

# aquaPELE: a Monte Carlo-based algorithm to sample the effects of buried water molecules in proteins

Martí Municoy<sup>1</sup>, Sergi Roda<sup>1</sup>, Daniel Soler<sup>2</sup>, Alberto Soutullo<sup>1</sup> and Víctor Guallar<sup>1,3,\*</sup>

<sup>1</sup>Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain

<sup>2</sup>Nostrum Biodiscovery, Jordi Girona 29, Nexus II D128, 08034, Barcelona, Spain

<sup>3</sup>ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

\* Corresponding author

e-mail: [victor.guallar@bsc.es](mailto:victor.guallar@bsc.es)

phone number: +34 93 413 7727

ORCIDs of authors:

Martí Municoy: 0000-0003-4399-153X

Alberto Soutullo: 0000-0002-8156-3766

Sergi Roda: 0000-0002-0174-7435

Víctor Guallar: 0000-0003-3274-2482

Daniel Soler: 0000-0003-3274-2482

## Abstract

Water is frequently found inside proteins, carrying out important roles in catalytic reactions or molecular recognition tasks. Therefore, computational models that aim to study protein-ligand interactions usually have to include water effects through explicit or implicit approaches in order to obtain reliable results. While full explicit models might be too computationally daunting for some applications, implicit models are normally faster but omit some of the most important contributions of water. This is the case of our in-house software, called PELE, which uses implicit models to speed up conformational explorations as much as possible; the lack of explicit water sampling, however, limits its model. In this work, we confront this problem with the development of aquaPELE. It is a new algorithm that extends the exploration capabilities while keeping efficiency as it employs a mixed implicit/explicit approach to also take into account the effects of buried water molecules. With an additional MC routine a set of explicit water molecules is perturbed inside protein cavities and their effects are dynamically adjusted to the current state of the system. As a result, this implementation can be used to predict the principal hydration sites or the rearrangement and displacement of conserved water molecules upon the binding of a ligand. We

benchmarked this new tool focusing on estimating ligand binding modes and hydration sites in cavities with important interfacial water molecules, according to crystallographic structures. Results suggest that aquaPELE sets a fast and reliable alternative for molecular recognition studies in systems with a strong water-dependency.

## Introduction

Water is one of the main contributors that determine the shape of a protein, thereby defining its function. It can also be found inside its cavities, taking part in catalytic reactions and molecular recognition tasks. Analysis over 392 high-resolution crystallographic structures revealed more than 1,000 interfacial water molecules, ie, water molecules that were bound to both the ligand and the protein.<sup>1</sup> They might stabilize specific binding modes of the ligand like in the L-arabinose binding protein, where water modulates the recognition of L-arabinose and D-galactose.<sup>2</sup> On the other hand, displacement of buried water molecules is an important driving force for the binding of a ligand. This is the case, for example, of a series of nonpeptide cyclic urea inhibitors that were rationally designed and exhibited an outstanding selectivity towards the HIV protease, after displacing a highly conserved water molecule from its binding site.<sup>3</sup>

The determination of whether a certain buried water is prone to be displaced or conserved is a challenging subject that has become an important piece in the computational predictions of protein-ligand interactions.<sup>4-10</sup> In many cases, buried water molecules are tightly bound through enthalpic effects, thereby compensating the entropic penalty. On the contrary, trapped water molecules surrounded by weaker interactions will be committed to be displaced by the ligand. Despite these extreme cases, ordinary situations are blurry and difficult to classify in one of the two categories.

In most protein-ligand docking software water effects are only considered through implicit models or embedded in empirical scoring functions.<sup>11</sup> We find several examples, however, where docking predictions are significantly enhanced when water sites are explicitly considered. For example, Graaf et al. reported up to 70% of improvement in the RMSD accuracy of different docking software once explicit water molecules were included in the binding site of 19 cytochrome P450 and 19 thymidine kinase crystal structures.<sup>12</sup> As a consequence, many approaches have emerged to predict water locations inside protein cavities. Experimental-based strategies were the first natural choice to estimate water sites. Methods like Consolv<sup>13</sup>, WaterScore<sup>14</sup>, HINT/Rank<sup>15</sup> or AcquaAlta<sup>16</sup> employ structural data as hydrogen bonds, temperature B-factors, solvent-contact surface areas and protein-ligand contacts to make such estimations.

However, the strong dependence on reliable experimental/structural data has compelled the scientific community to develop *in silico* methods to locate and study water sites in protein

structures.<sup>17</sup> For example, GRID<sup>18</sup> and 3D-RISM<sup>19</sup> exhibited a prominent ability to extract potential hydration sites from protein structures.<sup>20,21</sup> Moreover, docking-based methods such as FlexX<sup>22</sup>, SLIDE<sup>23</sup>, GOLD<sup>24</sup>, Glide<sup>25</sup>, FITTED<sup>26</sup>, WaterDock<sup>27,28</sup> and Rosetta<sup>29</sup> incorporated strategies to locate hydration sites and account for a possible protein-ligand mediation through explicit water molecules. Some of them also included specialized scoring functions to contemplate water effects such as the entropic benefits upon displacement. For example, an “on and off” strategy to account for water occupancy was added to GOLD in conjunction with an updated scoring function with entropic penalties.

Nevertheless, the sampling capabilities of docking methods are limited and might fail to fully capture the contribution of water dynamics to molecular recognition. Other alternatives use Monte Carlo (MC) or molecular dynamics (MD) simulations to extend the exploration. JAWS<sup>30</sup> and GCMC<sup>31</sup> employ a MC algorithm along with a grand canonical ensemble to easily hydrate solvent-inaccessible regions by creating and destroying explicit water molecules in cavities within a user-defined grid. Other methods like WaterMap<sup>32</sup>, STOW<sup>33</sup>, SPAM<sup>34</sup>, WATsite<sup>35</sup> and WATCLUST<sup>36</sup> rely on MD simulations to perform conformational explorations of hydrated protein complexes. Besides identifying hydration sites with the clustering of the resulting water locations, some of them can determine the entropic cost of trapping water in each cavity by applying the inhomogeneous fluid solvation theory (IFST)<sup>37</sup> or derived grid-based techniques<sup>38</sup>. IFST-based methods have extensively been used to model and study the effects of buried water during the last few years. Among them, WaterMap might be the most popular and it has been successfully applied to locate and classify hydration cavities according to their thermodynamic properties in several publications; obtaining good correlations between ligand affinities and free energies of displaced water molecules.<sup>39-41</sup>

All these methods take into account the dynamics of water to predict the location and the thermodynamic properties of buried water molecules. Notwithstanding, the direct observation of their diffusion and rearrangement induced by the entrance of a flexible ligand is still a challenge owing to the large computational cost associated with these kinds of studies.<sup>42</sup> In fact, current MC and MD methods are either too simplistic or require significant computation to accurately appreciate water displacement in a reasonable amount of time.

In light of this, we decided to exploit the algorithm of PELE as it is one of the best techniques to quickly map protein-ligand interactions with implicit solvents, largely used in explorations of ligand

binding and migration processes.<sup>43-45 46,47</sup> To do so, we have developed a new MC routine for water (waterMC) that introduces an efficient mixed explicit/implicit solvent model. Thus, it allows us to dynamically consider a vast amount of explicit hydration sites while using PELE's protein and ligand conformational analysis routines. We have tested this new implementation, named aquaPELE, in three different scenarios: to infer the location of interfacial water molecules, to predict the binding mode of a ligand and to assess the displacement of water upon ligand binding. Results suggest that aquaPELE sets a fast and reliable alternative to predict the nature and position of interfacial water molecules and to study the energetic impact of their displacement along ligand binding/unbinding.

# Methodology

## PELE

PELE, which stands for Protein Energy Landscape Exploration, is a heuristic MC procedure designed to map protein-ligand interactions. Each MC step involves a complex series of events, including ligand and protein perturbation, side-chain sampling and a soft (generous convergence condition) minimization, requiring between 30 seconds and a minute on a single computing core; typical simulations involve hundreds of steps and dozens of computing cores. Currently, the energy function includes the OPLS2005 or the parmBSC0 (AMBER) force field and two different implicit solvent models: VDGBNP<sup>48</sup> or OBC<sup>49</sup>, please refer to our recent book chapter for a comprehensive explanation of the method<sup>50</sup>. To date, when water molecules were considered important, they were manually added (or kept from the crystal structure) and only sampled through the minimization phases, being mostly constrained to its initial position.<sup>51–53</sup>

Exploration capabilities of PELE are enhanced with a clustering/spawning protocol called Adaptive-PELE.<sup>54</sup> This methodology relies on short classical PELE simulations, also called epochs, whose phase space is discretized with a clustering algorithm. Thus, it detects those regions that were less explored in a certain epoch and promotes their exploration in the forthcoming epoch by spawning more conformations in the corresponding space; additional bias based on a user defined metric (distances, ligand SASA, etc.) is also possible. In this way, this technique results in a performance gain of the exploration capabilities of PELE, especially when studying the entrance pathways of a ligand or predicting its binding mode.<sup>50,53</sup>

## Implicit and explicit solvent models

In order to speed up the conformational explorations as much as possible, PELE uses implicit models to account for the solvation energy. In these models, the solvation free energy is usually considered to rely exclusively on two different terms, electrostatic and nonpolar. In PELE, the estimation of the electrostatic term relies on the Surface Generalized Born (SGB) approximation to the Poisson-Boltzmann model.<sup>55</sup> On the other hand, the influence of nonpolar terms is also considered through a penalty function that depends on the solvent accessible surface area (SASA) of each atom.<sup>56</sup> Particularly, the implicit solvent model that has been employed to validate aquaPELE is one variant that introduces a variable dielectric constant (also referred to as

VDGBNP).<sup>48</sup> Other implemented variants of the SGB model like the simplified model published by Onufriev et. al. (OBC)<sup>49</sup> have not been extensively tested with aquaPELE yet. Similar results are expected, though, owing to their similarities.

Inner water molecules that are considered explicitly use the SPC model from Berendsen et. al.<sup>57</sup> It is a three-site rigid model that sets a point charge to each atom and global Lennard-Jones parameters to the oxygen atom. Thus, explicit water molecules can be employed to account for polar interactions such as hydrogen bonds. They can also be used to stabilize certain conserved conformations of proteins or to prevent the protein from collapsing or shrinking internal hydrophilic cavities due to a lack of polar interactions. Furthermore, apart from those molecules that are defined in the water region of aquaPELE, additional explicit water molecules can be included in the simulation. In this way, explicit effects of water can be statically applied into a specific location if water is expected to remain very conserved there.

The interference between SPC water molecules and the SGB solvent is expected to have no notable effects. On one hand, SPC model is known to perform well when characterizing the role of buried water molecules.<sup>58,59</sup> On the other hand, SGB model will treat inner explicit water molecules as part of the solute in the calculation of the hydration energy as long as they remain inside the protein surface that is estimated by the implicit model. Moreover, our waterMC implementation, see below, guarantees that water molecules will remain in close contact to the protein, eliminating dewetting instances. Similar hybrid implementations have been used before in quantum mechanical calculations<sup>60</sup> or in molecular mechanics<sup>61</sup> by defining full explicit solvation shells within an implicit model. However, our hybrid implementation centers on a small number of buried water molecules, instead of modelling full solvation shells.

