Predicting Positions of Bridging Water Molecules in Nucleic Acid-Ligand Complexes

Wanlei Wei,^a Jiaying Luo,^a Jérôme Waldispühl^b and Nicolas Moitessier^{*,a}

^a Department of Chemistry and ^b School of Computer Science, McGill University, 801 Sherbrooke St W, Montreal, QC, Canada H3A 0B8.

nicolas.moitessier@mcgill.ca

Abstract

Over the past two decades, interests in DNA and RNA as drug targets have been growing rapidly. Following the trends observed with protein drug targets, computational approaches for drug design have been developed for this new class of molecules. Our efforts towards the development of a universal docking program, FITTED, led us to focus on nucleic acids. Throughout the development of this docking program, efforts were directed towards displaceable water molecules which must be accurately located for optimal docking-based drug discovery. However, although there is a plethora of methods to place water molecules in and around protein structures, there is, to the best of our knowledge, no such fully automated method for nucleic acids, which are significantly more polar and solvated than proteins. We report herein a new method, SPLASH'EM (Solvation Potential Laid around Statistical Hydration on Entire Macromolecules) developed to place water molecules within the binding cavity of nucleic acids. This fast method was shown to have high agreement with water positions in crystal structures and will therefore provide essential information to medicinal chemists.

INTRODUCTION

While proteins have traditionally been the primary target for drug development, interests in nucleic acids as promising targets for drugs and molecular probes has been more recent.¹ The realization that targeting nucleic acids would allow the treatment of previously untreatable diseases has led pharmaceutical companies and academic groups to invest heavily in this area of research, colloquially known as the “RNA gold rush.”² Two landmark discoveries – Ribocil, an antibiotic agent recently discovered by Merck,³ and Branaplam, a drug currently undergoing phase II clinical trials by Novartis for spinal muscular atrophy^{2, 4} – have shown that drugs targeting nucleic acids could be both effective and selective to their targets.²

Ribocil, in particular, was shown to inhibit bacterial cell growth and bind to a regulatory region of a metabolically essential mRNA transcript.³ These non-coding regions, located in the upstream 5'-untranslated regions (UTRs), called RNA riboswitches, are often involved in the negative feedback loop of bacterial biosynthesis pathways of life-sustaining endogenous metabolites.⁵ Riboswitches are attractive targets for antibacterial agents as they are not present in mammalian cells. Consequently, small molecules targeting riboswitches, are a desperately needed repertoire of antibiotics, especially in the current era of increasing antibiotic resistance around the world. Aminoglycosides, in particular, exhibit favorable nucleic acid binding properties (eg. positively charged ammonium groups), which make them potent antibiotics against bacterial nucleic acids, along with other promising methods such as siRNA and antisense therapy.⁵ In general, a lower dosage of drugs is sufficient when targeting mRNA over proteins, mitigating the effects of drug toxicity (one mRNA gives rise to many copies of the protein). Recently, viral RNA has also been targeted, as reported by scientists from the Scripps Research Institute.⁶

DNA structural elements have also gained attention as attractive targets for drugs, such as Cisplatin, which crosslinks with DNA bases and induces cell death in various types of cancers,⁷ intercalating agents (e.g., Doxorubicin), alkylating agents (e.g., Mechlorethamine), and more recently G-quadruplex binders.⁸ The binding and stabilization of the latter by small molecules have been shown to prevent hyperactive telomerases from repairing the ends of chromosomes in cancer cells.⁹ This inhibition prevents cancer cells from being immortal, and is a promising area of drug discovery.

Nucleic acids, in contrast to proteins, possess shallow, water-permeable binding pockets. In addition, phosphate groups of nucleic acids and their corresponding counterions (e.g., Mg^{2+} , K^{+}) are highly charged and induce polarization upon both nearby water molecules and functional groups of drugs, significantly affecting the nucleic acid-drug complex stability. Modeling of these highly polarized molecules has been identified as a source of error during the simulation of biomacromolecules.^{10, 11} Thus, water molecules are expected to play a critical role in nucleic acid-drug complex formation as illustrated in Figure 1.

Several medicinal chemistry approaches have been envisioned to develop protein binding drugs. Amongst these, structure-based drug design has become an indispensable strategy for medicinal chemists. In fact, docking libraries of small molecules to proteins was instrumental in discovering potential drugs and several docking programs have been reported since DOCK was first reported and released.¹²

While docking methods for proteins have made significant advances, the current computational drug discovery methodologies targeting nucleic acid are inadequate. While docking libraries of small molecules to protein crystal structures has been instrumental in the discovery of potential drugs, nucleic acids have been largely overlooked by developers of docking programs with a few exceptions (e.g., RiboDock,¹³ Mordor¹⁴ and rDOCK¹⁵). In other situations, programs, initially developed for proteins, were retrained to work with RNA (ICM,¹⁶ DOCK6^{17, 18}).⁸ Due to the previously stated biochemical differences between nucleic acids and proteins, accurate placement of water molecules in the binding site is often crucial for successful *in silico* docking of ligands to nucleic acids. For a recent review on the field of docking to nucleic acids and the associated challenges, readers are referred to our recent review.⁸

Figure 1. Water molecule (red) interacting with U19, G27, U28 and Lividomycin, in the binding site of 16S-rRNA (PDB: 2ESJ).

