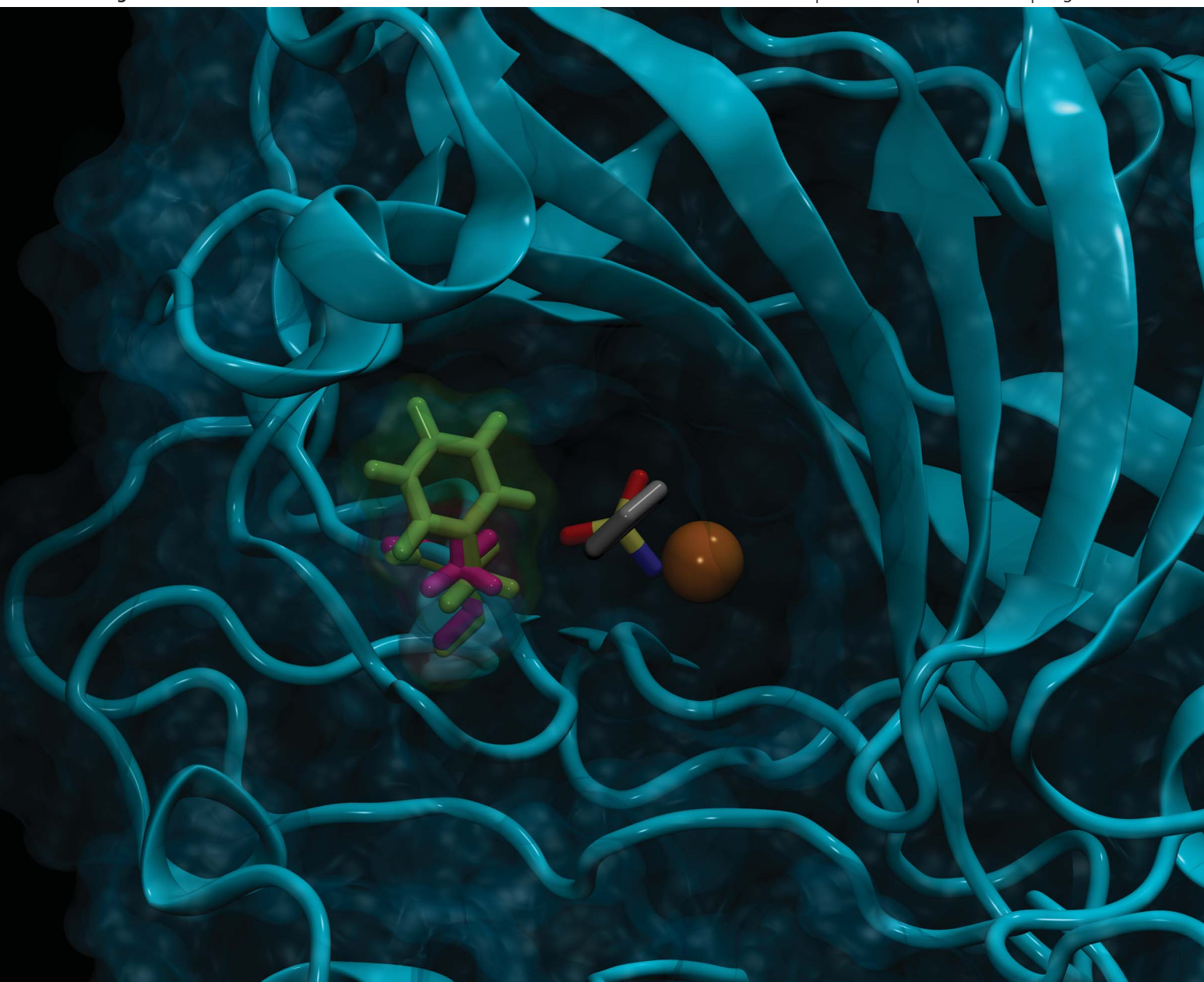


# Chemical Science

www.rsc.org/chemicalscience

Volume 3 | Number 3 | March 2012 | Pages 617–908



Winner of the ALPSP Award for Best New Journal 2011

ISSN 2041-6520

RSC Publishing

**EDGE ARTICLE**

Thomas R. Ward and Markus Meuwly *et al.*

Arylsulfonamides as inhibitors for carbonic anhydrase: prediction & validation



2041-6520 (2012) 3:3;1-U

Cite this: *Chem. Sci.*, 2012, **3**, 690

www.rsc.org/chemicalscience

# Arylsulfonamides as inhibitors for carbonic anhydrase: prediction & validation†

Maurus Schmid,<sup>ab</sup> Elisa S. Nogueira,<sup>a</sup> Fabien W. Monnard,<sup>a</sup> Thomas R. Ward<sup>\*a</sup> and Markus Meuwly<sup>\*b</sup>

Received 2nd September 2011, Accepted 14th November 2011

DOI: 10.1039/c1sc00628b

Arylsulfonamide derivatives are widely studied high affinity inhibitors of the isozyme human carbonic anhydrase II (hCA II). From molecular dynamics simulations and MM-GBSA calculations, reliable ( $R = 0.89$ ) relative binding free energies are determined for 17 previously experimentally characterized protein–ligand complexes. Decomposition of these energies led to the identification of critical amino acid residues with a significant contribution to the affinity towards the ligands. In particular, Leu198 was predicted as a key residue and was subjected to computational mutagenesis. This prediction was verified experimentally by producing hCA II mutants L198A, L198F and L198Q and determining the resulting affinities towards inhibitor **1**. The computed vs. experimental energies are in good agreement thus suggesting that the force field parameters reported herein are useful for the *in silico* design of a wider range of carbonic anhydrase inhibitors.

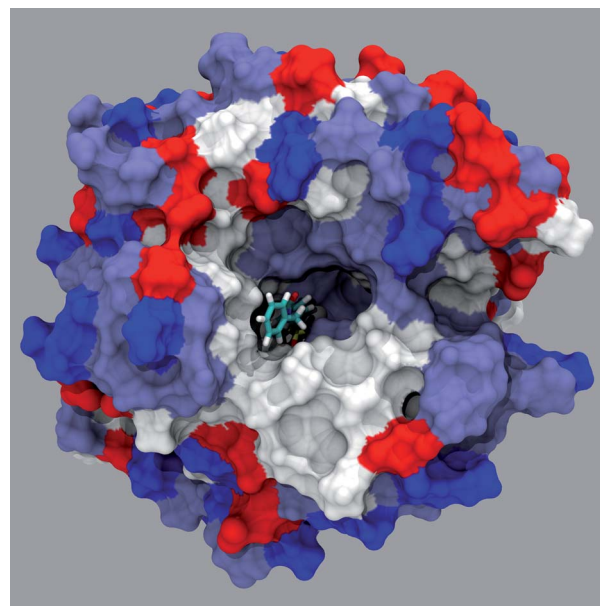
## 1 Introduction

Carbonic anhydrases (CA) are ubiquitous metalloenzymes that catalyze the reversible hydration of carbon dioxide with remarkable efficiency ( $k_{\text{cat}}/K_{\text{m}} \approx 1.5 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ ). CA isoforms are involved in various pathological processes including infections, tumorigenicity, osteoporosis, epilepsy, obesity, gluconeogenesis, lipogenesis, ureagenesis, or glaucoma. CAs have thus been the focus of many biophysical studies of protein–ligand interactions. Today, at least 25 clinically used drugs are known to display pronounced CA inhibitory properties.<sup>1</sup>

The active site of most CAs consists of a  $\text{Zn}(\text{His})_3$  moiety which is essential for catalysis. Thanks to the position of the Zn-cofactor in a cone-shaped cavity (15 Å deep and 15 Å wide at its mouth, see Fig. 1), this system lends itself ideally to the design of potent and selective inhibitors. Among these, arylsulfonamides, which bind tightly to the Zn ion at physiological pH (down to sub-nM), occupy a place of choice.<sup>2</sup>

Determining the binding free energy between proteins and ligands is a formidable and important task in improving ligands or to determine favourable interaction sites of ligands within proteins.<sup>3,4</sup> A range of computational methods are available to address this problem, ranging from free energy perturbation theory and thermodynamic integration to more approximate

procedures, including MM-GBSA. However, routine applications of these methods only has become possible lately and typically retrospective assessments are carried out instead of prospective ones. In the present work we demonstrate that it is possible to predict the effects of protein mutations on ligand binding affinities from atomistic simulation by validating the



**Fig. 1** Rendering of **8**  $\subset$  hCA II. The ligand is displayed as wireframe and the protein as solvent accessible surface (red: acidic, white: neutral, light blue: polar and dark blue: basic residues).

<sup>a</sup>Department of Chemistry, University of Basel, Spitalstrasse 51, CH-4056 Basel, Switzerland. E-mail: thomas.ward@unibas.ch

<sup>b</sup>Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland. E-mail: m.meuwly@unibas.ch

† Electronic supplementary information (ESI) available: force field parameters and their derivation, additional information on the experiments and complementary tables and figures can be found in the supporting information. See DOI: 10.1039/c1sc00628b

computed properties *vis-a-vis* experimentally determined binding constants.

