JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

Modeling of the Water Network at Protein−RNA Interfaces

Yiyu Li,[†] Brian T. Sutch,[†,∥] Huynh-Hoa Bui,[‡] Timothy K. Gallaher,[§] and Ian S. Haworth[†,*]

[†]Department of Pharmacology & Pharmaceutical Sciences, University of Southern California, Los Angeles, California 90089-9121, United States

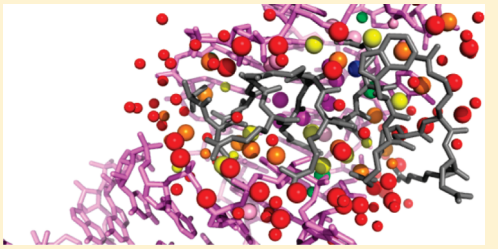
[‡]Isis Pharmaceuticals Inc., 1896 Rutherford Road, Carlsbad, California 92008-7326, United States

[§]Neumedicines Inc., 2708 Foothill Boulevard, Suite #403, La Crescenta, California 91214, United States

[∥]Ligand Insight, LLC, Post Office Box 5731, Pasadena, California 91117, United States

Supporting Information

ABSTRACT: Water plays an important role in the mediation of biomolecular interactions. Thus, accurate prediction and evaluation of water-mediated interactions is an important element in the computational design of interfaces involving proteins, RNA, and DNA. Here, we use an algorithm (WATGEN) to predict the locations of interfacial water molecules for a data set of 224 protein−RNA interfaces. The accuracy of the prediction is validated against water molecules present in the X-ray structures of 105 of these complexes. The complexity of the water networks is deconvoluted through definition of the characteristics of each water molecule based on its bridging properties between the protein and RNA and on its depth in the interface with respect to the bulk solvent. This approach has the potential for scoring the water network for incorporation into the computational design of protein−RNA complexes.



BACKGROUND

Water plays an important role in biomolecular association. Water molecules at protein−protein and protein−oligonucleotide interfaces form extensive hydrogen-bonded networks and facilitate the formation and dissociation of these complexes.[1−3] Water molecules influence molecular recognition, specificity, and affinity of protein complexes with other proteins[4−6] and with DNA,[7−10] RNA,[10,11] and sugars.[12] Water is also important in the interaction of proteins with small ligands,[13] and recognition of the hydration status of a protein binding pocket may guide optimal drug design.[14,15] It is also apparent that not all water molecules at an interface are equivalent: some show rapid exchange with the bulk, while others are relatively tightly bound,[3] and some water molecules bound to a protein surface prior to ligand binding are retained in complexes with small molecules[14] and proteins.[4] These observations suggest the need for a method of classification of water molecules in the complex network formed at protein interfaces.

Solvation of a protein surface has been predicted computationally on the basis of X-ray data,[16,17] molecular dynamics simulations,[3,18,19] and grid-based simulation.[20] Key "wet spots" have been identified as critical features of protein−protein interactions,[21] and the inclusion of water within docking algorithms is likely to increase the accuracy of predicted structures.[1] van Dijk and Bonvin[22] addressed this issue by mimicking the exclusion of water molecules during the formation of an interface, while the solvated rotamer approach of Jiang et al.[23] permits protein interface design with inclusion of water-mediated hydrogen bonds. We have also described an algorithm, WATGEN, for rapid solvation of a protein−protein interface using optimization of the hydrogen-bonded water−protein network, which we validated against data for water positions from X-ray structures.[24]

The structural details of protein−RNA interfaces are emerging, based on analyses[25−27] of the growing number of solved structures. Treger and Westhof[28] first pointed out the key role of water at the protein−RNA interface, and Bahadur et al.[26] proposed that an average protein−RNA interface contains 32 water molecules, on the basis of X-ray data. The detailed findings for protein−protein complexes and the emerging evidence for protein−RNA interfaces indicates that rational design of these interfaces must take into account the "fit" of the water network at the predicted interface. This requires prediction of this water network and classification of the binding properties of each water molecule in the network. In this study, we used the WATGEN algorithm to predict water networks at protein−RNA interfaces. On the basis of these results, we describe a method for classification of the water molecules in the network.