Throughout the development of our docking program, FITTED,¹⁹⁻²² conserved water molecules were found to be critical for optimal docking accuracy.²³ When we became interested in the docking of small

molecules to nucleic acids,²⁴ proper treatment of water molecules was also found to significantly improve docking performances.²⁵ In the past, the critical role of water molecules and the prevalence of proteins as drug targets prompted the development of several methods to predict the position and orientations of water molecules at the interface of protein-ligand complexes and/or to evaluate their free energy of binding. Examples include WaterMap, which relies on molecular dynamics (MD) simulations of solvated proteins;^{26, 27} JAWS, which estimates the free energy for each water molecule from Monte Carlo simulations;²⁸ statistical methods such as Szmap²⁹ and 3D-RISM;^{30, 31} docking water molecules to binding sites (e.g., WaterDock relying on AutoDock Vina³²) and empirical methods such as AQUARIUS which predicts solvent sites within a protein by mapping each amino acid to a data set of crystal structures, and others as was recently reviewed.³³ However, to the best of our knowledge, these methods were developed primarily for proteins and never applied to nucleic acids.

Recently, a report on the use of a 3D-RISM-derived method illustrated the challenges of placing water molecules in nucleic acids;³⁴ it had an accuracy (of water placement within 1 Å of crystallographic water molecules) below 40%, although testing was carried out on only a handful of structures. Herein, we present a methodology and available software for placing water molecules in the binding site of nucleic acid-ligand complexes, using a hybrid scoring function composed of statistical survey of water molecules in existing PDBs and a dedicated force field, developed specifically for water placement in nucleic acids.

THEORY AND IMPLEMENTATIONS

Challenges. At the outset of this project, we identified key challenges to address: (1) placing an explicit (3-atom) water molecule would require not only identification of its location but also description of its orientation; (2) the binding free energy of water molecules is a complex combination of entropy and enthalpy; (3) hydrogen bonds (H-bonds) are directional and sensitive to polarization; (4) water molecules could become highly polarized in such polar media, and electrostatic interactions would therefore be inaccurately computed using static point-charges; (5) the developed method should be available to medicinal chemists, fully automated, and user-friendly. In order to solve these challenges, a new strategy for water placement (Figure 2) in nucleic acids was developed, which will be addressed sequentially in the following sections.

Selection of a Suitable Water Model: United Field Water Molecules – Challenge #1. To address the first challenge, we have previously developed particle waters (PWs) which model the hydrogen bond donor (HBD) and acceptor (HBA) properties of water molecules in a single bead hence reducing the problem to only bead placement. The placement of PWs was originally implemented into PREPARE, a program previously developed to convert PDB files into useable structures (adding and optimizing hydrogen positions and water orientation).^{23, 35} In this early version, interactions between proteins and PWs were modeled using a Lennard-Jones potential which was trained to account for all the interactions (ie. electrostatics, van der Waals, and hydrogen bonds). However, this implementation was poor in accuracy (see results and discussion) and we have fully redesigned these particles to take into account additional physical phenomena, such as directionality of hydrogen bonding and polarization of water molecules. Additionally, these PWs were completely redesigned for nucleic acids.

Figure 2. Method flowchart designed for SPLASH'EM.

Statistics as a Scoring Method for Water Placement – Challenge #2. To address the difficult task of water placement in nucleic acids, statistics and molecular mechanics (MM) methods were used in combination. Statistics information could not only be used to place water molecules, but also to assign a probability to each position in space, which could be converted into a free energy of binding. As shown in Figure 2, statistics could provide key information on the likely positions of water molecules near bases, phosphates, sugars, and ligands. These locations of high occurrences would then be targeted to position water molecules.

Thus, our first task was to collect structural information on preferred water molecule positions. Statistical analysis around each of the nucleotides was carried out in over 4,100 crystal structures obtained from the Protein Data Bank (PDB), for a total of ca. 5.8M water molecules. Water distribution around each H-bond acceptor/donor was plotted and summarized into a series of graphs (Figure 3). This water distribution data was used as a starting point for our new program SPLASH'EM (Solvation Potential Laid around Statistical Hydration on Entire Macromolecules).

Figure 3. Statistics information collected on the occurrences of water molecules near: Adenine N1 (A), N3 (B), N6 (C), and N7 (D) at various angles (-90° to 90°) within the plane of the base. Statistics were similarly collected outside the plane at different “pitch” angles. In each voxel, frequencies are indicated as a total percentage of waters collected. Statistics for other hydrogen bond acceptors/ donors at various pitch angles could be found in the supplementary information.

As water molecules are often H-bonded to more than one residue (Figure 1), estimating the free energy of water molecules was achieved through the combination of the free energy distributions from each nearby polar atom (although we are aware that entropy is not additive).

Particle Water Force Field as a Scoring Method for Water Placement – Challenges #3 and #4.

While the PDB-derived statistics of water positions give an approximation of the likely locations of water molecules, it was believed that refinement may be needed. This refinement could be performed by using gradient descent optimization in order to position the water molecules at regions of energy minima. Thus, the free energy score obtained from statistics was complemented with MM, the latter of which precludes any clashes, considers any other interactions with the nucleic acids, and enables an energy optimization necessary to refine PW positions. A new function and corresponding parameters (force field, or FF) for computing the nucleic acid-PW and ligand-PW interaction energies were developed for use in energy evaluation and optimization of PWs including directionality of hydrogen bonding (challenges #3) and polarization (challenge #4).

Finding a Suitable Function and Parameters for Modelling Hydrogen Bonding between Nucleic Acids and PWs. Functions (Equation 1) and parameters were trained to reproduce the energy profile of the interaction between nucleic acids and water molecules at the MP2/6-31+G(d,p) level of theory, as

illustrated in Figure 4 and further discussed in the SI. This FF was named the Particle Water Force Field (PWFF).

$$E_P = \frac{A}{r^6} - \frac{B}{r^5} - \frac{C}{r^3} \quad (1)$$

Figure 4. Hydrogen bonding potential in QM (red) as compared to the developed PWFF (blue) for a linear distance scan for Ado-N1 (A) and Ado-H62 (B).