Even though QM/MM models have been applied to metal-containing proteins,<sup>5,6</sup> their computational requirements remain too high for routine usage in the design of metalloenzyme inhibitors. Thus molecular mechanics and empirical force fields are widely used for this purpose. Despite the wealth of structural, kinetic, thermodynamic and quantitative structure–activity relationship data on CA inhibitors,<sup>7</sup> detailed atomistic simulations for a broad range of systems containing Zn–arylsulfonamide binding motifs are rare. Also, a variety of scoring functions have been recently used to score protein–ligand complexes or involving carbonic anhydrase II,<sup>8</sup> including a full QM scoring function.<sup>9</sup> Furthermore, the role of active-site water molecules has been investigated computationally for which dedicated force field parameters were determined (see also supporting information†).<sup>10</sup> The hydrophobic interactions in hCA II have been studied very recently<sup>11</sup> and finally, a computational study of a known sulfonamide inhibitor for two isozymes II and VII of human carbonic anhydrases has been presented.<sup>12</sup>

Because for a wide range of chemically distinct ligands, binding affinities have been determined experimentally (also for selected mutations in the protein), we decided to use hCA II as the system of choice for the present study. The focus in this work is on first validating a simulation strategy for a given set of ligands with experimentally measured ligand binding affinities. For this purpose, we selected MM-GBSA to estimate ligand-binding free energies. To validate this approach, we compared the results with published biophysical data as well as with a simulation using a QM/MM implementation with the Self-consistent charge Density-Functional Tight-Binding (SCCDFTB<sup>13</sup>) method. Finally, having identified amino acid residues critical to binding of arylsulfonamides, hCA II point mutants were computationally investigated and binding free energies were determined. These predictions were compared with experimental biophysical data on the hCA II mutants expressed recombinantly in *E. coli*.

## 2 Methods

Atomistic simulations were carried out with NAMD<sup>14</sup> and CHARMM.<sup>15</sup> NAMD was used for all simulations with the CHARMM27<sup>16</sup> force field whereas CHARMM was used for analysis and mixed quantum mechanical/molecular mechanics (QM/MM) simulations (see below).

### 2.1 Molecular dynamics simulations

A specific ligand–protein complex was set up in the following way: When a crystal structure was available, the coordinates from the crystal structure were used for the ligand. When no crystal structure for the ligand ⊂ protein system was available, the minimized energy structure of the ligand (from density functional theory) was docked into the hCA II structure (PDB code 1G54). This structure contains a fluorinated derivative (the five H-atoms of the tailing phenyl are substituted) of **8** (see Fig. 2) as the ligand. To place the new ligand, it was superimposed on the sulfonamide and aryl group of the original ligand present in 1G54 to yield a minimal root mean square deviation (RMSD).

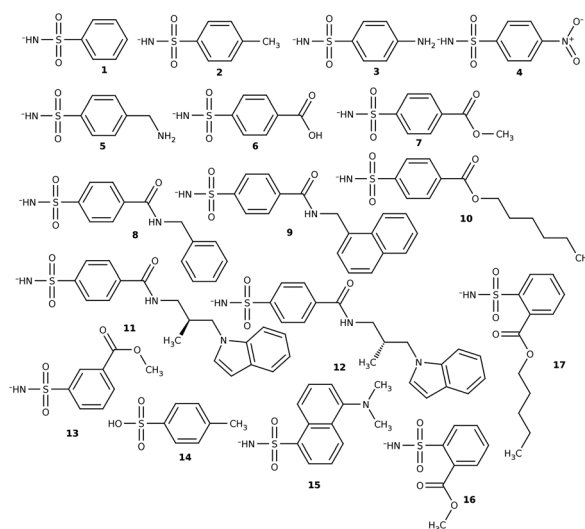


Fig. 2 hCA II inhibitors used in this study.

If the aryl substituents were unsymmetrical (*e.g.* the non *para*-substituted ligand **13**), two alternative structures were generated.

Next, the structure of the protein including the crystal waters and the ligand was minimized in CHARMM for 500 steps using Steepest Descent (SD) to eliminate unfavourable contacts. Then, the system was solvated in a cubic box with dimensions  $80 \times 80 \times 80 \text{ \AA}^3$  of previously equilibrated water molecules and periodic boundaries were applied. The resulting system consisted of approximately 50000 atoms, depending on the ligand and the number of water molecules added. If the total charge of the system was non-zero due to a charged ligand or charged residues, an adequate number of water molecules was replaced by sodium ions to render the system neutral.

After these setup steps performed in CHARMM, Molecular Dynamics (MD) simulations were performed in NAMD with a time step of  $\Delta t = 1 \text{ fs}$ . SHAKE<sup>17</sup> was used on the water molecules. For long-range electrostatic interactions the Particle Mesh Ewald (PME) method was used.<sup>18</sup> Cutoffs for van der Waals (vdW) interactions were  $12 \text{ \AA}$  and a switching scheme was used. Scaled 1–4 parameters were enabled for 1–4 interactions. First, the positions of the water molecules were minimized for 5000 steps of a Conjugate Gradient minimization while keeping the protein and ligand fixed, then the entire system was minimized for the same number of steps. Next, the water molecules were heated to 300 K in increments of 25 K for a total of 12000 steps using a Langevin *NPT* ensemble, keeping the protein and the ligand fixed. In a final step, the entire system was equilibrated for 1 ns in the *NPT* ensemble with all constraints removed and a Langevin dampening coefficient of  $5 \text{ ps}^{-1}$ . Pressure was controlled with the Nosé–Hoover Langevin piston pressure control. Production runs were typically carried out in blocks of 5 ns with a Langevin dampening coefficient of  $1 \text{ ps}^{-1}$ .

### 2.2 QM/MM simulations

To validate the force field parameters, mixed quantum mechanics/molecular mechanics (QM/MM) simulations were carried out using the Self-consistent charge Density-Functional Tight-Binding (SCCDFTB<sup>13</sup>) method implemented in

CHARMM. The QM part for these simulations included the ligand, the zinc atom and the three histidine sidechains bound to the zinc atom. As the sum of the formal charges of the QM atoms is +1, this charge was used for the QM region. Link atoms between the QM- and MM-part were located between the C $\alpha$  and the C $\beta$  of the histidine residues, see Fig. 3. The system was set up in an analogous fashion as described above. Due to the significantly increased computational cost, only one 8 ns simulation was performed and analyzed.

### 2.3 Analysis

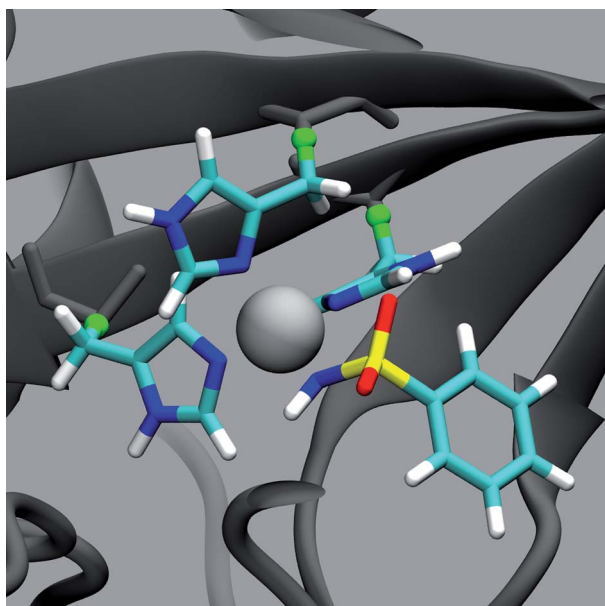
The binding free energy was calculated using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA<sup>19</sup>) approach. In MM-GBSA the ligand binding free energy  $\Delta G$  is decomposed in the following way:

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S_{\text{MM}} \quad (1)$$

Eqn (1) describes the computation of a binding free energy according to a thermodynamic cycle which includes the enthalpic ( $E_{\text{MM}}$ ) contribution for protein–ligand interactions in the gas phase and the desolvation free energies for the separated and combined protein–ligand complex, respectively.  $E_{\text{MM}}$  is the “gas phase” energy which is calculated with the CHARMM27<sup>16</sup> force field.  $E_{\text{MM}}$  contains all internal, electrostatic, and van der Waals energies and the nonbonded interactions are computed without cutoff.

The solvation part  $\Delta G_{\text{solv}}$  consists of a solvent–solvent cavity term ( $G_{\text{cav}}$ ), a solute–solvent van der Waals term ( $G_{\text{vdw}}$ ) and a solute–solvent electrostatic polarization term ( $G_{\text{pol}}$ ), *i.e.*

$$G_{\text{solv}} = G_{\text{cav}} + G_{\text{vdw}} + G_{\text{pol}} \quad (2)$$



**Fig. 3** Close-up view of the QM/MM interface of 1 c HCA II used for the SCCDFTB simulation. The atoms displayed as coloured wireframe and zinc atom (sphere) were included in the QM part. The dummy atoms (green spheres) form the link between the QM and the MM (black) part of the simulation.

