

# Free enthalpies of replacing water molecules in protein binding pockets

Sereina Riniker · Luzi J. Barandun ·  
François Diederich · Oliver Krämer ·  
Andreas Steffen · Wilfred F. van Gunsteren

Received: 25 May 2012 / Accepted: 27 November 2012 / Published online: 18 December 2012  
© Springer Science+Business Media Dordrecht 2012

**Abstract** Water molecules in the binding pocket of a protein and their role in ligand binding have increasingly raised interest in recent years. Displacement of such water molecules by ligand atoms can be either favourable or unfavourable for ligand binding depending on the change in free enthalpy. In this study, we investigate the displacement of water molecules by an apolar probe in the binding pocket of two proteins, cyclin-dependent kinase 2 and tRNA-guanine transglycosylase, using the method of enveloping distribution sampling (EDS) to obtain free enthalpy differences. In both cases, a ligand core is placed inside the respective pocket and the remaining water molecules are converted to apolar probes, both individually and in pairs. The free enthalpy difference between a water molecule and a CH<sub>3</sub> group at the same location in the pocket in comparison to their presence in bulk solution calculated from EDS molecular dynamics simulations corresponds to the binding free enthalpy of CH<sub>3</sub> at this location. From the free enthalpy difference and the enthalpy

difference, the entropic contribution of the displacement can be obtained too. The overlay of the resulting occupancy volumes of the water molecules with crystal structures of analogous ligands shows qualitative correlation between experimentally measured inhibition constants and the calculated free enthalpy differences. Thus, such an EDS analysis of the water molecules in the binding pocket may give valuable insight for potency optimization in drug design.

**Keywords** Water sites · Binding pocket · Molecular dynamics · Free energy · Enveloping distribution sampling

## Introduction

Ligand-protein binding is determined by the difference in free enthalpy [1] or Gibbs energy [2]  $\Delta G_{\text{ligand}}^{\text{bind}}$  between the ligand bound to the protein and unbound in aqueous solution, and is related to the experimentally measurable inhibition constant  $K_i$ , reflecting the binding affinity,

$$\Delta G_{\text{ligand}}^{\text{bind}} = G_{\text{ligand}}^{\text{bound}} - G_{\text{ligand}}^{\text{unbound}} = RT \ln(K_i), \quad (1)$$

where  $R$  is the gas constant and  $T$  the temperature. The free enthalpy of binding comprises enthalpic and entropic contributions  $\Delta G_{\text{ligand}}^{\text{bind}} = \Delta H_{\text{ligand}}^{\text{bind}} - T\Delta S_{\text{ligand}}^{\text{bind}}$ . The process of ligand binding involves desolvation of the ligand in solution, desolvation of the binding pocket in the protein, i.e. removal of the water molecules from it, and ligand-protein complexation,

$$\Delta G_{\text{ligand}}^{\text{bind}} = \Delta G_{\text{ligand}}^{\text{desolv}} + \Delta G_{\text{pocket}}^{\text{desolv}} + \Delta G_{\text{ligand+pocket}}^{\text{complex}} \quad (2)$$

When only a limited set of ligands is of interest, molecular dynamics (MD) simulations combined with methods to compute free enthalpy differences based on configurational

**Electronic supplementary material** The online version of this article (doi:10.1007/s10822-012-9620-8) contains supplementary material, which is available to authorized users.

S. Riniker · W. F. van Gunsteren (✉)  
Laboratory of Physical Chemistry, ETH Zurich, 8093 Zurich,  
Switzerland  
e-mail: wfvgn@igc.phys.chem.ethz.ch

L. J. Barandun · F. Diederich  
Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich,  
Switzerland  
e-mail: diederich@org.chem.ethz.ch

O. Krämer · A. Steffen  
Boehringer Ingelheim RCV GmbH & Co KG, Vienna, Austria  
e-mail: oliver.kraemer@boehringer-ingelheim.com

A. Steffen  
e-mail: andreas.steffen@boehringer-ingelheim.com

ensembles (e.g. thermodynamic integration [3] or enveloping distribution sampling (EDS) [4–7]) can be used to calculate the relative binding free enthalpy between two ligands *A* and *B*,  $\Delta G_{AB}^{bind}$ . Such calculations generally use the most detailed and accurate force fields available, do account for entropic contributions through conformational Boltzmann sampling, and include ligand desolvation contributions. However, these methods are computationally expensive and are thus often not suitable for industrial drug design where time is critical. In order to rapidly estimate  $\Delta G_{ligand}^{bind}$  values for large sets of ligands, severe approximations are often made [8, 9]. For example, the contributions from the desolvation of the pocket and the ligand are ignored, i.e.  $\Delta G_{pocket}^{desolv} = 0$  and  $\Delta G_{ligand}^{desolv} = 0$ , the motion of the protein and thus its entropic contribution is neglected, a simple, computationally cheap force field of limited accuracy is used, and average energies or enthalpies instead of free energies or free enthalpies are calculated.

In the process of desolvation of the binding pocket and ligand-protein complexation, water molecules in a binding pocket are displaced by the ligand. Such water molecules in the binding pocket can have a low entropy due to the orientational and positional restrictions imposed by the protein or have a high enthalpy due to a lack of polar interaction partners. If their free enthalpy is high, the binding affinity of a ligand will increase if it displaces such water molecules. From an analysis of the high or low  $\Delta G_{site}^{desolv}$  values of the sites occupied by water molecules and their replacement by atoms of a particular type, a relation to the binding free enthalpies of a large series of ligands could be established using an appropriate scoring function. This scoring function would map particular atom types at the various water replacement sites on the atom types and the structure of ligands. An essential condition for this idea to be fruitful is, however, that the analysis should distinguish the water molecule sites with a positive  $\Delta G_{site}^{desolv}$  from the ones important to a favourable interaction between the protein and the ligand, i.e. with a large negative  $\Delta G_{site}^{desolv}$ .

In recent years, the importance of considering the role of water during docking and structure-based drug design has been recognized [10–22]. In the WaterMap method [18, 23], the contributions  $\Delta G_{pocket}^{desolv}$  and  $\Delta G_{ligand}^{desolv}$  are included. The interaction characteristics of a binding pocket are analysed by performing MD simulations of the protein with its binding pocket filled with water molecules. From the configurational distributions of the water oxygens, a number of water sites are identified. For each of the water molecules the energy difference with a water in aqueous solution is calculated and its configurational entropy is estimated. These two quantities contribute to  $\Delta G_{site}^{bind}$  of that site. During a WaterMap simulation, the motion of the protein and its entropic contribution

are neglected, only water molecules are used to probe the binding pocket and the correlation between displacements at adjacent sites is not accounted for.

Here, we propose a method based on the technique of enveloping distribution sampling (EDS) that would not suffer from these approximations and calculates  $\Delta G_{ligand}^{bind}$  directly instead of building it from  $\Delta H$  and  $T\Delta S$  contributions. Thereby, the water molecules (W) in the binding pocket are perturbed to apolar probes, i.e. a CH<sub>3</sub> group (M), and the free enthalpy difference  $\Delta G_{MW}^{pocket}$  is calculated. In the case that the ligand is a CH<sub>3</sub> group, this quantity corresponds to the desolvation of the site and the ligand-protein complexation,  $\Delta G_{site}^{desolv} + \Delta G_{M+site}^{complex}$ . The same perturbation in aqueous solution  $\Delta G_{MW}^{unbound}$  corresponds to  $-\Delta G_M^{desolv}$ , and thus the free enthalpy of binding can be calculated as,

$$\begin{aligned}\Delta G_M^{bind} &= \Delta G_{MW}^{pocket} - \Delta G_{MW}^{unbound} \\ &= \Delta G_{site}^{desolv} + \Delta G_{M+site}^{complex} - (-\Delta G_M^{desolv}).\end{aligned}\quad (3)$$

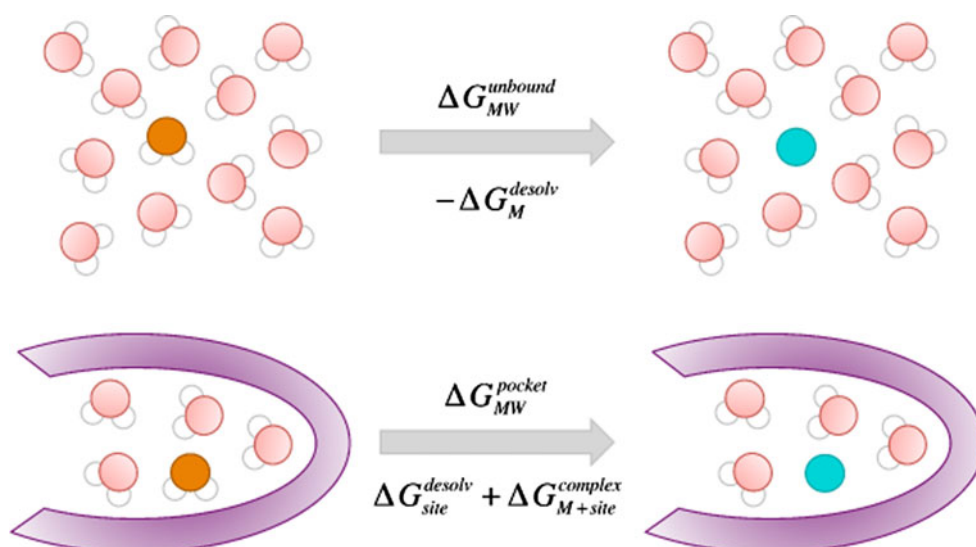
The method is shown schematically in Fig. 1. From a comparison of the enthalpy differences  $\Delta H_{MW}$  and free enthalpy differences  $\Delta G_{MW}$  of the apolar probe and the water molecule in the pocket and unbound in solution, the entropic contribution to binding can be obtained, i.e.  $T\Delta S_M^{bind} = \Delta H_M^{bind} - \Delta G_M^{bind}$ . To allow a mapping of the calculated free enthalpies of binding to sites, weak position restraints are applied to the water molecules and the CH<sub>3</sub> group. This affects the translational degrees of freedom but not the rotational ones as only the oxygen atoms of the water molecules are restrained. The force constant of the position restraints is chosen such that different configurations of the hydrogen-bonding network within the pocket and in bulk solution can be sampled. By applying the position restraints both inside the pocket and in bulk solution, the effect is largely cancelled in the resulting values of  $\Delta G_M^{bind}$ ,  $\Delta H_M^{bind}$  and  $T\Delta S_M^{bind}$ . The correlation between displacements at adjacent sites can be studied by fixing a selected site as CH<sub>3</sub> group and perturbing the surrounding water molecules. In this study, we place a ligand core in the binding pocket and investigate the displacement of the remaining water molecules using the technique sketched above in order to predict which ligand substituents would lead to increased binding affinities.

## Methods

### Proteins

The binding pockets of two proteins are investigated in this study, cyclin-dependent kinase 2 (CDK2) and tRNA-guanine transglycosylase (TGT). CDK2 is a kinase

**Fig. 1** Schematic representation of the perturbation process simulated in this study. Orange oxygen of the water molecule (W) of interest; cyan apolar atom (M) replacing it; purple protein; red other water molecules



involved in the cell cycle [24]. A large number of scaffolds have been proposed which bind at the adenosine 5'-triphosphate (ATP) binding pocket of the protein (for a selection see Refs. [25–29]). We consider a series of  $O^6$ -substituted guanines (Table 1) which probe the ATP-ribose pocket [27]. The binding pocket of CDK2 in complex with ligand **1** is shown in Fig. 2.