## aquaPELE benchmark

The performance of aquaPELE is evaluated in three different scenarios. First, we check its capacity of predicting interfacial water sites in HIV-1 protease (HIV1) and L-arabinose binding protein (ABP). These two systems have been confidently reported in the literature and show clear evidence on how interfacial water sites affect ligand binding modes. Thus, they are a convenient choice to validate the capabilities of our new implementation when predicting interfacial water locations. In this study, simulations start with the ligand in the crystallographic binding mode but

with water molecules at random initial positions around the ligand, to reduce bias to the extent possible.

Secondly, we analyze the performance of aquaPELE when performing water-wise induced-fit docking predictions. In these studies, we start a simulation with the ligand out of its crystallographic binding site and several explicit water molecules occupying random positions within the same binding cavity. For this test, we work with complexes that contain important interfacial water molecules mediating protein-ligand interactions like the Heat Shock Protein 90 (HSP90), the Bruton's Tyrosine Kinase (BTK) and the Major Urinary Protein 1 (MUP). Then, we estimate the preferential binding mode of the ligand along with its corresponding interfacial water locations.

Finally, we evaluate the capability to numerically estimate how favorable a water displacement is, in order to assess ligand optimization studies. We apply this study on the HSP90, the BTK and the Beta-secretase 1 (BACE), whose relative experimental binding free energies are available. For this inquiry, we use the same induced-fit docking protocol but some water molecules are kept fixed, out of the perturbation set of aquaPELE, in a crystallographic position. Fixed water molecules are selected to be near the common chemical core shared along all ligands in each congeneric series, thereby focusing on sampling the effects of water molecules only on regions with substantial chemical differences upon ligand binding.

All systems used in the benchmark are represented in Figure 1, where we show the principal interactions between the crystallographic ligand binding mode and the receptor. Interactions with interfacial water molecules are also displayed. Besides, experimental relative binding free energies are tagged in the corresponding ligand congeneric series, if available.

Structures from panels A and B correspond to two different inhibitors of HIV1, Kynostatin-272 and Mozenavir (PDB ids: 1HPX and 1DMP, respectively). Mozenavir was specifically designed to displace one structural water molecule of HIV1 (W301 in Figure 1A) thanks to its cyclic urea scaffold.<sup>62</sup> In the Kynostatin-272 crystal, there is a second water molecule (W608 in Figure 1A) interacting with the ligand but with a high exposure to solvent.

Structures from panels C and D belong to L-arabinose and D-galactose bound to ABP (PDB ids: 1ABE and 5ABP, respectively). The presence of the extra hydroxymethyl group in D-galactose

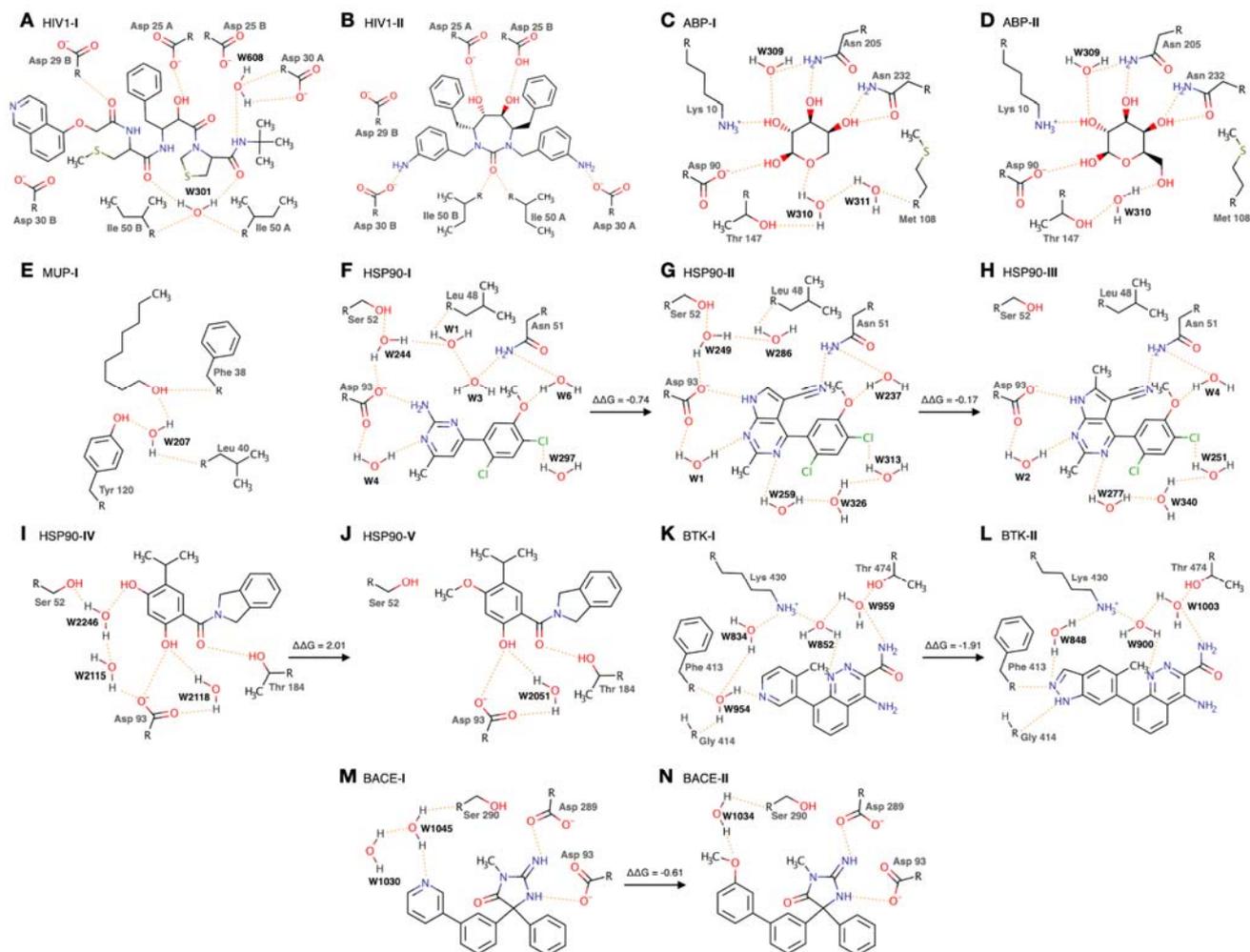
causes the displacement of W311 and the relocation of W310. A third water molecule, W309, is conserved in both ligands, interacting with one hydroxyl group.<sup>2</sup>

Structure from panel E illustrates the binding mode of nonanol in MUP (PDB id: 1ZNK). The binding site of MUP is of great hydrophobic nature with a localized polar cavity where W207 is found. Nonanol, which contains a single polar group, approaches the polar cavity through this polar group and establishes a hydrogen bond with W207.<sup>63</sup>

Structures from panels F-H and I-J show two different ligand congeneric series of HSP90 where several chemical groups are added to displace water molecules. One series is represented by panels F, G and H (PDB ids: 3RLP, 3RLQ, 3RLR). The extra pyrrole ring in G forces the displacement of W3 in F while it keeps important hydrogen bond interactions, resulting in an affinity boost. The subsequently methyl group addition entails the displacement of W249 and W286 from panel G and slightly improves the binding affinity.<sup>64</sup> In the second series, represented by panels I and J (PDB ids: 2XAB and 2XJG), the displacement of W2246 and W2115 from panel I is achieved by introducing an extra methyl group. However, the affinity of the resulting ligand is lowered as a consequence of disrupting this stabilizing water network.<sup>65</sup>

Structures from panels K and L stand for a ligand congeneric series in BTK. In this case, an indazole ring is responsible to displace W954 from panel K and maintain the stabilizing hydrogen bond with the receptor and W834. As a consequence, the compound from panel L displays a higher affinity towards BTK.<sup>66</sup>

Finally, structures from panels M and N show the displacement of one water molecule in BACE. In this case, W1045 from panel M is shifted after converting a pyridine ring to a phenyl with an additional hydroxymethyl group. This modification leads to a boost in affinity.<sup>67</sup>



**Figure 1.** Set of systems studied in the aquaPELE benchmark. It comprises binders to HIV1 (panels A and B), ABP (panels C and D), MUP (panel E), HSP90 (panels F to J), BTK (panels K and L) and BACE (panels M and N). Residues and water molecules are labelled according to the ids from corresponding crystallographic structures. Key hydrogen bonds involving interfacial water molecules are highlighted as seen in the crystallographic binding modes. The relative change in binding free energy is also reported for the series of congeneric ligands.

## Water clustering

An analysis method has been established to identify preferent hydration regions by tracking all the water locations that were obtained throughout the simulation. To achieve that, the three-dimensional space is discretized with a clustering algorithm that defines a set of centroids according to the water's oxygen atom coordinates. The analysis script uses the implementation of the Mean Shift algorithm<sup>68</sup> from the Scikit-learn library for Python<sup>69</sup>. In this way, given a set of points  $N(x)$ , and a specific window  $\lambda$ , it calculates the weighted mean,  $m(x)$ , through equation 1.

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x)x_i}{\sum_{x_i \in N(x)} K(x_i - x)} \quad (1)$$

The employed kernel function  $K(x_i - x)$  is a flat kernel with a fixed bandwidth which determines the size of the window that defines each cluster. Its function is represented in equation 2. Specifically, the bandwidth of clusters was set to  $1.5\text{\AA}$  as it is a reasonable value – according to the dimensions of the water molecule – to capture the different hydration sites.

$$k(x) = \begin{cases} 1 & \text{if } x \leq \lambda \\ 0 & \text{if } x > \lambda \end{cases} \quad (2)$$

By iteratively calculating the mean shift according to equation 1, cluster centroids are placed at high-density points given a fixed width, once convergence is reached. Therefore, each water location can be assigned to the closest cluster centroid and cluster populations are calculated by applying the simple ratio from equation 3.

$$P_i = \frac{N_i}{N_T} \quad (3)$$

$N_i$  is the number of times a perturbed water molecule has been found in cluster  $i$  and  $N_T$  is the total amount of sampled positions of water molecules in an aquaPELE simulation. As a consequence, clusters with high-density values can reveal strongly-conserved water regions, as the most visited locations of aquaPELE correspond to places where water molecules are energetically stabilized.

Besides, the prediction of hydration regions is also evaluated by considering highly-conserved water molecules in reference crystals. As a result, a number of matches between the corresponding water locations and reference positions is assigned to each snapshot. A match is conceived when a perturbed water molecule lies inside a sphere of radius  $1.5\text{\AA}$  centered on a crystallographic water. This information allows us to consider the role of water in each of the resulting conformations that aquaPELE samples.

