Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 15 (2007) 7865–7877
Bioorganic & Medicinal Chemistry

Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors

Anna Maria Ferrari, Gianluca Degliesposti, Miriam Sgobba and Giulio Rastelli*

Dipartimento di Scienze Farmaceutiche, Università di Modena e Reggio Emilia, Via Campi 183, 41100 Modena, Italy

Received 11 June 2007; revised 31 July 2007; accepted 3 August 2007
Available online 22 August 2007

Abstract—Among the available methods for predicting free energies of binding of ligands to a protein, the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) and molecular mechanics generalized Born surface area (MM-GBSA) approaches have been validated for a relatively limited number of targets and compounds in the training set. Here, we report the results of an extensive study on a series of 28 inhibitors of aldose reductase with experimentally determined crystal structures and inhibitory activities, in which we evaluate the ability of MM-PBSA and MM-GBSA methods in predicting binding free energies using a number of different simulation conditions. While none of the methods proved able to predict absolute free energies of binding in quantitative agreement with the experimental values, calculated and experimental free energies of binding were significantly correlated. Comparing the predicted and experimental ΔG of binding, MM-PBSA proved to perform better than MM-GBSA, and within the MM-PBSA methods, the PBSA of Amber performed similarly to Delphi. In particular, significant relationships between experimental and computed free energies of binding were obtained using Amber PBSA and structures minimized with a distance-dependent dielectric function. Importantly, while free energy predictions are usually made on large collections of equilibrated structures sampled during molecular dynamics in water, we have found that a single minimized structure is a reasonable approximation if relative free energies of binding are to be calculated. This finding is particularly relevant, considering that the generation of equilibrated MD ensembles and the subsequent free energy analysis on multiple snapshots is computationally intensive, while the generation and analysis of a single minimized structure of a protein–ligand complex is relatively fast, and therefore suited for high-throughput virtual screening studies. At this aim, we have developed an automated workflow that integrates all the necessary steps required to generate structures and calculate free energies of binding. The procedure is relatively fast and able to screen automatically and iteratively molecules contained in databases and libraries of compounds. Taken altogether, our results suggest that the workflow can be a valuable tool for ligand identification and optimization, being able to automatically and efficiently refine docking poses, which sometimes may not be accurate, and rank the compounds based on more accurate scoring functions.

1. Introduction

In the last decade, many computational tools have been developed and applied for structure-based drug discovery aimed at screening large databases of chemicals against biologically active targets. One of the more popular approaches is molecular docking, which is able to screen and rank a relatively large number of molecules into the active site of a target structure in a reasonable amount of time. Docking methods find favorable orientations of potential ligands into the active site of the target, and evaluate the electrostatic and steric complementarities of the ligand with the target in terms of energy scores derived from empirical force fields. Different docking algorithms have been developed, involving different orientation and conformation searching algorithms as well as different scoring functions.[1–4] As a matter of fact, while scoring functions have been improved in the last years by including additional energy contributions and/or refining parameters in the force-field equations, we are still far from having docking tools that quantitatively and accurately predict biologically active candidates in silico.[5]

One of the key energy contributions in estimating ligand-target affinity is desolvation energy, that is, the

Keywords: Drug design; MM-PBSA; Virtual screening; Aldose reductase.
* Corresponding author. Tel.: +39 059 2055145; fax: +39 059 2055131; e-mail: rastelli.giulio@unimore.it

0968-0896/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2007.08.019

energy required to desolvate the ligand and the target when the complex is formed. Although docking methods that include desolvation energies in their force field have been recently developed, desolvation energies are necessarily approximated and still unable to yield fair agreement with experimental data. On the other hand, the application of more accurate computational methods, like free energy perturbation,[6–8] thermodynamic integration,[9] MM-PBSA,[10,11] and others, is computationally intensive and therefore unsuitable for screening purposes. Typical applications involved the calculation of binding free energies in small series of congeneric compounds, a situation far from virtual screening in which a large number of diverse chemicals have to be processed.

