



Article

# A Fragmenting Protocol with Explicit Hydration for Calculation of Binding Enthalpies of Target-Ligand Complexes at a Quantum Mechanical Level

István Horváth <sup>1</sup>, Norbert Jeszenői <sup>2</sup>, Mónika Bálint <sup>3</sup>, Gábor Paragi <sup>4,5</sup>  and Csaba Hetényi <sup>3,\*</sup>

<sup>1</sup> Chemistry Doctoral School, University of Szeged, Dugonics tér 13, 6720 Szeged, Hungary

<sup>2</sup> Institute of Physiology, Medical School, University of Pécs, Szigeti út 12, 7624 Pécs, Hungary

<sup>3</sup> Department of Pharmacology and Pharmacotherapy, Medical School, University of Pécs, Szigeti út 12, 7624 Pécs, Hungary

<sup>4</sup> MTA-SZTE Biomimetic Systems Research Group, Dóm tér 8, 6720 Szeged, Hungary

<sup>5</sup> Institute of Physics, University of Pécs, Ifjúság útja 6, 7624 Pécs, Hungary

\* Correspondence: hetenyi.csaba@pte.hu

Received: 5 August 2019; Accepted: 4 September 2019; Published: 6 September 2019



**Abstract:** Optimization of the enthalpy component of binding thermodynamics of drug candidates is a successful pathway of rational molecular design. However, the large size and missing hydration structure of target-ligand complexes often hinder such optimizations with quantum mechanical (QM) methods. At the same time, QM calculations are often necessitated for proper handling of electronic effects. To overcome the above problems, and help the QM design of new drugs, a protocol is introduced for atomic level determination of hydration structure and extraction of structures of target-ligand complex interfaces. The protocol is a combination of a previously published program MobyWat, an engine for assigning explicit water positions, and Fragmenter, a new tool for optimal fragmentation of protein targets. The protocol fostered a series of fast calculations of ligand binding enthalpies at the semi-empirical QM level. Ligands of diverse chemistry ranging from small aromatic compounds up to a large peptide helix of a molecular weight of 3000 targeting a leukemia protein were selected for systematic investigations. Comparison of various combinations of implicit and explicit water models demonstrated that the presence of accurately predicted explicit water molecules in the complex interface considerably improved the agreement with experimental results. A single scaling factor was derived for conversion of QM reaction heats into binding enthalpy values. The factor links molecular structure with binding thermodynamics via QM calculations. The new protocol and scaling factor will help automated optimization of binding enthalpy in future molecular design projects.

**Keywords:** peptide; interaction; design; affinity; optimization; binding; water; structure; correlation

## 1. Introduction

Determination of structure and binding thermodynamics of target-ligand complexes is a key step in drug design [1]. Thermodynamic quantities can be measured by experimental methods such as isothermal titration calorimetry (ITC [2–11]). Experimental measurements are often restricted by the lack and high cost of pure and concentrated target (protein) samples. Molecular structures and binding thermodynamics can be also predicted [12–16] by fast and cheap molecular mechanics methods. At the same time, molecular mechanics has serious limitations of calculation of electronic effects in complex structures. Such effects are present in almost all intermolecular interactions including ‘exotic’ cases such as cation- $\pi$  interactions between aromatic and charged side-chains [4,17] or polarization effects at structural water molecules [18]. Quantum mechanical (QM) approaches can properly handle electronic

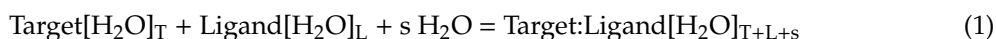
effects of intermolecular interactions. However, hydration and large size of target-ligand complexes impose further challenges on QM methods as detailed in the following paragraphs.

Hydration largely affects the structure and function of various biomolecules and their complexes [19,20]. Water molecules of the complex interface contribute to the stability and specificity of target-ligand interactions [21–28] by building hydrogen bonding networks [29,30], restraining interatomic distances, and filling cavities [19,31]. Despite their importance, determination of positions of interfacial water molecules is not trivial [32]. Available water positions have been determined mostly [33] by X-ray crystallography. However, even this well-established technique suffers from numerous limitations. Assignment of electron density peaks to possible interface water positions is still not a routine job due to inherent mobility of water and large number of degrees of freedom [34] and the quality of the structure depends on the solute size [35]. Protein hydration in the crystal is not the same as in solution [36] which is further complicated by cryo-artefacts [36]. Overfitting of electron density data and misleading identification of water sites were found to be a bad practice [25]. Other experimental techniques such as nuclear magnetic resonance spectroscopy or cryo-electron microscopy have produced a relatively small number of structures with water positions assigned. To overcome the above limitations of experimental methods, theoretical approaches were developed to help the assignment of water positions. These approaches either assign water positions based solely on solute structures [37] or involve calculation of dynamics [38–45] of water–water interactions. In the present study, a molecular dynamics-based method MobyWat [32,46] will be applied for completion of hydration structures of target-ligand interfaces.

Besides hydration, system size is another challenge of calculation of large complexes at the QM level. Such investigations would require large computer resources if the entire target molecule was calculated. A decomposition of the target-ligand complex into tractable sub-systems can handle this problem. There are at least two approaches to conduct such a decomposition. The first approach applies QM for the binding site and molecular mechanics simulations to the rest of the system [47–52]. Another branch of methods is based on skillful fragmentation of the target and applies QM for the sub-system of target fragments and the ligand. For example, Zhang and Zhang [53] developed a method for molecular fractionation where the protein is decomposed into individual capped fragments. They performed *ab initio* HF and DFT QM calculations for the target-ligand complexes. Nikitina et al. [54,55] cut the heavy atoms of the target at a distance equal or less than 5 Å from any heavy atom of the ligand. They also used structural water molecules determined by X-ray analysis, inserted new ones according to H-bonding valences of the solute molecules [54] and also proposed an iterative scheme [55] of *in silico* hydration. They developed correlations for binding enthalpy ( $\Delta H_b$ ) on sets of 8 [54], and 12 [55] complexes, respectively. The complexes included protein targets with small ligands of molecular weight (MW) up to 700 and the calculations were conducted at semi-empirical QM level using the PM3 parametrization. Dobes, Hobza et al. [56] investigated the small-molecule purine inhibitor Roscovitine in complex with cyclin-dependent kinase 2 at B3LYP/6–31G\*\* and MP2 levels of theory. They cut the chains of the kinase target into small fragments of a few amino acids at the  $C_\alpha$ -N bond. The peptide bond was maintained and they considered only amino acids and crystal water molecules located within 5 Å from the ligand.

Structure-based calculation of thermodynamic properties such as  $\Delta H_b$  is a central issue of engineering of efficient drug candidates. Enthalpic optimization of new lead molecules [57–59] is a successful pathway of drug design and requires determination or prediction of  $\Delta H_b$  of target-ligand complexes. Despite the need for  $\Delta H_b$  data, there are only a few QM studies on fragment-based calculation of target-ligand binding thermodynamics. Available studies of the previous paragraph mostly work with ligand molecules of moderate size. Complexes of large (peptidic) ligands with numerous hydration sites have not been studied extensively. Moreover, development of automated tools for extraction of structures of complex interfaces and a reliable hydration scheme would be also helpful for such fragment-based QM investigations.

A new protocol was introduced and tested in the present study to help the enthalpic design of drug candidates by answering the above challenges of automation of structure-based calculation of complexes of large ligands. For this purpose, an end-point approach was adopted for the calculation of  $\Delta H_b$  according to Equations (1) and (2). As the reaction occurs in a biological environment, T and L water molecules hydrate the target and the ligand, respectively. Waters can also remain bound to the partners ( $s = 0$ ), join the complex from the surrounding bulk ( $s > 0$ ) or leave ( $s < 0$ ) during ligand binding. The reaction heat ( $\Delta_r H$ ) of the binding process of Equation (1) can be calculated [14,15,54,55,60–62] according to Hess's law (Equation (2)), where  $\Delta_f H$  represents the calculated heat of formation of a reactant or a product as indicated in brackets.



$$\Delta_r H = \Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s) - \Delta_f H(\text{Target}) - \Delta_f H(\text{Ligand}) - s\Delta_f H(\text{H}_2\text{O}) \quad (2)$$

This end-point approach is simple and it has been successfully applied in previous publications [14,15,54,55,60–62]. In the present study, it was particularly useful for screening of various solvent models and conducting several trials in reasonable time. In the forthcoming sections, the fine-tuning of the corresponding protocol, and the development of a relationship between calculated reaction heats and experimental binding enthalpy values will be described.