The tRNA modifying enzyme TGT has been identified as a potential target in the bacterial diarrheal disease shigellosis [30]. Eukaryotic and prokaryotic TGT catalyze distinctively different substrates, thus allowing selectivity for the eukaryotic target. A series of high-affinity inhibitors of TGT (Table 2) has been synthesized with a *lin*-benzoguanine core and varying substituents which aim at replacing the function of a cluster of water molecules in the ribose-34 pocket observed in the crystal structure [31, 32]. The binding pocket of TGT with ligand **5** is shown in Fig. 2.

#### Free enthalpy calculation using enveloping distribution sampling

The free enthalpy difference  $\Delta G_{MW}$  for the transition between the water molecule (W) and the apolar probe  $\text{CH}_3$  (M) was calculated using EDS [4–6].

In EDS, a reference state  $R$  is sampled which is defined through the end states  $W$  and  $M$ . Two kinds of parameters, the smoothness parameter  $s$  and the energy offsets  $E_i^R$  were introduced to provide even sampling of both end states [4]. The potential energy term of the reference-state Hamiltonian  $H_R$  has the form,

$$V_R(\mathbf{r}^N) = -k_B T \ln \left[ e^{-s(V_W(\mathbf{r}^N) - E_W^R)/k_B T} + e^{-s(V_M(\mathbf{r}^N) - E_M^R)/k_B T} \right], \quad (4)$$

where a configuration of the system of  $N$  particles is denoted as  $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ . The reference-state

parameters  $s$  and  $E_i^R$  are determined in an automatic, iterative procedure [7]. The free enthalpy difference is then given by

$$\Delta G_{MW} = G_M - G_W = -k_B T \ln \left[ \frac{\langle e^{-(V_M - V_R)/k_B T} \rangle_R}{\langle e^{-(V_W - V_R)/k_B T} \rangle_R} \right], \quad (5)$$

where the ensemble average over the Boltzmann ensemble of configurations of the system defined by  $H_R$  is denoted by  $\langle \dots \rangle_R$ . The statistical uncertainties of the estimated free enthalpy differences  $\Delta G_{MW}$  are calculated as given in Ref. [4].

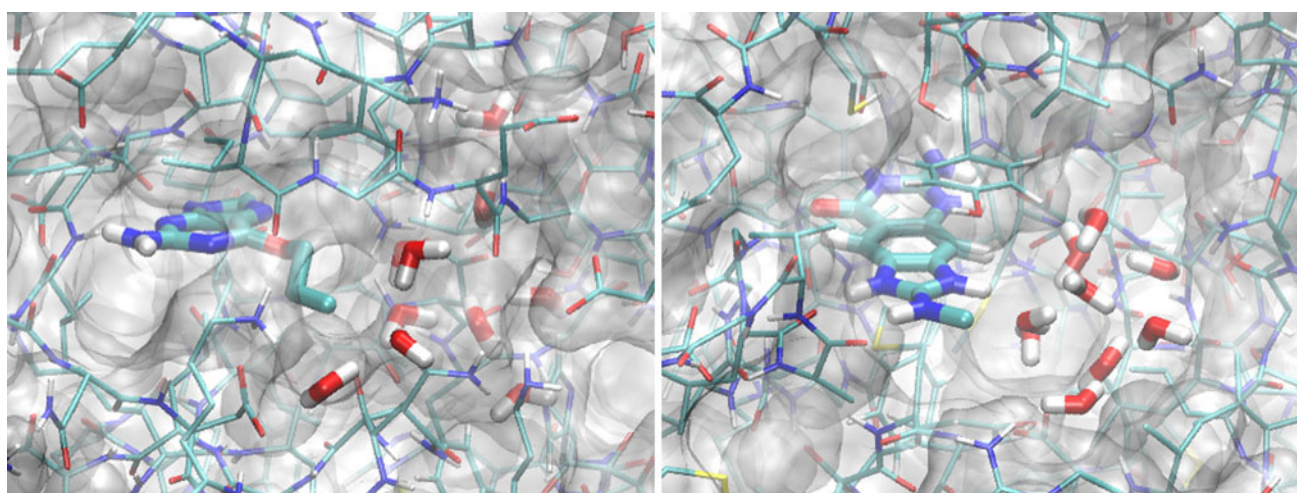
#### Simulation details

The MD simulations reported in this study were performed in explicit solvent at constant temperature and pressure using the GROMOS package of programs [33–35] and the thermodynamically calibrated GROMOS force field 53A6 [36]. The water molecules were modelled using the simple-point-charge (SPC) model [37]. The  $\text{CH}_3$  apolar probe had no charge and the van der Waals parameters of the  $\text{CH}_3$  united atom are given in the GROMOS force field 53A6 [36]. The temperature was maintained close to its reference value  $T = 300$  K by weak coupling to a temperature bath with a relaxation time of 0.1 ps [38]. The pressure was maintained close to its reference value  $P = 1.013$  bar (1 atm) by weak coupling to a pressure bath with a relaxation time of 0.5 ps and using the isothermal compressibility  $\kappa_T = 4.575 \times 10^{-4} (\text{kJ mol}^{-1} \text{nm}^{-3})^{-1}$ . Newton's equations of motion were integrated using the leap-frog scheme [39] with a time step of 2 fs. A reaction field force [40] was applied using the relative dielectric permittivity [41]  $\epsilon_{rf} = 66$ . For the non-bonded interactions, a twin cut-off method was used with a short-range cut-off radius of

**Table 1** Inhibitors of cyclin-dependent kinase 2 (CDK2) and  $IC_{50}$  or % inhibition values taken from Ref. [27]

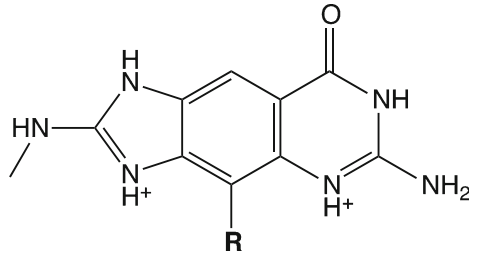
Ligand	R	$IC_{50}$ ( $\mu\text{M}$ ) or % inhibition at 100 $\mu\text{M}$ concentration	$\Delta G_{exp}^{bind}$ ( $\text{kJ mol}^{-1}$ )	PDB code [27]
1		$67 \pm 5$	-24.5	
2		$17 \pm 2$	-27.9	1E1V
3		$31 \pm 3$ %		1GZ8
4		$35 \pm 4$ %		1H0V

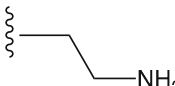
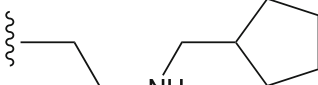
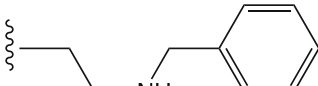
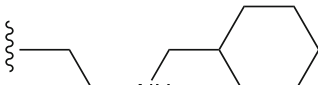
$IC_{50}$  values were converted to  $K_i$  values using the Cheng-Prusoff equation [50] with a substrate concentration  $[S] = 12.5 \mu\text{M}$  and  $K_m = 53 \mu\text{M}$  [25, 27], and subsequently  $\Delta G_{exp}^{bind}$  values were calculated using Eq. (1) at  $T = 300 \text{ K}$



**Fig. 2** Binding pockets of cyclin-dependent kinase 2 (CDK2) (*left*) and tRNA-guanine transglycosylase (TGT) (*right*). Protein residues are drawn in slim licorice mode with carbon atoms colored cyan, nitrogen atoms colored blue, oxygen atoms colored red and hydrogen

atoms in white. The ligand and the water sites used for the perturbations are shown in thick licorice mode. The solvent accessible surface of the protein is rendered as a transparent gray surface

**Table 2** Inhibitors of tRNA-guanine transglycosylase (TGT) and  $K_i$  values taken from Ref. [31]


Ligand	R	$K_i$ (nM)	$\Delta G_{exp}^{bind}$ (kJ mol <sup>-1</sup> )	PDB code [32]
5	H	58 ± 36	-41.6	
6		55 ± 3	-41.7	3GC5
7		2 ± 1	-50.0	3GE7
8		25 ± 2	-43.7	3GC4
9		4 ± 2	-48.2	3EOS

$\Delta G_{exp}^{bind}$  values were calculated using Eq. (1) at  $T = 300$  K

0.8 nm, an intermediate-range cut-off radius of 1.4 nm, and an update frequency of 5 time steps for the short-range pairlist and intermediate-range interactions. The bond lengths were constrained to the ideal values applying the SHAKE algorithm [42].

The initial structure of ligand **1** (Table 1) bound to CDK2 was derived from a crystal structure of monomeric CDK2 with ligand **2** (RCSB protein data bank [43] entry 1E1V). The initial structure of ligand **5** (Table 2) bound to TGT was derived from a crystal structure of TGT with 2,6-diamino-1,7-dihydro-8H-imidazo[4,5-g]quinazolin-8-one (RCSB protein data bank [43] entry 2Z7K). The first ten residues were not considered. Positions of atoms in missing residues were modelled using the web service ModLoop [44]. The protonation states were chosen to match a pH of 7 and hydrogen atoms were added to the X-ray structures according to standard geometric criteria. The histidine side chains were protonated at  $N_\delta$  or  $N_\epsilon$  depending on their hydrogen-bonding environment. Sodium or chlorine counter ions were added to neutralize the total charge of the protein. The parameters for ligand **1** (Table 1) and ligand **5** (Table 2) were derived from a combination of *ab initio* quantum-chemical calculations [45] and a comparison of

parameters of similar functional groups. The parameters are given in Tables S1 and S2 in supplementary material.

CDK2 was solvated in a rectangular periodic box with 16,485 water molecules and TGT likewise in a box with 15,782 water molecules. Initial velocities were randomly generated from a Maxwell-Boltzmann distribution at 60 K. During the equilibration, the temperature was increased stepwise from 60 to 300 K in steps of 60 K, after periods of 20 ps. Harmonic position restraints were initially applied to the protein atoms and were relaxed by lowering the force constant by a factor 10 after each period. The water sites in the binding pocket were selected in the final configuration of the equilibration simulation. During the EDS simulations, harmonic position restraints with a force constant of 500 kJ mol<sup>-1</sup> nm<sup>-2</sup> with respect to these sites were applied to the oxygen atoms of the ten (in the case of CDK2) and nine (in the case of TGT) selected water molecules. The EDS reference state parameters were obtained in an update simulation of 4–6 ns length with averaging periods of 0.2 ns using the scheme described in Ref. [7] and are listed in Tables S3 and S4 in supplementary material. Subsequently, production simulations were carried out for 4 ns. The perturbation of a single H<sub>2</sub>O

molecule to a CH<sub>3</sub> probe was performed for each selected water site in the binding pocket. In addition, the same perturbations were performed for the CDK2 system where sites 3, 4, and 9 were permanently occupied by a CH<sub>3</sub> probe, and for the TGT system where site 9 was permanently occupied by a CH<sub>3</sub> probe.

The perturbation of a single H<sub>2</sub>O molecule to a CH<sub>3</sub> probe in solution was performed using EDS, where the oxygen atoms of ten out of 1024 H<sub>2</sub>O molecules in a periodic box were positionally restrained with a force constant of 500 kJ mol<sup>-1</sup> nm<sup>-2</sup>. A production simulation of 3 ns was carried out. In addition, the same perturbation was performed with site 1 being permanently occupied by a CH<sub>3</sub> probe and the adjacent site 7 was perturbed.