The MM-PBSA method developed by Kollman and Case[10,11] associates molecular mechanics energies in vacuo and solvation free energies to estimate the ΔG of binding; it has been successfully applied to analyze the energetics of large collections of equilibrated snapshots of protein–ligand complexes collected during molecular dynamics simulations in water.[12–18] Calculating binding free energies using a single protein–ligand complex is considerably faster than generating and analysing an ensemble of equilibrated structures. However, the validity of this approach has not been explored in sufficient detail, since the few studies reported so far were limited to small series of congeneric compounds and a limited number of targets.[15,18,19]

In the present work, we attempt to set up and validate a simulation protocol able to predict relative binding free energies in reasonable agreement with experimental data using single minimized structures. Such a validation study is a first but essential step to support the usefulness and accuracy of these protocols as post-processing tools after docking screenings. At this aim, we extensively tested the performance of the MM-PBSA and MM-GBSA methods implemented in Amber and Delphi on a series of 28 inhibitors of aldose reductase spanning five orders of magnitude in biological activity. Compared to previous MM-PBSA applications reported in the literature,[12–19] this is, to our knowledge, the largest set of compounds used for predicting ligand affinities in silico with these methods. At this purpose, gas phase interaction energies were corrected with polar solvation free energies calculated by solving the Poisson–Boltzmann equation either with Delphi[20,21] or the PBSA module of Amber,[22] and by means of generalized Born (GBSA) methods. The nonpolar solvation free energies were estimated from accessible surface area calculations performed with Molsurf[23] and LCPO.[24] In order to investigate the ability in predicting free energies of binding, different combinations of methods, parameters, as well as different methods for generating the structures of the protein–ligand complexes have been explored.

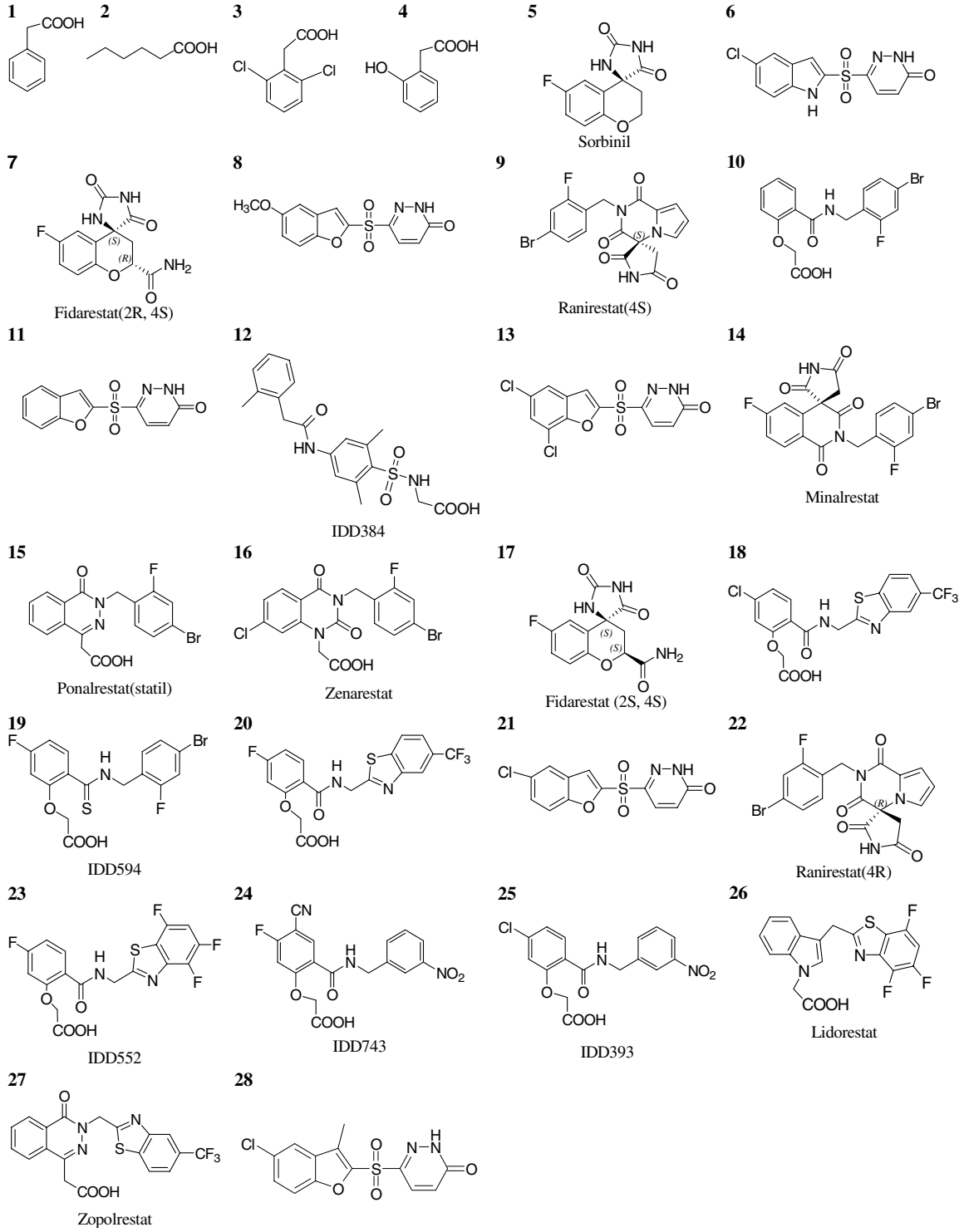
Finally, an automated procedure that integrates all the necessary steps required to generate structures and calculate free energies of binding of ligands in a database of compounds has been developed. The results prospect that our simulation protocol can be useful to predict relative binding free energies of structurally unrelated molecules in a given binding site, and hold promise as a post-processing method to more accurately score and rank hit compounds after docking screenings.