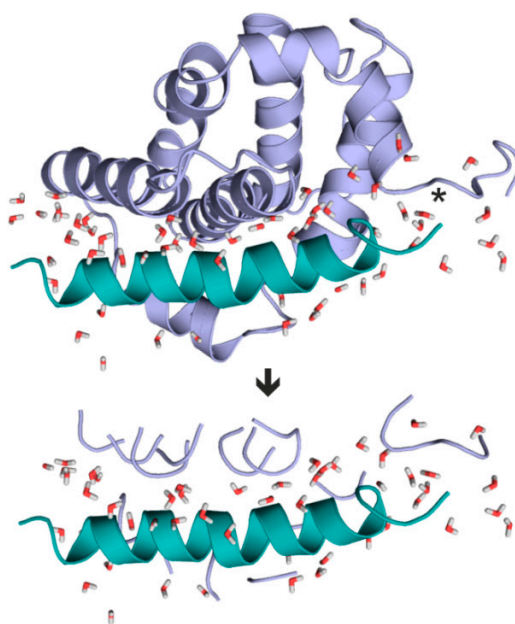
## 2. Results and Discussion

### 2.1. Fragmenter

As it was discussed in the Introduction, involving the entire target structure in a QM calculation is not feasible within a reasonable calculation time. Thus, QM calculation of the above  $\Delta_f H$  values (Equation (2)) necessitates an extraction of the interface region of the target-ligand complex. However, extraction of the complex interface and automated fragmenting of the target protein has no trivial solution. In the present study, a new protocol was elaborated including a fragmentation method, Fragmenter, to standardize the extraction of target-ligand interfaces (Figure 1). Fragmenter works on a complex structure including a target, a ligand and several water molecules. Amino acids of fragments are selected according to their intermolecular distance cut-off ( $d_{TL}$ , Table 1). A brief overview of Fragmenter and the data stream are sketched in Figures S1 and S2 and technical details are provided in Methods.

Fragmenter focuses on the neighboring parts of the target protein which have considerable interactions with the ligand and the interfacial water molecules. The whole ligand molecule and protein residues of interface regions of the complexes are extracted. The residues of the target molecule are preferably extracted as peptide fragments instead of single amino acids. The main goal is to obtain the shortest but continuous peptide chains from the target protein in a standardized way.

Thus, there is still a benefit of a considerably reduced target part, and continuity is also kept wherever it is possible. Parameter  $n$  specifies how many adjacent amino acids are added to the fragment chain of amino acids extracted according to  $d_{TL}$ . After some experimenting (Table S2), it was found that  $n = 0$  produces good correlations (as seen in the following sections), and it was not necessary to investigate  $n = 1$  for the systems of the present study. Fragmenter was implemented as a free web service (Figure S4). It provides the extracted complex interface structure (target fragments, ligand and water molecules) as an interactive image, also downloadable as PDB and Mopac input files from the 'results' tab (Figure S5) and also displays a list of estimates of per-residue intermolecular interaction energy ( $E_{inter}$ ) values to indicate unwanted close contacts.



**Figure 1.** Fragmenter extracts a hydrated interface (bottom) from the target-ligand complex (top). Target (fragments) and ligand are shown in light blue, and green, respectively. System 2roc contains the largest ligand investigated in the present study. In this example, Fragmenter extracted target residues with ( $d_{TL} = 5.0 \text{ \AA}$ ) considerably reducing the system size used for QM calculation. Interfacial water molecules ( $d_W = 5.0 \text{ \AA}$ , sticks) are also retained. Steps of extraction of target fragments are shown in atomic details for the C-terminal region (asterisk) in Figure S3 as an example. Fragmenter is available free of charge as a web service at [www.fragmenter.xyz](http://www.fragmenter.xyz).

## 2.2. Dry Systems and an Implicit Water Model

Having the Fragmenter protocol developed and implemented,  $\Delta_f H$  calculations of the (hydrated) target-ligand complex interfaces were conducted in a simplified and standardized way. Fragmenter was applied on all systems of Table 1 for extraction of the complex interfaces. All systems were prepared for Fragmenter using standard molecular mechanics energy minimization and explicit hydration protocols as described in Methods. The  $\Delta_f H$  values were calculated for the individual reactant (ligand and target fragments) and product (complex interface) structures, respectively. The calculations were performed at semi-empirical level using PM7 parameterization, with and without the Mozyme approach (Methods). The resulted, raw energy values are listed in Table S4.

Within the end-point approach (Introduction), calculation of  $\Delta_f H$  of the reaction participants (Equation (2)) and a linear scaling (Equation (3)) of  $\Delta_r H$  to known experimental  $\Delta H_b(\text{exp})$  values is necessary for calculation of  $\Delta H_b$ .

$$\Delta H_b(\text{exp})_i = \alpha \Delta_r H_i + \beta + \varepsilon_i = \Delta H_b(\text{calc})_i + \varepsilon_i, \text{ where } i = 1, 2, \dots, N \quad (3)$$

In the present study, 15 systems ( $N = 15$ ) of Table 1 were involved in the derivation of regression coefficients ( $\alpha$ ,  $\beta$ ) yielding  $\Delta H_b(\text{calc})$  values and residuals ( $\varepsilon$ ). Statistical parameters obtained for the dry complexes and various solvent models are listed in Table 2. Nine of the 15 systems with small ligands up to a MW of 550 were considered in a previous paper [55] as well. In the present study, additional six systems with large peptide ligands were included in the set as they often impose a challenge during lead optimizations due to their size and extensive hydration. Thus, the set of 15 systems involves various ligands with MW up to 3318, two orders of magnitude larger than the previous set. The experimental  $\Delta H_b$  values cover a wide range between  $-2.935$  and  $-15.5 \text{ kcal/mol}$  (Table 1).

**Table 1.** Target-ligand systems.

System <sup>a</sup>	Res <sup>b</sup> (Å)	Target	Ligand		Water Count			$\Delta H_b(\text{exp})$ <sup>d</sup>
			Name	MW <sup>c</sup>	Shell 1	Shell 2	Shell 3	kcal mol <sup>-1</sup>
3ptb_ben	1.7	beta-trypsin	benzamide	121.2	1	6	7	-4.507 [2]
3ptb_pme	1.7	beta-trypsin	p-methylbenzamide	135.2	1	5	6	-4.412 [2]
3ptb_pam	1.7	beta-trypsin	p-aminobenzamide	136.2	3	4	7	-6.417 [2]
3ptb_pmo	1.7	beta-trypsin	p-methoxybenzamide	151.2	1	6	7	-3.742 [2]
3ptb_pad	1.7	beta-trypsin	p-amidinobenzamide	164.2	2	8	10	-2.935 [2]
1k1l	2.5	bovine trypsin	NAPe-piperazine	467.6	5	10	15	-7.863 [4]
1k1m	2.2	bovine trypsin	NAP <sup>e</sup> -4-acetyl-piperazine	508.6	4	12	16	-8.222 [4]
1k1i	2.2	bovine trypsin	NAP <sup>e</sup> -D-pipecolic acid	508.6	2	13	15	-10.899 [4]
1k1j	2.2	bovine trypsin	NAP <sup>e</sup> -isopipecolic acid methyl ester	523.6	3	13	16	-9.465 [4]
1jyr	1.55	Grb2 SH2 domain	APS-PTR <sup>e</sup> -VNVQN	1069.0	1	14	15	-7.94 [6]
1rlq	NA	C-src tyrosine kinase SH3 domain	RALPPLPRY	1084.3	2	25	27	-10.2 [7]
2ke1	NA	autoimmune regulator	ARTKQTARKS	1150.3	12	15	27	-9.2 [8]
2bba	1.65	EphB4 receptor	NYLFSPNGPIARAW	1606.8	12	15	27	-15.5 [9]
1jgn	NA	human poly(A)-binding protein	VVKSNLNPNNAKEFVPGVKYGNL	2389.8	14	34	48	-14.8 [10]
2roc	NA	induced myeloid leukemia cell differentiation protein homolog	EEEWAREIGAQLRRIADDLNAQYERRM	3317.6	14	38	52	-14.3 [11]

<sup>a</sup> System codes are derived from the PDB identifiers, and abbreviated ligands names (where applicable).

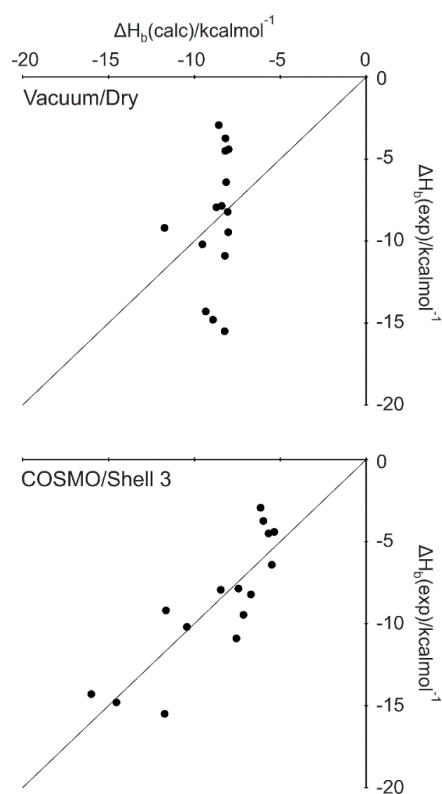
<sup>b</sup> Resolution (available for crystallographic structures). <sup>c</sup> Molecular weight. <sup>d</sup> Experimental binding enthalpy values are given at their original level of precision except those with three decimal digits converted from kJmol<sup>-1</sup>, where 1 J = 4.184 cal. Sources of values are indicated as references in superscript. <sup>e</sup> NAP: N-alpha-(2-naphthylsulfonyl)-N-(3-amidino-L-phenylalaninyl); PTR: o-phosphotyrosine.