### Analysis

The GROMOS analysis programs used for this study are described in detail in Ref. [46]. Energies were extracted using the program *ene\_ana*. The energy  $V_A$  in the ensemble  $A$ , i.e.  $\langle V_A \rangle_A$ , from configurations of ensemble  $R$  can be obtained by reweighting

$$\langle V_A \rangle_A = \frac{\langle V_A e^{-\beta(V_A - V_R)} \rangle_R}{\langle e^{-\beta(V_A - V_R)} \rangle_R} = \frac{\langle X \rangle_R}{\langle Y \rangle_R} \quad (6)$$

using the program *reweight*. The statistical uncertainty of the reweighted energy  $\langle V_A \rangle_A$  of state  $A$  was calculated analogously to the one for the free enthalpy difference [4],

$$\delta_{V_A}^2 = \frac{1}{N} \left( \frac{g_{XX}\sigma_{XX}}{\mu_X^2} + \frac{g_{YY}\sigma_{YY}}{\mu_Y^2} - 2 \frac{g_{XY}\sigma_{XY}}{\mu_X\mu_Y} \right) \quad (7)$$

where  $g$  denotes the statistical inefficiency calculated according to Eqs. (46)–(48) of Ref. [47],  $\sigma$  the covariance and  $\mu$  the average. The free enthalpy differences of the EDS simulations were calculated using the program *dfmult*. The error of differences of enthalpies and/or free enthalpies was obtained from a squared sum,

$$\delta_{AB} = \sqrt{\delta_A^2 + \delta_B^2} \quad (8)$$

Configurations and movies were extracted using the program *frameout*. Figures were generated using visual molecular dynamics (VMD) [48].

## Results and discussion

For the values of potential energies, free enthalpy differences, enthalpy differences and entropy differences the statistical uncertainty is given. However, as the true uncertainty is most likely larger, only differences larger than  $k_B T = 2.5$  kJ mol<sup>-1</sup> at 300 K are considered significant and will be discussed in detail.

Analysis of the free enthalpy differences and the enthalpic and entropic contributions

First the results for CDK2 are discussed. The results of the perturbation of a water molecule to a CH<sub>3</sub> group at the ten sites in the binding pocket of CDK2 are listed in Table 3. The corresponding values for the perturbation in bulk solution are given in the first row of Table 4. For each water site, the occupancy volumes are shown separately in Fig. 3 and for all ten water sites colored according to the value of  $\Delta G_M^{bind}$  in Fig. 4. A favourable water displacement, i.e. negative  $\Delta G_M^{bind}$ , is found for sites 3, 4 and 9 (Table 3). The displacement is in all three cases enthalpically favourable, i.e.  $\Delta H_M^{bind}$  is negative, and—to a lesser extent—entropically unfavourable, i.e.  $T\Delta S_M^{bind}$  is also negative but smaller. If a water molecule cannot form many hydrogen bonds in the binding pocket, its potential energy is likely to be less negative compared to the bulk, thus making a displacement enthalpically favourable, but at the same time the water molecule at such a location possesses more rotational freedom compared to the bulk, leading to a negative entropy difference for its displacement. Interesting observations can also be made for the water sites 1 and 10 where the displacement of water does not lead to a change in enthalpy but is entropically highly unfavourable. The replacement of water from site 2 by the apolar probe on the other hand is enthalpically highly unfavourable but entropically slightly favourable resulting in an overall unfavourable displacement. Such a site may be favourably occupied by a polar atom of a ligand. For an exact prediction, however, the calculation would need to be repeated with a polar probe instead of an apolar one. In general it is to be noted that the entropic contributions may vary between the apolar probe for which the calculations were performed and a real ligand atom depending on the rotational freedom accessible to the ligand atom.

The results of the perturbation of a water molecule to a CH<sub>3</sub> group at the nine sites in the binding pocket of TGT are listed in Table 5. The occupancy volumes are shown separately for each water site in Fig. 5 and for all water sites colored according to the value of  $\Delta G_M^{bind}$  in Fig. 6. Although the same force constant for the position restraining was used for both proteins, the water molecules in the binding pocket of TGT explored a much bigger volume. In some cases, only part of the simulation trajectory could be used for analysis in order to obtain meaningful results. For the water molecules at sites 4, 6, 7 and 9, only the first 2 ns were used for analysis, and for the water molecules at site 3 the first 2.4 ns of the simulation. Two sites, 4 and 9, were identified for a favourable replacement of water by an apolar probe, although the free enthalpy difference of site 4 is within the estimated overall uncertainty of 2.5 kJ mol<sup>-1</sup>. Both

**Table 3** Free enthalpy difference  $\Delta G_{MW}^{pocket}$ , enthalpy difference  $\Delta H_{MW}^{pocket}$  and entropy difference  $T\Delta S_{MW}^{pocket}$  of perturbing a single water (W) into a CH<sub>3</sub> group (M) inside the binding pocket of cyclin-dependent kinase 2 (CDK2). The binding free enthalpy  $\Delta G_M^{bind}$ ,binding enthalpy  $\Delta H_M^{bind}$  and binding entropy  $T\Delta S_M^{bind}$  of CH<sub>3</sub> are obtained by subtraction of the values in the binding pocket from the corresponding values in solution (first row of Table 4)

H <sub>2</sub> O site	$\Delta G_{MW}^{pocket}$	$\Delta H_{MW}^{pocket}$	$T\Delta S_{MW}^{pocket}$	$\Delta G_M^{bind}$	$\Delta H_M^{bind}$	$T\Delta S_M^{bind}$
1	52.4 ± 1.4	74.1 ± 1.1	21.7 ± 1.8	12.8 ± 1.5	0.4 ± 1.2	-12.4 ± 1.9
2	48.7 ± 1.9	84.4 ± 1.6	35.7 ± 2.4	9.1 ± 1.9	10.7 ± 1.6	1.6 ± 2.5
3	33.4 ± 0.5	59.3 ± 0.4	25.9 ± 0.6	-6.2 ± 0.6	-14.4 ± 0.6	-8.2 ± 0.8
4	37.2 ± 1.0	71.0 ± 0.8	33.8 ± 1.3	-2.4 ± 1.1	-2.7 ± 0.9	-0.3 ± 1.3
5	50.2 ± 1.0	82.2 ± 0.8	32.0 ± 1.3	10.6 ± 1.1	8.5 ± 0.9	-2.1 ± 1.3
6	40.9 ± 1.0	71.5 ± 0.8	30.6 ± 1.3	1.3 ± 1.1	-2.2 ± 0.9	-3.5 ± 1.3
7	39.7 ± 0.4	75.2 ± 0.4	35.5 ± 0.4	0.1 ± 0.6	1.5 ± 0.6	1.4 ± 0.8
8	40.8 ± 0.5	74.3 ± 0.4	33.5 ± 0.6	0.6 ± 0.6	1.2 ± 0.6	-0.6 ± 0.8
9	34.6 ± 0.9	53.4 ± 0.6	18.8 ± 1.1	-5.0 ± 1.0	-20.3 ± 0.7	-15.3 ± 1.2
10	49.4 ± 1.0	74.1 ± 0.8	24.7 ± 1.3	9.8 ± 1.1	0.4 ± 0.9	-9.4 ± 1.4

All values in kJ mol<sup>-1</sup>**Table 4** Free enthalpy difference  $\Delta G_{MW}^{unbound}$ , enthalpy difference  $\Delta H_{MW}^{unbound}$  and entropy difference  $T\Delta S_{MW}^{unbound}$  of perturbing a single water (W) into a CH<sub>3</sub> group (M) in solution using a force constant of 500 kJ mol<sup>-1</sup> nm<sup>-2</sup> for the position restraints, where no or one adjacent site is permanently transformed to a CH<sub>3</sub>

CH <sub>3</sub>	$\Delta G_{MW}^{unbound}$	$\Delta H_{MW}^{unbound}$	$T\Delta S_{MW}^{unbound}$
0	39.6 ± 0.4	73.7 ± 0.4	34.1 ± 0.6
1	38.1 ± 0.4	72.2 ± 0.3	34.1 ± 0.5

All values in kJ mol<sup>-1</sup>

displacements are enthalpically favourable and entropically neutral or slightly unfavourable. The enthalpy change upon displacement of water at sites 3 and 7 is also favourable, but in this case the unfavourable entropy change is larger, leading to a positive  $\Delta G_M^{bind}$ . Replacing the water molecules at sites 1, 5 and 8 by a CH<sub>3</sub> group on the other hand is enthalpically highly unfavourable, but to a lesser extent entropically favoured similar to site 2 in the case of CDK2. Again, such sites may be favourably occupied by a polar atom of a ligand. The water molecule at site 6, however, should not be displaced at all as this is both enthalpically and entropically unfavourable.

If solely the potential energies of the water sites in the binding pocket and in solution were compared (Table 6), only sites with an unfavourable enthalpy difference between pocket and solution, for example sites 3 and 9 in the case of CDK2 or site 9 in the case of TGT, would have been predicted favourable for displacement. Site 7 in the case of TGT, however, would have been predicted as favourable for displacement due to positive  $\Delta H_W^{pocket-unbound}$ , while the unfavourable entropic change upon displacement results in an unfavourable, positive  $\Delta G_M^{bind}$ . This illustrates the importance of considering both the enthalpic and entropic contributions of  $\Delta G$  in the analysis of water molecules in the binding pocket.