METHODS

Summary of WATGEN Algorithm. WATGEN has been described in detail and validated for solvation of protein−peptide interfaces.[24] Calculation of the water network at a protein−RNA interface was performed by the same method, with modifications only to add geometry and atom information for RNA. Briefly,

Received: March 9, 2011
Published: May 25, 2011

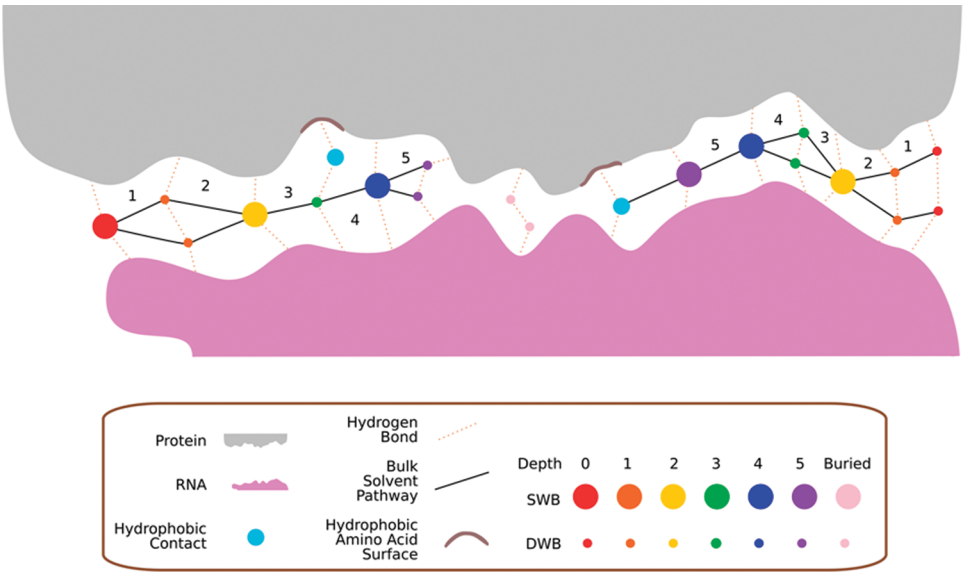


Figure 1. Classification of the water network at a protein−RNA interface. The protein is shown in gray and the RNA in violet. Water molecules fill the interface and are represented by spheres of different colors and radii. Numbers represent the depth of the water molecule from the bulk solvent (not depicted). WATGEN calculates the first level of water molecules, which are in direct contact with the bulk solvent. These are colored red for depth 0. Water molecules making contact with red (depth 0) water molecules are colored orange (depth 1) and this process is repeated until all water molecules are classified. Water molecules that do not make contact with the bulk solvent directly or indirectly via other water molecules are considered buried and are colored pink. WATGEN adds water molecules to empty regions of the interface while avoiding clashes, and these occasionally result in water contacts with hydrophobic surfaces. These water molecules are colored light blue. The sphere radius depends upon the hydrogen-bonding pattern. Single water bridges are shown as large spheres, double water bridges as small spheres, and hydrophobic bridges as intermediate spheres. Dashed lines represent hydrogen bonds between water molecules and RNA or protein. Black solid lines represent hydrogen bonds between water molecules.

after addition of hydrogen atoms to the protein and RNA in standard geometries, the water molecules at the interface are calculated in WATGEN in four steps:[24] (1) water sites defined by the oxygen atom of each water molecule are distributed around hydrogen-bonding centers (donors and acceptors) on both sides of the interface; (2) the oxygen sites are classified and may be combined on the basis of their distance to other oxygen sites or hydrogen-bonding sites in the interface; (3) the best sites are selected on the basis of maximization of potential hydrogen-bonding interactions (without hydrogen atoms at this stage) and minimization of van der Waals clashes; and (4) hydrogen atoms are added to each O atom in a geometry that optimizes the number of hydrogen-bonding contacts.