In PWFF, dispersion and Pauli repulsion terms were obtained from Lennard-Jones potential of Amber ff95. As PWs are neutral, the hydrogen bonding term implicitly accounted for electrostatic interactions. A Lennard-Jones 6-5/3 term was found to best reproduce this energy profile obtained through QM calculations (Equation 1).

Methodology Used to Develop the Particle Water Force Field. Optimizations were performed on entire nucleic acid fragments (ie. adenine, guanine, cytosine, thymine, uracil, ribose, deoxyribose, phosphate) in QM while the model for water molecules was derived from TIP3P.³⁶ To develop the PWFF, H-bonding energy profiles between nucleic acid fragments and water molecule were calculated using the previously described QM- and MM-levels of theory (Figure 5). To derive the H-bonding potential, the difference between the QM energy profile and the van der Waals MM energy profile was used (Equation 2). After attempting various Lennard-Jones potentials, including the well-known 12-10, 12-8, and many others, a 6-5/3 term was found to best reproduce the H-bond potential.

Figure 5. Model used for evaluating the energy at A) MP2/6-31+G (d,p) and B) MM, with united atom representation of the PW. Arrows depict the path taken. C) Energy as a function of distance, depicting the QM energy, van der Waals energy, and their difference.

$$E_{H-bond(PW)} = E_{QM} - E_{MM:LJ\ 12-6} \quad (2)$$

Incorporating Directionality of Hydrogen Bonding into Particle Water Force Field – Challenge

#3. In the past, attempts to include directionality of H-bond terms in some force fields were reported,³⁷⁻³⁹ although no clear improvements in performance were noted and these terms were subsequently discarded. However, these previous developments were largely tailored towards MD simulations and not towards water placement methods or *in silico* docking studies. To this end, the angular dependence of H-bonding with respect to simplified nucleic acid subunits were investigated by scanning water positions with respect to these subunits (Figure 6).

Figure 6. Simplified subunits used to probe the relationship between enthalpy and angles. Polar atoms of interest are in red, while angle bisectors are depicted by dashed lines.

An angular contribution function (Figure 7A) was next developed to account for the directionality of the H-bonding, penalizing deviations of H-bonding from ideal positions. For H-bond acceptors and donors, the ideal angle was designated at the positions of the electron lone pair(s) and directly opposite

the R-H bond, respectively. An example of the H-bonding potentials as a function of angles in both QM and the developed PWFF potentials for various chemical fragments are shown (Figure 7B-E).

The angular contribution function utilizes a normal distribution-like function, which is always bound between 0 and 1. At ideal angles, the angular contribution function yielded a value close to 1, which instates the full strength of the hydrogen bond. In contrast, when far from the ideal angle, the hydrogen bonding is essentially turned off by a value close to 0. The augmentation of PWFF with angular contribution function (PWFFa), although not perfect, reproduced the QM energy profile for these molecules, even at non-ideal H-bonding geometries. Equation 3, along with table 1 describes the PWFFa function form, along with its various associated parameters, the latter of which are the same as those used for PWFF.

$$E_p = \begin{cases} \underbrace{\left(\frac{A}{r^6} - \frac{B}{r^5} - \frac{C}{r^3}\right)}_{\text{distance contribution}} \cdot \underbrace{\left(e^{-(\alpha\theta)^{2n}}\right)}_{\text{angle contribution}} \\ \underbrace{\left(\frac{A}{r^6} - \frac{B}{r^5} - \frac{C}{r^3}\right)}_{\text{distance contribution}}, \text{ if } r < r_{\text{threshold}} \end{cases} \quad (3)$$

A potential ambiguity with this angular contribution function could arise if a PW was within steric clashing distance, but at a non-ideal angle. In order to prevent this from occurring, a threshold distance was given for different atoms. For hydrogen atoms, this value was 1.5 Å while all other heteroatoms were assigned a value of 2.5 Å. At distances less than these values, the full steric clash would be restored.

Figure 7. (A) Angular contribution function implemented into PWFF for various atom types. (B-E) Energy profile as a function of deviation from the angle bisector, obtained by $E_{QM-EVDW}$ compared to the developed PW potential for: (B) pyridine, (C) aniline, (D) tetrahydrofuran and (E) 1-methylpyridin-4(1H)-one.

Incorporating Polarization of Hydrogen Bonding into Particle Water Force Field – Challenge #4.

Polarization of water molecules in nucleic acids occurs more frequently than in proteins due to the anionic charge of the phosphate backbone and presence of counteractions such as Mg^{2+} and K^+ . Polarization has

been found to lead to artifacts during MD simulations of proteins,^{10,11} and methods to address this problem have been proposed.^{40, 41} However, it remains to be addressed for the more polar nucleic acids. From our quantum mechanics (QM) calculations, it was observed that a water molecule bridged by two polar atoms experienced an overall potential unequal to the sum of the two individual potentials (Figure 8).