### Hydrogen-bonding environment in the binding pocket

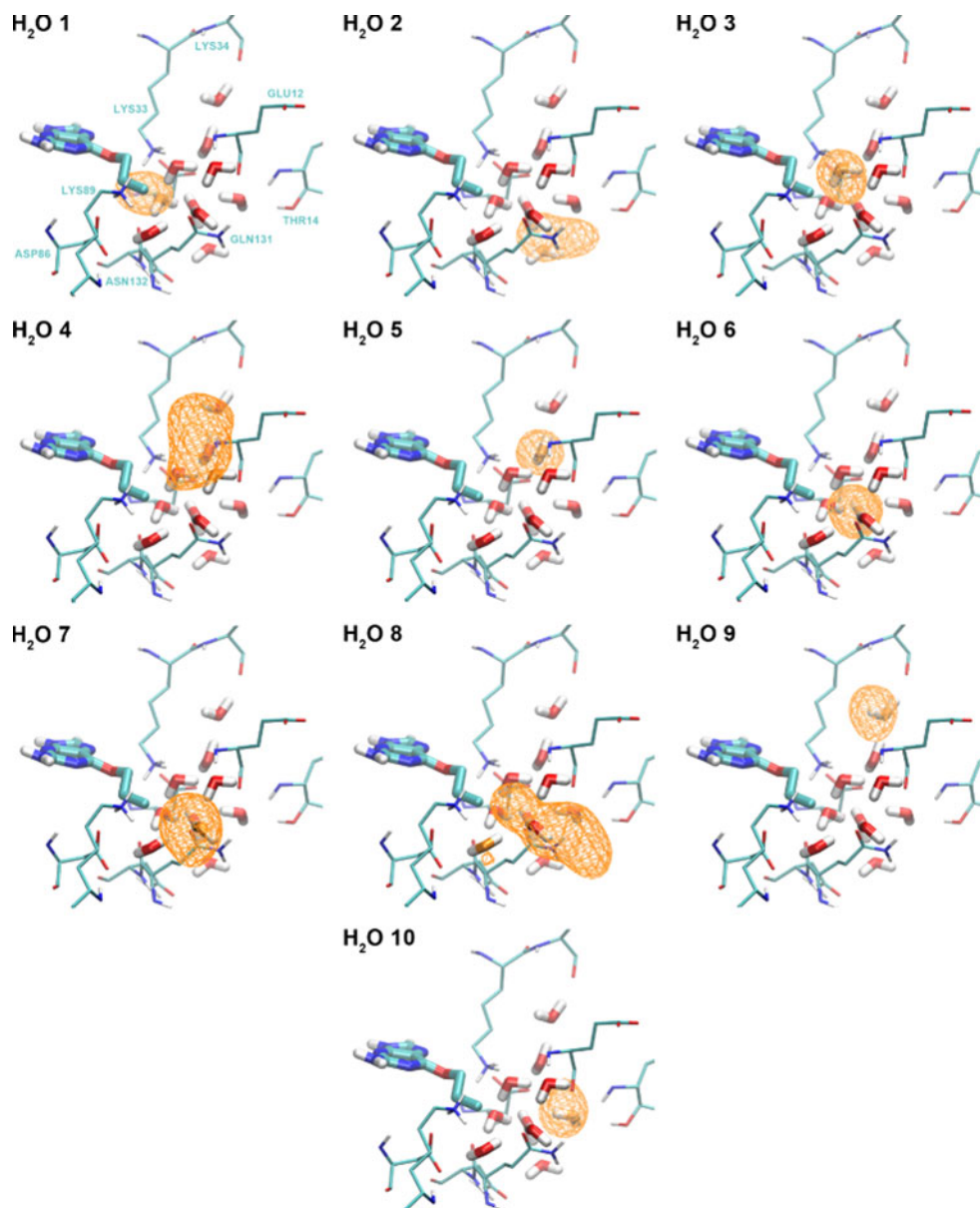
The hydrogen-bonding environments of the selected water sites in the initial configuration of the binding pocket of cyclin-dependent kinase 2 (CDK2) are shown in Fig. 7 and of tRNA-guanine transglycosylase (TGT) in Fig. 8.

In the case of CDK2, there is no apparent correlation between the number of hydrogen bonds observed in the initial configuration and the potential energies  $V_W^{pocket}$  of the water molecules obtained from the 4 ns simulations (Table 6). Water sites 5 and 9 for example which have both two hydrogen bonds have a difference in potential energy of 27.6 kJ mol<sup>-1</sup>. The water sites with four hydrogen bonds on the other hand, sites 1, 4 and 8, have—as expected—a potential energy rather close to the one in bulk solution, but this is also the case for water sites 6 and 7 with three hydrogen bonds and site 10 with only two hydrogen bonds. Water site 2 has with only 3 hydrogen bonds a potential energy much lower than the one in bulk solution. On the other hand, the water molecule at site 3 forms only one hydrogen bond and has a rather positive potential energy compared to the bulk.

There is also no apparent correlation between the number of hydrogen bonds in the initial configuration and the potential energies  $V_W^{pocket}$  of the water molecules obtained from the simulations in the case of TGT (Table 6). Water molecules with the same number of hydrogen bonds have different potential energies, for example water sites 6 and 9, and others with similar potential energies form different numbers of hydrogen bonds, for example water sites 7 and 9. The only water molecule, site 5, which forms four hydrogen bonds on the other hand also has the lowest potential energy.

These observations may be an indication that the analysis of the hydrogen-bonding network of a single

**Fig. 3** Occupancy volume (orange net) of the water sites in the binding pocket of cyclin-dependent kinase 2 (CDK2) during the simulations of 4 ns length. The cut-off was at 20 % of the maximum occupancy. Selected protein residues are drawn in slim licorice mode with carbon atoms colored cyan, nitrogen atoms colored blue, oxygen atoms colored red and hydrogen atoms in white. The ligand and the water sites used for the perturbations are shown in thick licorice mode. For the water molecule of interest, the oxygen atom is colored orange



configuration, e.g. a crystal structure, is not sufficient to predict the potential energy of a water molecule and subsequently the favourable or unfavourable nature of the displacement of this water molecule by an atom of a ligand.

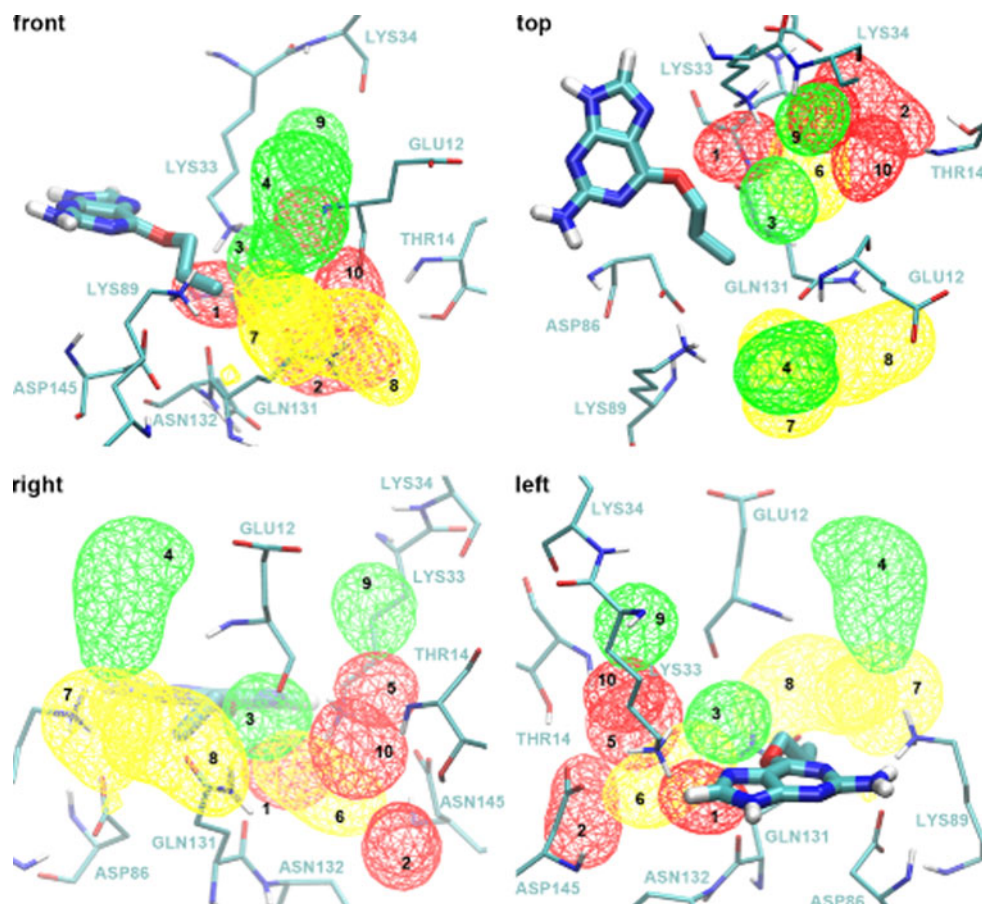
#### Overlay with inhibitor structures

The overlay of the occupancy volumes of the water sites with available crystal structures of ligands 2–4 of CDK2 are shown in Fig. 9. Ligand 2 has an inhibition concentration that is a factor three lower than that of ligand 1. As can be seen in the left panel in Fig. 9, the cyclohexyl ring of ligand 2 displaces water from site 3. Such a displacement is predicted to be favourable, mainly due to a negative enthalpy difference. If the water is displaced by something

polar on the other hand as in ligand 3 (middle panel in Fig. 9), the ligand is markedly less active. The overlay of ligand 4 (right panel in Fig. 9) shows the displacement of water from site 3 by the apolar side of the pyrrolidinone ring, but the oxygen of the amide is close to water at sites 4 and 7. Displacement of water from sites 4 and 7 by something polar should be either neutral or weakly unfavourable, which is probably not enough to explain the difference in activity between ligand 2 and ligand 4. However, the location of the amide oxygen of ligand 4 is suboptimal with respect to formation of hydrogen bonds with the backbone of GLU12 and the side chain of GLN131 which are hydrogen-bonded to water molecules 4 and 7 in the presence of ligand 1, making the replacement of ligand 2 by ligand 4 enthalpically more unfavourable.



**Fig. 4** Occupancy volumes of the water sites in the binding pocket of cyclin-dependent kinase 2 (CDK2) during the simulations of 4 ns length colored according to  $\Delta G_M^{bind}$ . Negative  $\Delta G_M^{bind}$  values are shown as *green nets*, neutral or slightly positive (up to  $2.5 \text{ kJ mol}^{-1}$ )  $\Delta G_M^{bind}$  values as *yellow nets* and positive  $\Delta G_M^{bind}$  values as *red nets*. The cut-off for the occupancy was at 20 % of the maximum occupancy. Selected protein residues are drawn in slim licorice mode with carbon atoms colored cyan, nitrogen atoms colored *blue*, oxygen atoms colored *red* and hydrogen atoms in *white*. The ligand is shown in thick licorice mode



**Table 5** Free enthalpy difference  $\Delta G_{MW}^{pocket}$ , enthalpy difference  $\Delta H_{MW}^{pocket}$  and entropy difference  $T\Delta S_{MW}^{pocket}$  of perturbing a single water (W) into a  $\text{CH}_3$  group (M) inside the binding pocket of tRNA guanine transglycosylase (TGT). The binding free enthalpy  $\Delta G_M^{bind}$ , binding