Protein−RNA Complexes. A set of 224 protein−RNA complexes (Supporting Information, Table S1) were solvated by use of the WATGEN algorithm. These complexes were derived from an initial set of 541 PDB files containing protein−RNA complexes, which were downloaded from the RCSB database[29] by use of the advanced search facility with "Macromolecule Type" set to "Contains Protein = Yes, RNA = Yes, DNA = No, DNA/RNA Hybrid = Ignore" and "Text Search" set to "PROTEIN OR RNA NOT RIBOSOME NOT RIBOSOMAL". These PDB files were reduced to 394 unique binary protein−RNA complexes (one protein and one RNA molecule) by use of the SIMA algorithm, which computes a similarity score for pairwise comparison of two protein−RNA interfaces.[30] Binary complexes from a single PDB file were then recombined in two stages to create contiguous interfaces: interacting RNA chains were identified and combined and then all protein chains with an interaction with the combined RNA chain were included. This gave 292 complexes, which were further reduced by elimination of 68 complexes with nonnatural RNA nucleotides at the interface (within 20 Å of any amino acid). This left 224 protein−RNA complexes for analysis in WATGEN. The RNA chain was specified as the "ligand" in WATGEN. Solvation was performed with a water box that extended 6 Å from the minimum and maximum x, y, and z coordinates of the ligand.[24]

The CPU time for computation of the 224 water networks was 2 h 3 min 53 s on a dual quadcore Intel Xeon 2.33 GHz (12GB memory) node on a Linux cluster. The average CPU time was 33.2 s/complex, and the average interface had 61.3 amino acids and 22.0 nucleotides. WATGEN can be used at http://rockscluster.hsc.usc.edu/research/software/watgen/watgen.html.

Algorithm Validation. WATGEN has been validated in detail for solvation of protein−peptide interfaces[24] and the core of the algorithm used here is unchanged. However, to show that WATGEN solvates protein−RNA interfaces accurately, we compared the predicted water networks with water sites determined by X-ray crystallography. Of the 224 protein−RNA complexes used in the study, 105 contain experimentally determined water sites, and these complexes were used in the comparison. Experimental water molecules were mapped to the closest predicted water site. The predictive power of the algorithm was assessed by determining the proportion of experimental waters with a proximal predicted water site. As a negative control, a randomized water network was calculated for each interface by placing the same number of predicted interface water molecules in the same spatial volume at grid points 2.6 Å apart.[24] Statistical analysis was performed via linear regression analysis and a Wilcoxon rank-sum test. All analyses were performed in SAS 9.2 (SAS Institute Inc., Cary, NC) with $P < 0.05$ considered to indicate significance.

Classification of Interfacial Water Molecules. An understanding of the complexity of the water network at a biomolecular