At each water position, the two greatest hydrogen bonding energies were selected, and polarization factors were assigned to these two interactions. Five distinct classes of polarization were created: 1) strongly stabilizing, 2) strongly destabilizing, 3) weakly stabilizing, 4) weakly destabilizing, and 5) unchanged. A polarization factor was introduced for each of these cases. The first case involved the interaction of oppositely-charged polar atoms, and was given a polarization factor of $1+\beta$. Similarly, the second case involved like-charged polar atoms interacting, and was given a polarization factor $1-\beta$. The third and fourth cases involved an uncharged polar atom and a charged polar atom. In both cases, the uncharged atom was given a polarization factor while the charged was not. More specifically, the third case involved a HBA-HBD+ or HBD-HBA- pair and was given a polarization factor of $1+\alpha$ while the third case involved a HBD-HBD+ or HBA-HBA- pairs and was given a polarization factor of $1-\alpha$. Finally, the polarization between uncharged groups were expected to be minimal and no polarization was given. Different sets of factors were evaluated by varying both α and β parameters between 0.0~0.9 at 0.05 increments with the criteria that $\alpha \leq \beta$ and the optimal values were found to be $\alpha = 0.2$ and $\beta = 0$. This polarizable force field was named polarizable PWFF with angle contribution function (pPWFFa), which is shown in equation 4.

Figure 8. A) H-Bond potentials between CH_3NH_4^+ and H_2O (in QM) when the latter is: not bridged (green), bridged by $(\text{CH}_3)_2\text{PO}_4^-$ (blue), and bridged by CH_3NH_4^+ (red). B) Corresponding schematics are shown (i-iii), respectively.

Water Placement Strategy. In our newly developed fully automated program (challenge #5), SPLASH'EM, water molecules are initially placed in the centers of the top 50 most populated grid points for each polar atom based on the previously collected statistical information obtained (Figure 2). At this point, each PW was assigned a binding free energy and force field energy based on its position relative to all polar atoms in the complex. PWs sterically clashing with the receptor and ligand were removed, along with PWs not located in the binding site. As our focus was on nucleic acid-ligand complexes, only bridging water molecules were considered (i.e. waters interacting with both the nucleic acid and a ligand). Finally, PWs are ranked energetically (force field and/or free energy), and placed from most to least stable, with

the criteria that a PW does not clash with a previously placed PW and is not located on the water-exposed surface.

Validation Dataset for Water Placement. As of December 2018, there were over 4,000 nucleic acid crystal structures available in the PDB. Of these, approximately 300 nucleic acids contain ligands and at least one water molecule. To compile a validation set for water placement, a further refinement of the nucleic acid crystal structures was performed. A filtering step by a threshold resolution of 2.6 Å was performed on these previously obtained structures, resulting in 91 PDBs, which contained a total of 12,955 water molecules. However, many of these water molecules were found in bulk water or away from the ligand binding site, and thus unimportant for medicinal chemistry applications. Only bridging waters, less than 3.5 Å away from both a nucleic acid polar atom and a ligand polar atom were kept, leading to 398 water molecules. Finally, in order to ensure that these water molecules, present in the original PDB structures were substantiated by electron density, the corresponding electron density data, containing the 2F0-Fc map was used as experimental evidence. To carry out this procedure, EDIA³⁶ calculation was performed on each crystal structure. EDIA calculates the experimental support for an atom by taking into account the electron density within its van der Waals radius, and allows comparison between different structures by normalization. Water molecules with EDIA score of less than 0.24 were discarded, as suggested by Rarey and co-workers, since these waters are less supported by electron density.^{26, 36} The final compiled validation set was made up of 91 crystal structures, containing a total of 230 water molecules.

Table 1. PW hydrogen bonding parameters.

Particle Water-Polar Atom Interactions		Distance Contribution			Angle Contribution	
		A	B	C	α	n
Adenine	N1	18310	7535	0	0.015	1
	N3	27792	11522	0	0.015	1
	H61	297	0	76	0.01	4
	H62	297	0	76	0.01	4
Guanine	N7	19187	7814	0	0.015	1
	H1	1431	886	0	0.01	4
	H21	163	0	48	0.01	4
	H22	163	0	48	0.01	4
	N3	14986	6170	0	0.015	1
	O6	25510	10500	0	0.015	2
	N7	20596	8008	0	0.015	1
Cytosine	O2	25590	10568	0	0.015	2
	N3	20043	8262	0	0.015	1
	H41	364	0	99	0.01	4
	H42	364	0	99	0.01	4
Thymine/ Uracil	O2	17551	7230	0	0.015	2
	H3	800	574	0	0.01	4
	O4	17311	7521	0	0.015	2
Deoxyribose	O4'	17311	7521	0	0.015	2
Ribose	O2'	14300	6128	0	0.015	2
	HO2'	300	0	85	0.01	4
	O4'	18899	8090	0	0.015	2
Phosphate	O3'	22280	8740	0	0.015	2
	O5'	22280	8740	0	0.015	2
	OP1	5000	0	525	0.019	1

	OP2	5000	0	525	0.019	1
Pyridine	N	22770	9419	0	0.015	1
Methylamine	HN	631	0	231	0.01	4
Methanol	HO	1490	0	965	0.01	4
	OH	16690	7020	0	0.015	2
Acetylamide	O	17810	7145	0	0.015	2
	HN	320	0	89.1	0.01	4
Formate	O	5383	0	615	0.015	2

Table 2. Polarization factors assigned to H-bond potentials.