2. Results and discussion

In order to test the performance of MM-PBSA and MM-GBSA in predicting ligand binding affinities, we have applied these methods to a series of known inhibitors of aldose reductase. ALR2 is a well-characterized enzyme with a large number of inhibitory activity data available in the literature[25–34] and plenty of crystal structures in complex with inhibitors, some of which at high and ultra-high resolution.[27,31,33–43] Our training set was composed of 28 inhibitors with measured activities and known binding modes (Scheme 1). These inhibitors were selected according to the highest possible diversity in structure, significant variation in inhibitory activity (five orders of magnitude variation in IC_{50}, corresponding to a $\sim$7 kcal/mol difference in free energy of binding), and availability of crystal structures.

To validate the methods and the procedure, we used different protocols to generate the structures of the ALR2-inhibitor complexes, and different methods to perform binding free energy estimates. Regarding structures, these were generated (a) in explicit water solvent with energy minimization of the complexes solvated in octahedral boxes of TIP3P water molecules, (b) in implicit solvent with GBSA energy minimization, (c) with a distance-dependent dielectric function, and (d) with a protocol consisting in distance-dependent dielectric energy minimization followed by molecular dynamics on the ligand and final re-minimization. While free energy predictions are usually made on large collections of equilibrated structures sampled during molecular dynamics in water, we were interested in testing whether a single minimized structure was a reasonable approximation for estimating free energies of binding. In fact, generating a single minimized structure of a protein–ligand complex and estimating the ΔG of binding from one structure is relatively fast, and therefore suited for high-throughput virtual screening studies. Regarding the various methods used for free energy calculations, the Amber force field was used to calculate gas-phase interaction energies. Delphi, Amber PBSA or GBSA were used for estimating polar solvation free energies, and Molsurf or LCPO were used for estimating nonpolar solvation free energies.

The ΔG_{bind} value of each inhibitor can be calculated according to the equation $\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S_{solute}$, where ΔE_{MM} is the molecular mechanics contribution expressed as the sum of the internal, electrostatics and van der Waals contributions to binding in vacuo; ΔG_{solv} is the solvation free energy contribution to binding expressed as the sum of polar and nonpolar solvation free energies ($\Delta G_{solv} = \Delta G_{psolv} + \Delta G_{npsolv}$, respectively); and $T\Delta S_{solute}$ is the contribution of solute entropy to binding.



Scheme 1. Chemical structures of the aldose reductase inhibitors of the training set.

2.1. Free energies of binding using structures minimized in explicit water solvent

Table 1 reports the energy contributions to the free energy of binding calculated after energy minimization of the ALR2-inhibitor complexes in explicit water. The gas-phase interaction energies (ΔE_{MM}) are not correlated with the experimental free energies of binding. This finding was expected, given that the inclusion of desolvation energies is crucial for predicting binding affinities in agreement with experiment. While the van der Waals (ΔE_{vdW}) component of the interaction energy in vacuo is always negative, thereby contributing favorably to binding, the electrostatic (ΔE_{el}) component is generally positive. However, positive values for ΔE_{el}, in this case, should not lead to the conclusion that the electrostatics

Table 1. Energy contributions (kcal/mol) to the free energy of binding between aldose reductase and the set of 28 inhibitors