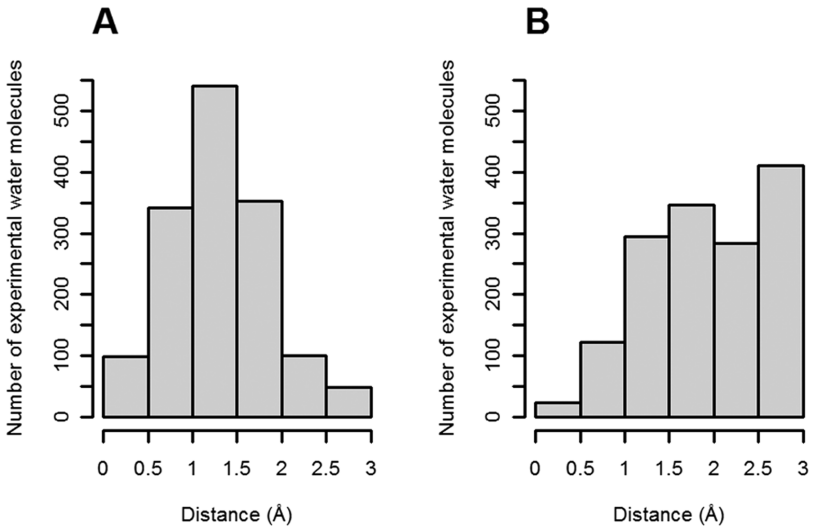


Figure 2. Numbers of experimental water molecules within a specific distance (0.5−3.0 Å) from predicted water sites in (A) the WATGEN-predicted network and (B) a random network. Data are calculated on the basis of O−O distances between experimental and predicted or random water sites. This figure was made with R 2.12.1.[31].

interface requires classification of the location of each water molecule. To establish this classification, we defined each water molecule on the basis of its interactions with the protein and RNA and its positional relationship with bulk water (Figure 1).

A water molecule forming hydrogen bonds with both protein and RNA is defined as a single-water bridge, and a water molecule with hydrogen bonds to one molecule (protein or RNA) of the interface and to another water that is in turn hydrogen-bonded to the other molecule (RNA or protein) is defined as participating in a double-water bridge. The WATGEN algorithm completes the water network by inserting water molecules into free space in the interface (as defined by the absence of clashes), which results in some water molecules having contacts with hydrophobic surfaces in the interface. Such water molecules are categorized as forming a single or double hydrophobic bridge (that is, at least one contact with the protein or RNA is not made through a hydrogen bond). Relatively few of these water molecules are present in the interface.

The position of a water molecule with respect to the bulk solvent is defined by two criteria (Figure 1). First, if the water molecule has a path to the bulk solvent (free space in the calculated system) that can be traced through hydrogen bonds, it is considered to be exchangeable with the bulk. The depth of each exchangeable water molecule is evaluated with respect to the bulk. Second, a water molecule may be trapped in a cavity with no hydrogen-bonded pathway to the bulk and is then defined to be buried.

RESULTS

Algorithm Validation. Of the 224 protein−RNA complexes solvated by WATGEN (Supporting Information, Table S1), a subset of 105 complexes with experimental water molecules was used to validate the algorithm. The predicted water networks for these complexes were compared with the experimental water sites. Of the 1481 experimental water sites, WATGEN predicts 981 (66.2%) and 1332 (89.9%) within 1.5 and 2.0 Å, respectively (Figure 2A). One difficulty with interpretation of these data is that the number of water molecules in the 105 predicted water networks is much higher than that found experimentally. We have shown[24] that WATGEN does not overpredict the number of water molecules, based on the less favorable energies of networks containing fewer or more water molecules compared to the networks obtained with the WATGEN conditions used in the present and previous work.[24]

The difference from experimental data is due to the difficulty of identifying electron density for relatively mobile water molecules in X-ray structures. However, the large number of predicted sites could account for the good agreement with fewer experimental sites simply by chance. To address this issue, we calculated a random water network for each complex comprising the same number of water molecules in the same interface volume. In the random networks, only 441 (29.8%) and 787 (53.1%) of water molecules were within 1.5 and 2.0 Å of the 1481 experimental sites, respectively (Figure 2B). For each experimental water site, we calculated the distances (O to O) to the closest water molecule in the random network and the closest water molecule in the WATGEN-predicted network. The median of these distances over the 1481 water sites was significantly smaller for the WATGEN-predicted network ($P < 0.001$ by Wilcoxon rank-sum test; Supporting Information, Figure S1).