Moving	Stationary	Polarization Factor
Weak Donor	Weak Donor	1
Weak Donor	Weak Acceptor	1
Weak Donor	Strong Donor	1 - α
Weak Donor	Strong Acceptor	1 + α
Weak Acceptor	Weak Donor	1
Weak Acceptor	Weak Acceptor	1
Weak Acceptor	Strong Donor	1 + α
Weak Acceptor	Strong Acceptor	1 - α
Strong Donor	Weak Donor	1
Strong Donor	Weak Acceptor	1
Strong Donor	Strong Donor	1 - β
Strong Donor	Strong Acceptor	1 + β
Strong Acceptor	Weak Donor	1
Strong Acceptor	Weak Acceptor	1
Strong Acceptor	Strong Donor	1 + β
Strong Acceptor	Strong Acceptor	1 - β

$$E_P = \begin{cases} \underbrace{\left(\frac{A}{r^6} - \frac{B}{r^5} - \frac{C}{r^3}\right)}_{\text{distance contribution}} \cdot \underbrace{\left(e^{-(\alpha\theta)^{2n}}\right)}_{\text{angle contribution}} \cdot \underbrace{f_{pol}}_{\text{polarization correction}} \\ \underbrace{\left(\frac{A}{r^6} - \frac{B}{r^5} - \frac{C}{r^3}\right)}_{\text{distance contribution}}, \text{ if } r < r_{threshold} \end{cases} \quad (4)$$

Results and Discussion

Assessing the Accuracy of Placed Water Molecules. Assessing the accuracy of water placement on nucleic acids is difficult primarily due to its dependence on X-ray crystal structures of nucleic acids. The structure is a static representation of a biological molecule in a non-physiological environment. Some waters may not be resolved due to their high mobility and low residence times at certain positions. However, placing a water at these locations should not necessarily be considered wrong. Any X-ray crystal structure is a model, and only one interpretation of the electron density. Consequently, it may suffer from

biases which were introduced during the refinement process. This is especially true if the crystal structure is low in resolution. In fact, in a few cases, the placement of water molecule would have been satisfied in more than one way. For example, during the model construction of the crystal structure of a DNA dodecamer complexed with propamide (PDB ID: 102D), a single water molecule (O36) was placed in the region of high electron density, which yielded an EDIA score of 0.59 (Figure 9). Alternatively, two water molecules could have been placed adjacent to the crystallographic water location at distances of approximately 1.7 Å and 1.6 Å away. At these two alternate locations, there is also support from electron density with EDIA scores of 0.24 and 0.30. In this case, the X-ray diffraction pattern caused by these two water molecules would have superposed and created the maximum electron density signal at the location of the crystallographic water. In physiological conditions, the movement of water molecules would probably have allowed its presence at both locations. Consequently, this presents a challenge in scoring of placed water molecules since these alternative positions are not considered successful by a first criterion for placement success (PW within 1 Å from crystallographic water). Using a different second criterion ($\text{EDIA} \geq 0.24$), these water molecules are considered successful.

Figure 9. Crystallographic water position (green circle) compared to possible alternative water positions (blue circles) overlaid with the 2F0-Fc electron density map. EDIA score is given adjacent to each water. (A). Placed waters are supported by electron density but are situated away (distances of 1.6 and 1.7 Å) from the crystallographic water. (B). Placed waters are not supported by electron density but are in close proximity to the crystallographic water (distance of 1.0 Å).

However, a disadvantage with the latter arises when water molecules which are in close proximity to a crystallographic water (less than 1 Å) is considered unsuccessful by electron density, as illustrated by another bridging water molecule (Figure 9B). In this case, the crystallographic water and the placed water had an EDIA score of 0.37 and 0.18, respectively. As a result, the placed water was not scored as successful. However, due to their close spatial proximity (distance of 1 Å apart), the placed water may still be useful in providing important information for *in silico* docking. As a result, neither of the criterion could satisfactorily measure the accuracy or precision. Even worse, the lack of an electron does not necessarily preclude the presence of water. As an example, a ligand bound to 2'-Deoxyguanosine riboswitch (PDB: 3SKI) was found to have two identical binding sites (Figure 10). In spite of this, one binding site lacked electron density for water molecules while the other showed the presence of two water molecules.

Consequently, until a robust scoring methodology could be envisioned, it is difficult to assess the true accuracy of a water placement methodology. Ultimately however, the accuracy of water placement methodologies depends on its utility towards a given problem (e.g., docking, MD, biocatalysis). Due to our interest in using water placement for docking, the accuracy of the developed water placement methodology was given with the first aforementioned criteria.

Figure 10. 2'-Deoxyguanosine bound to each binding site of the dimeric 2'-Deoxyguanosine riboswitch (PDB: 3SKI). Shown in A) and B) are the two co-crystallized ligands in each binding site, and their associated 2F0-Fc electron densities.

Accuracy of Various Placement Methods. In testing the newly developed water placement method, the list of 91 crystal structures from the PDB was included in the testing set, which contained a total of 230 bridging water molecules (SI). Testing was first carried out with random water placement as a reference, which correctly positioned a PW within 1 Å of experimentally observed water molecule, in nearly 1 of every 5 cases (Table 3). Additionally, water placement using our previous protocol, PREPARE,³⁵ (developed for proteins), was tested along with the scoring methods developed herein.

Table 3. PW placement accuracy of current developments

Placement method	Scoring method	Accuracy ^a
Random	N/A	19%
Prepare ³⁵	PREPARE ³⁵	30%
Statistics	Statistics	51%
Statistics	PWFF	51%
Statistics	PWFFa	56%
Statistics	pPWFFa	60%
Statistics	Splash'Em	62%

^a Number of crystallographic water with a PW within 1 Å.

Each distinct scoring method (free energy from statistics or FF energy) was tested individually and in combination to see how it would impact the accuracy of the methodology. In total, five different scoring criteria are reported, including: statistics, PWFF, PWFFa, pPWFFa, and SPLASH'EM, which incorporates all of the above.

Accuracy of PREPARE. Water placement using PREPARE³⁵ gave an accuracy of 30%. Based on recent evidences, it has been suggested that the water network surrounding ligands differ based on its identity, as was the case of aminoglycosides binding to RNA.²⁵ However, the water placement method in PREPARE did not take the influence of the ligand into consideration. To rectify this, each new developed placement methodology took the ligand into consideration.