The polarization term in MM-GBSA is calculated within the generalized Born approximation

$$G_{\text{pol}} = -166 \left( 1 - \frac{1}{\epsilon} \right) \sum \sum \frac{q_i q_j}{f_{\text{GB}}} \quad (3)$$

where  $\epsilon$  is the dielectric constant of water ( $\epsilon = 80$ ),  $q_i$  and  $q_j$  are the charges of atoms  $i$  and  $j$ .  $f_{\text{GB}}$  is an expression that depends on the Born atom radii  $\alpha_i$  and  $\alpha_j$  and distances  $r_{ij}$ .<sup>20</sup>

$$f_{\text{GB}} = r_{ij}^2 + \alpha_i \alpha_j \exp \left( \frac{-r_{ij}^2}{8\alpha_i \alpha_j} \right) \quad (4)$$

$G_{\text{vdw}}$  and  $G_{\text{cav}}$  are assumed to be linearly related to the solvent accessible surface area (SA):

$$G_{\text{cav}} + G_{\text{vdw}} = \sum \sigma_k SA_k \quad (5)$$

From the MD simulations snapshots were recorded every 10 ps and the above contributions to  $\Delta G$  were computed. Then, the energies of these snapshots were averaged to yield final energies and their fluctuations.

The total entropy  $S$  consists of translational, rotational and vibrational contributions

$$S_{\text{MM}} = S_{\text{trans}} + S_{\text{rot}} + S_{\text{vib}} \quad (6)$$

$S_{\text{trans}}$  and  $S_{\text{rot}}$  depend upon the mass and moments of inertia, whereas calculation of  $S_{\text{vib}}$  involves normal mode analysis for each frame, which is computationally very costly. It has been previously shown that the influence of  $S_{\text{vib}}$  on *relative* binding free energies  $\Delta\Delta G$  is in general small and does not affect the ranking of ligands.<sup>21–25</sup> Therefore, contributions to  $S_{\text{vib}}$  were only calculated every 25 frames. For this, the vibrational entropy for the entire system and for the protein was calculated in CHARMM using the standard normal mode module.

All individual energy contributions  $\epsilon_i$  are calculated according to the thermodynamic cycle<sup>22</sup> separately for the entire system, the protein and the ligand. Then the difference  $\Delta\epsilon_i$  between the protein–ligand complex and the separated system is computed for every component:

$$\Delta\epsilon_i = \epsilon_i^{\text{Prot-Lig}} - (\epsilon_i^{\text{Prot}} + \epsilon_i^{\text{Lig}}) \quad (7)$$

The total binding free energy  $\Delta G$  can be further decomposed into per-residue contributions  $\Delta G_i$  to the overall  $\Delta G = \sum_i^N \Delta G_i$ . Such a decomposition allows us to trace the changes between simulations back to individual residues. This is important in order to identify suitable mutations to enhance or decrease ligand-binding affinities.

Binding free energies from the SCC-DFTB simulations were calculated using the classical force field, as the QM energies are given only as total energies and no decomposition into electrostatic or vdW contributions is possible and the QM part consists of both, the ligand and parts of the protein.

### 2.4 Ligands examined

The ligands for which relative free energies of binding were determined consisted of sulfonamide ligands displayed in Fig. 2. Most of the ligands consisted of *para*-substituted

arylsulfonamides (**2** to **13**). In addition, two *ortho*-substituted (**16** and **17**) and one *meta*-substituted species were considered. DNSA (**15**) was included in the test set, as well as the sulfonic acid derivative **14**.

To validate the force field, the calculated binding free energies were compared to experimental data. For all ligands included in this study, experimental binding data ( $K_d$  or  $K_i$ ) is available.<sup>2</sup> For some, direct  $\Delta G$  measurements are available.<sup>2</sup> From  $K_d$ ,  $\Delta G$  can be calculated using

$$\Delta G = RT \ln(K_d). \quad (8)$$

The correlation between experimental and computed ( $\Delta G$  or  $\Delta E_{MM} + \Delta G_{solv}$ ) binding free energies was considered for both the complete set of ligands as well as the subset containing only the *meta*- and *para*-arylsulfonamides collected in Fig. 2.

### 2.5 Binding free energies to protein mutants

Binding free energies were computed for the set of ligands and WT hCA II as well as several mutants thereof. Mutants for which experimental data is available include F131V with ligand **8**, for which  $K_d$  has been measured.<sup>26</sup> In addition, the calculations suggested (see Results) that L198 contributes significantly to stabilizing the protein–ligand complex. Consequently, mutants L198A, L198F, L198Q were expressed recombinantly and the corresponding ligand binding free energies with ligand **1** were determined. The proteins were mutated *in silico* by substituting the leucine residue with either A, F, or Q, followed by a short optimization using CHARMM. The subsequent simulations and binding free energy calculations were performed as for the native system, described above.

### 2.6 Recombinant production of hCA II L198X mutants (X = A, F, Q)

Plasmid encoding hCA II and containing a T7 RNA polymerase promoter and an ampicillin resistance gene (pACA)<sup>27</sup> was a generous gift from Prof. Carol Fierke, Michigan University.<sup>28</sup> Ultra competent BL21(DE3)pLysE *E. coli* cells (produced in-house) were transformed by the plasmids containing the desired mutations. Transformed cells were plated on LB-Lennox agar plates containing ampicillin (50  $\mu\text{g mL}^{-1}$ ), chloramphenicol (34  $\mu\text{g mL}^{-1}$ ) and glucose (2% w/v), and incubated overnight at 37 °C.

One colony was chosen to inoculate 25 mL pre-culture of Luria-Bertoni (LB) medium (10 g  $\text{L}^{-1}$  tryptone, 5 g  $\text{L}^{-1}$  yeast extract, 10 g  $\text{L}^{-1}$  NaCl, 100  $\mu\text{g mL}^{-1}$  ampicillin and 34  $\mu\text{g mL}^{-1}$  chloramphenicol). Pre-culture was grown overnight at 37 °C and 250 rpm. 5 mL of the pre-culture was used to inoculate 1000 mL of induction media (20 g  $\text{L}^{-1}$  tryptone, 10 g  $\text{L}^{-1}$  yeast extract, 5 g  $\text{L}^{-1}$  NaCl, 0.36X M9 salts solution, 0.4% glucose, 60  $\mu\text{M}$  ZnSO<sub>4</sub>, 100  $\mu\text{g mL}^{-1}$  ampicillin and 34  $\mu\text{g mL}^{-1}$  chloramphenicol). Cells were grown at 37 °C, for 3 to 4 h or until  $A_{600} = 0.6$ – $0.8$ . Addition of isopropyl- $\beta$ -D-thiogalacto-pyranoside (IPTG, 250  $\mu\text{M}$  final concentration) and ZnSO<sub>4</sub> (450  $\mu\text{M}$  final concentration) induced protein expression and the temperature was lowered to 18 °C (to prevent formation of inclusion bodies). After overnight incubation at 18 °C the cells were harvested (5346  $\times g$ , for 15 min at 4 °C) and frozen at  $-20$  °C overnight. Cells were lysed by activating the gene encoding T7 lysozyme

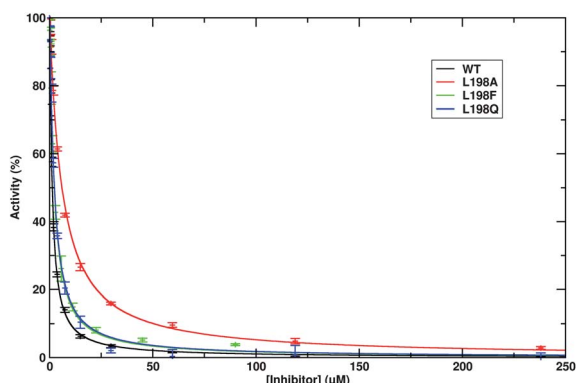
using three cycles of “freezing/thawing”. Cells were resuspended in a buffer containing 50 mM Tris-sulfate, pH 8.0, 50 mM NaCl, 10 mM EDTA, pH 8.0, 0.5 mM ZnSO<sub>4</sub>, and the protease inhibitor phenylmethanesulfonyl fluoride (PMSF, 10  $\mu\text{g mL}^{-1}$ ). Cell resuspension was incubated under vigorous shaking ( $\sim 250$  rpm) at room temperature for 1 h, DNase I (1  $\mu\text{g L}^{-1}$  final concentration) was added, and cells were left for another hour under the same conditions as previously described. The cellular remnants were centrifuged (12150  $\times g$ , for 45 min at 4 °C) and the cell debris was discarded. The supernatant, containing crude hCA II, was purified by affinity chromatography (4-amino-methylbenzene sulfonamide agarose). The column was first equilibrated with 5-column volumes (CVs) of activity buffer (50 mM Tris-sulfate, pH 8.0 and 0.5 mM ZnSO<sub>4</sub>). The protein was then loaded onto the column. The affinity gel was washed with 5-CVs of 50 mM Na<sub>2</sub>SO<sub>4</sub>/50 mM NaClO<sub>4</sub>/25 mM Tris, pH 8.8. The bound protein was eluted with 10-CVs of 200 mM NaClO<sub>4</sub>/100 mM NaAc, pH 5.6. Collected fractions were pooled and dialyzed at 4 °C against activity buffer for 24 h, deionized water for another 24 h, and finally against double-deionized (ddH<sub>2</sub>O) overnight. Protein was lyophilized and kept at 4 °C as a powder for further experiments. This procedure yielded 100–200 mg of >95% pure hCA II, as confirmed by SDS-PAGE. The molecular weight of the three mutants was confirmed by ESI-TOF mass spectroscopy (Bruker micrOTOF II, USA). 0.5 to 1.0 mg of lyophilized protein was dissolved in ddH<sub>2</sub>O to a final concentration of 1 mg  $\text{mL}^{-1}$  and further diluted in MS buffer (50% MeOH, 0.05% formic acid, pH 3.0–4.0, and 50% ACN, 0.1% acetic acid and 0.11% TFA, pH 0–1.0). Molecular weight was calculated using Bruker Daltonics DataAnalysis program (Bruker Daltonics, USA). Theoretical and experimental data were compared, and differences were considered not significant (in the range of 0.004 to 0.005%).