Overview of the Water Networks. Linear regression analysis showed that the number of predicted water molecules (*n*Wat in Supporting Information, Table S1) in the 224 protein−RNA interfaces is significantly related to the size of the interface: $p < 0.0001$ and $R^2 = 0.9854$ for the model

$$n\text{Wat} = -0.553 + 2.448n\text{Intfaa} + 3.315n\text{Intfnt} \quad (1)$$

where *n*Wat is the number of predicted water molecules in the protein−RNA interface, *n*Intfaa is the number of amino acids in the protein−RNA interface (the number of amino acids with direct or water-mediated interactions with the RNA), and *n*Intfnt is the number of nucleotides in the protein−RNA interface (the number of nucleotides with direct or water-mediated interactions with the protein) (Supporting Information, Figure S2). The estimated slopes of *n*Intfaa and *n*Intfnt are both significantly different from zero ($p < 0.0001$), with 95% confidence limits of (2.376, 2.521) and (3.003, 3.627), respectively, and the estimated

intercept is not significantly different from zero (p = 0.8524). This model suggests that the mean number of predicted water molecules in the protein−RNA interface increases by about 2.4 per increase of one amino acid in the interface and by about 3.3 per increase of one nucleotide in the interface. This regression model was generated from a high-quality data set in which duplicate interfaces were excluded.[30] The results of jackknife error analysis of the model are shown in Table S2 (Supporting Information). The parameter estimates from the predictive model are consistent with the jackknife estimates.

Classification of Water Molecules in the Network. Each water molecule in the predicted water networks was classified (Figure 1) on the basis of its bridging location between the protein and RNA molecules as a hydrogen-bonded single-water bridge (SWB), as part of a hydrogen-bonded double-water bridge (DWB), or as a water molecule with contact with a hydrophobic surface (HPHOB). Water molecules that fulfilled two or more of these definitions were classified on the basis of their most direct interaction (i.e., SWB > DWB > HPHOB). This classification is shown for the 224 protein−RNA interfaces in Table S1 (Supporting Information). As a percentage of the total number of water molecules in each complex, the interfaces contained 33.7% ± 5.4% SWB, 62.3% ± 5.3% DWB, and 4.1% ± 2.7% HPHOB water molecules.

Each water molecule was further classified with respect to its depth from the bulk solvent. Water molecules with a direct hydrogen-bonded path to the bulk were considered to be solvent-accessible (that is, exchangeable with bulk). The depth was then defined by the number of intervening waters between the water molecule and the bulk; therefore, a water molecule in direct hydrogen-bonded contact with the bulk solvent had a depth of 0. Water molecules with no direct hydrogen-bonded path to the bulk were classified as buried. These classifications were overlaid on the SWB, DWB, and HPHOB classifications, as shown for the 224 interfaces in Table S3 (Supporting Information). For SWB molecules, 46.5% ± 12.2% and 34.3% ± 7.4% were at depths of 0 and 1, respectively, and for DWB molecules, 79.7% ± 8.6% and 11.7% ± 4.8% were at these respective depths. Thus, most interface water molecules are close to the bulk solvent.

Images of the Water Network. An example of a classified water network is shown for the 1ULL_B_A interface in Figure 3. This image shows the complex of the HIV-1 rev peptide bound in the major groove of the stem of an RNA hairpin[32] solvated by WATGEN. The classification procedure identified water molecules forming single-water bridges at a depth of five water molecules (six hydrogen bonds) from bulk water and also identified several buried water molecules. The detailed three-dimensional nature of the network is better appreciated by viewing electronically. The computed water networks for the 224 interfaces are available for download (PDB file with a PyMol[33] script) at http://rockscluster.hsc.usc.edu/research/software/watgen/watgen.html.

DISCUSSION

Water is a critical component of biomolecular interfaces, and inclusion of solvation effects is required in rational design and affinity prediction for these interfaces. Accurate placement of water molecules in experimentally determined interfaces is a first step, and the results shown here demonstrate that this can be achieved by the WATGEN algorithm. The derived parameters for the predicted water networks give an indication of the acceptable features that are required for a network at a designed interface. These parameters are likely to be representative of a typical protein−RNA interface, since they were derived from a data set of 224 unique complexes. Therefore, these data can make a significant contribution to design of new biomolecular interfaces for diagnostic and therapeutic purposes.