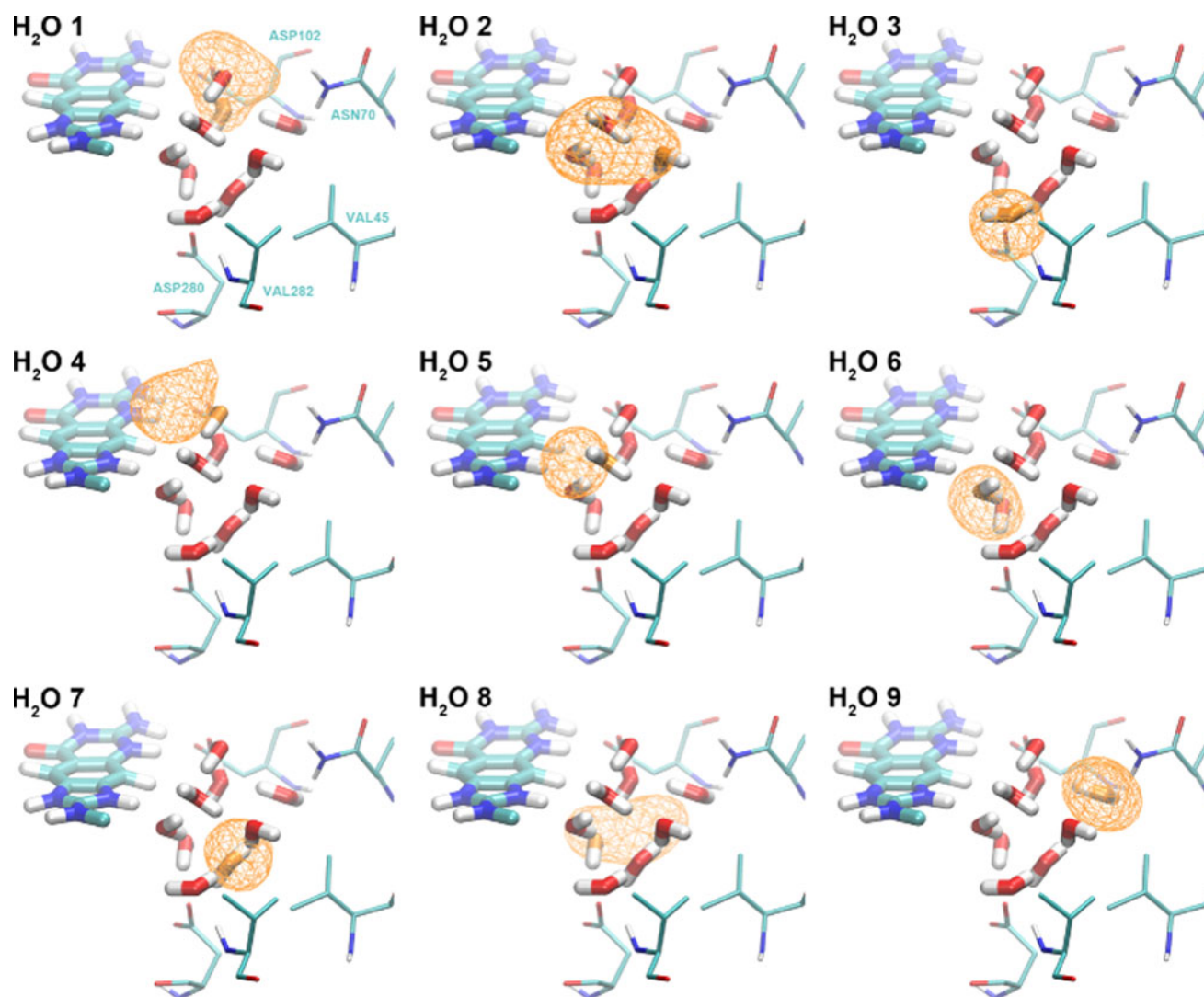
enthalpy  $\Delta H_M^{bind}$  and binding entropy  $T\Delta S_M^{bind}$  of  $\text{CH}_3$  are obtained by subtraction of the values in the binding pocket from the corresponding values in solution (first row in Table 4)

$\text{H}_2\text{O}$ site	$\Delta G_{MW}^{pocket}$	$\Delta H_{MW}^{pocket}$	$T\Delta S_{MW}^{pocket}$	$\Delta G_M^{bind}$	$\Delta H_M^{bind}$	$T\Delta S_M^{bind}$
1	$41.7 \pm 1.9$	$83.7 \pm 1.6$	$42.0 \pm 2.5$	$2.1 \pm 1.9$	$10.0 \pm 1.6$	$7.9 \pm 2.5$
2	$41.7 \pm 0.6$	$77.4 \pm 0.4$	$35.7 \pm 0.7$	$2.1 \pm 0.7$	$3.7 \pm 0.6$	$1.6 \pm 0.9$
3	$42.0 \pm 1.0$	$70.3 \pm 0.8$	$28.3 \pm 1.3$	$2.4 \pm 1.1$	$-3.4 \pm 0.9$	$-5.8 \pm 1.3$
4	$37.5 \pm 1.8$	$71.4 \pm 1.4$	$33.9 \pm 2.3$	$-2.1 \pm 1.8$	$-2.3 \pm 1.5$	$-0.2 \pm 2.3$
5	$52.1 \pm 2.3$	$90.6 \pm 1.8$	$38.5 \pm 2.9$	$12.5 \pm 2.3$	$16.9 \pm 1.8$	$4.4 \pm 2.9$
6	$55.7 \pm 1.7$	$81.5 \pm 1.4$	$25.8 \pm 2.2$	$16.1 \pm 1.7$	$7.8 \pm 1.5$	$-8.3 \pm 2.3$
7	$42.6 \pm 1.7$	$68.7 \pm 1.4$	$26.1 \pm 2.2$	$3.0 \pm 1.7$	$-5.0 \pm 1.5$	$-8.0 \pm 2.3$
8	$44.2 \pm 1.6$	$82.0 \pm 1.2$	$37.8 \pm 2.0$	$4.6 \pm 1.6$	$8.3 \pm 1.3$	$3.7 \pm 2.1$
9	$36.0 \pm 1.1$	$68.9 \pm 0.9$	$32.9 \pm 1.3$	$-3.6 \pm 1.2$	$-4.8 \pm 1.0$	$-1.2 \pm 1.6$

All values in  $\text{kJ mol}^{-1}$

The overlay of the occupancy volumes of the water sites with available crystal structures of ligands **6–9** of TGT are shown in Fig. 10. The primary ammonium group of ligand **6** is located near the water sites 5 and 8. As discussed before, the displacement of water molecules from these sites by an apolar probe is enthalpically

highly unfavourable but to a lesser extent entropically favourable. The displacement of water from these sites by a polar group of a ligand is expected to be enthalpically neutral and possibly entropically favourable. The latter effect does not seem to occur as the binding affinities of ligand **5** and ligand **6** are similar (Table 2).



**Fig. 5** Occupancy volume (*orange net*) of the water sites in the binding pocket of tRNA-guanine transglycosylase (TGT) during the simulations of 4 ns length. The cut-off was at 20 % of the maximum occupancy. Selected protein residues are drawn in slim licorice mode with carbon atoms colored cyan, nitrogen atoms colored *blue*, oxygen

atoms colored *red* and hydrogen atoms in *white*. The ligand and the water sites used for the perturbations are shown in thick licorice mode. For the water molecule of interest, the oxygen atom is colored *orange*

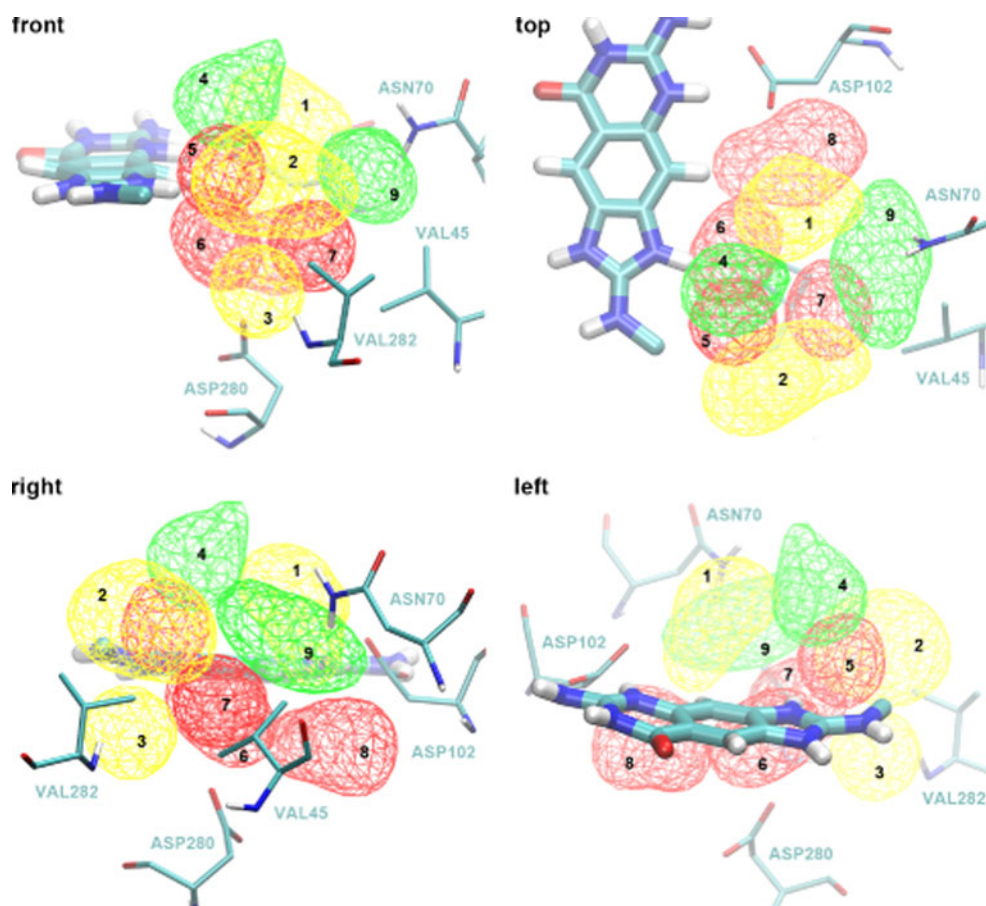
However, the measured binding affinity of ligand **5** has a large uncertainty, and the rotational freedom of the ligand substituent may be different than for the apolar probe. The cyclopentyl group of ligand **7** and the cyclohexyl group of ligand **9** displace the water at site 9 which is predicted to be favourable. Indeed, both ligands have a similar binding affinity which is approximately a factor 20 smaller than the binding affinity of ligand **5**. The phenyl group of ligand **8** on the other hand displaces also the water molecule at site 1 which is slightly unfavourable, but becomes significantly unfavourable if correlations are taken into account (see “[Correlation between replacements at adjacent water sites](#)” section). This is reflected by a weaker binding affinity compared to ligand **5** (Table 2).

#### Correlation between replacements at adjacent water sites

For the water sites which are favourable for replacement by a methyl group, the correlation between water displacements at adjacent sites was investigated, i.e. the effect an apolar probe at one site has on the  $\Delta G_M^{bind}$  of its neighbouring sites. Thereby, a site that is favourable for water replacement by an apolar probe was occupied by a CH<sub>3</sub> group and the adjacent water sites were perturbed. The results are listed in Table 7. The corresponding values for the perturbation in bulk solution are given in the second row of Table 4.

Three water sites in the binding pocket of CDK2 were identified as favourable for displacement of water by an apolar probe. Replacing water at site 3 by a CH<sub>3</sub> group, the positive

**Fig. 6** Occupancy volume of the water sites in the binding pocket of tRNA-guanine transglycosylase (TGT) during the simulation of 4 ns length colored according to  $\Delta G_M^{bind}$ . Negative  $\Delta G_M^{bind}$  values are shown as green nets, neutral or slightly positive (up to  $2.5 \text{ kJ mol}^{-1}$ )  $\Delta G_M^{bind}$  values as yellow nets and positive  $\Delta G_M^{bind}$  values as red nets. The cut-off for the occupancy was at 20 % of the maximum occupancy. Selected protein residues are drawn in slim licorice mode with carbon atoms colored cyan, nitrogen atoms colored blue, oxygen atoms colored red and hydrogen atoms in white. The ligand is shown in thick licorice mode



**Table 6** Potential energy  $V_W^{pocket}$  of a water molecule at the ten and nine sites in the binding pockets of cyclin-dependent kinase 2 (CDK2) and tRNA-guanine transglycosylase (TGT) and enthalpy difference  $\Delta H_W^{pocket-unbound}$  for a water molecule at a site in the pocket and in solution