### 2.7 Experimental affinity determination for **1** $\subset$ hCA II L198X (X = A, F, Q)

All steady-state measurements<sup>28,29</sup> were performed in Tris-sulfate buffered solution (25 mM, pH 8.0) in presence of 5% DMSO at 25 °C.<sup>30</sup> The organic solvent ensures the solubility of the substrate (*p*-nitrophenyl acetate) as well as of the ligand used as inhibitor. The initial rates of the enzyme-catalyzed activity were measured by following the hydrolysis of the chromogenic substrate, *p*-nitrophenyl acetate, at 348 nm (25 measurements over a period of 35 min). Experiments were carried out in triplicate for each inhibitor. Kinetic measurements were performed in a total reaction volume of 300  $\mu\text{L}$  (in Tris-sulfate buffered solution), containing 0.5 mM *p*-nitrophenyl acetate and different concentrations of inhibitors.

The initial rates of enzyme catalysis were determined using the linear maximum slopes (first 10 min, 5 points) of the reaction traces measured by the plate reader. For comparison of the inhibition data, the initial rates were translated into % activity as a function of the inhibitor concentration (see Fig. 4). The inhibition data were analyzed *via* eqn (9)<sup>29</sup> using gnuplot (Version 4.2, least-squares method)

$$v = \frac{v_o K_i}{K_i + \{[I]_t - 0.5(A - \sqrt{A^2 - 4[I]_t[E]_t})\}} \quad (9)$$



**Fig. 4** Steady-state kinetic data for the inhibition of hCA II variants. The solid smooth lines represent the best fits of the data according to eqn (9).

$$A = [I]_t + [E]_t + K_i$$

with  $v_0$  being the initial velocity of the enzyme-catalyzed reaction in the absence of inhibitor,  $K_i$  the inhibition constant,  $[E]_t$  the total concentration of the enzyme and  $[I]_t$  the total concentration of the inhibitor.

### 3 Results and discussion

#### 3.1 Validation of the computational approach

The benzene sulfonamide ligand **1** was selected for an in-depth validation of the computational strategy, in particular of the force field (see supporting information). Various geometrical and energetic properties of the ligand-protein system are considered and analyzed in detail below.

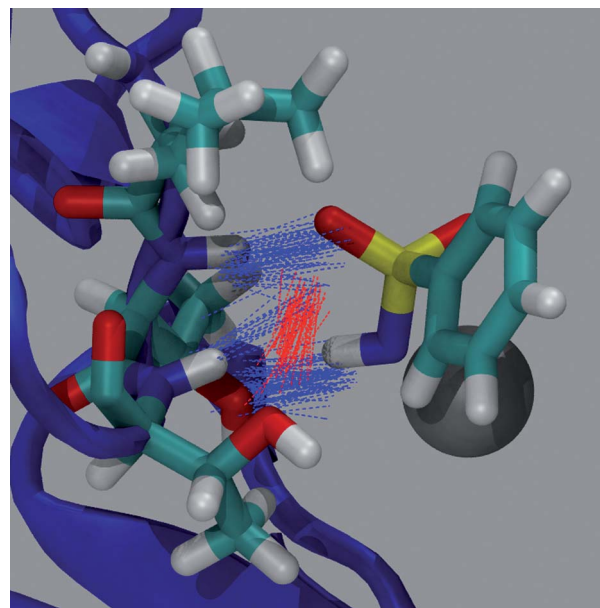
**Structural aspects.** To assess the structural integrity of the protein–ligand complex, the root mean square deviation (RMSD) along the trajectory relative to the starting structure was considered. For the RMSD analysis, the translational and rotational degrees of freedom of the entire protein were removed by aligning the snapshots to the initial protein structure. This starting structure was also taken as reference for the subsequent RMSD calculations. The RMSD of the coordinates was stable for the entire simulation time which was 32.5 ns. The RMSD of all protein atoms fluctuates around 1.5 Å, while for the backbone and the sidechains of the entire protein, it is around 1 Å and 2 Å, respectively. The RMSD for the ligand itself fluctuates considerably more than the RMSD of the protein atoms. Whereas it ranges from 0.5 Å to 2 Å for the ligand, the protein structure does not fluctuate by more than 0.5 Å.

The backbone atoms of the residues forming the binding pocket are located roughly 7.5 Å from the ligand. As the backbone atoms are stabilized by the secondary structure (see supporting information Figure S3†), they fluctuate less than the sidechain atoms and the RMSD for protein atoms within 7.5 Å of the ligand is lower than the RMSD for protein atoms 5 or 10 Å away. The RMSD was stable for all ligands and simulations with RMSDs in the same range as above.

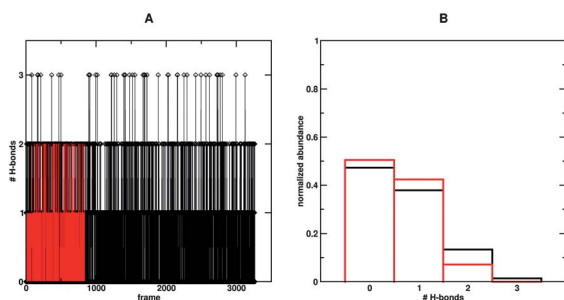
**H-Bonds.** Hydrogen bonds can contribute considerably to protein–ligand interactions.<sup>24</sup> Therefore, it is of interest to

examine how many H-bonds can and actually are formed between the protein and the ligand and how they are maintained throughout an MD simulation. In the simulation, a total of four distinct hydrogen bonds are observed during the simulation for ligand **1**: The most stable H-bond (as judged from the probability distribution) is the one between the sulfonamide hydrogen  $H_N$  and oxygen OG1 of residue Thr199. An additional H-bond can form between the sulfonamide oxygen  $O_S$  and hydrogen  $H_N$  of Thr199. Two further H-bonds can form between  $H_N$  or  $HG1$  of residue Thr200 and the same oxygen  $O_S$  as before (see Fig. 5). The maximum number of H-bonds observed at once between the ligand and the protein is 3. However, this only occurred for short periods of time and the majority of snapshots displays between 0 and 2 H-bonds (see Fig. 6). Occasionally, a water molecule was located between the sulfonamide nitrogen  $N_S$  and  $O_S$  of residue Thr200. But this occurred only rarely and for short periods of time ( $\leq 50$  ps).

**Free enthalpy as a function of time.** Binding free energies from MM-GBSA can be calculated either from sufficiently long single trajectories or from several short, independent trajectories.<sup>19,23,31–33</sup> It was previously reported that a minimal simulation time is required to obtain stable, albeit not necessarily converged, binding free energies.<sup>24</sup> In the present work, it was found that during the first nanosecond  $\langle \Delta E_{MM} + \Delta G_{solv} \rangle_{1ns} = -17.6$  kcal mol<sup>-1</sup>. After 5 ns, this decreased to  $\langle \Delta E_{MM} + \Delta G_{solv} \rangle_{5ns} = -13.8$  kcal mol<sup>-1</sup> and for the entire trajectory  $\langle \Delta E_{MM} + \Delta G_{solv} \rangle_{32.5ns} = -12.8$  kcal mol<sup>-1</sup>. Instantaneous values for  $\Delta E_{MM} + \Delta G_{solv}$  varied between  $-40$  and  $+10$  kcal mol<sup>-1</sup> with a standard deviation of  $\sigma_{\Delta E_{MM} + \Delta G_{solv}} = 7.6$  kcal mol<sup>-1</sup> (see supporting information†). The largest contributions to  $\Delta E_{MM} + \Delta G_{solv}$  arise from  $G_{pol}$  and  $E_{MM,elstat}$  which approximately compensate each other as  $\langle G_{pol} \rangle$  is positive and  $\langle E_{MM,elstat} \rangle$  is negative.

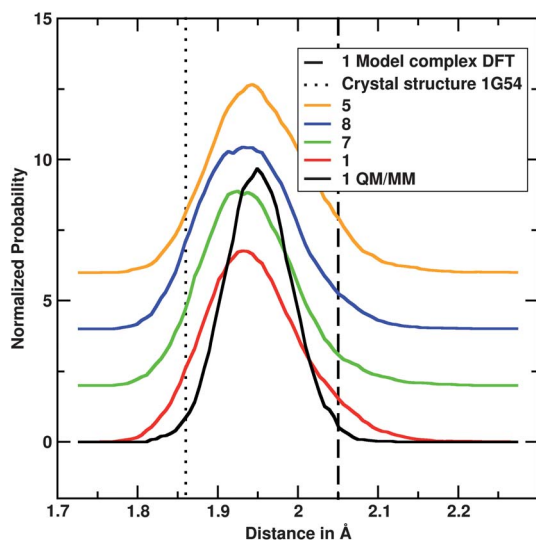


**Fig. 5** H-bonding pattern between ligand **1** and hCA II residues Thr199 and Thr200 for 330 snapshots. Red and blue highlight the H-bonds to oxygens and nitrogens respectively.