In the first step of the present investigations, no solvent models were applied ( $s = 0$  in Equations (1) and (2)). That is, dry input structures without explicit water molecules were calculated in vacuo. The complete lack of water models resulted no correlation between the calculated and experimental  $\Delta H_b$  values (column Vacuum/Dry in Table 2, Figure 2). The application of an implicit water model (CONductor-like Screening MOdel, COSMO [63]) increased the correlation (column COSMO/Dry in Table 2). However, this correlation can still be improved as reflected by the cross-validation. In general, the use of COSMO proved advantageous if compared with the vacuum/dry results (Table 2). There was a single case of System 2ke1 where  $\Delta H_b(\text{exp})$  could not be converted to 298.15 K and the original value at 296.15 K (Table S3) was used for the regressions of Table 2. To check the influence of this data point on the results, linear regressions were performed without System 2ke1, as well. The statistical parameters showed (Table S5) that leaving out System 2ke1 did not improve the results in vacuo and COSMO yields considerable correlation.

**Table 2.** Per-system residuals ( $\epsilon$ ) and statistical parameters of linear regressions obtained with different water models.

System	Vacuum				COSMO				
	Dry	Shell 1	Shell 2	Shell 3	Dry	Shell 1	Shell 2	Shell 3	Shell 3 <sup>b</sup>
	$ \epsilon ^a$								
3ptb_ben	3.70	3.18	2.90	1.93	3.73	2.33	2.45	1.18	0.85
3ptb_pme	3.60	3.09	3.05	2.03	0.53	1.12	2.48	0.95	1.22
3ptb_pam	1.74	1.29	0.88	0.03	0.01	0.02	0.44	0.92	3.03
3ptb_pmo	4.45	3.91	3.71	2.64	3.59	3.25	3.32	2.24	0.33
3ptb_pad	5.65	5.11	5.04	4.25	3.03	3.12	4.32	3.22	1.37
1k1l	0.56	0.39	0.10	0.10	2.59	1.05	0.56	0.43	1.74
1k1m	0.16	0.56	0.36	0.79	2.71	1.17	0.75	1.50	3.12
1k1i	2.67	3.19	2.95	3.51	2.80	3.62	2.88	3.34	4.60
1k1j	1.43	1.81	1.60	2.10	0.57	2.60	1.47	2.32	3.75
1jyr	0.78	0.28	0.53	0.09	1.73	0.36	0.40	0.53	0.36
1rlq	0.67	0.64	0.27	0.33	0.37	2.13	0.61	0.24	0.16
2ke1	2.54	4.67	4.66	5.28	4.38	4.21	5.73	2.46	2.89
2bba	7.25	6.38	7.21	6.23	4.34	2.13	6.58	3.77	3.31
1jgn	5.88	5.24	4.75	2.56	1.61	0.27	3.25	0.25	1.36
2roc	4.96	4.11	3.74	1.26	2.76	1.24	3.05	1.71	3.92
R <sup>2</sup>	0.06	0.18	0.19	0.44	0.51	0.65	0.33	0.73	0.93
R <sup>2</sup> (cv) <sup>c</sup>	0.00	0.01	0.02	0.22	0.34	0.54	0.07	0.65	0.91
F	0.81	2.77	3.14	10.20	13.46	24.28	6.36	34.55	179.66
RMSE <sup>a</sup>	4.02	3.76	3.72	3.10	2.90	2.45	3.40	2.17	2.65
t <sub><math>\alpha</math></sub>	0.90	1.66	1.77	3.19	3.67	4.93	2.52	5.88	13.40
t <sub><math>\beta</math></sub>	-5.56	-5.04	-4.68	-3.90	-2.18	-3.99	-4.24	-2.81	-

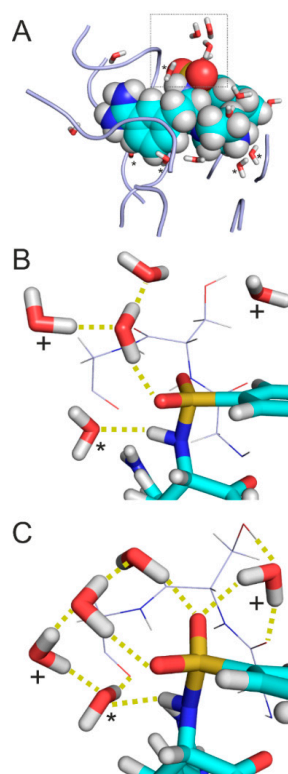
<sup>a</sup> Unit: kcalmol<sup>-1</sup>. <sup>b</sup> Linear regression with  $\beta = 0$  (last column), and  $\beta \neq 0$  (other columns). <sup>c</sup> Leave-one-out cross-validated coefficient of determination.

**Figure 2.** Correlation plots obtained without (Vacuum/Dry) and with (COSMO/Shell3) the hybrid water model.

### 2.3. Explicit Hydration and a Hybrid Model

A systematic investigation on explicit hydration was conducted to further improve the correlations of the previous section. It is challenging to give a straightforward definition for the origin of water molecules in the complexes, and prediction of ligand-bound water molecules is rather uncertain due to the relatively small binding surface of ligands. Thus,  $T = L = 0$  was set and all interface water molecules were considered as if they had originated from the surrounding bulk solvent ( $s > 0$ , in Equations (1) and (2)). Hydration structure of the target-ligand complex was built up by the MobyWat method [32] and extracted by Fragmenter as part of the interfaces. MobyWat can produce complete, void-free hydration structures of complex interfaces. This is guaranteed by a soaking step during the systematic evaluation of a series of snapshots of molecular dynamics simulations accounting for water–water interactions besides solute–water ones. Thus, MobyWat can find all experimental reference water positions in many cases [32] and assign water positions not detectable by experimental [25,33,34,64,65] measurements.

In the present study, three shells were defined according to  $d_w$  (Table 1 and Table S1) using interfacial water molecules (Figure 3A). Shell 1 contains water molecules closest to the solutes ( $d_w = 3.5 \text{ \AA}$ ). Shell 2 holds waters with intermediate positions ( $3.5 \text{ \AA} < d_w < 5.0 \text{ \AA}$ ). Shell 3 consists of all interfacial water molecules of Shells 1 and 2 with a  $d_w = 5.0 \text{ \AA}$ .



**Figure 3.** Extracted complex interface of System 1k11. (A) Initial structure equipped with water molecules and energy-minimized at the molecular mechanics level. Target fragments and ligand are shown in ribbon and space filling representations, respectively. Water molecules in Shell 1 ( $d_w = 3.5 \text{ \AA}$ , sticks marked with asterisk) are positioned close to the solute partners and play a bridging role. The rest of Shell 3 ( $d_w = 5.0 \text{ \AA}$ ) waters belong to Shell 2 (sticks without asterisks) and located at the edges of the interface, close to the bulk. Shell 3 = Shell 1 + Shell 2. (B) A rotated close-up of the box in Panel A showing the surrounding of the sulphonyl group of the ligand (sticks) and the neighboring residues  $G_{216}S_{G_{218}}$  of the target (lines) where the numbering follows that of the crystallographic structure (PDB ID 1k11). Hydrogen bonds are marked with yellow dotted lines. (C) Structure in Panel B after relaxation at semi-empirical level using PM7 parameterization and Mozyme. Water molecules with a displacement above  $1.5 \text{ \AA}$  after relaxation are marked with crosses.

The use of explicit water molecules in vacuum improved the 'dry' correlations to an  $R^2$  of 0.44 (all systems, column Vacuum/Shell 3 in Table 2) and 0.65 (without System 2ke1, Table S5). Cross-validation indicates that this improvement of correlation is robust only without System 2ke1. At this point, it seemed reasonable to check whether a hybrid model using both implicit and explicit hydration further improves the correlation. Indeed, the hybrid model (column COSMO/Shell 3 in Table 2) provided the best agreement between calculated and experimental  $\Delta H_b$  with an  $R^2$  of 0.73 (Figure 2) using all interfacial water molecules. Notably, Shell 1 waters also yielded considerable correlation ( $R^2$  of 0.65). In both cases, the correlations survived the challenge of cross-validation. The intermediate water positions alone (Shell 2) yielded a stable correlation only without System 2ke1 (Table S5).

To investigate the effect of ligand size and target diversity on the stability of the above correlation (COSMO/Shell 3), the set of systems in Table 1 was split into two sub-sets according to ligand MW. The first sub-set contains nine systems with small ligands of MW < 600. All these ligands have a common target, beta trypsin. The second sub-set contained six systems with large ligands of MW > 1000 and various targets. Linear regressions were performed separately for the two sub-sets and  $\Delta H_b(\text{calc})$  values were calculated by the two regression equations, respectively. Overall statistical parameters obtained (Table S6) were comparable to those of the regression for all systems (column COSMO/Shell 3 in Table 2) detailed above. Thus, stability of the correlations is not influenced by ligand size and target diversity of the systems in the case of the hybrid model.