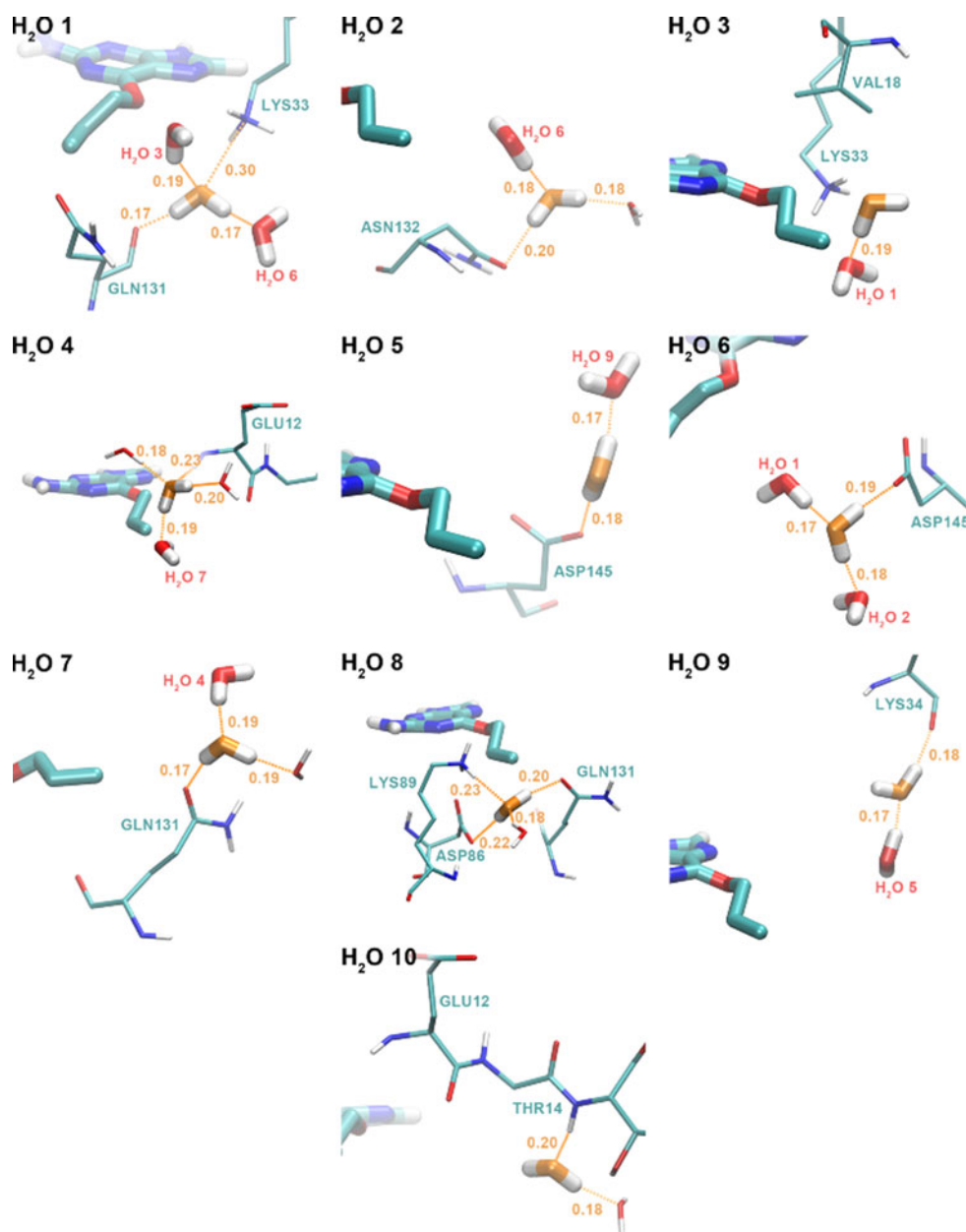
The potential energy of a water molecule at a site in solution is  $V_W^{unbound} = -84.1 \pm 0.2 \text{ kJ mol}^{-1}$ . All values in  $\text{kJ mol}^{-1}$

H <sub>2</sub> O site	CDK2		TGT	
	$V_W^{pocket}$	$\Delta H_W^{pocket-unbound}$	$V_W^{pocket}$	$\Delta H_W^{pocket-unbound}$
1	$-86.6 \pm 0.7$	$-2.5 \pm 0.7$	$-99.6 \pm 1.3$	$-15.5 \pm 1.3$
2	$-100.0 \pm 1.1$	$-15.9 \pm 1.1$	$-89.8 \pm 0.2$	$-5.7 \pm 0.3$
3	$-73.7 \pm 0.2$	$10.4 \pm 0.3$	$-85.0 \pm 0.8$	$-0.9 \pm 0.8$
4	$-84.7 \pm 0.6$	$-0.6 \pm 0.6$	$-85.9 \pm 1.4$	$-1.8 \pm 1.4$
5	$-96.0 \pm 0.7$	$-11.9 \pm 0.7$	$-102.4 \pm 0.3$	$-18.3 \pm 0.4$
6	$-86.7 \pm 0.6$	$-2.6 \pm 0.6$	$-92.2 \pm 0.2$	$-8.1 \pm 0.3$
7	$-87.9 \pm 0.3$	$-3.8 \pm 0.4$	$-80.9 \pm 0.4$	$3.2 \pm 0.4$
8	$-87.7 \pm 0.3$	$-3.6 \pm 0.4$	$-95.6 \pm 0.4$	$-11.5 \pm 0.4$
9	$-68.4 \pm 0.5$	$15.7 \pm 0.5$	$-81.5 \pm 0.8$	$2.6 \pm 0.8$
10	$-88.5 \pm 0.6$	$-4.4 \pm 0.6$		

$\Delta G_M^{bind}$  values of the adjacent water sites 1, 5 and 10 are reduced, i.e. less positive, because the enthalpy difference has changed from unfavourable to favourable due to the neighbouring apolar site. Interestingly, the effect of the adjacent apolar probe is very different for water site 6 where  $\Delta G_M^{bind}$  was in the neutral range with a favourable enthalpy change and an unfavourable entropy change. With the neighbouring apolar probe, the enthalpy difference becomes unfavourable and thus  $\Delta G_M^{bind}$  becomes clearly positive. Replacing water by

an apolar probe at site 4, the effect on  $\Delta G_M^{bind}$  of water site 7 is within the uncertainty of the calculation and for water 8, there is a shift from neutral to weakly unfavourable. The most significant effect of correlation between replacements of water by an apolar moiety at adjacent sites can be observed for sites 5 and 9. If at site 9 water is replaced by an apolar probe,  $\Delta G_M^{bind}$  at site 5 changes from highly unfavourable to weakly favourable which is mainly due to the displacement becoming enthalpically favourable.

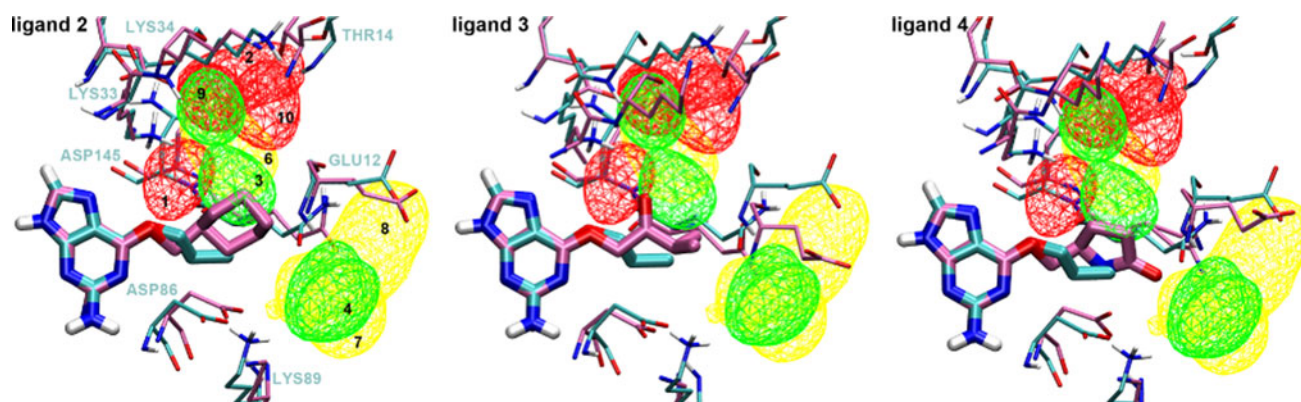
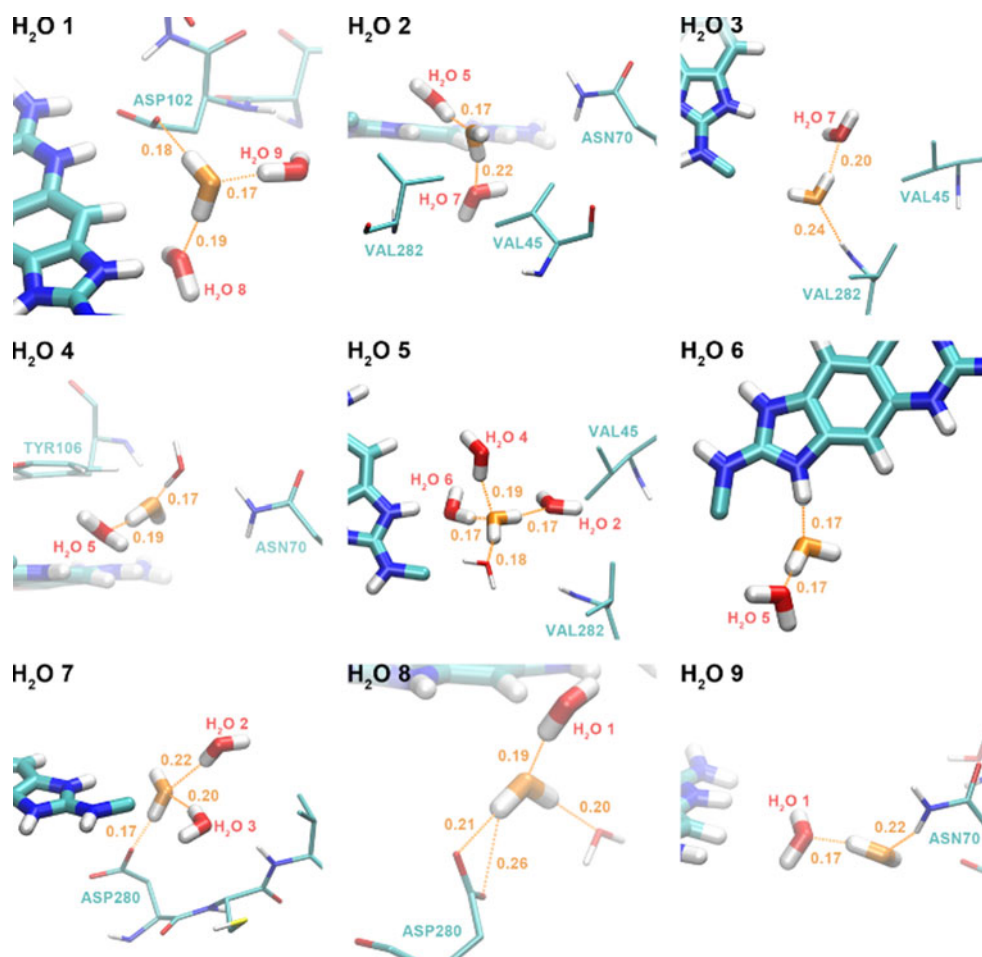
**Fig. 7** Hydrogen-bonding environment of the water sites in the binding pocket of cyclin-dependent kinase 2 (CDK2). Selected protein residues and selected solvent molecules are drawn in slim licorice mode with carbon atoms colored cyan, nitrogen atoms colored blue, oxygen atoms colored red and hydrogen atoms in white. The ligand and the water sites used for the perturbations are shown in thick licorice mode. For the water molecule of interest, the oxygen atom is colored orange. Atom distances in nm



In the case of TGT, the correlation between the displacements of water at site 9 and displacements at neighbouring sites was studied (Table 7). Again, only part of the simulation trajectory could be used in some cases for analysis in order to obtain meaningful results. For the water molecules at sites 4, 7 and 8, only the first 2 ns of simulation were used for analysis. The effect on  $\Delta G_M^{bind}$  at site 7 and its enthalpic and entropic contributions is small. In the case of water site 2,  $\Delta G_M^{bind}$  increases slightly and the enthalpic contribution becomes more unfavourable while the entropic contribution becomes more favourable.  $\Delta H_M^{bind}$  of site 8 becomes also more positive, but here the entropic contribution becomes less favourable leading to a larger

increase in  $\Delta G_M^{bind}$ .  $\Delta G_M^{bind}$  of site 1 on the other hand increases significantly, i.e. becomes more unfavourable, with the enthalpy change becoming less unfavourable but the entropic contribution changing from favourable to unfavourable. This observation may explain the rather large change in binding affinity of a factor ten between ligand 8 compared to ligand 9. The water molecule at site 1 is displaced by ligand 8 but not by ligand 9. If the correlation between site 9 and site 1 is not taken into account,  $\Delta G_M^{bind}$  is only slightly unfavourable and may not be enough to explain the factor ten. For the water at site 4 lastly,  $\Delta H_M^{bind}$  changes to unfavourable and  $T\Delta S_M^{bind}$  changes to favourable leading to an similar value for  $\Delta G_M^{bind}$ .

