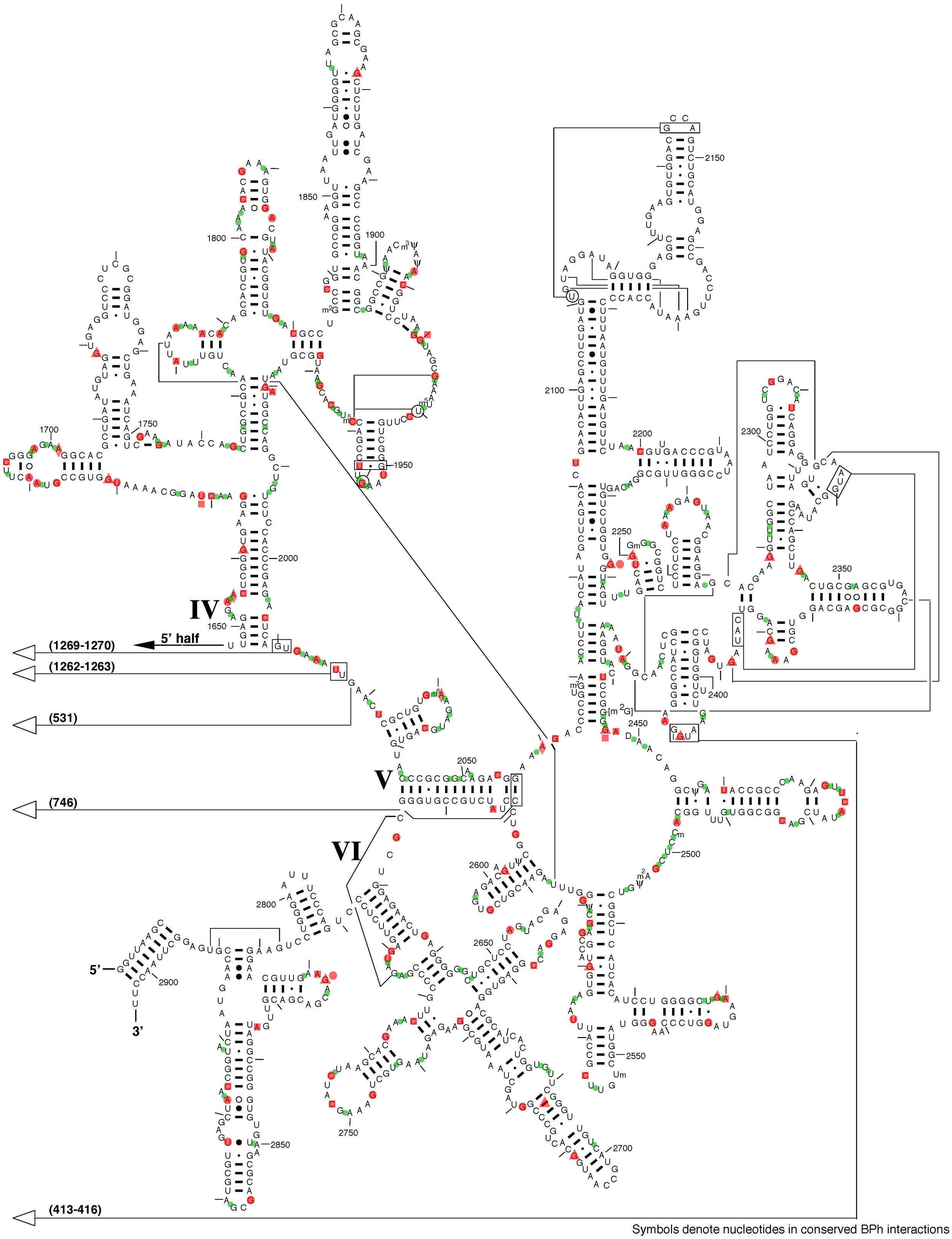


Symbols denote nucleotides in conserved BPh interactions in the *E.c.* and *T.th.* 3D structures and the edge used in the interaction:

- ● Watson-Crick edge
- ■ Hoogsteen edge
- ▲ Sugar edge
- ♦ 2BPh (Adenine: Watson-Crick or Sugar Edge)
- • Phosphate acceptor position

Supplemental Materials S8. BPh interactions conserved between *E.c.* and *T.th.* rRNA 3D structures mapped on the 2D structure of *E.c.* 23S rRNA – 5’ end (3). Red symbols were used to denote the edge used by each base donor (circle for Watson-Crick edge, square for Hoogsteen edge, triangle for Sugar edge and diamond for the Adenine 2BPh which straddles the WC and Sugar edges). The 1BPh interactions that are conserved at the base pair level are marked by red triangles placed between the bases forming the WC base pair. Green circles denote the locations of phosphate acceptors. They are placed on the 5' side of the letter for the corresponding nucleotide.