Accuracy of Statistics as Water Placement Methods. Statistics as a scoring criterion has many advantages, including the implicit incorporation of polarization, directionality of H-bonding, and free energies. Using this method for water placement gave a significantly improved accuracy of 51%. The additivity of free energies is a challenge in this approach, as the relationship for sum of entropy has not been fully elucidated. For this reason, it is difficult to predict the relative stability between two potential water positions. Additionally, the lack of available ligand statistics forced us to use approximate free energies, which used similar nucleic acid atom types, leading to possible errors in free energies.

Accuracy of Particle Water Force Field as Water Placement Method. The three variations of the developed Particle Water Force Field (PWFF, PWFFa, and pPWFFa) yielded accuracies of 51%, 56%, and 60%, respectively. We were very pleased that with each additional chemical principle incorporated into the force field (ie. angles and polarization), significant increases in accuracy were observed.

The use of PWFFa over PWFF eliminated weak H-bond interactions (Figure 11A), which were at large angular deviations from the ideal geometry. For example, during the placement of a water molecules in the binding site of an RNA riboswitch-SAM - ligand complex (PDB: 5FJC), PWFF placed a water molecule 2.3 Å away from the crystallographic position (Figure 11A). At this position, it was within H-bonding distances to two polar atoms: ribose-O2' and Ado-N3. Chemically however, the geometry at this position did not allow for a strong hydrogen bonding to ribose-O2', which was not taken into account by PWFF. On the other hand, PWFFa and pPWFFa were able to take directionality of H-bonding into consideration and greatly penalize this interaction. Consequently, an adjacent location, 0.4 Å away from the crystallographic water, was correctly chosen (Figure 11).

In spite of the successes of the angular correction function in PWFFa, there were still a number of water molecules which were incorrectly predicted—one reason being that it was unable to account for the nuance behaviors of the free energy surface, including polarization. pPWFFa, in contrast to PWFFa, identified destabilized interactions due to polarization (Figure 11B). In other cases, pPWFFa also identified stabilized interactions (ie. water bridged by a strong donor and a strong acceptor).

Figure 11. Shown are crystallographic water (red) and PWs placed using PWFF (green), PWFFa (blue), and pPWFFa (orange). H-bonds are denoted by black lines while weak H-bonds are denoted by orange line. Comparison of (A) PWFF to PWFFa and (B) PWFFa to pPWFFa.

In another example, during the placement of water molecules of an RNA-Paromomycin complex (Figure 11B; PDB 2PWT), PWFF incorrectly selected the water position which emphasized interactions with two phosphate groups, which had the deepest innate H-bonding potentials. However, this location was expected to be destabilized since both groups are strong hydrogen bond acceptors. Consequently, the crystallographic water molecule was found away from this position, and instead H-bonded to a nearby hydroxyl hydrogen of the modified Paromomycin ligand although still H-bonded to one of the phosphate groups. This crystallographic water position was correctly predicted by pPWFFa, but not PWFFa.

Originally, it was expected that the hydroxyl hydrogen of the modified Paromomycin would increase its hydrogen bonding potential in the presence of a strong acceptor (phosphate group). However, during the search for a suitable polarization factor, negligible changes in accuracy was observed for incorporating polarization of weak HBA and HBD. As a result, polarization for these types of interactions were not included in pPWFFa. More specifically, it was found that polarization factors of $\alpha = 0.2$, $\beta = 0$ yielded optimal accuracies for water placement. This meant that the polarization factor for strongly stabilizing, strongly destabilizing, weakly stabilizing, and weakly destabilizing was 1.2, 0.8, 1, and 1, respectively (Table 2). In the future, a more encompassing polarization scheme could be devised, such as varying the degree of polarization, based on distance. However, the polarization scheme described in this method was a proof-of-principle showing that its consideration is an important aspect which must be considered when attempting to accurately predict water molecule positions. Overall, in both cases (ie. moving from PWFF

to PWFFa and PWFF to pPWFFa), the more advanced FF correctly identified the crystallographic water positions.

Accuracy of a Hybrid Scoring Method for Water Placement. We next envisioned that the two energy evaluation methods could be combined into a single metric, with the hopes of further increasing the accuracy of water placement. More specifically, a large number of overlapping PWs were placed, scored with pPWFFa, and two alternative populations of non-clashing sets of PWs extracted. These two distinct populations were subsequently combined, and the approximate free energy (from statistics), was used to select the final list of PWs. This method produced an accuracy of 62% on the selected 230 crystal water molecules, which was the final version implemented into SPLASH'EM itself integrated into our drug discovery platform FORECASTER (challenge #5).³⁵ In essence, SPLASH'EM employs a similar method to hybrid scoring functions developed for molecular docking, where two or more methods are used together in hopes of more accurately identifying the correct binding mode of a ligand in the receptor. In the case of water placement, the combination of approximate free energy and force field energy gave an increased performance over either of these methods used alone.

Conclusion

In conclusion, we showed that the accuracy of water placement in nucleic acid structures is difficult to obtain due to the insufficient number of nucleic acid-ligand 3D structures. In addition, due to poor resolution, subjectivity during model construction, and a lack of electron density in some regions, crystal structures themselves are problematic. Ultimately, the accuracy of a given water placement method rests with its utility in a given problem.