**Fig. 8** Hydrogen-bonding environment of the water sites in the binding pocket of tRNA-guanine transglycosylase (TGT). Selected protein residues and selected solvent molecules are drawn in slim licorice mode with carbon atoms colored *cyan*, nitrogen atoms colored *blue*, oxygen atoms colored *red* and hydrogen atoms in *white*. The ligand and the water sites used for the perturbations are shown in thick licorice mode. For the water molecule of interest, the oxygen atom is colored *orange*. Atom distances in nm



**Fig. 9** Overlay of the crystal structures of 2–4 with ligand 1 and the occupancy volumes of the water sites in the binding pocket of cyclin-dependent kinase 2 (CDK2) colored according to  $\Delta G_M^{bind}$ . Negative  $\Delta G_M^{bind}$  values are shown as *green nets*, neutral or slightly positive (up to  $2.5 \text{ kJ mol}^{-1}$ )  $\Delta G_M^{bind}$  values as *yellow nets* and positive  $\Delta G_M^{bind}$

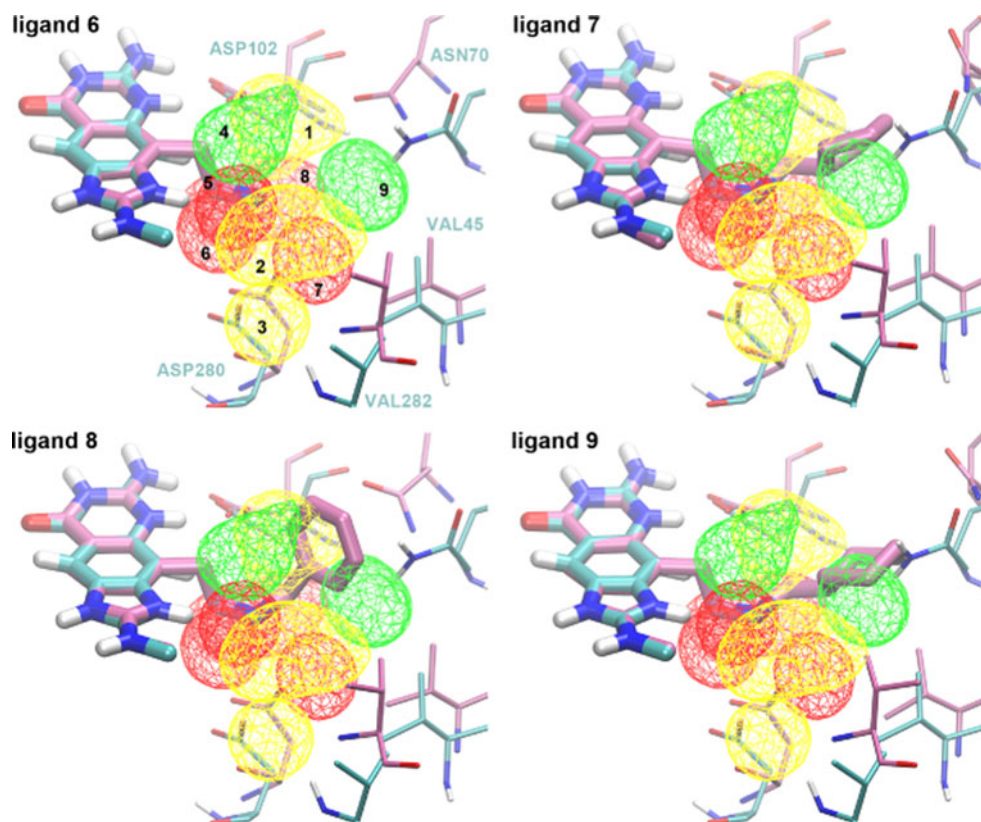
values as *red nets*. The cut-off for the occupancy was at 20 % of the maximum occupancy. Selected protein residues are drawn in slim licorice mode with carbon atoms colored *cyan*, nitrogen atoms colored *blue*, oxygen atoms colored *red* and hydrogen atoms in *white*. The ligands are shown in thick licorice mode

#### Computational effort

For the systems used in this study with 52,547 (CDK2) and 50,980 atoms (TGT), the computational effort was

26 h per ns using 8 cores on an AMD Opteron 6147 compute cluster, where each node contains two 12-core CPUs. The computational effort, however, depends largely on the simulation software in general and the

**Fig. 10** Overlay of the crystal structures of ligands **6** - **9** with ligand **5** and the occupancy volumes of the water sites in the binding pocket of tRNA-guanine transglycosylase (TGT) colored according to  $\Delta G_M^{bind}$ . Negative  $\Delta G_M^{bind}$  values are shown as *green nets*, neutral or slightly positive (up to  $2.5 \text{ kJ mol}^{-1}$ )  $\Delta G_M^{bind}$  values as *yellow nets* and positive  $\Delta G_M^{bind}$  values as *red nets*. The cut-off for the occupancy was at 20 % of the maximum occupancy. Selected protein residues are drawn in slim licorice mode with carbon atoms colored *cyan*, nitrogen atoms colored *blue*, oxygen atoms colored *red* and hydrogen atoms in *white*. The ligands are shown in thick licorice mode



**Table 7** Free enthalpy difference  $\Delta G_{MW}^{pocket}$ , enthalpy difference  $\Delta H_{MW}^{pocket}$  and entropy difference  $T\Delta S_{MW}^{pocket}$  of perturbing a single water (W) into a  $\text{CH}_3$  group (M) inside the binding pocket where one neighbouring  $\text{CH}_3$  group is present. Data for the binding pocket of cyclin-dependent kinase 2 (CDK2) and tRNA-guanine transglycosylase (TGT) are shown. The

binding free enthalpy  $\Delta G_M^{bind}$ , binding enthalpy  $\Delta H_M^{bind}$  and binding entropy  $T\Delta S_M^{bind}$  of  $\text{CH}_3$  are obtained by subtraction of the values in the binding pocket from the corresponding values in solution (second row in Table 4)

Protein	$\text{CH}_3$ site	$\text{H}_2\text{O}$ site	$\Delta G_{MW}^{pocket}$	$\Delta H_{MW}^{pocket}$	$T\Delta S_{MW}^{pocket}$	$\Delta G_M^{bind}$	$\Delta H_M^{bind}$	$T\Delta S_M^{bind}$
CDK2	3	1	$47.6 \pm 1.3$	$68.4 \pm 1.1$	$20.8 \pm 1.7$	$9.5 \pm 1.4$	$-3.8 \pm 1.1$	$-13.3 \pm 1.8$
		5	$44.5 \pm 1.1$	$70.3 \pm 0.9$	$25.8 \pm 1.4$	$6.4 \pm 1.2$	$-1.9 \pm 0.9$	$-8.3 \pm 1.5$
		6	$45.3 \pm 1.0$	$74.3 \pm 0.8$	$29.0 \pm 1.3$	$7.2 \pm 1.1$	$2.1 \pm 1.9$	$-5.1 \pm 1.3$
		10	$42.4 \pm 1.4$	$68.4 \pm 1.1$	$26.2 \pm 1.8$	$4.1 \pm 1.5$	$-3.8 \pm 1.1$	$-7.9 \pm 1.9$
	4	7	$38.8 \pm 0.6$	$75.1 \pm 0.5$	$36.3 \pm 0.8$	$0.7 \pm 0.7$	$2.9 \pm 0.6$	$2.2 \pm 0.9$
		8	$40.9 \pm 1.5$	$78.4 \pm 1.2$	$37.5 \pm 1.9$	$2.8 \pm 1.6$	$6.2 \pm 1.2$	$3.4 \pm 2.0$
TGT	9	5	$36.6 \pm 1.8$	$65.3 \pm 1.4$	$28.7 \pm 2.3$	$-1.5 \pm 1.8$	$-6.9 \pm 1.4$	$-5.4 \pm 2.3$
		1	$45.3 \pm 1.3$	$74.9 \pm 1.0$	$29.6 \pm 1.6$	$7.2 \pm 1.4$	$2.7 \pm 1.0$	$-4.5 \pm 1.7$
	9	2	$42.4 \pm 1.3$	$78.7 \pm 1.1$	$36.3 \pm 1.7$	$4.3 \pm 1.4$	$6.5 \pm 1.1$	$2.2 \pm 1.8$
		4	$35.5 \pm 1.7$	$78.3 \pm 1.4$	$42.8 \pm 2.2$	$-2.6 \pm 1.7$	$6.1 \pm 1.4$	$8.7 \pm 2.2$
		7	$40.8 \pm 1.5$	$64.6 \pm 1.2$	$23.8 \pm 1.9$	$2.7 \pm 1.6$	$-7.6 \pm 1.2$	$-10.3 \pm 2.0$
8	8	$49.5 \pm 1.9$	$85.6 \pm 1.6$	$36.1 \pm 2.5$	$11.4 \pm 1.9$	$13.4 \pm 1.6$	$2.0 \pm 2.5$	

All values in  $\text{kJ mol}^{-1}$

amount of explicit water in the simulation box, the scheme for calculating the long-range non-bonded interactions, the cut-off radius for the non-bonded interactions, the algorithm to generate the pair-list and the implementation of calculating solvent-solvent

interactions, in particular. In this study, we use the reaction-field method with a cut-off radius of 1.4 nm for the non-bonded interactions, a grid-cell based pair-list algorithm [49], and a solvent-solvent loop with hard-coded parameter values.

## Summary and conclusion

The free enthalpy change of replacing a water molecule by an apolar probe at ten or nine sites in the binding pocket of a protein together with its enthalpic and entropic contributions were investigated using the examples of CDK2 and TGT. The binding pocket of each protein was partly filled by a ligand core and the remaining water molecules were perturbed one by one or in pairs to an apolar CH<sub>3</sub> moiety using the EDS method to obtain the corresponding change in free enthalpy. The water molecules were kept in a small volume by applying position restraints to allow a mapping of the resulting free enthalpy differences on to water sites in the binding pocket. Through the comparison to the same perturbation in the unbound state of a water molecule, the binding free enthalpy of the apolar probe at a specific site in the binding pocket was calculated.