## Ligand binding modes filtering

The estimation of the most prominent binding modes relies on the energies calculated at the OPLS2005 level of theory. Firstly, the most significant snapshots are selected by taking those that fall under the best 15% in total energy. Then, an all-atom Mean Shift clusterization is applied over the filtered ligand structures to determine the most representative binding modes, following equations 1 and 2 and using a bandwidth from 5 to 10Å, depending on the size of the ligand. Clusters with populations below 0.01 (according to equation 3) are discarded to avoid the selection of dissonant poses.

Subsequently, we look at the interaction energy of each filtered snapshot to select the most remarkable binding modes. PELE calculates the interaction energy through equation 4.

$$E_{interaction} = E_{complex} - (E_{protein} + E_{ligand}) \quad (4)$$

It consists in subtracting the energy of the unbound state of the protein,  $E_{protein}$ , and the isolated ligand,  $E_{ligand}$ , from the total energy of the complex,  $E_{complex}$ . This metric shows clear evidence about whether the ligand is stabilized in a certain pose or not. Contrarily, the total energy of the system is usually more noisy because it not only depends on the binding mode of the ligand but on the overall conformational state of the system.

Furthermore, we apply the water clustering method introduced above considering all the snapshots that belong to each cluster. Thus, for each selected binding mode, we also obtain an estimation of the preferential hydration sites.

## System preparation

Crystal structures were retrieved from the Protein Data Bank (PDB) server. Missing atoms and side chains were corrected with Maestro and its 3D Builder.<sup>70</sup> Connectivity and protonation states were determined at pH 7 with the Protein Preparation Wizard from Schrödinger<sup>71</sup>, and PROPKA<sup>72</sup>. Furthermore, OPLS2005 force field was used to parameterize protein residues and Schrodinger's `ffld_server` was employed to parameterize non-standard compounds<sup>73</sup>. Rotamer libraries for ligands were generated by defining a central core of fixed atoms and all the branches that rotate with respect to the core. The core was selected to contain a centered set of atoms with the lowest

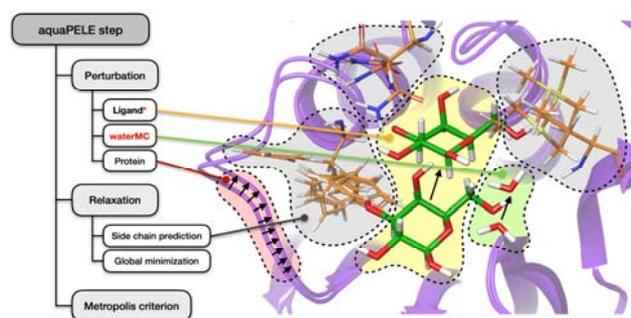
rotatable-bond eccentricity, thereby ensuring that the resulting rotamer branches contain as few consecutive rotatable bonds as possible. PELE uses these rotamer libraries in the side chain prediction algorithm.

Initial structures for PELE simulations were taken from x-ray crystals. In those cases where simulations started from a non-crystallographic conformation, a preliminar Adaptive-PELE simulation was carried out to find the exit pathway of the ligand from its crystallographic binding mode. Subsequently, one of the intermediate conformations along the resulting path was selected as the starting point of the simulation. To reduce even more the influence of the starting structure on the final simulation, a 10-step equilibration run was performed in all the cases. Finally, the structures obtained from the equilibration were clustered in order to obtain the 10 most representative conformations that were used as input for the production simulations.

## Results and discussion

### aquaPELE

aquaPELE has been developed by implementing a new internal PELE module called waterMC. The diagram in Figure 2 presents the main stages that are executed at each iterative MC step of aquaPELE and highlights in red the changes introduced in this new implementation. Thus, changes affect the perturbation algorithm with minor modifications to the ligand perturbation and the new waterMC routine, which is responsible for exploring the conformational space of a selection of explicit water molecules inside protein cavities (in a confined region). The remaining stages, protein perturbation, side chain prediction and global minimization, are kept unchanged from the standard implementation of PELE and they are described elsewhere.<sup>43,50,74</sup>

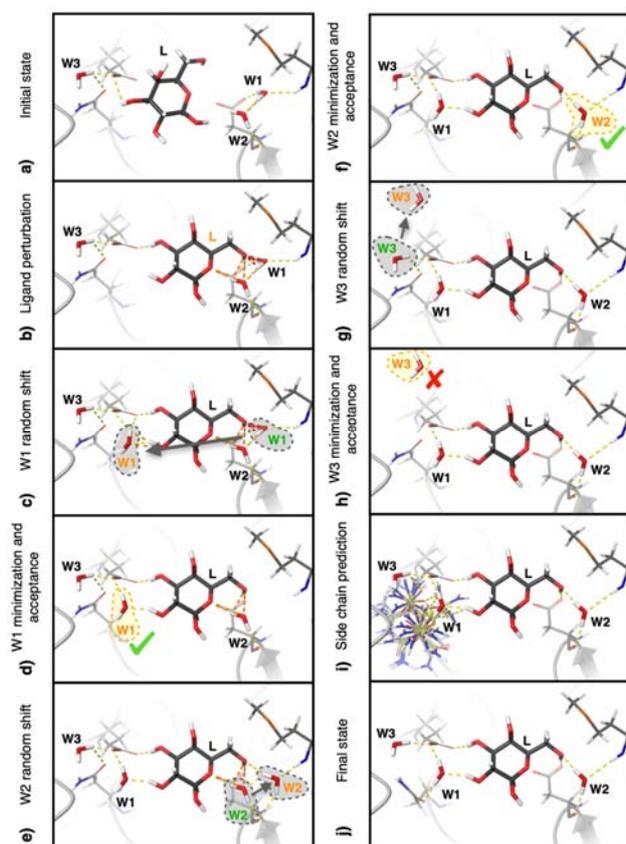


**Figure 2.** Main stages of the aquaPELE step. Changes introduced in the new implementation are highlighted in red in the diagram on the left. The image on the right exemplifies the main changes that are applied to the system throughout an aquaPELE step. It starts with the perturbation stage where the ligand and water molecules are translated and rotated to available new locations, if any, highlighted with yellow and green shadows, respectively. The protein backbone is also perturbed allowing it to follow its principal vibrational normal modes, highlighted with a pink shadow. It continues with the relaxation stage with the side chain prediction algorithm, depicted by grey areas, and a global minimization which is not represented but affects all the atoms in the complex.

The general strategy of waterMC within aquaPELE is illustrated in Figure 3. In the first place, it is convenient to distinguish the two different MC algorithms aquaPELE depends on. There is the outer MC step which is in control of exploring a new global conformation of the system applying the steps outlined in Figure 2. On the other side, there is the inner MC step which is named waterMC and explores new water locations following the steps from panels c-h in Figure 3. In this section, we will focus on the latter since it is the central core of aquaPELE. We also introduce the modifications applied to the ligand perturbation algorithm.

During the initialization, a selection of molecules is defined so as to constitute the water region (WR) of the algorithm. It is an immutable set that includes all the explicit water molecules that will be perturbed throughout the simulation. Notice that not all the water molecules that belong to the

WR are necessarily perturbed in each MC step of aquaPELE; instead, the user can choose a maximum number of water molecules to perturb per step. The resulting subset is referred to as water perturbation subset (WPS). In this way, a fixed number of water molecules is selected to constitute the WPS at each step by looping over all the WR molecules according to the user criterion. In the particular case of Figure 3, there are three water molecules defined in the WPS, labelled as W1, W2 and W3. Thus, all of them will be perturbed in the forthcoming MC step.



**Figure 3.** Illustration of aquaPELE's ligand perturbation step, applied to ligand L, along with the main stages of the forthcoming waterMC algorithm, applied to water molecules W1, W2, and W3, and the final side chain prediction. First, the ligand is perturbed in **b**, ignoring clashes with W1, W2, and W3 (as they all belong to the WPS). Next, W1, W2, and W3 are perturbed in **c-h** leading to a final state, **j**, through a side chain prediction step in **i**. waterMC sequentially applies a random shift (**c**, **e**, and **g**), minimization and acceptance verdict (**d**, **f**, and **h**) to W1, W2, and W3. During the random shift, one water molecule is taken from its initial location (highlighted in grey with a green label) to a new steric-free location (highlighted in grey with an orange label). Minimized water molecules are highlighted in yellow and the acceptance result is represented with a green tick if the new location is accepted or a red cross in case it is rejected.

An aquaPELE step still starts with the perturbation of the ligand, labelled as L in Figure 3b. The algorithm consists of applying a stochastic translation and rotation to the ligand (employing similar formulas as in equations 5 and 6, explained below). This search iteratively runs until reaching a new conformation with no clashes between ligand core atoms and the protein backbone. The remaining clashes that involve rotatable groups are subsequently relieved by a quick side chain

repositioning (including both ligand and receptor dihedrals). To this end, it employs only a steric evaluation along the rotamer sampling of a defined local region, until all steric clashes are alleviated (or 10,000 rotamer moves are satisfied). Normally, this method (ligand translation, rotation and side chain repositioning) is run multiple times (trials) to finally select the most stabilized perturbation through a total energy evaluation of the final pose in each trial. Importantly, water molecules defined in the WPS are now ignored in the computation of the total energy. In this way, the evaluation of the new conformations of the ligand does not depend on the location of the water molecules that are going to be perturbed in this same aquaPELE step, enhancing both the sampling of the ligand and the water molecules. As a result of this, a possible accepted pose for the ligand could be as in Figure 3b where there are strong clashes between the ligand and one water molecule, W1, that belongs to the WPS. Subsequently, during the upcoming waterMC step, the affected WPS molecules will have their chance to go to a new clash-free location.

waterMC runs after the ligand perturbation. Its algorithm relies on an inner MC loop with three principal stages: perturbation, minimization and acceptance evaluation. These three steps are individually applied to all the WPS molecules of the current aquaPELE step. Firstly, a random translation and rotation are applied to the first WPS molecule with the aim of finding a suitable new position. The translation is defined through a random three-dimensional unitary vector and a euclidean norm which is encompassed within a range that is calculated as in equation 5.  $T$  is the median of this range and has a default value of 5Å which allows the WPS molecule to escape from sterically trapped locations. However, lower values can be set to achieve a more local exploration.

$$||x|| = rand(0.75, 1.25) \cdot T \quad (5)$$

Apart from translating each WPS molecule, waterMC applies a random three-dimensional rotation to them. Random rotation angles within the ranges in equation 6 are assigned on roll,  $\phi$ , pitch,  $\psi$ , and yaw,  $\theta$ , axes.

$$\begin{aligned} \phi &\in [-\pi, \pi] \\ \psi &\in [-\pi, \pi] \\ \theta &\in [-\pi/2, \pi/2] \end{aligned} \quad (6)$$

This action is applied several times (up to a few thousands of trials) until finding a location without significant steric clashes; a certain overlap factor, however, is allowed between the WPS molecule and its neighboring atoms. This water perturbation phase is represented in Figure 3 with a grey arrow that goes from the initial position, in green, to the new location, in orange. Once a new location is found, a local truncated-Newton minimization, represented in Figure 3 with a yellow box, is applied only to the WPS molecule to find a more suitable orientation for it, thereby stabilizing its new pose. To this end, a single truncated-Newton minimization iteration of 30 inner Newton steps is employed to move the water molecule to the closest minimum in a fast manner. Besides, the minimization only evaluates the degrees of freedom of the current WPS molecule and the rest of the atoms in the complex are frozen.