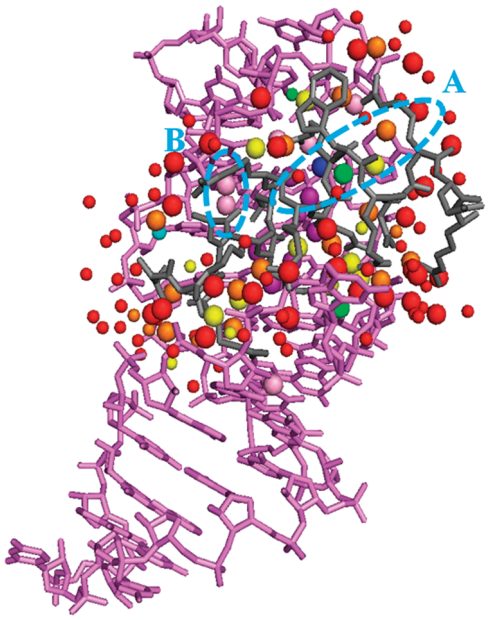


Figure 3. WATGEN solvation of an RNA aptamer complexed with HIV-1 rev peptide (1ULL_B_A). Water molecules are shown as spheres based on the position of the O atom, with the color and size corresponding to the schematic image shown in Figure 1. Ellipse A identifies a pathway of hydrogen-bonded water molecules from red (depth 0, direct interaction with the bulk) to orange, yellow, green, blue, and violet (depth 5). Ellipse B identifies two buried water molecules (shown in pink) that are not connected to bulk solvent directly or indirectly via hydrogen bonding to other water molecules.

Determination of protein−RNA complexes via docking is currently a major challenge. A recent successful approach in this area has been reported by de Vries et al.[34] using the HADDOCK algorithm developed by the same authors.[35] We envisage two potential uses of WATGEN in docking of RNA molecules to proteins (or vice versa). First, the solvated docking approach used by van Dijk and Bonvin[22] may be facilitated by analysis of WATGEN-computed water networks to identify highly conserved water positions that can be included in docking protocols, without dependence on X-ray data for the water location. Second, we have described a motif-based approach to deconvolution of a protein−RNA interface,[30] and establishment of a database of these motifs may allow knowledge-driven docking of RNA molecules to proteins, initially in the absence of water. Predicted structures can then by solvated by use of WATGEN, with the goal of obtaining solvated interfaces with similar properties to those found for experimentally determined protein−RNA interfaces. Computation of the number of water molecules for a particular interface can be achieved by use of eq 1 above. This equation may reflect a fundamental property of solvation of a protein−RNA interface and is likely to be useful for evaluation of simulation and docking results.

The classification of the water network described here provides a further basis for detailed analysis of solvated protein−RNA

complexes determined biophysically or by docking. First, it allows a complete description of the water network in qualitative terms, with identification of potentially important water molecules that form key single-water bridges. Second, since each water molecule has a specific classification, a numerical score for the whole network can be assigned to describe the quality of interface solvation. Enthalpically, the score for each molecule should reflect the assumption that a water molecule should maintain at least four hydrogen bonds in the complex, based on pure water and also on the finding in Schlessman et al.[36] of 3−5 hydrogen bonds for internal water molecules in proteins. Entropically, a penalty can be applied to unfavorable water molecules that are buried in a cavity or positioned deeper from the bulk water. Development of a validated scoring system will require further analysis of the energetics of the water networks.

We briefly note a few technical aspects of WATGEN. First, the algorithm requires that one of the molecules of the interface is defined as the "ligand", since the "water box" within which the calculation is performed is defined with respect to the ligand. In this study, we defined the RNA molecule as the ligand in each complex, but the calculations can equally be performed with the protein defined as the ligand. We performed this set of calculations (data not shown) and found that the algorithm predicts networks that do not differ significantly from the networks obtained with RNA as the ligand, in terms of the numbers of water molecules and the average root-mean-square deviation (rmsd) of the water sites. Second, in this work we used X-ray structures downloaded from the RCSB, since we assume that this is the most likely source of structures for most users. However, these structures often contain breaks in the protein or RNA chain, and computational "repair" of these breaks may be appropriate for use of the algorithm for specific complexes. Third, we excluded nonnatural nucleotides from the present data set, but these will be included in a future version of the program.