**Fig. 6** Evolution of the H-bonding pattern between ligand **1** and protein (A) and corresponding histogram (B). Values for the MM-Simulation (black), values for the QM/MM simulation (red).

**Comparison between MM and QM/MM.** The MM simulations (including the force field parametrizations) can be validated to some extent by comparing them with trajectories based on the more elaborate QM/MM approach. The simulations are compared by, *e.g.*, examining geometrical parameters including bond lengths, or by analysing the calculated binding free energy. The distance  $r_{\text{Zn},\text{N}_5}$  between the zinc and the sulfonamide-nitrogen  $\text{N}_5$  is compared between the QM/MM and the classical MM simulation. A normalized histogram for the probability  $p(r_{\text{Zn},\text{N}_5})$  of the bond lengths for all the snapshots displays a slightly wider distribution for the force field simulations compared to QM/MM. The maximum  $p_{\text{max}}(r_{\text{Zn},\text{N}_5})$  is shifted by  $+0.025 \text{ \AA}$  in the QM/MM simulations. In the X-ray structure, the distance is  $r_{\text{Zn},\text{N}_5} = 1.86 \text{ \AA}$ , whereas the DFT optimized value for the model complex is  $2.05 \text{ \AA}$  (see Fig. 7). The RMSD of all backbone atoms in the QM/MM simulation is around  $1 \text{ \AA}$  which is comparable to the MM simulation (see supporting information Figure S1†). The RMSD for the ligand fluctuates more in the QM/MM compared to the MM simulation. As above, all RMSDs are reported relative to the starting structure, which is the same for the protein and ligand atoms in MM and QM/MM. The number of hydrogen bonds between the ligand and the protein is, on average, slightly smaller in the QM/MM compared



**Fig. 7** Graphical summary of computed Zn- $\text{N}_5$  sulfonamide distances for inhibitors **1**, **5**, **7**, **8**.

to the MM simulations: there are at most 2 hydrogen bonds at once, never 3 as in the MM simulation (see Fig. 6).

$\langle \Delta E_{\text{MM}} + \Delta G_{\text{solv}} \rangle$  calculated for the trajectory generated with QM/MM was slightly lower (QM =  $-9.3$  vs. MM =  $-12.8 \text{ kcal mol}^{-1}$ ) than that from the empirical force field. However, the difference is still well within the expected error bars of  $\sigma \approx 7 \text{ kcal mol}^{-1}$ .

### 3.2 Simulations for all ligands

After validating the force field and establishing that the computational strategy is robust and meaningful, a broader assessment of the binding free energies for all compounds displayed in Fig. 2 was performed. The primary aim is to follow a simulation strategy which is sufficiently robust to also *predict* the effect of modifications either on the ligand or on the protein on the resulting protein–ligand interaction. Before addressing this last point, a thorough investigation of the 17 compounds is presented. Where necessary, specific ligands are discussed in more detail.

The distribution of the distance between the zinc and nitrogen  $\text{N}_5$  the ligand is very similar for all ligands except for the sulfonic acid derivative **14**. In this case, the conformationally averaged O–Zn distance is  $1.76 \text{ \AA}$  compared to the  $1.95 \text{ \AA}$  for  $\text{N}_5$ . For most ligands reported in Fig. 2, a total of 20 to 35 ns were simulated. The results are collected in Table 1. As observed for the model inhibitor **1**, between 10 and 15 ns of simulations are necessary to obtain stable binding free energies. Values for  $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  range from  $-6.4 \text{ kcal mol}^{-1}$  for ligand **14** to  $-23.9 \text{ kcal mol}^{-1}$  for ligand **15**. The standard deviations  $\sigma_{\Delta E_{\text{MM}} + \Delta G_{\text{solv}}}$  varied from 6.1 to  $8.3 \text{ kcal mol}^{-1}$ . The calculated  $\Delta G$  are between  $10.8 \text{ kcal mol}^{-1}$  for **14** and  $-6.4 \text{ kcal mol}^{-1}$  for **15**.

For ligand **10**, several calculations with simulation times of 5 ns were performed to investigate the sensitivity to the initial conditions. The calculated binding free enthalpy ranges from  $-14.6$  to  $-17.9 \text{ kcal mol}^{-1}$  with an average of  $-16.1 \text{ kcal mol}^{-1}$ . The ensemble average for a 25 ns simulation is bracketed by these values ( $-16.5 \text{ kcal mol}^{-1}$ ). Thus the starting conditions do not greatly influence the simulation results.

For arylsulfonamide ligands with *ortho*-substituents (ligands **16** and **17**), two independent starting structures were generated because no X-ray structure is available. The *ortho*-substituents lead to unstable simulations with significant rearrangement of the ligand position or the surrounding residues. For DNSA (**15**), a commonly used competing ligand for hCA II in fluorescence assays,<sup>34</sup> the crystal structure shows an unusual binding mode with the large aromatic moiety rotated in the hydrophobic subpocket.<sup>35</sup> Two conformations were therefore considered in hCA II: one similar to the X-ray structure (conformation A) and one similar to the other arylsulfonamides, *i.e.* rotated by 180 degrees (conformation B). It was found that conformation A is stabilized relative to conformation B:  $-23.9 \text{ kcal mol}^{-1}$  and  $-15.2 \text{ kcal mol}^{-1}$  for A and B respectively. The hydrophobic interactions within the subpocket appear to be over-estimated compared to other inhibitors where this subpocket is not filled as tightly.

As some of the experimentally measured  $K_{\text{d}}$ -values were determined by different groups, a range of values is available for them. In such cases the average value was used to assess the correlation with the computed data. For ligand **1**, the published

**Table 1** Summary of binding data for ligands **1–18** with hCA II

Ligand	Calculated						Experimental <sup>a</sup>						
	$\Delta E_{\text{MM}} + \Delta G_{\text{solv}} / \text{kcal mol}^{-1}$	$\sigma$	$\Delta G / \text{kcal mol}^{-1}$	$-T\Delta S / \text{kcal mol}^{-1}$	$t_{\text{sim}} / \text{ps}$	$K_{\text{d}} / \text{nM}$	$\Delta G_{\text{Kd}} / \text{kcal mol}^{-1}$	$\Delta G_{\text{obs}} / \text{kcal mol}^{-1}$	$\Delta H_{\text{obs}} / \text{kcal mol}^{-1}$	$-T\Delta S_{\text{obs}} / \text{kcal mol}^{-1}$	$\Delta G_{\text{ArSO}_2\text{NH}^-} / \text{kcal mol}^{-1}$	$\Delta H_{\text{ArSO}_2\text{NH}^-} / \text{kcal mol}^{-1}$	$-T\Delta S_{\text{ArSO}_2\text{NH}^-} / \text{kcal mol}^{-1}$
<b>1</b>	-12.84	±7.64	2.33	15.17	32500	850 <sup>b</sup>	-8.33	-9.1	-10.9	1.8	-13.5	-13.1	-0.4
<b>1.scc</b> <sup>c</sup>	-9.33	±8.09	6.93	16.26	8000	850 <sup>b</sup>	-8.33						
<b>2</b> <sup>f</sup>	-13.30	±7.98	3.57	13.96	30000	82	-9.72	-9.7	-10.8	1.1	-14.2	-12.1	-2.1
<b>3</b>	-10.01	±6.21	7.35	17.36	35000	13000 <sup>b</sup>	-6.70	-6.6	-7.7	1.1	-11.1	-9.3	-1.8
<b>4</b>	-11.72	±6.11	4.94	16.54	25000	63	-9.88	-10.3	-9.5	-0.8	-13.9	-14.2	0.3
<b>5</b> <sup>f</sup>	-9.04	±6.16	7.10	16.14	25000	36000	-6.10	-6.1	-2.4	-3.7	-8.4	-6.5	-1.9
<b>6</b> <sup>f</sup>	-11.05	±6.98	4.85	15.90	25000	270	-9.01	-8.6	-9.6	1.0	-12.6	-10.5	-2.1
<b>7</b>	-14.14	±7.53	2.96	17.09	25000	10	-10.97						
<b>8</b>	-18.27	±6.64	2.33	20.60	25000	1.1	-12.29						
<b>9</b>	-20.71	±6.98	0.09	20.80	25000	0.6	-12.65						
<b>10</b>	-16.49	±6.63	3.22	19.71	25000	0.41	-12.88						
<b>10.1</b> <sup>d</sup>	-17.94	±7.04	0.64	18.58	5000	0.41	-12.88						
<b>10.2</b> <sup>d</sup>	-14.56	±6.51	4.03	18.59	5000	0.41	-12.88						
<b>10.3</b> <sup>d</sup>	-16.85	±5.58	3.92	20.78	5000	0.41	-12.88						
<b>10.4</b> <sup>d</sup>	-15.96	±7.54	2.64	18.60	5000	0.41	-12.88						
<b>10.5</b> <sup>d</sup>	-15.94	±9.28	4.11	20.04	5000	0.41	-12.88						
<b>11</b>	-23.79	±7.04	-2.21	21.60	25000	0.23	-13.22						
<b>12</b>	-18.04	±6.80	4.57	22.61	25000	0.03	-14.44						
<b>13.a</b>	-14.56	±6.78	2.73	17.30	25000	700	-8.44						
<b>13.b</b> <sup>e,h</sup>	-14.04	±6.18	1.88	15.92	5000	700	-8.44						
<b>14</b>	-6.36	±6.36	10.82	17.18	30000	460000	-4.58						
<b>15.a</b> <sup>f</sup>	-23.88	±7.26	-6.38	17.50	25000	250	-9.06	-8.8	-5.7	-3.1	-12.9	-8.2	-4.7
<b>15.b</b> <sup>e,f</sup>	-15.19	±7.75	2.96	18.15	25000	250	-9.06	-8.8	-5.7	-3.1	-12.9	-8.2	-4.7
<b>16.a</b>	-14.04	±8.32	1.72	15.77	20000	39000	-6.05						
<b>16.b</b> <sup>e</sup>	-6.20	±7.16	11.80	18.00	20000	39000	-6.05						
<b>17.a</b> <sup>g</sup>	-16.47	±7.21	4.38	20.85	20000	660	-8.48						
<b>17.b</b> <sup>e,g</sup>	-27.07	±7.62	-6.47	20.60	5000	660	-8.48						