We succeeded in developing SPLASH'EM, a program to identify and place water molecules (in the form of PWs) in nucleic acids on the order of seconds. This method was able to correctly identify as many as 62% of water molecules in nucleic acids, laying the foundation for water placement for these macromolecules. SPLASH'EM utilizes a hybrid scoring function composed of statistical data of water occurrences for each polar atom of nucleic acids and ligands; and a newly developed pPWFFa FF to describe H-bonding of polar atoms and water molecules. Associated parameters for the FF were developed, which is novel in that it adds H-bond directionality and polarization to our previously developed united field water model, Particle Waters. pPWFFa is expected to be useful with *in silico* docking of nucleic acids when incorporated into FITTED. Overall, SPLASH'EM has been shown, at the time of publication, to be the most accurate water placement tool developed for nucleic acids and available for use by chemists.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: TBD
Detailed methodology and relevant examples.

ACKNOWLEDGMENT

We thank the Canadian Institutes of Health Research (CIHR, BOP-149429) and Genome Canada (BCB 2015) for funding. WW thanks the Fonds de Recherche Nature et Technologies du Quebec for a graduate studentship.

REFERENCES

1. Wang, M. L.; Yu, Y. Y.; Liang, C.; Lu, A. P.; Zhang, G., Recent Advances in Developing Small Molecules Targeting Nucleic Acid. *Int. J. Mol. Sci.* **2016**, *17*, 24.
2. Cross, R., The RNA drug hunters Academics, biotech start-ups, and big pharma companies are designing small molecules that target RNA. *Chemical & Engineering News* **2017**, *95*, 16-18.
3. Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.; Murgolo, N.; Barbieri, C. M.; Mann, P. A.; Carr, D.; Xia, E.; Zuck, P.; Riley, D.; Painter, R. E.; Walker, S. S.; Sherborne, B.; de Jesus, R.; Pan, W. D.; Plotkin, M. A.; Wu, J.; Rindgen, D.; Cummings, J.; Garlisi, C. G.; Zhang, R. M.; Sheth, P. R.; Gill, C. J.; Tang, H. F.; Roemer, T., Selective small-molecule inhibition of an RNA structural element. *Nature* **2015**, *526*, 672-677.
4. Palacino, J.; Swalley, S. E.; Song, C.; Cheung, A. K.; Shu, L.; Zhang, X. L.; Van Hoosear, M.; Shin, Y.; Chin, D. N.; Keller, C. G.; Beibel, M.; Renaud, N. A.; Smith, T. M.; Salcius, M.; Shi, X. Y.; Hild, M.; Servais, R.; Jain, M.; Deng, L.; Bullock, C.; McLellan, M.; Schuierer, S.; Murphy, L.; Blommers, M. J. J.; Blaustein, C.; Berenshteyn, F.; Lacoste, A.; Thomas, J. R.; Roma, G.; Michaud, G. A.; Tseng, B. S.; Porter, J. A.; Myer, V. E.; Tallarico, J. A.; Hamann, L. G.; Curtis, D.; Fishman, M. C.; Dietrich, W. F.; Dales, N. A.; Sivasankaran, R., SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat Chem Biol* **2015**, *11*, 511-+.
5. Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.; Murgolo, N.; Barbieri, C. M.; Mann, P. A.; Carr, D.; Xia, E.; Zuck, P.; Riley, D.; Painter, R. E.; Walker, S. S.; Sherborne, B.; de Jesus, R.; Pan, W.; Plotkin, M. A.; Wu, J.; Rindgen, D.; Cummings, J.; Garlisi, C. G.; Zhang, R.; Sheth, P. R.; Gill, C. J.; Tang, H.; Roemer, T., Selective small-molecule inhibition of an RNA structural element. *Nature* **2015**, *526*, 672-677.
6. Childs-Disney, J. L.; Tran, T.; Vummidi, B. R.; Velagapudi, S. P.; Haniff, H. S.; Matsumoto, Y.; Crynen, G.; Southern, M. R.; Biswas, A.; Wang, Z.-F.; Tellinghuisen, T. L.; Disney, M. D., A Massively Parallel Selection of Small Molecule-RNA Motif Binding Partners Informs Design of an Antiviral from Sequence. *Chem* **2018**, *4*, 2384-2404.
7. Dasari, S.; Tchounwou, P. B., Cisplatin in cancer therapy: molecular mechanisms of action. *Eur. J. Pharmacol.* **2014**, *0*, 364-378.
8. Luo, J.; Wei, W.; Waldispühl, J.; Moitessier, N., Challenges and Current Status of Computational Methods for Docking Small Molecules to Nucleic Acids. *Eur. J. Med. Chem.* **2019**, in press.
9. Collie, G. W.; Parkinson, G. N., The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.* **2011**, *40*, 5867-92.
10. Meng, E. C.; Cieplak, P.; Caldwell, J. W.; Kollman, P. A., Accurate Solvation Free-Energies of Acetate and Methylammonium Ions Calculated with a Polarizable Water Model. *J. Am. Chem. Soc.* **1994**, *116*, 12061-12062.
11. Guallar, V.; Jarzecki, A. A.; Friesner, R. A.; Spiro, T. G., Modeling of ligation-induced helix/loop displacements in myoglobin: Toward an understanding of hemoglobin allostery. *J. Am. Chem. Soc.* **2006**, *128*, 5427-5435.
12. Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R., Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br. J. Pharmacol.* **2008**, *153*, S7-S26.
13. Morley, S. D.; Afshar, M., Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock®. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 189-208.
14. Guilbert, C.; James, T. L., Docking to RNA via Root-Mean-Square-Deviation-Driven Energy Minimization with Flexible Ligands and Flexible Targets. *J. Chem. Inf. Model.* **2008**, *48*, 1257-1268.
15. Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D., *PLoS Comp. Biol.* **2014**, *10*, e1003571.
16. Stelzer, A. C.; Frank, A. T.; Kratz, J. D.; Swanson, M. D.; Gonzalez-Hernandez, M. J.; Lee, J.; Andricioaei, I.; Markovitz, D. M.; Al-Hashimi, H. M., Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* **2011**, *7*, 553.
17. Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D., DOCK 6: Combining techniques to model RNA–small molecule complexes. *RNA* **2009**, *15*, 1219-1230.