It is clear that water plays an essential role within biomolecular interfaces. Where once it was assumed that the effect of water was secondary to the interaction, it is now apparent that water is part of the interaction itself. The WATGEN algorithm provides a rapid method for accurate solvation of protein−RNA and protein−protein interfaces. Further development of a scorable hydration model will provides the basis for use of the water network as an important component of interface design.

■ ASSOCIATED CONTENT

Supporting Information. Three tables and two figures, giving more information about statistics and classification of water molecules at 224 protein−RNA interfaces. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail ihaworth@usc.edu; tel (323) 442-3310; fax (323) 442-1390.

■ ACKNOWLEDGMENT

Computation for the work described in this paper was supported by the University of Southern California Center for High-Performance Computing and Communications (www.usc.edu/hpcc).

■ REFERENCES

(1) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573–581.

(2) Levy, Y.; Onuchic, J. N. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389–415.

(3) Makarov, V.; Pettitt, B. M.; Feig, M. Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment. *Acc. Chem. Res.* **2002**, *35*, 376–384.

(4) Reichmann, D.; Phillip, Y.; Carmi, A.; Schreiber, G. On the contribution of water-mediated interactions to protein-complex stability. *Biochemistry* **2008**, *47*, 1051–1060.

(5) Yokota, A.; Tsumoto, K; Shiroishi, M.; Kondo, H; Kumagai, I. The role of hydrogen bonding via interfacial water molecules in antigen-antibody complexation. *J. Biol. Chem.* **2003**, *278*, 5410–5418.

(6) Hayes, J. M.; Skamnaki, V. T.; Archontis, G.; Lamprakis, C.; Sarrou, J.; Bischler, N.; Skaltsounis, A. L.; Zographos, S. E.; Oikonomakos, N. G. Kinetics, in silico docking, molecular dynamics, and MM-GBSA binding studies on prototype indirubins, KT5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: the role of water molecules examined. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 703–719.

(7) Luo, X.; Lv, F.; Pan, Y.; Kong, X.; Li, Y; Yang, Q. Structure-based prediction of the mobility and disorder of water molecules at protein-DNA interface. *Protein Pept. Lett.* **2011**, *18*, 203–209.

(8) Robinson, C. R.; Sligar, S. G. Changes in solvation during DNA binding and cleavage are critical to altered specificity of the EcoRI endonuclease. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2186–2191.

(9) Duan, J.; Nilsson, L. The role of residue 50 and hydration water molecules in homeodomain DNA recognition. *Eur. Biophys. J.* **2002**, *31*, 306–316.

(10) Sonavane, S; Chakrabarti, P. Cavities in protein-DNA and protein-RNA interfaces. *Nucleic Acids Res.* **2009**, *37*, 4613–4620.

(11) Castrignano, T; Chillemi, G.; Varani, G.; Desideri, A. Molecular dynamics simulation of the RNA complex of a double-stranded RNA-binding domain reveals dynamic features of the intermolecular interface and its hydration. *Biophys. J.* **2002**, *83*, 3542–3552.

(12) Nurisso, A.; Blanchard, B.; Audfray, A.; Rydner, L.; Oscarson, S.; Varrot, A.; Imberty, A. Role of water molecules in structure and energetics of *Pseudomonas aeruginosa* lectin i interacting with disaccharides. *J. Biol. Chem.* **2010**, *285*, 20316–20327.

(13) de Beer, S. B. A.; Vermeulen, N. P. E.; Oostenbrink, C. The role of water molecules in computational drug design. *Curr. Topics Med. Chem.* **2010**, *10* (1), 55–66.

(14) Homans, S. W. Water, water everywhere—except where it matters?. *Drug Discovery Today* **2009**, *37*, 4613–4620.