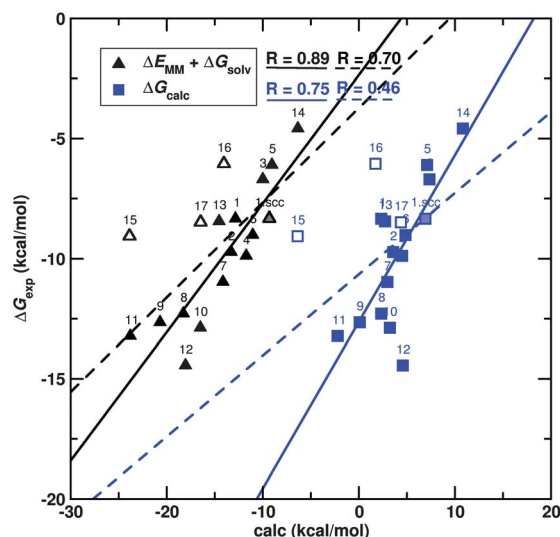
<sup>a</sup> Values from ref. 2. <sup>b</sup> Averaged. <sup>c</sup> SCCDFTB calculation. <sup>d</sup> Different random seed. <sup>e</sup> Alternative starting structure. <sup>f</sup> Experimental decomposed energies from bovine carbonic anhydrase II. <sup>g</sup> Unstable simulation. <sup>h</sup> Becomes the same structure as the other simulation.

$K_{\text{d}}$ s range from 200 to 1500 nM.<sup>2</sup> The average value  $K_{\text{d}} = 850$  nm was used. This translates into a difference in  $\Delta G$  of  $\approx 1$  kcal mol<sup>-1</sup>. In the following, this difference is assumed uniformly as the error in the experiments.

Correlations between ( $\Delta G_{\text{calc}}$ ,  $\Delta G_{\text{exp},K_{\text{d}}}$ ) and ( $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$ ,  $\Delta G_{\text{exp},K_{\text{d}}}$ ) are considered in more detail in the following. This is motivated by the fact that entropic corrections were computed somewhat less rigorously and because previous MM-GBSA studies showed that  $-T\Delta S$  had no profound influence on the correlation and did not change the ranking of ligands.<sup>24</sup> Furthermore, experimental data suggests that  $-T\Delta S$  is similar for all ligands and therefore leads to a constant offset.<sup>2</sup>

From the correlation between computed and experimentally measured binding free energies (Fig. 8) modest  $R$ -values are computed ( $R = 0.70$  for  $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  vs.  $\Delta G_{\text{exp},K_{\text{d}}}$ ;  $R = 0.46$  for  $\Delta G_{\text{calc}}$  vs.  $\Delta G_{\text{exp},K_{\text{d}}}$ ). Inspection of the arylsulfonamide structures used in this study reveals three outliers: the *ortho*-substituted ligands **16** and **17** impose steric constraints in the proximity of the Zn-atom which are absent in the *para*- and *meta*-arylsulfonamides **1–14**. Another outlier is the naphthyl derivative **15**. It is interesting to note that the binding free energy and enthalpy calculated for the structure docked in a similar way as the other ligands (conformation B) is much closer to the correlation than the conformation similar to the X-ray structure<sup>35</sup> (conformation A). Without these outliers, correlation increases to  $R = 0.89$  ( $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  vs.  $\Delta G_{\text{exp},K_{\text{d}}}$ ) and  $R = 0.75$  ( $\Delta G_{\text{calc}}$  vs.  $\Delta G_{\text{exp},K_{\text{d}}}$ ), respectively.

Additional correlations between  $\Delta G_{\text{calc}}$  and  $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  with  $\Delta G_{\text{obs}}$ ,  $\Delta H_{\text{obs}}$ ,  $\Delta G_{\text{ArSO}_2\text{NH}^-}$  and  $\Delta H_{\text{ArSO}_2\text{NH}^-}$  were also considered.  $\Delta H_{\text{obs}}$  are the experimentally determined enthalpies



**Fig. 8** Correlation between calculated  $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  (black triangles) and  $\Delta G$  (blue squares) and experimental  $\Delta G$  for all ligands (See Fig. 2 for ligand structures). Empty symbols display the outliers and dashed lines the correlation for the outliers. The half-filled symbols (**1.scc**) display the QM/MM results.

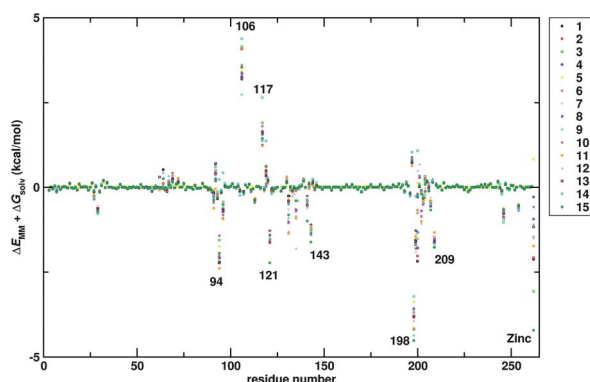


derived either from temperature dependent  $\Delta G$ -measurements using the van't Hoff equation  $\left(\ln K_d = \frac{\Delta H^\circ}{RT} - \frac{\Delta S^\circ}{R}\right)$  or using isothermal titration calorimetry.  $\Delta G_{\text{obs}}$  are their associated total binding free energies.  $\Delta G_{\text{ArSO}_2\text{NH}^-}$  and  $\Delta H_{\text{ArSO}_2\text{NH}^-}$  are energies corrected to take into account the deprotonation of the sulfonamide group upon binding to the protein.<sup>2</sup> As all theoretical values are calculated for the deprotonated species,  $\Delta G_{\text{ArSO}_2\text{NH}^-}$  and  $\Delta H_{\text{ArSO}_2\text{NH}^-}$  should correlate better with  $\Delta G_{\text{calc}}$  than  $\Delta G_{\text{obs}}$  and  $\Delta H_{\text{obs}}$ , respectively. It is found that for the experimental enthalpies the  $R$ -values remain essentially unchanged, *i.e.*  $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  vs.  $\Delta H_{\text{obs}}$  and  $\Delta H_{\text{ArSO}_2\text{NH}^-}$  (both  $R = 0.84$ ) and  $\Delta G_{\text{calc}}$  vs.  $\Delta H_{\text{obs}}$  and  $\Delta H_{\text{ArSO}_2\text{NH}^-}$  (both  $R = 0.83$ ). However, when considering binding free energies,  $R$ -values increase from  $R = 0.78$  ( $\Delta G_{\text{calc}}$  vs.  $\Delta G_{\text{obs}}$ ) to  $R = 0.89$  ( $\Delta G_{\text{ArSO}_2\text{NH}^-}$  vs.  $\Delta G_{\text{obs}}$ ) and from  $R = 0.83$  ( $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  vs.  $\Delta G_{\text{obs}}$ ) to  $R = 0.89$  ( $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  vs.  $\Delta G_{\text{ArSO}_2\text{NH}^-}$ ).