#### 2.4. Scaling Factor

The above COSMO/Shell 3 model with  $\beta \neq 0$  in Equation (3) is significant and robust regarding its overall regression parameters. However, the  $t_\beta$  value (Table 2) indicates that the level of significance of regression coefficient  $\beta$  is moderate ( $p = 0.015$ ). Thus, a linear regression with  $\beta = 0$  was also developed and the corresponding statistical parameters are listed in the last column of Table 2 (Table S7). In this way, a model of high significance ( $p < 0.01$ ) of all parameters was obtained and Equation (3) was simplified. The resulting Equation (4) includes only the value of regression coefficient  $\alpha$ , which serves as a single, unit-independent scaling factor for conversion of calculated  $\Delta_r H$  into  $\Delta H_b$ .

$$\Delta H_b = 0.031 (\pm 0.002) \Delta_r H \quad (4)$$

A similar value of 0.032 ( $\pm 0.002$ ) was obtained for the scaling factor if System 2ke1 was not involved in the regression. Via QM calculations, this factor serves as a direct link between molecular structure and binding thermodynamics of molecular complexes.

#### 2.5. Case Studies on Hydration Structures

In two-thirds of the 15 systems, application of Shell 1 or 3 explicit water molecules resulted in the decrease of residuals (COSMO models in Table 2). Shell 2 waters have similar effect in one-third of the cases. For example, in the case of System 1k11, the residuals decreased from 2.59 (dry) to 0.43 (Shell 3,  $\beta \neq 0$ ) and 1.74 kcal/mol (Shell 3,  $\beta = 0$ , Table 2), and a similar trend can be observed for the vacuum values.

In the interface of System 1k11 extracted after molecular mechanics energy-minimization (Figure 3A), Shells 1 and 2 contain 5 and 10 water molecules, respectively (Table 1). The water molecules of Shell 1 (Figure 3A) are located at the bottom of the interface bridging between the target and ligand (solute) partners. Shell 2 waters mostly occur at the opening of the interface towards the bulk (right side of Figure 3A) waters/region. As it was expected, large clusters of waters gathered around charged or polar groups. For example, the sulfonyl group (Figure 3B) of the ligand is surrounded by a group of water molecules, and only one of them belongs to Shell 1. No interactions were observed between the waters and the closest target fragment ( $G_{216}SG_{218}$ ).

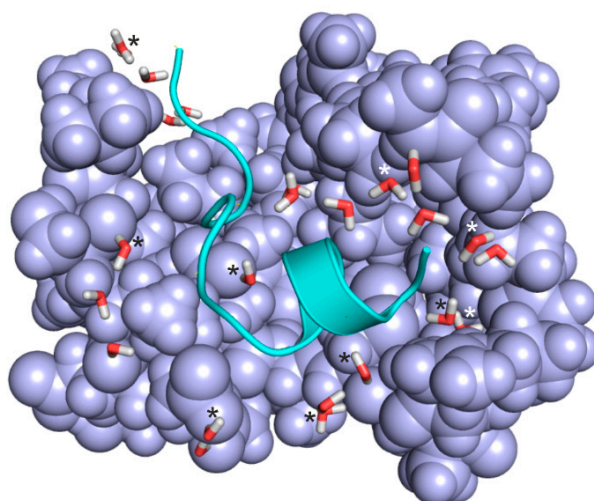


During semi-empirical QM relaxation (Figure 3C), positions and orientations of some water molecules were changed. For example, two water molecules (marked with crosses in Figure 3C) were shifted by 3.2 and 1.8 Å. The orientation of Shell 1 water molecule (marked with asterisk in Figure 3C) was changed to interact with the target fragment. Such changes resulted in an extensive H-bonding network of water molecules stabilizing the target-ligand interaction around the sulfonyl group. Formation of new hydrogen bonds imply that some of the Shell 2 water molecules became Shell 1 (not marked in Figure 3C).

While the hydration structure underwent a remarkable transformation during semi-empirical QM relaxation, the conformation of the target fragment was preserved. The above example of System 1k1l (Figure 3) showed how water molecules in the different shells contribute to the completion of the target-ligand interface structure and a consequent decrease in residuals of calculated  $\Delta H_b$ .

Besides small, rigid ligands like the phenylalanine derivative of System 1k1l, large peptide ligands were also involved in the present study. For example, System 2bba (Figure 4) has a penta-decapeptide ligand (Table 1) and a relatively extensive hydration structure of 27 water molecules in the extracted interface. In the case of 2bba, the largest decrease from 4.34 to 2.13 kcal/mol of the residual (COSMO models in Table 2) was obtained with Shell 1 water molecules. A detailed overview of the hydration structure shows that water molecules of Shell 1 (asterisks in Figure 4) mostly positioned at the bottom of the binding pocket and play a bridging role between the target and ligand partners. In this case, application of Shell 2 waters in addition to Shell 1 ones was not beneficial as they increased the residual. However, the final residual with Shell 3 is still below the dry model.

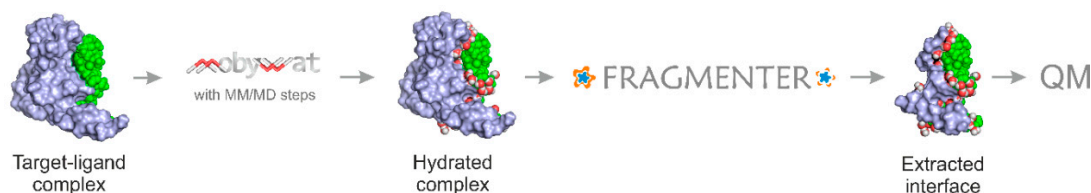
Beyond bridging and space filling roles presented in Figures 3 and 4, interfacial hydration also exerts a shielding effect [66] on target-ligand intermolecular interactions, as well. Despite the importance of the hydration structure, crystallography often does not supply crucial water positions or erroneously assigns waters in close contact (see also Introduction). This leads to limitations of the use of experimental complex structures in drug design.



**Figure 4.** Extracted complex interface of System 2bba after relaxation at semi-empirical level using PM7 parameterization and Mozyme. Target fragments and ligand are shown in light blue space filling and green cartoon representations, respectively. Water molecules (sticks) in Shell 1 are marked with asterisks. Non-marked waters belong to Shell 2.

The present study has overcome such limitations of experimental determination of hydration structures, and calculation of  $\Delta H_b$  was possible using complete interfacial hydration structures resulted exclusively by MobyWat calculations (see Methods). Besides hydration structures, missing ligand positions of four Systems (3ptb\_pad, 3ptb\_pam, 3ptb\_pme, 3ptb\_pmo) were also produced by computational modeling. Thus, modeling provided atomic resolution data reliably completing

experimental structures and yielding robust correlations of the present study. Notably, modeling steps (Figure 5) of building the hydration structure and the full complex require only moderate computational resources, and can be accomplished on a single workstation. With the application of a parallelized MD engine and a supercomputing facility, the calculation time can be reduced to a couple of hours. The Fragmenter step takes some seconds.



**Figure 5.** Modeling steps used for preparation of hydrated target-ligand interface for QM calculations shown on the example of System 1jgn. The procedure starts from the complex of the target (grey) and ligand (green) molecules. Program MobyWat [32,46] provides accurate hydration structure with MD-based calculations of positions of water molecules (red and white) in the interface. MobyWat is downloadable free of charge at [www.mobywat.com](http://www.mobywat.com). Finally, Fragmenter, a web service extracts the complex interface with considerably reduced target part for subsequent use in QM calculations. Fragmenter was introduced in the present study and available at [www.fragmenter.xyz](http://www.fragmenter.xyz).

### 3. Methods

#### 3.1. Preparation of Complexes

The primary input structures of all systems (Table 1) were obtained from the Protein Databank (PDB [67]). All crystallographic water molecules were removed. Missing atoms of solute side chains (both protein and ligand) were reconstructed with Swiss PDB Viewer [68]. In the case of missing terminal and non-terminal amino acids, acetyl and amide capping groups were added with the Schrödinger Maestro program package v. 9.6 [69] to the N- and C-terminus, respectively. In cases of homodimer structures, chain A was selected for calculations.

#### 3.2. Parameters of Non-Amino Acid Ligands

For non-standard (non-amino-acid) ligands or residues molecular mechanics force field parameters were obtained from the GAFF force field [70]. Considering a non-standard residue, it was first capped on both terminals, with Ace- and -NHMe groups and pre-minimized with PC Model 9 [71] using MMFF94 force field [72]. Subsequently, semi-empirical quantum mechanics optimization was performed with MOPAC-2009 [73] using the PM6 parameterization with a 0.001 gradient norm [74]. In all cases, the force constant matrices were positive definite. Then, the completely minimized molecules were uploaded to RED server [75] to perform ab initio geometry optimization to obtain partial charges by RESP-A1B charge fitting (compatible with the AMBER99SB-ILDN force field). The calculations were performed with the Gaussian09 software [76], using HF/6-31G\* split valence basis set [77]. The caps on the termini were excluded from charge derivation, charge restraints were applied on these atoms. Normal mode analysis was performed using GAMESS [78] to ensure that the final geometry is in energy minimum. Bond stretching, angle bending, and torsional parameters were assigned with the parmchk utility of AmberTools 1.5 [79] and used together with the partial charges to build GROMACS [80,81] residue topology entries for the non-standard residues.