(15) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129*, 2577–2587.

(16) Durchschlag, H.; Zipper, P. Modeling the hydration of proteins: prediction of structural and hydrodynamic parameters from X-ray diffraction and scattering data. *Eur. Biophys. J.* **2003**, *32*, 487–502.

(17) Pitt, W. R.; Goodfellow, J. M. Modelling of solvent positions around polar groups in proteins. *Protein Eng.* **1991**, *4*, 531–537.

(18) Virtanen, J. J.; Makowski, L.; Sosnick, T. R.; Freed, K. F. Modeling the hydration layer around proteins: HyPred. *Biophys. J.* **2010**, *99*, 1611–1619.

(19) Geroult, S.; Hooda, M.; Virdee, S.; Waksman, G. Prediction of solvation sites at the interface of Src SH2 domain complexes using molecular dynamics simulations. *Chem. Biol. Drug Des.* **2007**, *70*, 87–99.

(20) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2009**, *113* (40), 13337–13346.

(21) Teyra, J.; Pisabarro, M. T. Characterization of interfacial solvent in protein complexes and contribution of wet spots to the interface description. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 1087–1095.

(22) van Dijk, A. D.; Bonvin, A. M. Solvated docking: introducing water into the modelling of biomolecular complexes. *Bioinformatics* **2006**, *22*, 2340–2347.

(23) Jiang, L.; Kuhlman, B.; Kortemme, T.; Baker, D. A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 893–904.

(24) Bui, H. H.; Schiewe, A. J.; Haworth, I. S. WATGEN: an algorithm for modeling water networks at protein-protein interfaces. *J. Comput. Chem.* **2007**, *28*, 2241–2251.

(25) Lewis, B. A.; Walia, R. R.; Terribilini, M.; Ferguson, J.; Zheng, C.; Honavar, V.; Dobbs, D. PRIDB: a protein-RNA interface database. *Nucleic Acids Res.* **2011**, *39*, D277–D282.

(26) Bahadur, R. P.; Zacharias, M.; Janin, J. Dissecting protein-RNA recognition sites. *Nucleic Acids Res.* **2008**, *36*, 2705–2716.

(27) Ellis, J. J.; Broom, M.; Jones, S. Protein-RNA interactions: structural analysis and functional classes. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 903–911.

(28) Treger, M.; Westhof, E. Statistical analysis of atomic contacts at RNA-protein interfaces. *J. Mol. Recognit.* **2001**, *14*, 199–214.

(29) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(30) Sutch, B. T.; Chambers, E. J.; Bayramyan, M. Z.; Gallaher, T. K.; Haworth, I. S. Similarity of protein-RNA interfaces based on motif analysis. *J. Chem. Inf. Model.* **2009**, *49*, 2139–2146.

(31) R Development Core Team. *R: A language and environment for statistical computing*; Foundation for Statistical Computing, Vienna, Austria, 2011.

(32) Ye, X.; Gorin, A.; Ellington, A. D.; Patel, D. J. Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nat. Struct. Biol.* **1996**, *3*, 1026–1033.

(33) *The PyMOL Molecular Graphics System*, version 1.3; Schrödinger, LLC, 2010; http://www.pymol.org/ (accessed April 22, 2011).

(34) de Vries, S. J.; Melquiond, A. S.; Kastritis, P. L.; Karaca, E.; Bordogna, A.; van Dijk, M.; Rodrigues, J. P.; Bonvin, A. M. Strengths and weaknesses of data-driven docking in critical assessment of prediction of interactions. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 3242–3249.

(35) de Vries, S. J.; van Dijk, M.; Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **2010**, *5*, 883–897.

(36) Schlessman, J. L.; Abe, C.; Gittis, A.; Karp, D. A.; Dolan, M. A.; García-Moreno, E. B. Crystallographic study of hydration of an internal cavity in engineered proteins with buried polar or ionizable groups. *Biophys. J.* **2008**, *94*, 3208–3216.