Finally, the Metropolis criterion is employed to decide whether to keep the new location or return to the initial state. Specifically, it is applied by comparing the total energies of the system before and after running waterMC individually to each water molecule in the current WPS. Then, the overall process is applied to the next water molecule defined in the WPS, if any. For instance, this procedure admits accepting the move of one WPS molecule while rejecting the rest depending on how the energy changes upon each move. In this way, all water molecules defined in the WPS have one attempt, in the current aquaPELE step, to individually proceed throughout the perturbation, minimization and acceptance stages described above (as seen in Figure 3 for W1, W2, and W3). Results after the minimization and the acceptance criterion are represented in Figure 3, with a green tick in case it is accepted and a red cross if it is rejected.

Since the rest of the protein is frozen at this stage, reasonable acceptance for the waterMC algorithm is obtained only if working with (Metropolis) temperatures around 5000K, which admit energetic increments of 7 kcal/mol – approximately the energetic penalty of losing two hydrogen bonds – with a probability of 50%. Thus, a high acceptance temperature accepts new potential locations where water molecules might not be energetically stabilized at first. Nevertheless, their energy is substantially lowered during the forthcoming relaxation stage, involving the protein (and ligand) side chain prediction and the overall minimization. This phenomenon is seen in Figure 3 with W1. Initially, W1 is shifted to a new place in 3c and 3d where it can only participate in two hydrogen bonds. However, after running the side chain prediction in 3i it can arrange a third hydrogen bond with a neighboring glutamine residue, thereby stabilizing the new pose.

The exploration of waterMC can be restricted to a specific region for each WR molecule through the definition of perturbation boxes. Thus, a set of perturbation boxes can be defined and assigned to explicit water molecules so that each one is perturbed in a specific region. Each box is defined as a geometrical sphere by specifying a center and a radius. They should be wisely chosen to ensure that WR molecules explore the desired cavity and can not reach a vast area of the bulk solvent; waterMC performance drastically drops when this happens.

Although waterMC can be set to perturb multiple explicit water molecules at the same time, it is important to include a definite number of them. When having too few explicit water molecules, aquaPELE might fail to estimate both the correct binding mode of the ligand and the principal water sites, as all the required interactions might not be fully captured. On the other hand, a high number of water molecules can be problematic too, since each explicit water molecule added to the system introduces new degrees of freedom, thereby increasing the exploration cost. Having access to crystallographic information might help in choosing the number of water molecules needed to recognize a ligand in a certain system. Otherwise, performing several attempts with a varying number of water molecules can help us to find the best configuration. In anycase, the algorithm works better in the 1-5 range of water molecules.

Generally, the waterMC algorithm takes between 4-10 additional seconds in Intel Xeon Platinum processors each time it is executed within an aquaPELE step. The computational time is affected by the size of the system and the number of explicit water molecules that are perturbed. Although no modifications were added to other stages of the PELE step, such as the side chain prediction or the minimization, their computational times could be indirectly affected by the increment on the degrees of freedom when explicit water molecules are added to the system. However, the average time of a whole aquaPELE step is around one minute per processor, which allows to achieve speeds of 2 steps per second when running the software in parallel with 128 processors.

Note that aquaPELE simulations use a flexible model for protein residues, ligands and water molecules. Therefore, depending on the amount of conformational space to explore, a full simulation can take from less than half an hour to 1-2 days, when running in our standard setup of 128 computing cores. Short simulations of 2,000 MC steps might be enough to predict interfacial water sites when the ligand is already positioned in a valid binding mode. Longer simulations are required when more complex studies, such as water rearrangement upon binding mode change or ligand migration, are chased. In these cases, more than 100,000 MC steps might

be needed, leading to simulation times between 1 and 2 days. Thus, aquaPELE sets a fast alternative for these case studies that represent a challenge for standard unbiased MD simulations where large conformational changes like water rearrangement or ligand migration are difficult to simulate.

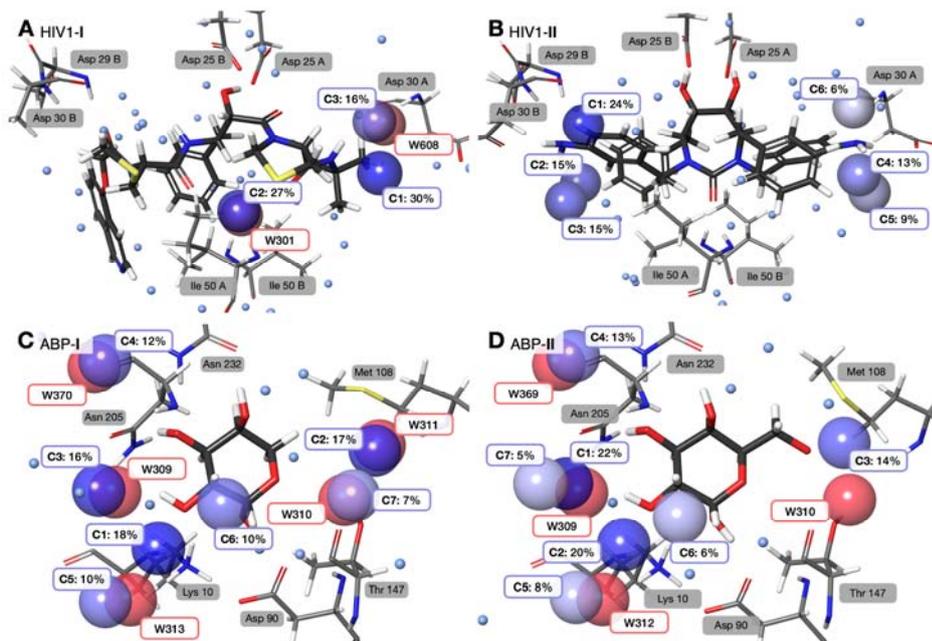
## Prediction of interfacial water sites

Interfacial water sites were estimated for both HIV1 and ABP systems throughout an aquaPELE simulation of 400 steps and 128 processors (a total of 50,800 MC steps, taking into account 127 productive processors and one that only manages the parallel interface and does not produce any trajectory). The ligand started from the crystallographic binding mode and several water molecules were randomly placed at clash-free cavities within the binding site. In order to enhance the flexibility of the system, local perturbations were also applied to the ligand, through small translations and rotations. Side-chain sampling and protein perturbation were also enabled. To reduce the bias towards the initial conformation of the system, the first 200 steps were discarded and the water clustering method was applied only to the last 200 steps of the simulation. So, the number of MC steps used for production was actually half the overall value, a total of 25,400 steps.

Water sites resulting from the binding of the two inhibitors of HIV1 were predicted with aquaPELE by including three explicit water molecules around the binding site (they are represented in Figure S1). They were allowed to move inside a spherical box of 8Å that covered the entire binding site of HIV1. As a result, a set of water clusters (depicted by small light blue spheres in Figure 4) was obtained within the targeted cavity. Clusters with populations below 5% were considered negligible and discarded. According to this criteria, three clusters can be highlighted in Figure 4A (bigger blue spheres in Figure 4). Two of them, C2 and C3, strongly overlap with crystallographic water locations, W301 and W608 (big red spheres in Figure 4). The third cluster, C1, is established next to W608, in a region which is accessible to the bulk of solvent. In the case of Mozenavir, the population distribution displays a broader spread along the binding site of HIV1, as seen in panel B. Water can barely access buried cavities such as the former W301 due to the presence of the cyclic urea of Mozenavir. As a consequence, explicit water molecules are found mostly occupying external solvent exposed locations.

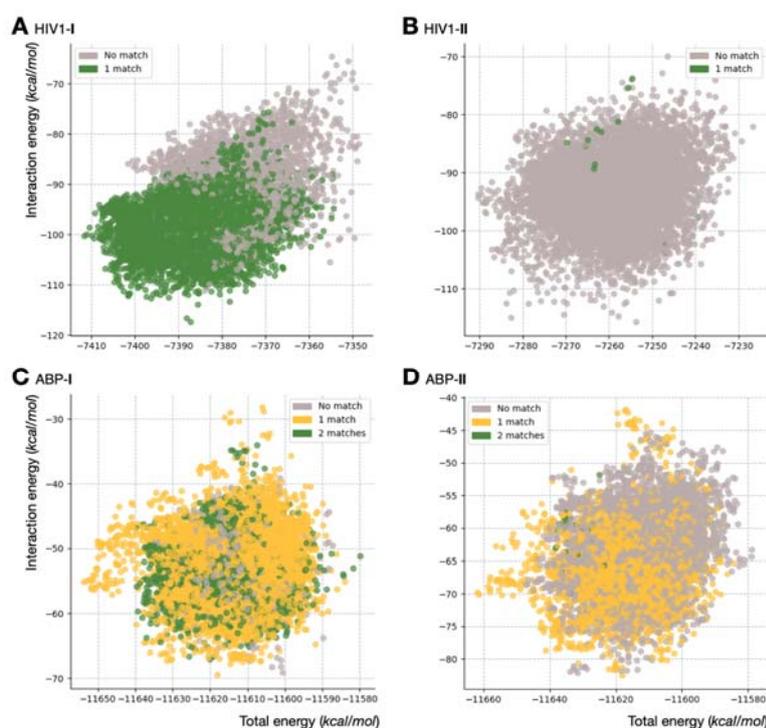
When an aquaPELE simulation with no inhibitor was set to perturb four explicit water molecules in the binding site of HIV1, three important water clusters were obtained (represented in Figure S2). Two of them are next to the crystallographic water location W301 from Figure 1A and the third is nearby Asp 25 A and Asp 25 B. This cavity belongs to the catalytic region of the enzyme and the presence of water is required for the reaction to take place.<sup>75</sup> However, once an inhibitor is bound to HIV1, this region is no longer occupied by water and, therefore, no clusters are predicted by aquaPELE at that region; as seen in panels A and B from Figure 4.

The ABP system has a more occluded binding site with few clash-free cavities. However, five explicit water molecules were inserted in this case because it is a particularly polar region where many hydration sites are expected (they are represented in Figure S3). This time, waterMC was configured with a spherical box of 7Å, centered on the binding site of ABP. Clusters C3, C7, C2, C4 obtained with L-arabinose in Figure 4C totally match with water sites W309, W310, W311 and W370, respectively. Clusters C1 and C5 are both near W313, which could indicate a more mobile water molecule in this place that even populates a nearby cluster C6. When L-galactose replaces L-arabinose almost the same clustering pattern is achieved (Figure 4D). The major difference is seen in the region occupied by the extra hydroxymethyl group of L-galactose. In contrast with the previous C2 and C7 clusters that match with W310 and W311 in panel C, only a single cluster, labelled as C3, is obtained near W310 in panel D. This difference shows how the presence of the hydroxymethyl group displaces one water molecule from the neighboring cavity.