#### 3.3. Calculation of Interfacial Hydration Structure

MobyWat [46] predictions along with GROMACS MD simulations were used for calculation of water positions in the target-ligand complex. A uniform procedure was followed based on Method 3 of a previous study [32] briefly described in the following points. An overview of the modeling steps described in the forthcoming sections is provided in a flow chart of Figure 5.

### 3.4. Molecular Mechanics Energy-Minimization during MobyWat Predictions

For pre-MD minimization, the target or complex structure was placed in a cubic box using a distance criterion of 1 nm between the solute and the box. Void spaces of the box were filled up by explicit TIP3P water molecules [82] with the standard gmx solvate routine of GROMACS. Counter-ions (sodium or chloride) were added to neutralize the system. A uniform procedure was applied in all cases prior to the MD steps, including a steepest descent (sd) followed by a conjugate gradient (cg) step. Exit tolerance levels were set to  $10^3$  and  $10 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-1}$  while maximum step sizes were set to 0.5 and 0.05 nm, respectively. Position restraints were applied on solute heavy atoms at a force constant of  $10^3 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ . All calculations were performed with programs of the GROMACS software package [81], using the AMBER99SB-ILDN force field [83]. The above energy-minimization was performed twice, once for the target and once for the re-assembled target-ligand complex (see below).

### 3.5. Molecular Dynamics of the Protein Target

After energy-minimization, 5-ns-long NPT MD simulations were carried out with a time step of 2 fs. For temperature-coupling the velocity rescale [84] and the Parrinello–Rahman algorithm were used. Solute and solvent were coupled separately with a reference temperature of 300 K and a coupling time constant of 0.1 ps. Pressure was coupled by the Parrinello–Rahman algorithm [85–87] and a coupling time constant of 0.5 ps, compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  and reference pressure of 1 bar. Particle Mesh-Ewald summation was used for long range electrostatics. Van der Waals and Coulomb interactions had a cut-off at 11 Å. Coordinates were saved at regular time-intervals of 1 ps yielding  $1.001 \times 10^3$  frames. Position restraints were applied on solute heavy atoms at a force constant of  $10^3 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ . Periodic boundary conditions were treated before analysis to make the solute whole and recover hydrated solute structures centered in the box. Each frame was fit to the original protein crystal structure using C $\alpha$  atoms. The final trajectory including all atomic coordinates of all frames was converted to portable binary files. The target structure, and the surrounding (surface) water molecules were extracted as the last frame of the 5-ns-long MD simulation. At this point, there is a difference between the present study and Method 3 applied previously [32]. In Method 3, surface water molecules had been provided by MobyWat using 1-ns-long MD simulation. In the present study, the final frame of a 5-ns-long MD simulation was applied.

### 3.6. Re-Assembly of the Target-Ligand Complex

The target-ligand complex was re-assembled. For this, the target part of the holo and the hydrated apo systems were fitted on the top of each-other and the ligand was used together with the hydrated target (soaking), and interfacial water molecules were extracted. A water molecule was considered interfacial if intermolecular distance was smaller than/equal to a pre-defined maximal threshold ( $d_{\text{max}}$ ) of 5 Å for both the ligand and target partners. Water molecules conflicting with the ligand structure were excluded using the editing mode of MobyWat at a minimum distance limit ( $d_{\text{min}}$ ) of 1.75 Å prior the second MD simulation.

### 3.7. Molecular Dynamics of the Target-Ligand Complex

The MD simulation protocol described above for protein targets was performed for the re-assembled target-ligand complex structure, as well. In this case, all frames of the final trajectory of the target-ligand complex (in a water box) were used in the next step for production of interfacial water positions.

### 3.8. Production of Interfacial Water Positions

After the MD simulation of the target-ligand complex, MobyWat prediction of interfacial water positions was performed with  $d_{\text{max}}$ , clustering and prediction tolerances of 5.0, 3.0, and 3.0 Å, respectively. The MER clustering algorithm of MobyWat was applied. At this point, the present

procedure differs from Method 3 [32]. As a result, a list of predicted water oxygen positions was produced by MobyWat in PDB format.

### 3.9. Molecular Mechanics Energy-Minimization after MobyWat

The MobyWat-supplied oxygen atoms of predicted water positions were equipped with hydrogen atoms and energy minimization was performed for the hydrated complexes. A four-step protocol was applied for energy minimization of complexes with predicted water positions following an sd-cg-sd-cg pattern with parameters of sd and cg methods described above. During the first two steps, all solute heavy atoms and the oxygen of the predicted interfacial water molecules were position restrained and bulk waters and ions were released. In the last two steps, position restraints were not applied on predicted waters, only solute heavy atoms were position restrained. Other details were the same as described in Section 3.4 above.

### 3.10. Extraction of Target-Ligand Interfaces by Fragmenter

Fragmenter automatically extracts target-ligand interfaces of large complexes and is freely available as a web service at [www.fragmenter.xyz](http://www.fragmenter.xyz). Algorithm details and connections between input, algorithm, implementation, and output scripts are presented in Figures S1 and S2. In brief, the extraction is based on the selection determined by the target-ligand ( $d_{TL}$ ) and the water-solute ( $d_w$ ) distances as well as the inter-residual distance ( $n$ ). In the main loop (Figure S1), a target amino acid residue is extracted if it has at least one heavy atom with  $d_{cls} \leq d_{TL}$ , where  $d_{cls}$  is the spatial distance measured between the closest heavy atoms of the actual target and ligand molecules. The maximal distance allowed between the closest heavy atoms of the target and the ligand ( $d_{TL}$ ) can be provided by the user and a default value is set to 3.5 Å. The same distance between solute partners and water molecules ( $d_w$ ) is also defined and applied for extraction of interfacial waters. Connecting amino acids and terminating groups are also inserted. The length of fragment peptides is influenced by the maximal inter-residual topological distance ( $n$ ) of the target. Parameter  $n$  specifies how many adjacent amino acids are added to the fragment chain of amino acids extracted above by the  $d_{TL}$  criterion. If  $n > 0$ , then the fragment was grown by adding  $n$  connecting amino acid residues. If  $n = 0$  only amino acids with  $d_{cls} \leq d_{TL}$  are added to the fragment chain. If  $n = 1$ , the sequential first neighbors are also attached to the terminus (termini) of the fragment chain, even if the attached amino acids have a  $d_{cls} > d_{TL}$ , etc. (Table S2).

#### 3.10.1. Input

The actual content of the query form of the 'submit' tab of the web interface (Figure S4) is saved as a single input file (project\_ID.inp) generated according to a template inputfile.inp (Figure S6). This template contains the system variables, php path, the path of the createqinput.sh, and the template for the input parameters from the website. The 'submit' tab allows setting distance ( $d_{TL}$ ,  $d_w$ , and  $n$ ) and other parameters of Table S1. Fragmenter offers an option to freeze (restrain) atomic positions by labeling certain groups of heavy atoms such as backbone  $C_\alpha$ -atoms, heavy atoms, all heavy atoms in the Mopac input file. The definition of these restraints, additional Mopac parameters, and other administrative details are also collected in the project\_ID.inp file. The latter parameters include the path and the file name of the complex structure, the process name (for the SLURM workload manager), the path of the php executable and Fragmenter scripts the mopac license file (for SLURM) and the path of the mopac and php executable. Using the above path and file information, setting of system variables is performed by script genqinput.sh. The user does not need to care about server configuration (e.g., server specific php executable path), it is stored on the server. Clicking on the 'submit' button the script calculate.php checks the integrity of the complex structure (Figures S1 and S2) by a PDB to PDB file conversion using OpenBabel [88]. In the case of conversion errors Fragmenter terminates and the errors are displayed on a separate page. Then it collects and transforms the input parameters from inputfile.inp and from the site from the user for the script createqinput.sh and calls

createqinput.sh. Script genqinput.sh requires only one input file (project\_ID.inp), which contains all necessary parameters for the run (Figure S6).

### 3.10.2. Main Algorithm

Having all input data in project\_ID.inp, script createqinput.sh calls script fragment.php, the main engine of Fragmenter and creates the output files using other php classes (point.php, atom.php, charge.php, ligand.php). Among the classes (i) atom.php represents the atom objects with coordinates, type; (ii) module charge.php calculates the charge the ligand and generated fragment chains; (iii) module ligand.php handles a ligand object, contains the atoms, bonds, it reads and writes the pdb files; (iv) Point.php is a small class reserved for the coordinates of the atoms. Utils.php collects technical parameters, for example operating system dependent information, config file handling, etc.

Script fragment.php (Figure S2) includes steps for input processing, fragmenting, and working with output files. Target and ligand objects are obtained from the input steps, target, ligand residues, and water molecules are detected based on their chain IDs and residue types (WAT, SOL, H2O), respectively. Accordingly, the input structure is split into ligand, target and water molecules, the residues are sorted by their residue IDs and only heavy atoms of the target are examined. In the main loop of fragment.php, the target amino acid residues are selected according to dTL and n by ligand.php and point.php. Single residue-gaps are excluded by connecting two neighboring fragments by selecting the connecting residue, as well.