23S *Escherichia coli* - 3' end

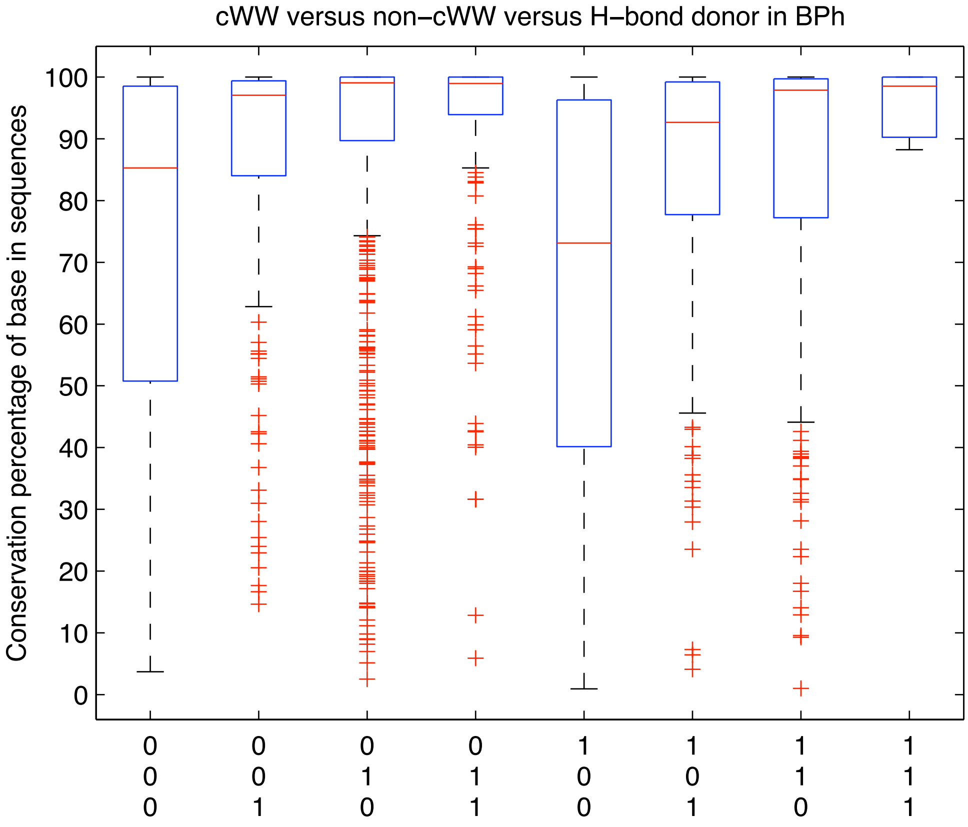


Symbols denote nucleotides in conserved BPh interactions in the *E.c.* and *T.th.* 3D structures and the edge used in the interaction:

- ● Watson-Crick edge
- ■ Hoogsteen edge
- ▲ Sugar edge
- ◆ 2BPh (Adenine: Watson-Crick or Sugar Edge)
- ● Phosphate acceptor position

Supplemental Materials S9. BPh interactions conserved between *E.c.* and *T.th.* rRNA 3D structures mapped on the 2D structure of *E.c.* 23S rRNA – 3' end (3). Red symbols were used to denote the edge used by each base donor (circle for Watson-Crick edge, square for Hoogsteen edge, triangle for Sugar edge and diamond for the Adenine 2BPh which straddles the WC and Sugar edges). The 1BPh interactions that are conserved at the base pair level are marked by red triangles placed between the bases forming the WC base pair. Green circles denote the locations of phosphate acceptors. They are placed on the 5' side of the letter for the corresponding nucleotide.

Supplement to Section 5.3 – Base conservation. We can model the effect on base conservation of interactions made by a nucleotide in another way. **Supplemental Materials S10** presents box-plots of the conservation percentages for eight disjoint subsets of nucleotides. The subsets are defined by three binary attributes: (1) whether the nucleotide makes a cWW pair (0 or 1 in the first row below the box-plots), (2) whether the nucleotide makes at least one non-cWW pair (0 or 1 in the second row), and (3) whether the nucleotide is the H-bond donor in at least one BPh interaction (0 or 1 in the third row). These attributes define eight disjoint sets of nucleotides from 3D structures of *E.c.* 16S and 23S rRNA. Comparing box-plots 1 against 2, 5 against 6, and 7 against 8, we see that in each case, the subset with at least one BPh interaction has a higher median conservation percentage, higher first quartile, and higher minimum. Comparing box-plot 3 against 4, the medians are both high and roughly the same, but the first quartile is higher for the case with the BPh interaction. Comparing box-plot 5 against 6 shows the most dramatic effect. These are nucleotides which make one cWW pair and no non-cWW pair. Those which also make a BPh interaction (box-plot 6) are much more conserved than those which do not (box-plot 5).



Supplemental Materials S10. Base conservation percentage for various disjoint subsets of nucleotides in E.c. rRNA 3D structures (2avy and 2aw4), as calculated from sequence alignments. Each box-plot has three numbers below it. The first is the number of cWW base pairs the nucleotide makes, the second is whether (1) or not (0) the nucleotide makes one or more non-cWW base pairs, and the third is whether (1) or not (0) the base is an H-bond donor in any BPh interaction. The horizontal red line in each box-plot is the median conservation percentage for the subset of nucleotides having these three characteristics. Horizontal blue lines indicate the first and third quartile. Horizontal black lines are placed at 1.5 times the inter-quartile range above and below the nearest quartiles, but not above or below the maximum and minimum of the data points. Horizontal red lines indicate outliers, data points beyond the 1.5 times the inter-quartile range from the nearest quartiles.

REFERENCES

1. No, K.T., Nam, K.-Y. and Scheraga, H.A. (1997) Stability of Like and Oppositely Charged Organic Ion Pairs in Aqueous Solution. *Journal of the American Chemical Society*, **119**, 12917-12922.
2. Spackova, N. and Sponer, J. (2006) Molecular dynamics simulations of sarcin-ricin rRNA motif. *Nucleic Acids Res*, **34**, 697-708.
3. Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M. *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC bioinformatics*, **3**, 2.