**Figure 4.** Superposition of crystallographic water locations (big red spheres) with water sites predicted with waterMC (big blue spheres). All water clusters obtained throughout the simulation, regardless of their population, are illustrated with smaller spheres colored in light blue. The binding mode of the ligand and some relevant residues are represented from the corresponding crystallographic structure of each system. They belong to Kynostatin-272 bound to HIV1 in panel A, Mozenavir bound to HIV1 in panel B, L-arabinose bound to ABP in panel C and L-galactose bound to ABP in panel D. Crystallographic water sites and waterMC clusters along with their population percentages are labeled on top of each corresponding sphere. Intensity of blue spheres agrees with population percentages, lighter blue for low populations and darker blue for high populations.

Figure 5 displays the energetic profiles for the simulations from above. Clear evidence about the crucial mediation of W301 to the binding mode of Kynostatin-272 is seen in Figure 5A as water prevails in this location at low protein-ligand interaction energies (bottom left corner). However, water occupancy in the original cavity of W301 diminishes drastically when Mozenavir is bound to HIV1 instead, as seen in Figure 5B and in concurrence with literature.<sup>62</sup> In the case of ABP, the presence of water is generally reduced at W310 and W311 cavities when comparing Figure 5C with Figure 5D as a consequence of the extra hydroxymethyl group of galactose. Thus, while several conformations of arabinose could host two water matches at the same time, we only detect a single match in most of the conformations of galactose, in agreement with experimental evidence.<sup>2</sup>



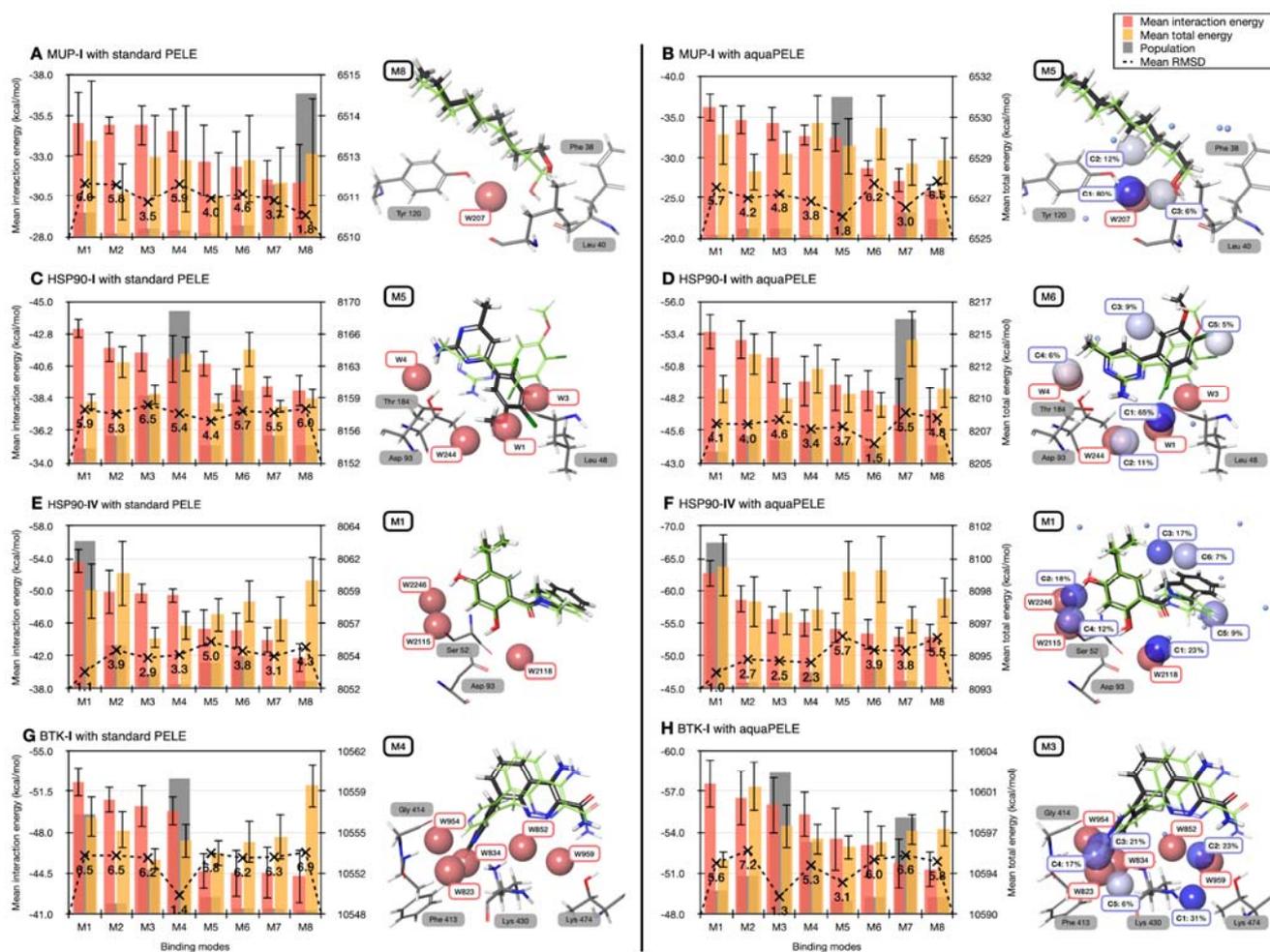
**Figure 5.** Energetic profiles obtained from aquaPELE simulations colored by the number of matches with crystallographic water locations. Plots A and B belong to HIV1 and water matches with the crystallographic W301 location are represented. Plots C and D belong to ABP and water matches with crystallographic W310 and W311 locations are represented. A match is considered when one water molecule is inside the spherical region of radius 1.5Å centered on the corresponding water location.

Although we opted for employing a high number of MC steps to ensure convergence, shorter simulations can be used in this case because of the low exploration requirements of this first application. For instance, Figure S4 displays the results that were obtained for the HIV1 system with a tenth part of steps (a total of 2,540 steps); the strong agreement displayed between the results in the long (Figures 4A and 4B) and short runs (Figures S4A and S4B), demonstrate this statement. Under these circumstances, this study can be conducted in less than half an hour.

### Water-wise induced fit docking

In this study, we examined the capacity of aquaPELE to predict the right conformation for a ligand and those interfacial water molecules involved in its recognition. In this case, initial conformations of ligands were located outside the crystallographic binding site, following the protocol introduced in the Methodology section. Figure S5 displays the structures that resulted from the equilibration run and which were used as input in the forthcoming production simulations. Then, the Adaptive-PELE algorithm was employed to run 20 epochs of 50 steps each on 128 processors (a total of 127,000 MC steps). A variable number of explicit water molecules (from 1 to 3, depending on the size of the binding site) were included at clash-free initial positions and perturbed inside a spherical box that encompassed the full binding site. Apart from analyzing them with aquaPELE, each system was also studied with the standard PELE methodology, which uses a pure implicit solvent model.

Ligand binding modes and water sites were estimated using the binding modes filtering and water clustering algorithms explained above. Figure 6 includes the eight clusters with lower interaction energies and the corresponding binding modes with lowest RMSD values for MUP-I, HSP90-I, HSP90-IV and BTK-I. The selected binding modes and the predicted water sites are compared with crystallographic structures by superimposing them in each structural panel in Figure 6. The results obtained in a fifth system, HSP90-II, are included in Figure S6. Besides, more details about the energetic profiles obtained in each simulation are displayed at Figures S7-S9.



**Figure 6.** Energetic diagrams of the eight ligand clusters with lower interaction energies and selected low-RMSD binding modes that were obtained for each system in the water-wise induced fit docking. Energetic diagrams display mean interaction and total energies (left axis values and coral vertical bars, and right axis values and amber vertical bars, respectively) with their corresponding standard deviation (black error bars). Relative population values and RMSD with respect to crystallographic structures are also shown (grey vertical bars in the background and black crosses in the foreground dashed line, respectively). Ligand clusters in the energetic diagram are sorted by their mean interaction energies from lowest to highest. Besides, the superposition of crystallographic binding modes and water locations (thin light green tubes and big red spheres, respectively) with selected binding modes and water sites obtained with aquaPELE (thick dark grey tubes and big blue spheres, respectively) is displayed. Represented binding modes belong to the structures with the lowest total energy of each low-RMSD cluster. All water clusters obtained throughout the simulation, regardless of their population, are illustrated with smaller spheres colored in light blue. Crystallographic water sites and waterMC clusters along with their population percentages are labeled on top of each corresponding sphere. Color intensity of blue spheres agrees with population percentages, lighter blue for low populations and darker blue for high populations. RMSD values are calculated in Å.

As seen in the energetic diagrams of Figure 6, the population of each cluster is generally the best hint towards identifying the conformation with the lowest RMSD value with respect to crystallographic reference. Despite the presence of interfacial water, standard PELE could detect one binding mode with an RMSD below 2Å in 4 out of the 5 systems that were tested. However, aquaPELE not only could detect the fifth binding mode in HSP90-I (structural panel in Figure 6D),

but it significantly strengthened the population of the low-RMSD cluster in HSP90-II (from A to B in Figure S6 and from E to F in Figures S7-S9). Besides, it energetically stabilizes the correct binding mode with respect to the other candidates thanks to the explicit consideration of key hydrogen bonds with interfacial water molecules. For instance, this situation is clearly observed in MUP (from A to B in Figure 6 and from A to B in Figures S7-S9), and BTK (from G to H in Figure 6 and from I to J in Figures S7-S9).

Structural panels A and B from Figure 6 show the representative structures that belong to the binding modes with the highest populations for MUP. According to its crystallographic structure, this system has a single hydration site which is highly localized in a polar cavity of the binding site, labelled as W207. The panel in 6B also shows the water clusters that resulted from employing aquaPELE to perturb 1 water molecule along the simulation. Among them, cluster C1 has the highest population and strongly matches with W207. Only when a water molecule is explicitly located at W207, the orientation of the alcohol group of nonanol can correctly be predicted .

Structural panels C and D from Figure 6 show the low-RMSD binding modes obtained in HSP90-I. In this case, aquaPELE exhibited a larger capacity to identify the proper binding mode thanks to the inclusion of 3 explicit water molecules. Contrarily, standard PELE could not detect the correct pose in the top 8 binding mode selection due to either an insufficient energetic stabilization or a too-low cluster population. HSP90-II, represented in panels A and B from Figure S6, belong to the same congeneric series as HSP90-I. In this case, standard PELE could detect successfully the correct binding mode owing to the displacement of W3 from structural panels 6C and 6D, and the affinity enhancement that it entails. Despite not bringing any structural enhancements to the binding mode prediction, running aquaPELE with 3 explicit water molecules in HSP90-II significantly increased the population in the correct cluster, as seen when comparing panels A and B in Figure S6. Furthermore, when looking at the water clusters predicted with aquaPELE for both systems, we see a sturdy agreement with crystallographic sites. The only undetected crystallographic site is W3, revealing that it might not be a key water molecule for the recognition of the right binding mode. In fact, W3 is displaced in HSP90-II leading to a strong boost in the binding affinity of the ligand. Besides, the population of cluster C1 in Figure 6D strongly diminishes in the homologous cluster C6 in Figure S6B, which suggests that the clustering algorithm might group together water molecules from both cavities W3 and W1. Thus, the displacement of W3 in HSP90-II is manifested in the decrease of population from C1 to C6. On the other hand, clusters

C3 and C5 in Figure 6D and C2, C4, C5 and C7 in Figure S6B belong to solvent accessible cavities where water presence is expected.