Having all target fragments produced in the previous steps, each of them are terminated by a uniform procedure as represented by the cycle of fragment.php (Figures S1 and S2). In the case of a free N-terminus a protonated amino group is built automatically by adding hydrogen atoms in a correct geometry. Similarly, in the case of a free C-terminus, a carboxylate anion is left unchanged. After merging ligand and water molecules with the target fragments into a new PDB file, the total charge of the complex is calculated and stored in the remark section of the file. For all cut target chains, Ac- (at N-terminus) and -NHMe (at C-terminus) blocking groups are built on both or non-free ends using atoms of previous and/or next amino acids of the chain and adding three hydrogen atoms to the methyl group (Figure S3). Following the generation of all fragments, the interface water molecules are extracted according to the intermolecular distance cut-off ( $d_w$ , Table S1). After extraction of the water molecules, their total net charges are calculated by charge.php using individual charges of amino acids (Table S3) at pH 7. Special care was taken for disulfide bridges between side-chains of cysteine amino acids. Following the main loop (Figure S1), Cys residues connected via disulfide bridges are also selected and added to the fragments. Total net charge (Table S3) of the target fragments is calculated. In the case of disulfide bridges or protonated sulfhydryl group the charge of Cys is automatically set to zero, otherwise  $-1$ . The charge of His is calculated according to the protonation state of the imidazole ring ( $-1, 0, +1$ ).

### 3.10.3. Target-Ligand Intermolecular Interaction Energy

Fragmenter calculates target-ligand intermolecular interaction energy ( $E_{\text{inter}}$ ) for the extracted interface, which is expressed as the sum of Lennard-Jones (LJ) and Coulomb potentials (Equation (5)). For both the LJ and Coulomb potentials, Amber force field parameters are used [83,89]. A per-residue list of the  $E_{\text{inter}}$  is printed in the 'results' table. The list can be used for identification of target residues colliding with the ligand as large  $E_{\text{inter}}$  values. In such cases, further MM energy-minimization may be required to achieve a complex structure appropriate for QM investigations.

$$E_{\text{inter}} = E_{\text{LJ}} + E_{\text{Coulomb}} = \sum_{i,j}^{N_T N_L} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \right] \quad (5)$$

$$A_{ij} = \epsilon_{ij} R_{ij}^{12}; B_{ij} = 2\epsilon_{ij} R_{ij}^6; R_{ij} = R_i + R_j; \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$$

where,  $\epsilon_{ij}$  is the potential well depth at equilibrium between the *i*th (ligand) and *j*th (target) atoms;  $\epsilon_0$  is the permittivity of vacuum;  $\epsilon_r = 1$ , relative permittivity;  $R_{ij}$  is the inter-nuclear distance at equilibrium between *i*th (ligand) and *j*th (target) atoms;  $q$  is the partial charge of an atom;  $r_{ij}$  is the actual distance between the *i*th (ligand) and *j*th (target) atoms;  $N_T$  is the number of target atoms;  $N_L$  is the number of ligand atoms.

#### 3.10.4. Output

Fragmenter stores the output files in an output directory and provides a download link to all of them in the 'results' site (Figure S5). The ligand, the selected water molecules and the target fragments are downloadable as a complex in PDB format. The final charge of the complex is stored in the remark section of the PDB file. Fragmenter also provides additional separate PDB files and also converts them into downloadable Mopac input files. These files include the structures of the complex with/without water molecules, separate ligand, or target fragments. The 'output' tab also features the fragmented complex in a small window and it can be manipulated by the user. The visualization and rotation is performed by JSmol [90] implemented in the web page.

#### 3.11. Calculation of Heats of Formation

Mopac 2012 [91] was used for structural relaxation and calculation of heats of formation of the extracted complex structures, separate ligand, water molecules, and target fragments. Hamiltonian of the Parametric Method number 7 (PM7 [92]) was applied. The exit criterion of the energy-minimization was defined as a gradient norm of 1.0. The value was set according to the instructions of the Mopac Support Team, and it is a magnitude smaller than the value of 10 suggested by the Manual [92]. There were only four vacuum calculations where the final gradient norm was slightly higher than 1, and the largest one of the four was 2.5. To reduce computational cost, the localized molecular orbital approach of Mozyme [93] was applied. Total net charges of the molecules were calculated from individual net charges of the amino acids (Table S3). The charges were indicated in the command line, checked manually and automatically with keyword GEO-OK. To prevent unwanted termination of calculations, keyword PREC was applied. Eigenvector following [94] was used as a default geometry optimization. Molecular mechanics correction to peptide bonds was applied by keyword MMOK. Except the cases of in vacuo calculations, the COSMO (COnductor-like ScreeningMOdel) model [63] was used. For this, a value of 78.3 was set at the EPS key word which is the dielectric constant of water at 293.15 K and 101325 Pa.  $\Delta_f H$  of water was calculated with the above keywords in vacuum and using the COSMO model, respectively. In the cases of four systems, integrity of disulphide bridges was conserved by restraining the coordinates of S atoms during COSMO calculations.

#### 3.12. Statistics

Simple linear regressions were performed between calculated  $\Delta_r H$  and experimental  $\Delta H_b(\text{exp})$  values in all cases of Table 2 and Tables S4–S7.  $\Delta H_b(\text{exp})$  values were obtained from various publications as listed in Table 1. Statistical parameters of the regressions including regression coefficients ( $\alpha$  and  $\beta$  in Equation (3)), coefficients of determination ( $R^2$ ), *t*-values, *F*-values, residuals and root mean square error (RMSE) values are listed in Table 2. Leave-one-out cross-validated  $R^2$  values were also calculated to check the stability of the correlations. Significance values of regression coefficients mentioned in the main text were calculated by two-sided *t*-test. For correlation plots,  $\Delta H_b(\text{calc})$  values were calculated using  $\Delta_r H$  values and the regression coefficients (Equation (3)).

## 4. Conclusions

Structure-based calculation of binding thermodynamics is challenging at the QM level. To overcome the limitations of system size and hydration, a new protocol was introduced combining a MobyWat-based prediction of hydration structure with Fragmenter, a tool designed for extraction of the target-ligand interface with peptide fragments representing the target molecule. The protocol

allowed fast QM calculations on a series of target-ligand interfaces with systematically adjusted hydration models. High correlations were achieved with a hybrid model involving a shell of explicit water molecules of calculated positions and the implicit solvation method COSMO. At semi-empirical QM level, and PM7 parameterization, a single, statistically significant scale factor was obtained for conversion of calculated reaction heats into experimental binding enthalpy values. The results of the present study will be particularly helpful in enthalpic optimization of drugs and in the molecular design of stable complexes and new ligands, in general. Further development and tests of the protocol have been also initiated for applications at the highest level of QM theory.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/18/4384/s1>.

**Author Contributions:** Conceptualization, C.H.; Methodology, N.J., M.B., I.H., and C.H.; Software, I.H. and C.H.; Validation, I.H., N.J., and M.B.; Formal analysis, I.H., and C.H.; Investigation, I.H., N.J., and M.B.; Data curation, C.H., G.P., M.B., and I.H.; Writing, C.H., I.H., M.B., N.J., and G.P.; Visualization, C.H. and I.H.; Supervision, C.H.; Project administration, C.H.; Funding acquisition, C.H. and I.H.

**Funding:** This work was supported by the Hungarian National Research, Development and Innovation Office (K123836). This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. We acknowledge the grant of computer time from CSCS Swiss National Supercomputing Centre, and the Governmental Information Technology Development Agency, Hungary. The University of Pécs is acknowledged for a support by the 17886-4/23018/FEKUTSTRAT excellence grant. C.H.'s work was supported by a grant co-financed by Hungary and the European Union (EFOP-3.6.2-16-2017-00008).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

COSMO	Conductor-like screening model
$E_{\text{inter}}$	Intermolecular interaction energy
ITC	Isothermal titration calorimetry
LJ	Lennard-Jones
MD	Molecular dynamics
MM	Molecular mechanics
MW	Molecular weight
PDB	Protein Databank
PM7	Parametric method 7
QM	Quantum mechanical
RMSE	Root mean square error

## References

1. Bajusz, D.; Ferenczy, G.; Keseru, G. Structure-Based Virtual Screening Approaches in Kinase-Directed Drug Discovery. *Curr. Top. Med. Chem.* **2017**, *17*, 2235–2259. [[CrossRef](#)]
2. Talhout, R.; Engberts, J.B. Thermodynamic analysis of binding of p-substituted benzamidines to trypsin. *Eur. J. Biochem.* **2001**, *268*, 1554–1560. [[CrossRef](#)]
3. Sleight, S.H.; Seavers, P.R.; Wilkinson, A.J.; Ladbury, J.E.; Tame, J.R.H. Crystallographic and Calorimetric Analysis of Peptide Binding to OppA Protein. *J. Mol. Biol.* **1999**, *291*, 393–415. [[CrossRef](#)]
4. Dullweber, F.; Stubbs, M.; Musil, D.; Stürzebecher, J.; Klebe, G. Factorising ligand affinity: A Combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J. Mol. Biol.* **2001**, *313*, 593–614. [[CrossRef](#)]
5. Palencia, A.; Cobos, E.S.; Mateo, P.L.; Martinez, J.C.; Luque, I. Thermodynamic Dissection of the Binding Energetics of Proline-rich Peptides to the Abl-SH3 Domain: Implications for Rational Ligand Design. *J. Mol. Biol.* **2004**, *336*, 527–537. [[CrossRef](#)]
6. McNemar, C.; Snow, M.E.; Windsor, W.T.; Prongay, A.; Mui, P.; Zhang, R.; Durkin, J.; Le, H.V.; Weber, P.C. Thermodynamic and structural analysis of phosphotyrosine polypeptide binding to Grb2-SH2. *Biochemistry* **1997**, *36*, 10006–10014. [[CrossRef](#)]