18. Philips, A.; Milanowska, K.; Łach, G.; Bujnicki, J. M., LigandRNA: computational predictor of RNA–ligand interactions. *RNA* **2013**, 19, 1605-1616.
19. Moitessier, N.; Pottel, J.; Therrien, E.; Englebienne, P.; Liu, Z.; Tomberg, A.; Corbeil, C. R., Medicinal Chemistry Projects Requiring Imaginative Structure-Based Drug Design Methods. *Acc. Chem. Res.* **2016**, 49, 1646-1657.
20. Corbeil, C. R.; Englebienne, P.; Moitessier, N., Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, 47, 435-449
21. Lawandi, J.; Toumieux, S.; Seyer, V.; Campbell, P.; Thielges, S.; Juillerat-Jeanneret, L.; Moitessier, N., Constrained Peptidomimetics Reveal Detailed Geometric Requirements of Covalent Prolyl Oligopeptidase Inhibitors. *J. Med. Chem.* **2009**, 52, 6672-6684
22. Pottel, J.; Therrien, E.; Gleason, J. L.; Moitessier, N., Docking ligands into flexible and solvated macromolecules. 6. Development and application to the docking of HDACs and other zinc metalloenzymes inhibitors. *J. Chem. Inf. Model.* **2014**, 54, 254-265.
23. Therrien, E.; Weill, N.; Tomberg, A.; Corbeil, C. R.; Lee, D.; Moitessier, N., Docking Ligands into Flexible and Solvated Macromolecules. 7. Impact of Protein Flexibility and Water Molecules on Docking-Based Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2014**, 54, 3198-3210.
24. Kieltyka, R.; Englebienne, P.; Fakhoury, J.; Autexier, C.; Moitessier, N.; Sleiman, H. F., A platinum supramolecular square as an effective G-quadruplex binder and telomerase inhibitor. *J. Am. Chem. Soc.* **2008**, 130, 10040-10041
25. Moitessier, N.; Westhof, E.; Hanessian, S., Docking of aminoglycosides to hydrated and flexible RNA. *J. Med. Chem.* **2006**, 49, 1023-1033
26. Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A., Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl Acad. Sci. USA* **2007**, 104, 808-813.
27. Wei, W.; Gault, J. W.; Monard, G., Pretransfer Editing in Threonyl-tRNA Synthetase: Roles of Differential Solvent Accessibility and Intermediate Stabilization. *ACS Catalysis* **2017**, 7, 3102-3112.
28. Michel, J.; Tirado-Rives, J.; Jorgensen, W. L., Prediction of the Water Content in Protein Binding Sites. *J. Phys. Chem. B* **2009**, 113, 13337-13346.
29. Bayden, A. S.; Moustakas, D. T.; Joseph-McCarthy, D.; Lamb, M. L., Evaluating Free Energies of Binding and Conservation of Crystallographic Waters Using SZMAP. *J. Chem. Inf. Model.* **2015**, 55, 1552-1565.
30. Kovalenko, A.; Hirata, F., Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach. *Chem. Phys. Lett.* **1998**, 290, 237-244.
31. Kovalenko, A.; Hirata, F., Self-consistent description of a metal–water interface by the Kohn–Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **1999**, 110, 10095-10112.
32. Ross, G. A.; Morris, G. M.; Biggin, P. C., Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. *PLoS ONE* **2012**, 7, e32036.
33. Nittinger, E.; Flachsenberg, F.; Bietz, S.; Lange, G.; Klein, R.; Rarey, M., Placement of Water Molecules in Protein Structures: From Large-Scale Evaluations to Single-Case Examples. *J. Chem. Inf. Model.* **2018**, 58, 1625-1637.
34. Giambaşu, G. M.; Case, D. A.; York, D. M., Predicting Site-Binding Modes of Ions and Water to Nucleic Acids Using Molecular Solvation Theory. *J. Am. Chem. Soc.* **2019**.
35. Therrien, E.; Englebienne, P.; Arrowsmith, A. G.; Mendoza-Sanchez, R.; Corbeil, C. R.; Weill, N.; Campagna-Slater, V.; Moitessier, N., Integrating medicinal chemistry, organic/combinatorial chemistry, and computational chemistry for the discovery of selective estrogen receptor modulators with FORECASTER, a novel platform for drug discovery. *J. Chem. Inf. Model.* **2012**, 52, 210-224.
36. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, 79, 926-935.

37. Lii, J. H.; Allinger, N. L., Directional Hydrogen-Bonding in the Mm3 Force-Field .1. *J. Phys. Org. Chem.* **1994**, 7, 591-609.
38. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., Dreiding - a Generic Force-Field for Molecular Simulations. *J. Phys. Chem.* **1990**, 94, 8897-8909.
39. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, 4, 187-217.
40. Farahvash, A.; Leontyev, I.; Stuchebrukhov, A., Dynamic and Electronic Polarization Corrections to the Dielectric Constant of Water. *J. Phys. Chem. A* **2018**, 122, 9243-9250.
41. Das, A. K.; Demerdash, O. N.; Head-Gordon, T., Improvements to the AMOEBA Force Field by Introducing Anisotropic Atomic Polarizability of the Water Molecule. *J. Chem. Theory Comput.* **2018**, 14, 6722-6733.