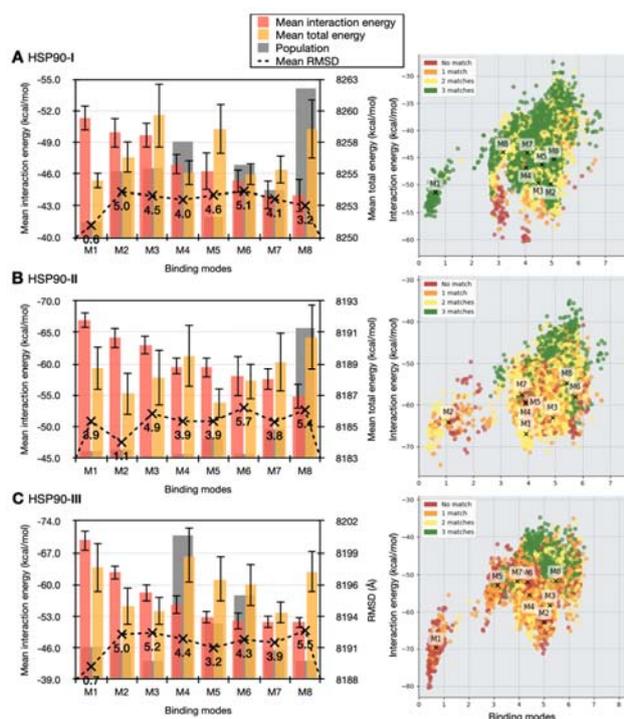
Structural panels E and F from Figure 6 display the binding modes that resulted from the HSP90-IV system. Structurally, there are no significant differences between standard PELE and aquaPELE, which indicates that PELE does not require explicit water molecules to allocate correctly the ligand. However, with the inclusion of 3 explicit water molecules, aquaPELE perfectly allocates clusters C1, C2 and C4 in crystallographic positions. C2 and C4 create a water network that bridges Asp93 and Ser52 with one alcohol group from the dihydroxybenzene ring. On the other hand, C1 interacts with Asp93, the second alcohol group of the dihydroxybenzene ring and the carbonyl group of the ligand. Water molecules are also expected to occupy clusters C3, C5 and C6 since they are located at solvent-accessible cavities.

Finally, structural panels G and H from Figure 6 correspond to the BTK system. Structures predicted with standard PELE and aquaPELE have great resemblance. In this case, aquaPELE detected 5 prominent water clusters after perturbing 2 explicit water molecules. All clusters were located in cavities with crystallographic water molecules. C1 is outside the binding site of BTK, in another inner cavity that hosts 4 water molecules in the crystallographic structure (not represented in Figure 6H). Cluster C2 is located in between W852 and W959 and acts as a hydrogen bond bridge between the ligand and the protein, stabilizing nitrogen atoms from the cinnoline ring and the terminal amide group. Although clusters C3, C4 and C5 are found in the same subcavity as W834, W823 and W954, their positions do not exactly match leading to the assumption of belonging to a very dynamic water network. Besides, a very flexible and unspecific water network meets with the fact that the binding modes predicted with both standard PELE and aquaPELE are strongly alike.

When looking at the energetic profiles that were obtained in the congeneric series of HSP90-I and HSP90-II (panels D and F, respectively, from Figures S8 and S9), we also detect a certain agreement between the relative depth of low-RMSD minima and their relative binding free energies, reported experimentally. The minimum in HSP90-II is lower and more populated than the minimum in HSP90-I, in consonance with the fact that the water displacement in HSP90-II is energetically favoured. The capacity of predicting the energetic loss or gain upon water displacement is analyzed in the third application of aquaPELE.

## Water displacement assessment

The ultimate application of aquaPELE is to estimate whether the displacement of a water molecule is energetically favourable or not. For this test, the binding of congeneric ligands was simulated, starting from non-binding conformations (using the same protocol explained in the Methodology section), with the same conditions as before (20 epochs of 50 steps in 128 processors, a total of 127,000 MC steps). Common water molecules within congeneric series were kept fixed and aquaPELE was only employed to perturb explicit water molecules in regions affected by their structural differences and crystallographic information previously known. Interfacial water molecules that were constrained are, according to Figure 1, W4 and W6 in HSP90-I, W1 and W237 in HSP90-II, W2 and W4 in HSP90-III, W2118 in HSP90-IV, W2051 in HSP90-V, W852 and W959 in BTK-I, W900 and W1003 in BTK-II, and no interfacial water molecule was constrained in BACE-I and BACE-II. On the other hand, aquaPELE was employed to perturb 2 explicit water molecules in HSP90-IV, HSP90-V, BTK-I, BTK-II, BACE-I and BACE-II and 3 of them were perturbed in HSP90-I, HSP90-II and HSP90-III. Take a look at Figure S10 to locate those water molecules that were constrained in the simulations (purple labels) and those that were tracked in the computation of crystallographic matches (green labels). Besides, some structural water molecules that did not belong to the binding site were also included. The resulting energetic profiles that belong to the congeneric series of HSP90-I, HSP90-II and HSP90-III are represented in Figure 7. For more details about the whole set of energetic profiles and clusters, please, refer to Figures S11-S14 at the Supporting information.



**Figure 7.** Energetic diagrams and energetic profiles of HSP90-I, HSP90-II and HSP90-III with the eight clusters with lower interaction energies obtained in the aquaPELE simulations for water displacement assessment. Mean interaction and total energies are included (left axis values and coral vertical bars, and right axis values and amber vertical bars, respectively) with their corresponding standard deviations (black error bars). Relative population values are also represented with grey vertical bars in the background. Mean RMSD values according to the crystallographic binding mode are displayed on top of the bars with a dashed black line and the corresponding value. Clusters are sorted by their mean interaction energies from lowest to highest. Energetic profiles represent interaction energies with respect to the corresponding RMSD difference according to reference crystals for each sampled structure. Points are classified according to the number of matches found in each structure. A match is considered when one explicit water molecule perturbed with waterMC is located inside a 1.5Å-radius sphere centered on a crystallographic interfacial water. Water sites that were tracked are W1, W3 and W244 in Figure 1F.

Figure 7 displays the number of explicit water molecules perturbed with waterMC that match with crystallographic water sites for HSP90-I, HSP90-II and HSP90-III. Water sites that were tracked are W1, W3 and W244 from Figure 1F. Thus, it is reasonable that we find 3 matches in almost all snapshots that belong to the low-RMSD cluster, M1, from HSP90-I, since all these water sites are occupied by water according to reference. Then, in HSP90-II we generally find 2 matches in the low-RMSD cluster, M2, owing to the displacement of one water molecule, W3 in Figure 1F, upon the binding of the ligand, as seen in the crystal. Finally, in HSP90-III we find no matches in the low-RMSD cluster, M1, which also agrees with the fact that, in this case, the ligand displaces the three water molecules that are being tracked. Apart from the agreement seen in the number of matches, we should also remark the trend that the depth of the low-RMSD minimum experiences along the congeneric series. When going from HSP90-I to HSP90-III, as a higher amount of water molecules is displaced, the low-RMSD minimum takes more negative values and becomes the

absolute minimum in the energetic profile, agreeing with the boost in the affinity of the ligands as experimental results point out.<sup>64</sup>

A similar behavior is seen in Figures S11 and S13 for the other systems. In HSP90-IV and HSP90-V, we tracked W2115 and W2246 sites from Figure 1I and, generally, 2 and 0 matches were obtained at the corresponding low-RMSD minimum, respectively, in agreement with reference.<sup>65</sup> In BTK-I and BTK-II, we tracked W834 and W954 sites from Figure 1K and only 1 match seems to be enough in order to identify the correct binding mode in BTK-I. On the other hand, no matches with former water sites were found in the low-RMSD cluster, M1, from BTK-II. Therefore, the displacement of W954 was predicted in agreement with reference but the crystallographic W848 site does not seem a key water site for the recognition of the correct binding mode in both BTK-I and BTK-II.<sup>66</sup> Lastly, in BACE-I and BACE-II, we tracked W1030 and W1045 sites from Figure 1M. While we see some snapshots with 2 matches with reference in the low-RMSD cluster, M6, from BACE-I, snapshots in cluster M1 from BACE-II have either 0 or 1 matches. In this way, it evidences that one water molecule gets displaced upon the binding of the ligand of BACE-II, also in concurrence with reference.<sup>67</sup>

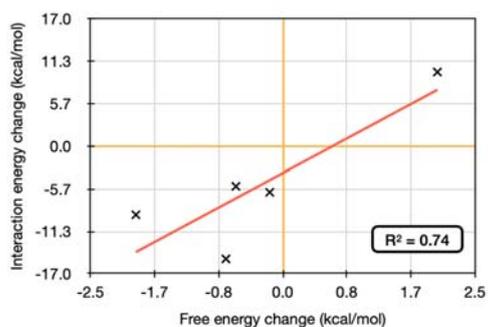
Finally, a representative interaction energy was obtained for each low-RMSD cluster, taking the 5th percentile (although similar results have been obtained in Table S1 and Figure S15 when selecting clusters according to the mean interaction energy predicted by aquaPELE instead of the crystallographic RMSD). Then, relative energetic differences between congeneric ligands were calculated and compared with experimental binding free energies (Table 1). Figure 8 shows the linear regression that was obtained from this comparison, which displays a significant correlation between both values.

Note that the energetic values computed with PELE estimate the intermolecular interaction energies between the ligand and the surrounding protein and water atoms, as presented in equation 4. Therefore, this metric only displays the energetic stability of a certain ligand conformation assuming unvarying global protein effects, and the consequences of water displacement are not directly considered. The prediction of binding free energies would require a more accurate approach, looking at the Boltzmann probabilities of each confirmation throughout the MC simulation and estimating the entropic term of displaced water molecules. Consequently, far from predicting the magnitude of binding free energies, the applicability of this method is limited, aiming to be a fast tool to distinguish favourable and unfavourable water displacements.

Despite the mentioned limitations and the small number of points, the resulting trend looks promising as it clearly differentiates non-favorable water displacements, such as the transition from HSP90-IV to HSP90-V, from favorable water displacements.

Initial structure	Final structure	Change of aquaPELE's interaction energies (kcal/mol)	Change of experimental binding free energies (kcal/mol)
HSP90-I	HSP90-II	-15.0	-0.74
HSP90-II	HSP90-III	-6.1	-0.17
HSP90-IV	HSP90-V	9.9	2.01
BTK-I	BTK-II	-9.1	-1.91
BACE-I	BACE-II	-5.3	-0.61

**Table 1.** Change of aquaPELE's interaction energies and experimental binding free energies for each transition of the congeneric series that were studied.