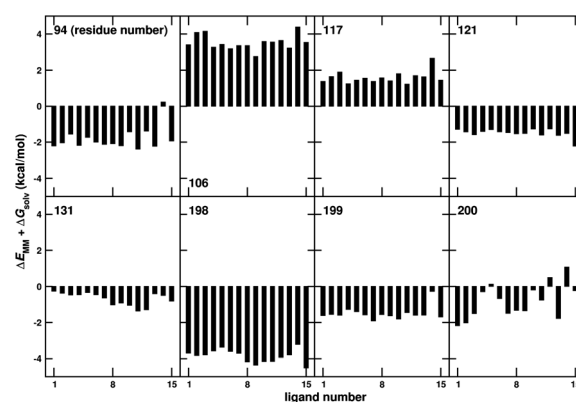
### 3.3 Per-residue decomposition of $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$

It is valuable to trace back the origin of the binding free energy to individual residues. Such information can be used to rationally design mutants for which favourable binding can be perturbed by site directed mutagenesis. Alternatively, the information also provides the basis for tailoring the ligand to the protein to identify which residues play an important role in binding. For the following, refer to Fig. 9 and 10. For all ligands, the dominant stabilizing contribution to the binding free enthalpy can be traced back to residue Leu198. This hydrophobic interaction amounts to almost 4 kcal mol<sup>-1</sup>. The hydrogen bonds between Thr199 and Thr200 and the sulfonamide moiety add another 2 kcal mol<sup>-1</sup> each. Residue His94, which is one of the histidines binding the zinc to the protein, interacts mostly in a nonpolar fashion and provides another 2 kcal mol<sup>-1</sup>.

For Thr200 there is quite a large difference depending on which ligand is considered. For the majority of ligands, the interaction with Thr200 due to the H-bond is stabilizing but for some of them (ligands **4**, **5**, **10**, **15**) it is either  $\approx 0$  or even destabilizing (ligands **12** and **14**). For Phe131 which is located in the mouth of the binding site, the length of the ligand correlates with the magnitude of the influence. Residues Ser29, Arg246 and Arg254 which are not located in the binding site, add 1 kcal mol<sup>-1</sup> of stabilization each. Residues Glu106 and Glu117 lead to an



**Fig. 9** Per-residue free energies ( $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$ ) for ligand C-hCA II combinations.



**Fig. 10** Free energies ( $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$ ) of the ligands (for selected residues see Fig. 2 for ligand structures).

overall destabilization of  $\approx 4$  kcal mol<sup>-1</sup> and 2 kcal mol<sup>-1</sup> respectively.

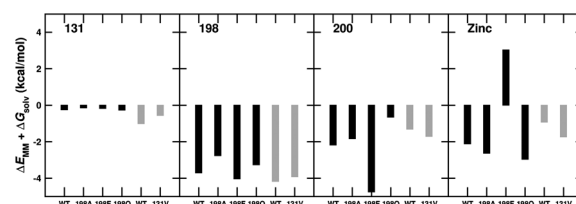
The influence of the zinc atom changes strongly from one ligand to the other. This effect is related to the distance between the zinc and the sulfonamide nitrogen (see Fig. 7), which, as a result of the other interactions, changes slightly for each ligand. For **14**, with an oxygen binding to the zinc instead of a nitrogen, Glu117 and His119 are much more destabilizing than in the sulfonamides. His94 and Thr199 even lost their stabilizing effect. Especially for **15.a** Val121 plays an important role in stabilization.

In summary, several mutations may be envisaged, especially the residues identified in Fig. 11. For most of the residues identified here, mutation studies have been carried out to investigate either the catalytic function or the binding of the metal cofactor to the protein.<sup>36–38</sup> For F131V, binding constants for arylsulfonamide **8** have been determined.<sup>26</sup>

As the residue which contributes most to the stabilizing interactions is Leu198, we anticipated that mutation at this site may have a significant effect on the resulting affinities. Another candidate would be Glu106, but it is not located directly in the binding site. For Thr200, the difference between ligands are most apparent. Mutations at this position might influence each ligand in a different way. Residue Phe131 is interesting only for ligands with a long tail as it is remote from the binding site.

### 3.4 Computational and experimental mutation study

**Simulation.** As experimental data for F131V with ligand **8** is available,<sup>26</sup> it was one of the mutations selected for simulation.



**Fig. 11** Contribution to the free energy ( $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$ ) for selected residues as a function of ligand (**1** black bars, **8** grey bars) and mutant. (Note the favourable interaction between **1** and residue Thr200 in mutant L198F which is compensated by a strong destabilizing interaction between **1** and Zn.)

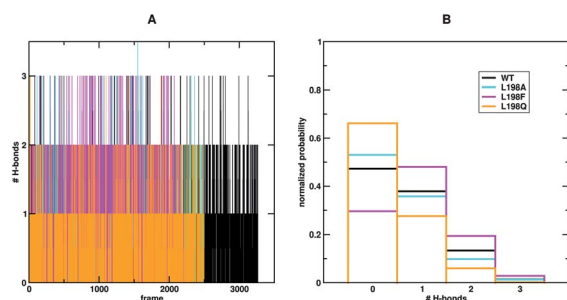
Other mutations concerned Leu198 as this residue has the largest influence on the binding free energy (Fig. 9). The molecular dynamics simulations were performed in a similar way to the ones for the wild-type protein. Again, the RMSDs were stable for all simulations. The number of H-bonds between the protein and the ligand is quite different for L198F and L198Q compared to the WT (see Fig. 12). Whereas there are more H-bonds for L198F (average 0.95) than in the WT (average 0.69), for L198Q (average 0.4) less hydrogen bonding is observed. Fluctuations in  $\langle \Delta E_{\text{MM}} + \Delta G_{\text{solv}} \rangle$  for the simulations involving the mutants are similar to those for the WT.

Comparing the binding free energy between ligand **1** and WT or the L198F mutant, respectively, a differential stabilization of  $\Delta \Delta E_{\text{MM}}^{\text{WT} \rightarrow \text{L198F}} + \Delta \Delta G_{\text{solv}}^{\text{WT} \rightarrow \text{L198F}} = -0.5 \text{ kcal mol}^{-1}$  is found (see Table 2). Considering individual per residue contributions (see Fig. 11), the change in  $\Delta \Delta E_{\text{MM}} + \Delta \Delta G_{\text{solv}}$  at position 198 is essentially 0 whereas Thr200 differentially stabilizes the ligand by more than 2 kcal mol<sup>-1</sup>. On the other hand, the zinc atom differentially destabilizes the ligand by more than 3 kcal mol<sup>-1</sup>. No differential changes are found, e.g., at position F131. Therefore, by mutating L198 the contribution of other residues surrounding the active site can be affected in pronounced ways. As will be seen below, this does, however, not affect the correlation between computed and measured ligand binding affinities and suggests, that MM-GBSA simulations are able to capture such effects. When comparing WT with L198Q a differential destabilization of 2 kcal mol<sup>-1</sup> is found. Again,  $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  at position 198 is essentially unchanged, but for Thr200 and the zinc an opposite effect than above can be observed: Thr200

becomes less stabilizing by 2 kcal mol<sup>-1</sup>, the zinc stabilizes by 1 kcal mol<sup>-1</sup>. For L198A, which replaces all favourable or unfavourable interactions with the amino acid side chain due to mutation to Ala, the changes for residue 198 are larger than for residue 200. Removing the sidechain from residue 198 leads to subtle changes in the entire binding region and the overall destabilization by  $\approx 2 \text{ kcal mol}^{-1}$  is the sum of several small contributions. With a valine instead of the phenylalanine at position 131, the stabilization of **8**  $\subset$  protein for this residue decreases. But already in the wild-type, it is not very large. Finally, its total  $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  is larger ( $-21.8 \text{ kcal mol}^{-1}$ ) compared to the WT ( $-18.3 \text{ kcal mol}^{-1}$ ). The difference cannot be traced back to a single residue but from a sum of several subtle changes.

**Experimental dissociation constants.** Having identified by computation position L198 as critical in terms of energetic contribution in the affinity of benzenesulfonamide **1** for WT hCA II, three mutants were designed and produced recombinantly in *E. coli*: L198A, L198F and L198Q. The WT hCA II and the three mutants were purified by affinity chromatography and characterized by ESI and SDS-PAGE. Next, the corresponding affinities were determined using the *p*-nitrophenyl acetate hydrolysis assay.<sup>28</sup> This straightforward assay yielded  $K_i = 1100 \pm 40 \text{ nM}$  for **1**  $\subset$  WT hCA II. This value lies well within the published data ranging from 200–1500 nM.<sup>2</sup> The experimentally measured inhibition constants of **1** for the L198X mutants are thus  $5500 \pm 270 \text{ nM}$  for L198A,  $1700 \pm 130 \text{ nM}$  for L198F and  $1800 \pm 100 \text{ nM}$  for L198Q respectively.

**Correlation.** The experimentally determined binding free energies of the L198X mutants correlate very well with the computed binding free energies (see Fig. 13). Although the differences between the mutants are slightly over-estimated, the calculated energies predict the experimental values quite well. Inclusion of these data into the correlation, affords a correlation coefficient ( $R = 0.86$  for  $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$ , i.e. 0.03 lower than without the mutants, Fig. 13). As illustrated for the L198X mutants, we have demonstrated that it is possible to correctly predict the influence of mutations at key positions in hCA II from atomistic simulations. Although the influence of the point mutations may, at first, seem subtle compared to the range of the ligands tested, this difference translates into an order of magnitude difference for the corresponding  $K_i$ s between the WT hCA II and mutant L198Q.