7. Wang, C.; Pawley, N.H.; Nicholson, L.K. The role of backbone motions in ligand binding to the c-Src SH3 domain. *J. Mol. Biol.* **2001**, *313*, 873–887. [[CrossRef](#)]
8. Org, T.; Chignola, F.; Chignola, F.; Hetényi, C.; Gaetani, M.; Rebane, A.; Liiv, I.; Maran, U.; Mollica, L.; Bottomley, M.J.; et al. The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep.* **2008**, *9*, 370–376. [[CrossRef](#)]
9. Chrencik, J.E.; Brooun, A.; Recht, M.; Kraus, M.L.; Koolpe, M.; Kolatkar, A.; Bruce, R.H.; Martiny-Baron, G.; Widmer, H.; Pasquale, E.B.; et al. Thermodynamic and structural analysis of phosphotyrosine polypeptide binding to Grb2-SH2. *Biochemistry* **2006**, *14*, 321–330.
10. Kozlov, G.; De Crescenzo, G.; Lim, N.S.; Siddiqui, N.; Fantus, D.; Kahvejian, A.; Trempe, J.-F.; Elias, D.; Ekiel, I.; Sonenberg, N.; et al. Structural basis of ligand recognition by PABC, a highly specific peptide-binding domain found in poly(A)-binding protein and a HECT ubiquitin ligase. *EMBO J.* **2004**, *23*, 272–281. [[CrossRef](#)]
11. Day, C.L.; Smits, C.; Fan, F.C.; Lee, E.F.; Fairlie, W.D.; Hinds, M.G. Structure of the BH3 domains from the p53-inducible BH3-only proteins Noxa and Puma in complex with Mcl-1. *J. Mol. Biol.* **2008**, *380*, 958–971. [[CrossRef](#)]
12. Lu, Z.; Zhang, Y. Interfacing ab initio Quantum Mechanical Method with Classical Drude Oscillator Polarizable Model for Molecular Dynamics Simulation of Chemical Reactions. *J. Chem. Theory Comput.* **2008**, *4*, 1237–1248. [[CrossRef](#)]
13. Vanommeslaeghe, K.; Guvench, O.; MacKerell, A.D. Molecular Mechanics. *Curr. Pharm. Des.* **2014**, *20*, 3281–3292. [[CrossRef](#)]
14. Ganoth, A.; Friedman, R.; Nachliel, E.; Gutman, M. A Molecular Dynamics Study and Free Energy Analysis of Complexes between the Mlc1p Protein and Two IQ Motif Peptides. *Biophys. J.* **2006**, *91*, 2436–2450. [[CrossRef](#)]
15. Fenley, A.T.; Muddana, H.S.; Gilson, M.K. Entropy—Enthalpy transduction caused by conformational shifts can obscure the forces driving protein—Ligand binding. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 20006–20011. [[CrossRef](#)]
16. Huggins, D.J. Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. *Biophys. J.* **2015**, *108*, 928–936. [[CrossRef](#)]
17. Dougherty, D.A. Cation- $\pi$  Interactions Involving Aromatic Amino Acids. *J. Nutr.* **2007**, *137* (Suppl. S1), 1504–1508. [[CrossRef](#)]
18. Majumdar, S.; Maiti, S.; Ghosh Dastidar, S. Dynamic and Static Water Molecules Complement the TN16 Conformational Heterogeneity Inside the Tubulin Cavity. *Biochemistry* **2016**, *55*, 335–347. [[CrossRef](#)]
19. Poole, P.L.; Finney, J.L. Hydration-induced conformational and flexibility changes in lysozyme at low water content. *Int. J. Biol. Macromol.* **1983**, *5*, 308–310. [[CrossRef](#)]
20. Jukič, M.; Konc, J.; Gobec, S.; Janežič, D. Identification of Conserved Water Sites in Protein Structures for Drug Design. *J. Chem. Inf. Model.* **2017**, *57*, 3094–3103. [[CrossRef](#)]
21. Choi, H.; Kang, H.; Park, H. New solvation free energy function comprising intermolecular solvation and intramolecular self-solvation terms. *J. Chem.* **2013**, *5*, 5–8. [[CrossRef](#)]
22. Cui, G.; Swails, J.M.; Manas, E.S. SPAM: A Simple Approach for Profiling Bound Water Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 5539–5549. [[CrossRef](#)]
23. García-Sosa, A.T.; Stuart, F.C.; Mancera, R.L. Including Tightly-Bound Water Molecules in de Novo Drug Design. Exemplification through the in Silico Generation of Poly(ADP-ribose)polymerase Ligands. *J. Chem. Inf. Model.* **2005**, *45*, 624–633. [[CrossRef](#)]
24. Huggins, D.J.; Tidor, B. Systematic placement of structural water molecules for improved scoring of protein—Ligand interactions. *Protein Eng. Des. Sel.* **2011**, *24*, 777–789. [[CrossRef](#)]
25. Ladbury, J.E. Just add water! The effect of water on the specificity of proteinligand binding sites and its potential application to drug design. *Cell Chem. Biol.* **1996**, *3*, 973–980.
26. Lloyd, D.G.; García-Sosa, A.T.; Alberts, I.L.; Todorov, N.P.; Mancera, R.L. The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models. *J. Comp. Aided Mol. Des.* **2004**, *18*, 89–100. [[CrossRef](#)]
27. Michel, J.; Tirado-Rives, J.; Jorgensen, W.L. Energetics of Displacing Water Molecules from Protein Binding Sites: Consequences for Ligand Optimization. *J. Am. Chem. Soc.* **2009**, *131*, 15403–15411. [[CrossRef](#)]
28. Zheng, Z. Computational Modeling of Solvent Effects on Protein-Ligand Interactions Using Fully Polarizable Continuum Model and Rational Drug Design. *Commun. Comput. Phys.* **2013**, *13*, 31–60. [[CrossRef](#)]



29. Kunstmann, S.; Gohlke, U.; Broecker, N.K.; Roske, Y.; Heinemann, U.; Santer, M.; Barbirz, S. Solvent Networks Tune Thermodynamics of Oligosaccharide Complex Formation in an Extended Protein Binding Site. *J. Am. Chem. Soc.* **2018**, *140*, 10447–10455. [[CrossRef](#)]
30. Brysbaert, G.; Blossey, R.; Lensink, M.F. The Inclusion of Water Molecules in Residue Interaction Networks Identifies Additional Central Residues. *Front. Mol. Biosci.* **2018**, *5*, 88. [[CrossRef](#)]
31. Quiococho, F.A.; Wilson, D.K.; Vyas, N.K. Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **1989**, *340*, 404–407. [[CrossRef](#)]
32. Jeszenői, N.; Bálint, M.; Horváth, I.; van der Spoel, D.; Hetényi, C. Exploration of interfacial hydration networks of target—Ligand complexes. *J. Chem. Inf. Model.* **2016**, *56*, 148–158. [[CrossRef](#)]
33. Afonine, P.V.; Grosse-Kunstleve, R.W.; Adams, P.D.; Urzhumstev, A. Bulk-solvent and overall scaling revisited: Faster calculations, improved results. *Acta Cryst. D Biol. Cryst.* **2013**, *69 Pt 4*, 625–634. [[CrossRef](#)]
34. Badger, J. Modeling and refinement of water molecules and disordered solvent. *Meth. Enzymol.* **1997**, *277*, 344–352.
35. Finney, J.L. The organization and function of water in protein crystals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1977**, *278*, 3–32. [[CrossRef](#)]
36. Halle, B. Protein hydration dynamics in solution: A critical survey. *Philos. Trans. R. Soc. Lond. B* **2004**, *359*, 1207–1224. [[CrossRef](#)]
37. Elsässer, S.J.; Huang, H.; Lewis, P.W.; Chin, J.W.; Allis, C.D.; Patel, D.J. DAXX envelops a histone H3.3–H4 dimer for H3.3-specific recognition. *Nature* **2012**, *491*, 560–565. [[CrossRef](#)]
38. Abel, R.; Young, T.; Farid, R.; Berne, B.; Friesner, R. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831. [[CrossRef](#)]
39. Pearlstein, R.; Hu, Q.; Zhou, J.; Yowe, D.; Levell, J.; Dale, B.; Kaushik, V.; Daniels, D.; Hanrahan, S.; Sherman, W.; et al. New hypotheses about the structure-function of proprotein convertase subtilisin/kexin type 9: Analysis of the epidermal growth factor-like repeat A docking site using WaterMap. *Proteins* **2010**, *78*, 2571–2586. [[CrossRef](#)]
40. Vukovic, S.; Brennan, P.E.; Huggins, D. Exploring the role of water in molecular recognition: Predicting protein ligandability using a combinatorial search of surface hydration sites. *J. Phys. Condens. Matter.* **2016**, *28*, 344007. [[CrossRef](#)]
41. Hylsová, M.; Carbain, B.; Fanfrlik, J.; Lepsik, M. Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: Inhibition of CDK2 by new pyrazolo [1,5-a] pyrimidines. *Eur. J. Med. Chem.* **2017**, *126*, 1118–1128. [[CrossRef](#)]
42. García-Sosa, A.T.; Mancera, R.L. Free Energy Calculations of Mutations Involving a Tightly Bound Water Molecule and Ligand Substitutions in a Ligand-Protein Complex. *Mol. Inf.* **2010**, *29*, 589–600. [[CrossRef](#)]
43. Lee, S.H.; Rossky, P.J. A comparison of the structure and dynamics of liquid water at hydrophobic and hydrophilic surfaces—A molecular dynamics simulation study. *J. Chem. Phys.* **1994**, *100*, 3334–3345. [[CrossRef](#)]
44. Watanabe, G.; Nakajima, D.; Hiroshima, A.; Suzuki, H.; Yoneda, S. Analysis of water channels by molecular dynamics simulation of heterotetrameric sarcosine oxidase. *Biophys. Physicobiol.* **2015**, *12*, 131–137. [[CrossRef](#)]
45. Patodia, S.; Bagaria, A.; Chopra, D. Molecular Dynamics Simulation of Proteins: A Brief Overview. *J. Phys. Chem. Bioph.* **2014**, *4*, 1. [[CrossRef](#)]
46. Jeszenői, N.; Horváth, I.; Bálint, M.; van der Spoel, D.; Hetényi, C. Mobility-based prediction of hydration structures of protein surfaces. *Bioinformatics* **2015**, *31*, 1959–1965. [[CrossRef](#)]
47. Field, M.J.; Bash, P.A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comp. Chem.* **1990**, *11*, 700–733. [[CrossRef](#)]
48. Hayik, S.A.; Dunbrack, R.; Merz, K.M. A Mixed QM/MM Scoring Function to Predict Protein-Ligand Binding Affinity. *J. Chem. Theory Comp.* **2010**, *6*, 3079–3091. [[CrossRef](#)]
49. Menikarachchi, L.C.; Gascón, J.A. QM/MM Approaches in Medicinal Chemistry Research. *Curr. Top. Med. Chem.* **2010**, *10*, 46–54. [[CrossRef](#)]
50. Van der Kamp, M.W.; Mulholland, A.J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728. [[CrossRef](#)]
51. Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249. [[CrossRef](#)]