**Figure 8.** Linear regression between experimental relative binding free energies and difference of aquaPELE's interaction energies. The coefficient of determination  $R^2$  of such relationship is displayed in the graph.

## Conclusions

This work introduces a new algorithm called waterMC to extend the perturbation capabilities of PELE by explicitly sampling the effects of buried water in combination with a general implicit solvent model. The design of waterMC allowed us to incorporate an extra stage inside the MC core of PELE to perturb within a geometric sphere a set of explicit water molecules selected by the user. We could evaluate this new approach, referred to as aquaPELE, by employing filtering and clustering techniques to estimate the principle ligand binding modes along with the required interfacial hydration sites. Therefore, the capabilities of aquaPELE were tested when applying it to solve three different inquiries.

First, aquaPELE was used to predict the interfacial hydration sites around the crystallographic binding mode of a ligand. When evaluated on two prominent systems, HIV and ABP, it exhibited a good agreement with water molecules found in their crystals. Despite the simplicity of this first application, it was a handy approach that let us validate, structurally and energetically (as shown in Figures 4 and 5, respectively) the accuracy of our hybrid solvent model.

We further broaden the applicability of our technique by performing water-wise induced fit docking simulations. Taking advantage of the exploration power that Adaptive-PELE manifests when traversing the protein energy landscape, we studied the rearrangement of the networks of buried water upon the entrance of the ligand. As a consequence, we could incorporate the sampling of explicit water effects while migrating the ligand towards the binding site and predicting its principal binding modes. In this case, the standard version of PELE displayed a significant capacity to find a correct ligand binding mode among the best ones, despite the lack of explicit water effects. However, the inclusion and sampling of explicit water molecules stabilized the energy of the correct binding modes with respect to the others. Furthermore, it allowed to accurately locate the principal hydration regions in accordance with each binding mode, which is an important feature for drug design studies.

Finally, we decided to test aquaPELE to assess the displacement of interfacial water according to the differences resulting from energetic profiles. Although underestimating the entropic contribution, we could find a notorious relationship between the representative interaction energy of each selected binding mode and the corresponding relative experimental binding free energies. Among the five transitions that were defined, in four of them the displacement of water increased the binding affinity of the ligand and only from HSP90-IV to HSP90-V resulted in a non-favourable

process. These results fitted perfectly with experimental observations. Moreover, the comparison between interaction energies from aquaPELE and experimental free energies, represented in Figure 8, exhibited a noteworthy correlation. Thus, our approach could be a potential tool to differentiate between favourable and unfavourable water-displacement events in a fast manner. However, more effort is required to rank the entropic penalties of buried water molecules and consider their effects when performing such predictions.

## Acknowledgments

This work has been supported by a predoctoral fellowship from the Government of Catalonia (2019FI\_B\_00154 to MM). We also thank Robert Soliva, Suwipa Saenoon and Yvonne Westermaier for early discussions in the study.

## Supporting information

Fifteen figures with structures, energetic plots, and diagrams, and one table. This information is available free of charge via the Internet at <http://pubs.acs.org>

## References

- (1) Lu, Y.; Wang, R.; Yang, C.-Y.; Wang, S. Analysis of Ligand-Bound Water Molecules in High-Resolution Crystal Structures of Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2007**, *47* (2), 668–675.
- (2) Quioco, F. A.; Wilson, D. K.; Vyas, N. K. Substrate Specificity and Affinity of a Protein Modulated by Bound Water Molecules. *Nature* **1989**, *340* (6232), 404–407.
- (3) Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as HIV Protease Inhibitors. *Science* **1994**, *263* (5145), 380–384.
- (4) Karplus, P. A.; Andrew Karplus, P.; Faerman, C. Ordered Water in Macromolecular Structure. *Current Opinion in Structural Biology*. 1994, pp 770–776. [https://doi.org/10.1016/s0959-440x\(94\)90178-3](https://doi.org/10.1016/s0959-440x(94)90178-3).
- (5) Dunitz, J. D. The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science* **1994**, *264* (5159), 670.
- (6) Poornima, C. S.; Dean, P. M. Hydration in Drug Design. 1. Multiple Hydrogen-Bonding Features of Water Molecules in Mediating Protein-Ligand Interactions. *J. Comput. Aided Mol. Des.* **1995**, *9* (6), 500–512.
- (7) Poornima, C. S.; Dean, P. M. Hydration in Drug Design. 2. Influence of Local Site Surface Shape on Water Binding. *J. Comput. Aided Mol. Des.* **1995**, *9* (6), 513–520.
- (8) Poornima, C. S.; Dean, P. M. Hydration in Drug Design. 3. Conserved Water Molecules at the Ligand-Binding Sites of Homologous Proteins. *J. Comput. Aided Mol. Des.* **1995**, *9* (6), 521–531.
- (9) Ladbury, J. E. Just Add Water! The Effect of Water on the Specificity of Protein-Ligand Binding Sites and Its Potential Application to Drug Design. *Chem. Biol.* **1996**, *3* (12), 973–980.
- (10) Ball, P. Biophysics: More than a Bystander. *Nature* **2011**, *478* (7370), 467–468.
- (11) Onufriev, A. V.; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.
- (12) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. E. Binding Mode Prediction of Cytochrome p450 and Thymidine Kinase Protein-Ligand Complexes by Consideration of Water and Rescoring in Automated Docking. *J. Med. Chem.* **2005**, *48* (7), 2308–2318.
- (13) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting Conserved Water-Mediated and Polar Ligand Interactions in Proteins Using a K-Nearest-Neighbors Genetic Algorithm. *J. Mol. Biol.* **1997**, *265* (4), 445–464.
- (14) García-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: A Novel Method for Distinguishing between Bound and Displaceable Water Molecules in the Crystal Structure of the Binding Site of Protein-Ligand Complexes. *J. Mol. Model.* **2003**, *9* (3), 172–182.

- (15) Amadasi, A.; Spyraakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Mapping the Energetics of Water–Protein and Water–Ligand Interactions with the “Natural” HINT Forcefield: Predictive Tools for Characterizing the Roles of Water in Biomolecules. *Journal of Molecular Biology*. 2006, pp 289–309. <https://doi.org/10.1016/j.jmb.2006.01.053>.
- (16) Rossato, G.; Ernst, B.; Vedani, A.; Smiesko, M. AcquaAlta: A Directional Approach to the Solvation of Ligand-Protein Complexes. *J. Chem. Inf. Model.* **2011**, *51* (8), 1867–1881.
- (17) Beer, S. de; de Beer, S.; Vermeulen, N.; Oostenbrink, C. The Role of Water Molecules in Computational Drug Design. *Current Topics in Medicinal Chemistry*. 2010, pp 55–66. <https://doi.org/10.2174/156802610790232288>.
- (18) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28* (7), 849–857.
- (19) Kovalenko, A.; Hirata, F. Three-Dimensional Density Profiles of Water in Contact with a Solute of Arbitrary Shape: A RISM Approach. *Chemical Physics Letters*. 1998, pp 237–244. [https://doi.org/10.1016/s0009-2614\(98\)00471-0](https://doi.org/10.1016/s0009-2614(98)00471-0).
- (20) Henchman, R. H.; McCammon, J. A. Extracting Hydration Sites around Proteins from Explicit Water Simulations. *J. Comput. Chem.* **2002**, *23* (9), 861–869.
- (21) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. Locating Missing Water Molecules in Protein Cavities by the Three-Dimensional Reference Interaction Site Model Theory of Molecular Solvation. *Proteins* **2007**, *66* (4), 804–813.
- (22) Rarey, M.; Kramer, B.; Lengauer, T. The Particle Concept: Placing Discrete Water Molecules during Protein-Ligand Docking Predictions. *Proteins* **1999**, *34* (1), 17–28.
- (23) Schnecke, V.; Kuhn, L. A. Virtual Screening with Solvation and Ligand-Induced Complementarity. *Virtual Screening: An Alternative or Complement to High Throughput Screening?*. pp 171–190. [https://doi.org/10.1007/0-306-46883-2\\_10](https://doi.org/10.1007/0-306-46883-2_10).
- (24) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Modeling Water Molecules in Protein-Ligand Docking Using GOLD. *J. Med. Chem.* **2005**, *48* (20), 6504–6515.
- (25) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196.
- (26) 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* (2), 435–449.
- (27) 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* (3), e32036.
- (28) Sridhar, A.; Ross, G. A.; Biggin, P. C. Waterdock 2.0: Water Placement Prediction for Holo-Structures with a Pymol Plugin. *PLoS One* **2017**, *12* (2), e0172743.

- (29) Lemmon, G.; Meiler, J. Towards Ligand Docking Including Explicit Interface Water Molecules. *PLoS One* **2013**, *8* (6), e67536.
- (30) 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.
- (31) Ross, G. A.; Bodnarchuk, M. S.; Essex, J. W. Water Sites, Networks, And Free Energies with Grand Canonical Monte Carlo. *J. Am. Chem. Soc.* **2015**, *137* (47), 14930–14943.
- (32) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130* (9), 2817–2831.
- (33) Li, Z.; Lazaridis, T. Computing the Thermodynamic Contributions of Interfacial Water. *Methods Mol. Biol.* **2012**, *819*, 393–404.
- (34) Cui, G.; Swails, J. M.; Manas, E. S. SPAM: A Simple Approach for Profiling Bound Water Molecules. *J. Chem. Theory Comput.* **2013**, *9* (12), 5539–5549.
- (35) Hu, B.; Lill, M. A. WATsite: Hydration Site Prediction Program with PyMOL Interface. *J. Comput. Chem.* **2014**, *35* (16), 1255–1260.
- (36) López, E. D.; Arcon, J. P.; Gauto, D. F.; Petruk, A. A.; Modenutti, C. P.; Dumas, V. G.; Marti, M. A.; Turjanski, A. G. WATCLUST: A Tool for Improving the Design of Drugs Based on Protein-Water Interactions. *Bioinformatics* **2015**, *31* (22), 3697–3699.
- (37) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *The Journal of Physical Chemistry B*. 1998, pp 3531–3541. <https://doi.org/10.1021/jp9723574>.
- (38) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137* (4), 044101.
- (39) Beuming, T.; Farid, R.; Sherman, W. High-Energy Water Sites Determine Peptide Binding Affinity and Specificity of PDZ Domains. *Protein Sci.* **2009**, *18* (8), 1609–1619.
- (40) Myrianthopoulos, V.; Kritsanida, M.; Gaboriaud-Kolar, N.; Magiatis, P.; Ferandin, Y.; Durieu, E.; Lozach, O.; Cappel, D.; Soundararajan, M.; Filippakopoulos, P.; Sherman, W.; Knapp, S.; Meijer, L.; Mikros, E.; Skaltsounis, A.-L. Novel Inverse Binding Mode of Indirubin Derivatives Yields Improved Selectivity for DYRK Kinases. *ACS Med. Chem. Lett.* **2013**, *4* (1), 22–26.
- (41) Horbert, R.; Pinchuk, B.; Johannes, E.; Schlosser, J.; Schmidt, D.; Cappel, D.; Totzke, F.; Schächtele, C.; Peifer, C. Optimization of Potent DFG-in Inhibitors of Platelet Derived Growth Factor Receptor $\beta$  (PDGF-R $\beta$ ) Guided by Water Thermodynamics. *J. Med. Chem.* **2015**, *58* (1), 170–182.
- (42) Bodnarchuk, M. S. Water, Water, Everywhere... It's Time to Stop and Think. *Drug Discov. Today* **2016**, *21* (7), 1139–1146.
- (43) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Journal of Chemical Theory and Computation*. 2005, pp 1304–1311. <https://doi.org/10.1021/ct0501811>.