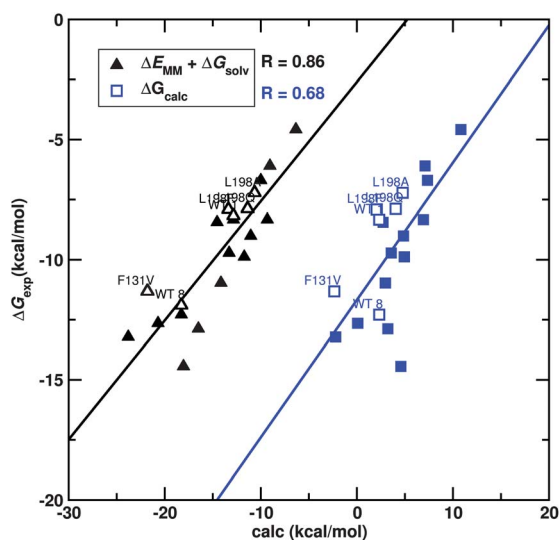


**Fig. 12** Evolution of the H-bonding pattern between ligand **1** and hCA II isoform (A) and corresponding histogram for WT hCA II and mutants at position Leu198 (B).

**Table 2** Calculated and experimental thermodynamic data

Ligand	Mutant	Calculated					Experimental	
		$\Delta E_{\text{MM}} + \Delta G_{\text{solv}} / \text{kcal mol}^{-1}$	$\sigma$	$\Delta G / \text{kcal mol}^{-1}$	$-T\Delta S / \text{kcal mol}^{-1}$	$t_{\text{sim}} / \text{ps}$	$K_i / \text{nM}$	$\Delta G_{K_i} / \text{kcal mol}^{-1}$
<b>1</b>	<b>WT</b>	-12.84	$\pm 7.64$	2.33	15.17	32500	1100	-8.17
<b>1</b>	<b>L198A</b>	-10.67	$\pm 7.57$	4.76	15.42	25000	5500	-7.22
<b>1</b>	<b>L198F</b>	-13.37	$\pm 7.34$	2.06	15.42	22000	1700	-7.91
<b>1</b>	<b>L198Q</b>	-11.38	$\pm 6.95$	4.04	15.42	25000	1800	-7.88
<b>8</b>	<b>WT</b>	-18.27	$\pm 6.64$	2.33	20.60	25000	2.1 <sup>a</sup>	-11.90
<b>8</b>	<b>F131V</b>	-21.80	$\pm 7.75$	-2.35	19.44	25000	5.6 <sup>a</sup>	-11.32

<sup>a</sup> Values from ref. 26.



**Fig. 13** Correlation between calculated  $\Delta E_{\text{MM}} + \Delta G_{\text{solv}}$  (black triangles) and  $\Delta G$  (blue squares) and experimental  $\Delta G$  for WT hCA II (full symbols) and hCA II mutants (empty symbols): **1** C L198Q; **1** C L198A; **1** C L198F; **8** C F131V (see Fig. 2 for ligand structures).

#### 4 Conclusion

The present work establishes that atomistic simulations for ligand binding in hCA II with validated force fields and sufficiently long conformational sampling allow to reliably rank ligands and predict the effects of mutations on ligand binding affinities. Ligand binding free energies based on MM-GBSA yielded a correlation of  $R = 0.89$  calculated vs. computed binding free energies between inhibitors **1–14** and hCA II. This corroborates earlier efforts for ligand-binding interactions in HIV-I protease which yielded a correlation of  $R = 0.93$ .<sup>24</sup> To single out important residues, the binding free energies were decomposed with respect to hCA II individual aminoacids. This led to the identification of Leu198 as a key residue contributing  $-3.2$  to  $-4.5$  kcal mol<sup>-1</sup> (see Fig. 11) to the interaction between benzenesulfonamide **1** and hCA II. Based on these considerations, three hCA II mutants (L198A, L198F and L198Q) were expressed, purified and tested for their affinity towards benzenesulfonamide **1**. Based on the present results, current efforts are directed towards designing *in silico* specific inhibitors towards carbonic anhydrase isoforms which are overexpressed in certain forms of cancer (e.g. hCA IX, hCA XII).

In conclusion, the combined computational and experimental approach to better characterize arylsulfonamide-based ligands interacting with hCA II shows considerable potential to extend the research towards ligand design along similar lines. The experimentally determined binding free energies show good correlation with the predicted energies, thus suggesting that the computational strategy (force field parameters, MD simulations in explicit solvation and of sufficient length, MM-GBSA and decomposition of free energies) presented herein may be widely applicable.

#### Acknowledgements

Financial support from the SNF (Grant PDFMP2 127457, Grant 200021-117810 (to MM)) and the Marie Curie Training

Network (Biotrains FP7-ITN-238531) is gratefully acknowledged. F. M. is recipient of a Novartis PhD fellowship. We thank Prof. C. A. Fierke (U of Michigan) for providing us with the hCA II plasmid.

#### References

- C. T. Supuran, *Nat. Rev. Drug Discovery*, 2008, **7**, 168–181.
- V. M. Krishnamurthy, G. K. Kaufman, A. R. Urbach, I. Gitlin, K. L. Gudiksen, D. B. Weibel and G. M. Whitesides, *Chem. Rev.*, 2008, **108**, 946–1051.
- W. L. Jorgensen, *Science*, 2004, **303**, 1813–1818.
- W. L. Jorgensen, *Acc. Chem. Res.*, 2009, **42**, 724–733.
- D. Hartsough and K. Merz, *J. Phys. Chem.*, 1995, **99**, 11266–11275.
- U. Ryde, *Curr. Opin. Chem. Biol.*, 2003, **7**, 136–142.
- G. Melagraki, A. Afantitis, H. Sarimveis, O. Igglessi-Markopoulou and C. T. Supuran, *Bioorg. Med. Chem.*, 2006, **14**, 1108–1114.
- R. Wang, Y. Lu, X. Fang and S. Wang, *J. Chem. Inf. Model.*, 2004, **44**, 2114–2125.
- K. Raha and K. M. Merz, *J. Am. Chem. Soc.*, 2004, **126**, 1020–1021.
- D. Lu and G. A. Voth, *Proteins: Struct., Funct., Genet.*, 1998, **33**, 119–134.
- P. W. Snyder, J. Mecinovic, D. T. Moustakas, S. W. Thomas, M. Harder, E. T. Mack, M. R. Lockett, a. Heroux, W. Sherman and G. M. Whitesides, *PNAS*, 2011, 1–6.
- J.-L. Zhang, Q.-C. Zheng and H.-X. Zhang, *Comput. Biol. Chem.*, 2011, **35**, 50–56.
- M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B: Condens. Matter*, 1998, **58**, 7260–7268.
- J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.
- B. R. Brooks, R. E. Bruccoleri, D. J. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
- A. D. MacKerell Jr., C. L. Brooks III, L. Nilsson, B. Roux, Y. Won and M. Karplus, in *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, ed. P. v. R. Schleyer *et al.*, John Wiley & Sons: Chichester, 1998, vol. 1, pp. 271–277.
- J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127–6129.
- M. S. Lee, M. Feig, F. R. Salsbury and C. L. Brooks, *J. Comput. Chem.*, 2003, **24**, 1348–1356.
- W. Wang and P. a. Kollman, *J. Mol. Biol.*, 2000, **303**, 567–582.
- H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891–913.
- V. Zoete, M. Meuwly and M. Karplus, *Proteins: Struct., Funct., Bioinf.*, 2005, **61**, 79–93.
- H. Thorsteinsdottir, T. Schwede, V. Zoete and M. Meuwly, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 407–423.
- T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.
- C.-Y. Kim, J. S. Chang, J. B. Doyon, T. T. Baird, C. A. Fierke, A. Jain and D. W. Christianson, *J. Am. Chem. Soc.*, 2000, **122**, 12125–12134.
- F. Studier and B. Moffatt, *J. Mol. Biol.*, 1986, **189**, 113–130.
- S. K. Nair, T. L. Calderone, D. W. Christianson and C. A. Fierke, *J. Biol. Chem.*, 1991, **266**, 17320–17325.
- D. K. Srivastava, K. M. Jude, A. L. Banerjee, M. Haldar, S. Manokaran, J. Kooren, S. Mallik and D. W. Christianson, *J. Am. Chem. Soc.*, 2007, **129**, 5528–5537.
- D. Vullo, M. Franchi, E. Gallori, J. Antel, A. Scozzafava and C. Supuran, *J. Med. Chem.*, 2004, **47**, 1272–1279.
- S. Genheden and U. Ryde, *J. Comput. Chem.*, 2010, **31**, 837–846.
- H. Gouda, I. Kuntz, D. Case and P. Kollman, *Biopolymers*, 2003, **68**, 16–34.

- 
- 33 H. Gohlke and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 238–250.
- 34 A. L. Banerjee, M. Swanson, B. C. Roy, X. Jia, M. K. Haldar, S. Mallik and D. K. Srivastava, *J. Am. Chem. Soc.*, 2004, **126**, 10875–10883.
- 35 S. Nair, D. Elbaum and D. Christianson, *J. Biol. Chem.*, 1996, **271**, 1003–1007.
- 36 S. Nair and D. Christianson, *Biochemistry*, 1993, **32**, 4506–4514.
- 37 S. K. Nair, J. F. Krebs, D. W. Christianson and C. A. Fierke, *Biochemistry*, 1995, **34**, 3981–3989.
- 38 S. Z. Fisher, C. Tu, D. Bhatt, L. Govindasamy, M. Agbandje-McKenna, R. McKenna and D. N. Silverman, *Biochemistry*, 2007, **46**, 3803–3813.