52. Vidossich, P.; Magistrato, A. QM/MM Molecular Dynamics Studies of Metal Binding Proteins. *Biomolecules* **2014**, *4*, 616–645. [[CrossRef](#)]
53. Zhang, D.W.; Zhang, J.Z.H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–Molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605. [[CrossRef](#)]
54. Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. Semiempirical calculations of binding enthalpy for protein–ligand complexes. *Int. J. Quantum Chem.* **2004**, *97*, 747–763. [[CrossRef](#)]
55. Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. Mixed Implicit/Explicit Solvation Models in Quantum Mechanical Calculations of Binding Enthalpy for Protein–Ligand Complexes. *Int. J. Quantum Chem.* **2006**, *106*, 1943–1963. [[CrossRef](#)]
56. Dobeš, P.; Otyepka, M.; Strnad, M.; Hobza, P. Interaction Energies for the Purine Inhibitor Roscovitine with Cyclin-Dependent Kinase 2: Correlated Ab Initio Quantum-Chemical, DFT and Empirical Calculations. *Chemistry* **2006**, *12*, 4297–4304. [[CrossRef](#)]
57. Freire, E. Do Enthalpy and Entropy Distinguish First in Class from Best in Class? *Drug Discov. Today* **2008**, *13*, 869–874. [[CrossRef](#)]
58. Ferenczy, G.G.; Keserü, G.M. Thermodynamics guided lead discovery and optimization. *Drug Discov. Today* **2010**, *15*, 919–932. [[CrossRef](#)]
59. Hann, M.M.; Keserü, G.M. Finding the sweet spot: The role of nature and nurture in medicinal chemistry. *Nat. Rev. Drug. Discov.* **2012**, *11*, 355–365. [[CrossRef](#)]
60. Zhang, D.W.; Xiang, Y.; Zhang, J.Z.H. New Advance in Computational Chemistry: Full Quantum Mechanical ab Initio Computation of Streptavidin–Biotin Interaction Energy. *J. Phys. Chem. B* **2003**, *107*, 12039–12041. [[CrossRef](#)]
61. Zhang, D.W.; Xiang, Y.; Gao, A.M.; Zhang, J.Z.H. Quantum mechanical map for protein–ligand binding with application to  $\beta$ -trypsin/benzamidine complex. *J. Chem. Phys.* **2004**, *120*, 1145–1148. [[CrossRef](#)]
62. Brown, S.; Shirts, M.; Mobley, D. Free-energy calculations in structure-based drug design. *Drug Des. Struct. Ligand Based Approaches* **2010**, *2010*, 61–86.
63. Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc. Perk. Trans.* **1993**, *2*, 799–805. [[CrossRef](#)]
64. Weichenberger, C.X.; Afonine, P.V.; Kantardjieff, K.; Rupp, B. The solvent component of macromolecular crystals. *Acta Cryst. D Biol. Cryst.* **2015**, *71*, 1023–1038. [[CrossRef](#)]
65. Halle, B. Biomolecular cryocrystallography: Structural changes during flash-cooling. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4793–4798. [[CrossRef](#)]
66. Schmidtke, P.; Barril, X.; Luque, F.J.; Murray, J.B. Shielded hydrogen bonds as structural determinants of binding kinetics: Application in drug design. *J. Am. Chem. Soc.* **2011**, *133*, 18903–18910. [[CrossRef](#)]
67. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]
68. Guex, N.; Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714–2723. [[CrossRef](#)]
69. *Schrödinger Release 2019-3: Maestro*; Schrödinger, LLC: New York, NY, USA, 2019.
70. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [[CrossRef](#)]
71. Gille, A.L.; Dutmer, B.C.; Gilbert, T.M. PCMODEL 9.2. *J. Am. Chem. Soc.* **2009**, *131*, 5714. [[CrossRef](#)]
72. Halgren, T. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519. [[CrossRef](#)]
73. Stewart, J.J.P. *MOPAC2009, 2009*; Steward Computational Chemistry: Colorado Springs, CO, USA, 2008.
74. Stewart, J.J.P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213. [[CrossRef](#)]
75. Vanqualef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J.C.; Cieplak, P.; Dupradeau, F.Y. R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011**, *39*, W511–W517. [[CrossRef](#)]
76. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.A.; et al. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, USA, 2009.
77. Krishnan, R.; Binkley, J.S.; Seeger, R.; Pople, J.A. Self-consistent molecular-orbital methods. XX. Basis set for correlated wave-functions. *J. Chem. Phys.* **1980**, *72*, 650–654. [[CrossRef](#)]

78. Schmidt, M.W.; Baldrige, K.K.; Boatz, J.A.; Elbert, S.T.; Gordon, M.S.; Jensen, J.H.; Koseki, S.; Matsunaga, N.; Nguyen, K.A.; Su, S.J.; et al. General atomic and molecular electronic-structure system. *J. Comp. Chem.* **1993**, *14*, 1347–1363. [[CrossRef](#)]
79. Case, D.; Darden, T.; Cheatham Iii, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K. *AmberTools, 15*; Amber; University of California: San Francisco, CA, USA, 2015.
80. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25. [[CrossRef](#)]
81. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M.R.; Smith, J.C.; Kasson, P.M.; van der Spoel, D.; et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854. [[CrossRef](#)]
82. 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. [[CrossRef](#)]
83. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958. [[CrossRef](#)]
84. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. [[CrossRef](#)]
85. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald—An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [[CrossRef](#)]
86. Nose, S.; Klein, M.L. Constant pressure molecular-dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076. [[CrossRef](#)]
87. Parrinello, M.; Rahman, A. Polymorphic transitions in single-crystals—A new molecular-dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190. [[CrossRef](#)]
88. O’Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [[CrossRef](#)]
89. Wang, J.; Cieplak, P.; Li, J.; Cai, Q.; Hsieh, M.J.; Luo, R.; Duan, Y. Development of polarizable models for molecular mechanical calculations. 4. van der Waals parametrization. *J. Phys. Chem. B* **2012**, *116*, 7088–7101. [[CrossRef](#)]
90. Hanson, R.M.; Prilusky, J.; Renjian, Z.; Nakane, T.; Sussman, J.L. JSmol and the Next-Generation Web-Based Representation of 3D Molecular Structure as Applied to Proteopedia. *Isr. J. Chem.* **2013**, *53*, 207–216. [[CrossRef](#)]
91. Stewart, J.J.P. Openmopac Online Manual. 2016. Available online: <http://www.openmopac.net> (accessed on 1 March 2013).
92. Stewart, J.J.P. Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1–32. [[CrossRef](#)]
93. Stewart, J.J.P. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int. J. Quant. Chem.* **1996**, *58*, 133–146. [[CrossRef](#)]
94. Baker, J. An Algorithm for the Location of Transition States. *J. Comp. Chem.* **1986**, *7*, 385. [[CrossRef](#)]