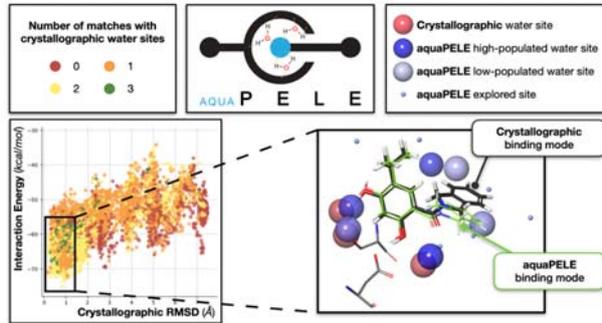
- (44) Guallar, V.; Lu, C.; Borrelli, K.; Egawa, T.; Yeh, S.-R. Ligand Migration in the Truncated Hemoglobin-II from Mycobacterium Tuberculosis: The Role of G8 Tryptophan. *J. Biol. Chem.* **2009**, *284* (5), 3106–3116.
- (45) Lucas, M. F.; Guallar, V. An Atomistic View on Human Hemoglobin Carbon Monoxide Migration Processes. *Biophys. J.* **2012**, *102* (4), 887–896.
- (46) Kotev, M.; Lecina, D.; Tarragó, T.; Giral, E.; Guallar, V. Unveiling Prolyl Oligopeptidase Ligand Migration by Comprehensive Computational Techniques. *Biophys. J.* **2015**, *108* (1), 116–125.
- (47) Kopečná, J.; de Vaca, I. C.; Adams, N. B. P.; Davison, P. A.; Brindley, A. A.; Neil Hunter, C.; Guallar, V.; Sobotka, R. Porphyrin Binding to Gun4 Protein, Facilitated by a Flexible Loop, Controls Metabolite Flow through the Chlorophyll Biosynthetic Pathway. *Journal of Biological Chemistry*. 2015, pp 28477–28488. <https://doi.org/10.1074/jbc.m115.664987>.
- (48) Zhu, K.; Shirts, M. R.; Friesner, R. A. Improved Methods for Side Chain and Loop Predictions via the Protein Local Optimization Program: Variable Dielectric Model for Implicitly Improving the Treatment of Polarization Effects. *J. Chem. Theory Comput.* **2007**, *3* (6), 2108–2119.
- (49) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55* (2), 383–394.
- (50) Gilabert, J. F.; Lecina, D.; Estrada, J.; Guallar, V. Monte Carlo Techniques for Drug Design: The Success Case of PELE. *Biomolecular Simulations in Structure-Based Drug Discovery*. 2018, pp 87–103. <https://doi.org/10.1002/9783527806836.ch5>.
- (51) Hosseini, A.; Alibés, A.; Noguera-Julian, M.; Gil, V.; Paredes, R.; Soliva, R.; Orozco, M.; Guallar, V. Computational Prediction of HIV-1 Resistance to Protease Inhibitors. *J. Chem. Inf. Model.* **2016**, *56* (5), 915–923.
- (52) Perez, C.; Soler, D.; Soliva, R.; Guallar, V. FragPELE: Dynamic Ligand Growing within a Binding Site. A Novel Tool for Hit-To-Lead Drug Design. *J. Chem. Inf. Model.* **2020**, *60* (3), 1728–1736.
- (53) Gilabert, J. F.; Grebner, C.; Soler, D.; Lecina, D.; Municoy, M.; Gracia Carmona, O.; Soliva, R.; Packer, M. J.; Hughes, S. J.; Tyrchan, C.; Hogner, A.; Guallar, V. PELE-MSM: A Monte Carlo Based Protocol for the Estimation of Absolute Binding Free Energies. *J. Chem. Theory Comput.* **2019**, *15* (11), 6243–6253.
- (54) Lecina, D.; Gilabert, J. F.; Guallar, V. Adaptive Simulations, towards Interactive Protein-Ligand Modeling. *Sci. Rep.* **2017**, *7* (1), 8466.
- (55) Ghosh, A.; Rapp, C. S.; Friesner, R. A. Generalized Born Model Based on a Surface Integral Formulation. *The Journal of Physical Chemistry B*. 1998, pp 10983–10990. <https://doi.org/10.1021/jp982533o>.
- (56) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP Hydration Free Energy Model Based on the Surface Generalized Born Solvent Reaction Field and Novel Nonpolar Hydration Free Energy Estimators. *J. Comput. Chem.* **2002**, *23* (5), 517–529.

- (57) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. *The Jerusalem Symposia on Quantum Chemistry and Biochemistry*. 1981, pp 331–342. [https://doi.org/10.1007/978-94-015-7658-1\\_21](https://doi.org/10.1007/978-94-015-7658-1_21).
- (58) Nguyen, T. T.; Viet, M. H.; Li, M. S. Effects of Water Models on Binding Affinity: Evidence from All-Atom Simulation of Binding of Tamiflu to A/H5N1 Neuraminidase. *ScientificWorldJournal* **2014**, 2014, 536084.
- (59) Hu, Z.; Jiang, J. Assessment of Biomolecular Force Fields for Molecular Dynamics Simulations in a Protein Crystal. *Journal of Computational Chemistry*. 2009, p NA – NA. <https://doi.org/10.1002/jcc.21330>.
- (60) Nikitina, E.; Sulimov, V.; Grigoriev, F.; Kondakova, O.; Lushekina, S. Mixed Implicit/explicit Solvation Models in Quantum Mechanical Calculations of Binding Enthalpy for Protein–ligand Complexes. *International Journal of Quantum Chemistry*. 2006, pp 1943–1963. <https://doi.org/10.1002/qua.20943>.
- (61) Brancato, G.; Rega, N.; Barone, V. A Hybrid Explicit/implicit Solvation Method for First-Principle Molecular Dynamics Simulations. *J. Chem. Phys.* **2008**, 128 (14), 144501.
- (62) Hodge, C. N.; Aldrich, P. E.; Bacheler, L. T.; Chang, C. H.; Eyermann, C. J.; Garber, S.; Grubb, M.; Jackson, D. A.; Jadhav, P. K.; Korant, B.; Lam, P. Y.; Maurin, M. B.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Reid, C.; Sharpe, T. R.; Shum, L.; Winslow, D. L.; Erickson-Viitanen, S. Improved Cyclic Urea Inhibitors of the HIV-1 Protease: Synthesis, Potency, Resistance Profile, Human Pharmacokinetics and X-Ray Crystal Structure of DMP 450. *Chem. Biol.* **1996**, 3 (4), 301–314.
- (63) Malham, R.; Johnstone, S.; Bingham, R. J.; Barratt, E.; Phillips, S. E.; Laughton, C. A.; Homans, S. W. Strong Solute-Solute Dispersive Interactions in a Protein-Ligand Complex. 2005. <https://doi.org/10.2210/pdb1znk/pdb>.
- (64) Kung, P.-P.; Sinnema, P.-J.; Richardson, P.; Hickey, M. J.; Gajiwala, K. S.; Wang, F.; Huang, B.; McClellan, G.; Wang, J.; Maegley, K.; Bergqvist, S.; Mehta, P. P.; Kania, R. Design Strategies to Target Crystallographic Waters Applied to the Hsp90 Molecular Chaperone. *Bioorg. Med. Chem. Lett.* **2011**, 21 (12), 3557–3562.
- (65) Woodhead, A. J.; Angove, H.; Carr, M. G.; Chessari, G.; Congreve, M.; Coyle, J. E.; Cosme, J.; Graham, B.; Day, P. J.; Downham, R.; Fazal, L.; Feltell, R.; Figueroa, E.; Frederickson, M.; Lewis, J.; McMenamin, R.; Murray, C. W.; O'Brien, M. A.; Parra, L.; Patel, S.; Phillips, T.; Rees, D. C.; Rich, S.; Smith, D.-M.; Trewartha, G.; Vinkovic, M.; Williams, B.; Woolford, A. J.-A. Discovery of (2,4-Dihydroxy-5-Isopropylphenyl)-[5-(4-Methylpiperazin-1-ylmethyl)-1,3-Dihydroisoindol-2-yl]methanone (AT13387), a Novel Inhibitor of the Molecular Chaperone Hsp90 by Fragment Based Drug Design. *J. Med. Chem.* **2010**, 53 (16), 5956–5969.
- (66) Smith, C. R.; Dougan, D. R.; Komandla, M.; Kanouni, T.; Knight, B.; David Lawson, J.; Sabat, M.; Taylor, E. R.; Vu, P.; Wyrick, C. Fragment-Based Discovery of a Small Molecule Inhibitor of Bruton's Tyrosine Kinase. *Journal of Medicinal Chemistry*. 2015, pp 5437–5444. <https://doi.org/10.1021/acs.jmedchem.5b00734>.
- (67) Cumming, J. N.; Smith, E. M.; Wang, L.; Misiaszek, J.; Durkin, J.; Pan, J.; Iserloh, U.; Wu, Y.; Zhu, Z.; Strickland, C.; Voigt, J.; Chen, X.; Kennedy, M. E.; Kuvelkar, R.; Hyde, L. A.;

Cox, K.; Favreau, L.; Czarniecki, M. F.; Greenlee, W. J.; McKittrick, B. A.; Parker, E. M.; Stamford, A. W. Structure Based Design of Iminohydantoin BACE1 Inhibitors: Identification of an Orally Available, Centrally Active BACE1 Inhibitor. *Bioorg. Med. Chem. Lett.* **2012**, *22* (7), 2444–2449.

- (68) Comaniciu, D.; Meer, P. Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002, pp 603–619. <https://doi.org/10.1109/34.1000236>.
- (69) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (70) Schrödinger, LLC, New York, NY, 2019. *Maestro*.
- (71) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27* (3), 221–234.
- (72) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.
- (73) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26* (16), 1752–1780.
- (74) Cossins, B. P.; Hosseini, A.; Guallar, V. Exploration of Protein Conformational Change with PELE and Meta-Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (3), 959–965.
- (75) Liu, H.; Müller-Plathe, F.; van Gunsteren, W. F. A Combined Quantum/classical Molecular Dynamics Study of the Catalytic Mechanism of HIV Protease. *J. Mol. Biol.* **1996**, *261* (3), 454–469.

# Table of Contents / Abstract Graphics



For Table of Contents Only