The analysis of the free enthalpy differences allowed the identification of water molecules that can favourably be replaced by the apolar probe. Three such water sites, 3, 4 and 9, were found in the binding pocket of CDK2 and two, sites 4 and 9, in the binding pocket of TGT. In all five cases, the displacement was enthalpically favourable and to a lesser extent entropically unfavourable. Other sites showed also a favourable enthalpy change upon replacement by the apolar probe, for example site 6 in the case of CDK2 and site 7 in the case of TGT, but here the unfavourable entropy difference outweighed the favourable enthalpy change. It is therefore important to consider not only the enthalpic contribution but also the entropic contribution in the analysis of water molecules in the binding pocket.

The occupancy volumes of the position restrained water molecules were overlaid with available crystal structures of analogues of the simulated ligand. A good qualitative correlation between the calculated  $\Delta G_M^{bind}$  values of the water molecules displaced by the ligands and the experimentally measured binding affinities or inhibition constants was found. The displacements of water molecules by an apolar part of the ligand which were predicted to be favourable were correlated with a tighter binding of the ligand, whereas the displacement of such a water molecule by a polar atom of the ligand correlated with a less tight binding. In addition, the displacement of a water molecule by a polar atom of the ligand which was predicted to be unfavourably replaced by an apolar probe correlated with a tighter binding of the ligand.

The analysis showed that the calculation of the binding free enthalpy and its enthalpic and entropic contributions of an apolar probe in the binding pocket of a protein may provide valuable insights for drug design, by guiding the placement and atom composition of ligand substituents. A mere analysis of the enthalpic difference of the water molecule in the binding pocket and in bulk solution or the hydrogen-bonding environment of the water molecules in

the binding pocket is not sufficient in some cases and thus may have led the drug designer astray.

**Acknowledgments** This work was financially supported by the National Center of Competence in Research (NCCR) in Structural Biology and by grant number 200020-137827 of the Swiss National Science Foundation, and by grant number 228076 of the European Research Council, which is gratefully acknowledged.

## References

- IUPAP (1978) Symbols, units and nomenclature in physics. Phys A 93A:1
- IUPAC (1988) Quantities, units and symbols in physical chemistry. Blackwell Scientific Publications, Oxford
- Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys 3:300–313
- Christ CD, van Gunsteren WF (2008) Multiple free energies from a single simulation: extending enveloping distribution sampling to non-overlapping phase-space distributions. J Chem Phys 128: 174112
- Christ CD, van Gunsteren WF (2009) Simple, efficient, and reliable computation of multiple free energy differences from a single simulation: a reference Hamiltonian parameter update scheme for enveloping distribution sampling (EDS). J Chem Theory Comput 5:276–286
- Christ CD, van Gunsteren WF (2009) Comparison of three enveloping distribution sampling Hamiltonians for the estimation of multiple free energy differences from a single simulation. J Comput Chem 31:1664–1679
- Riniker S, Christ CD, Hansen N, Mark AE, Nair PC, van Gunsteren WF (2011) Comparison of enveloping distribution sampling and thermodynamic integration to calculate binding free energies of phenylethanolamine N-methyltransferase inhibitors. J Chem Phys 135:024105
- Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 72:1047–1069
- Gilson MK, Zhou H-X (2007) Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct 36:21–42
- Poornima CS, Dean PM (1995) Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. J Comput Aided Mol Des 9:500–512
- Poornima CS, Dean PM (1995) Hydration in drug design. 2. Influence of local site surface shape on water binding. J Comput Aided Mol Des 9:513–520
- Poornima CS, Dean PM (1995) Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. J Comput Aided Mol Des 9:521–531
- Minke WE, Dillers DJ, Hol WGJ, Verlinde CLMJ (1999) The role of waters in docking strategies with incremental flexibility for carbohydrate derivatives: heat-labile enterotoxin, a multi-valent test case. J Med Chem 42:1778–1788
- Mancera RL (2002) De novo ligand design with explicit water molecules: an application to bacterial neuraminidase. J Comput Aided Mol Des 16:479–499
- de Graaf C, Pospisil P, Pos W, Folkers G, Vermeulen NPE (2005) Binding mode prediction of cytochrome P450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking. J Med Chem 48:2308–2318

16. Verdonk ML, Chessari G, Cole JC, Hartshorn MJ, Murray CW, Nissink JWM, Taylor RD, Taylor R (2005) Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem* 48:6504–6515
17. Barillari C, Taylor J, Viner R, Essex JW (2007) Classification of water molecules in protein binding sites. *J Am Chem Soc* 129: 2577–2587
18. Abel R, Young T, Farid R, Berne BJ, Friesner RA (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J Am Chem Soc* 130:2817–2831
19. Michel J, Tirado-Rives J, Jorgensen WL (2009) Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J Am Chem Soc* 131:15403–15411
20. Beuming T, Farid R, Sherman W (2009) High-energy water sites determine peptide binding affinity and specificity of PDZ domains. *Protein Sci* 18:1609–1619
21. Robinson DD, Sherman W, Farid R (2010) Understanding kinase selectivity through energetic analysis of binding site waters. *Chem Med Chem* 5:618–627
22. Pearlstein RA, Hu Q-Y, Zhou J, Yowe D, Levell J, Dale B, Kaushik VK, Daniels D, Hanrahan S, Sherman W, Abel R (2010) 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* 78:2571–2586
23. Wang L, Berne BJ, Friesner RA (2011) Ligand binding to protein-binding pockets with wet and dry regions. *Proc Natl Acad Sci USA* 108:1326–1330
24. Levkau B, Koyama H, Raines EW, Clurman BE, Herren B, Orth K, Roberts JM, Ross R (1998) Cleavage of p21Cip1/Waf1 and p27Kip1 mediates apoptosis in endothelial cells through activation of Cdk2: role of a caspase cascade. *Mol Cell* 1:553–563
25. Arris CE, Boyle FT, Calvert AH, Curtin NJ, Endicott JA, Garman EF, Gibson AE, Golding BT, Grant S, Griffin RJ, Jewsbury P, Johnson LN, Lawrie AM, Newell DR, Noble MEM, Sausville EA, Schultz R, Yu W (2000) Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. *J Med Chem* 43:2798–2804
26. Bramson HN, Corona J, Davis ST, Dickerson SH, Edelstein M, Frye SV, Gampe RT Jr, Harris PA, Hassell A, Holmes WD, Hunter RN, Lackey KE, Lovejoy B, Luzzio MJ, Montana V, Rocque WJ, Rusnak D, Shewchuk L, Veal JM, Walker DH, Kuyper LF (2001) Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J Med Chem* 44:4339–4358
27. Gibson AE, Arris CE, Bentley J, Boyle FT, Curtin NJ, Davies TG, Endicott JA, Golding BT, Grant S, Griffin RJ, Jewsbury P, Johnson LN, Mesguiche V, Newell DR, Noble MEM, Tucker JA, Whitfield HJ (2002) Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with *O*<sup>6</sup>-substituted guanine derivatives. *J Med Chem* 45:3381–3393
28. Chu X-J, DePinto W, Bartkovitz D, So S-S, Vu BT, Packman K, Lukacs C, Ding Q, Jiang N, Wang K, Goelzer P, Yin X, Smith MA, Higgins BX, Chen Y, Xiang Q, Moliterni J, Kaplan G, Graves B, Lovey A, Fotouhi N (2006) Discovery of [4-amino-2-(1-methanesulfonylpiperidin-4-ylamino)pyrimidin-5-yl](2,3-difluoro-6-methoxyphenyl)methanone (R547), a potent and selective cyclin-dependent kinase inhibitor with significant *in vivo* antitumor activity. *J Med Chem* 49:6549–6560
29. McIntyre NA, McInnes C, Griffiths G, Barnett AL, Kontopidis G, Slawin AMZ, Jackson W, Thomas M, Zheleva DI, Wang S, Blake DG, Westwood NJ, Fischer PM (2010) Design, synthesis, and evaluation of 2-methyl- and 2-amino-*N*-aryl-4,5-dihydrothiazolo-[4,5-*h*]quinazolin-8-amines as ring-constrained 2-anilino-4-(thiazol-5-yl)pyrimidine cyclin-dependent kinase inhibitors. *J Med Chem* 53:2136–2145
30. Okada N, Sasakawa C, Tobe T, Yamada M, Nagai S, Talukder KA, Komatsu K, Kanegasaki S, Yoshikawa M (1991) Virulence-associated chromosomal loci of *Shigella flexneri* identified by random Tn5 insertion mutagenesis. *Mol Microbiol* 5:187–195
31. Kohler PC, Ritschel T, Schweizer WB, Klebe G, Diederich F (2009) High-affinity inhibitors of tRNA-guanine transglycosylase replacing the function of a structural water cluster. *Chem Eur J* 15:10809–10817
32. Ritschel T, Kohler PC, Neudert G, Heine A, Diederich F, Klebe G (2009) How to replace the residual solvation shell of polar active site residues to achieve nanomolar inhibition of tRNA-guanine transglycosylase. *ChemMedChem* 4:2012–2023
33. Schmid N, Christ CD, Christen M, Eichenberger AP, van Gunsteren WF (2012) Architecture, implementation and parallelization of the GROMOS software for biomolecular simulation. *Comp Phys Comm* 183:890–903
34. Riniker S, Christ CD, Hansen HS, Hünenberger PH, Oostenbrink C, Steiner D, van Gunsteren WF (2011) Calculation of relative free energies for ligand-protein binding, solvation and conformational transitions using the GROMOS software. *J Phys Chem B* 115:13570–13577
35. The GROMOS package of programs can be obtained from <http://www.gromos.net>
36. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 25:1656–1676
37. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In: Pullmann B (eds) Intermolecular forces. Reidel, Dordrecht, pp 331–342
38. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690
39. Hockney RW (1970) The potential calculation and some applications. *Methods Comput Phys* 9:136–210
40. Tironi I, Sperb R, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. *J Chem Phys* 102:5451–5459
41. Glättli A, Daura X, van Gunsteren WF (2002) Derivation of an improved SPC model for liquid water: SPC/A and SPC/L. *J Chem Phys* 116:9811–9828
42. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J Comput Phys* 23:327–341
43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
44. Fiser A, Sali A (2003) ModLoop: automated modeling of loops in protein structures. *Bioinformatics* 12:2500–2501
45. Malde AK, Zuo L, Breeze M, Stroet M, Poger D, Nair PC, Oostenbrink C, Mark AE (2011) An automated force field topology builder (ATB) and repository: Version 1.0. *J Chem Theory Comput* 7:4026–4037
46. Eichenberger AP, Allison JR, Dolenc J, Geerke DP, Horta BAC, Meier K, Oostenbrink C, Schmid N, Steiner D, Wang D, van Gunsteren WF (2011) The GROMOS++ software for the analysis of biomolecular simulation trajectories. *J Chem Theory Comput* 7:3379–3390
47. Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA (2007) Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J Chem Theory Comput* 3:26–41



48. Humphrey W, Dalke A, Schulten K (1996) VMD—visual molecular dynamics. *J Mol Graph* 14:33–38
49. Christen M, Hünenberger PH, Bakowies D, Bürgi R, Geerke DP, Heinz TN, Kastenholz MA, Kräutler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF (2005) The GRMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* 26:1719–1751
50. Cheng Y, Prusoff WH (1973) Relationship between the inhibition constant ( $K_i$ ) and the concentration of inhibitor which causes 50 % inhibition ( $IC_{50}$ ) of an enzymatic reaction. *Biochem Pharmacol* 22:3099–3108