$n°$	IC_{50} (μM)	ΔG_{expt}[a]	ΔE_{el}	ΔE_{vdW}	ΔE_{MM}[b]	ΔG_{psolv}[c] Delphi	ΔG_{psolv}[c] PB	ΔG_{psolv}[c] GB	ΔG_{npsolv}[d] Molsurf	ΔG_{npsolv}[e] Molsurf Delphi	ΔG_{npsolv}[f] LCPO	$\Delta G'_{bind}$[g] Delphi	$\Delta G'_{bind}$[g] PBSA	$\Delta G'_{bind}$[g] GBSA
1	96	−5.5	−0.8	−14.0	−14.8	23.3	14.5	5.2	−3.0	−3.1	−3.0	5.3	−3.3	−12.6
2	68.2	−5.7	−4.9	−13.6	−18.5	23.5	14.4	7.2	−2.9	−3.1	−2.9	2.0	−6.9	−14.1
3	4.4	−7.3	−2.1	−20.3	−22.4	26.1	16.5	5.3	−3.3	−3.4	−3.3	0.3	−9.2	−20.4
4	3.5	−7.4	3.7	−17.0	−13.3	11.7	2.0	2.8	−2.9	−3.1	−2.9	−4.6	−14.1	−13.4
5	2	−7.8	8.9	−29.5	−20.6	20.1	11.1	0.8	−3.7	−3.7	−3.7	−4.2	−13.2	−23.5
6	1.1	−8.1	29.5	−43.7	−14.2	7.8	−0.3	−4.4	−4.8	−4.5	−4.7	−10.8	−19.2	−23.4
7	0.57	−8.5	−10.0	−32.3	−42.4	40.2	26.4	15.2	−4.1	−4.0	−4.1	−6.1	−20.0	−31.3
8	0.23	−9.1	23.0	−45.8	−22.8	17.8	6.8	1.3	−5.0	−4.7	−5.0	−9.7	−21.0	−26.4
9	0.19	−9.2	41.0	−46.9	−5.9	3.6	−7.2	−23.8	−5.5	−5.0	−5.5	−7.3	−18.6	−35.2
10	0.176	−9.2	16.9	−42.4	−25.5	17.7	6.2	−7.6	−5.5	−5.0	−5.5	−12.8	−24.8	−38.5
11	0.15	−9.3	30.9	−40.7	−9.8	4.0	−2.0	−5.9	−4.6	−4.3	−4.6	−10.2	−16.3	−20.2
12	0.108	−9.5	−18.2	−35.3	−53.5	51.8	41.6	30.3	−4.8	−4.5	−4.9	−6.3	−16.7	−28.0
13	0.087	−9.6	27.6	−46.8	−19.2	11.3	−0.5	−5.1	−4.9	−4.6	−4.9	−12.5	−24.5	−29.2
14	0.073	−9.7	29.0	−51.3	−22.4	13.7	4.2	−14.2	−5.8	−5.3	−5.8	−14.0	−23.9	−42.4
15	0.0565	−9.9	12.8	−42.6	−29.8	25.3	12.7	−4.7	−5.5	−5.1	−5.6	−9.5	−22.6	−40.0
16	0.044	−10.0	15.0	−46.2	−31.2	21.9	12.0	−5.3	−5.8	−5.3	−5.7	−14.6	−24.9	−42.2
17	0.035	−10.2	−13.0	−31.2	−44.2	36.3	21.7	14.1	−4.0	−4.0	−4.1	−11.8	−26.5	−34.1
18	0.035	−10.2	7.2	−53.5	−46.2	29.3	14.3	0.3	−6.2	−5.6	−6.2	−22.5	−38.1	−52.2
19	0.03	−10.3	21.6	−43.8	−22.3	9.8	−0.9	−12.1	−5.7	−5.2	−5.7	−17.7	−28.8	−40.1
20	0.029	−10.3	6.7	−51.9	−45.3	29.4	15.7	1.4	−6.1	−5.5	−6.1	−21.4	−35.7	−50.0
21	0.025	−10.4	32.6	−41.7	−9.1	−3.3	−10.1	−10.7	−4.6	−4.4	−4.6	−16.8	−23.8	−24.4
22	0.015	−10.7	29.0	−48.7	−19.7	11.0	2.1	−14.5	−5.4	−5.0	−5.4	−13.6	−23.0	−39.6
23	0.011	−10.9	10.9	−51.0	−40.1	20.4	8.3	−5.3	−5.8	−5.3	−5.8	−25.0	−37.6	−51.2
24	0.007	−11.1	17.3	−47.9	−30.6	16.9	3.4	−6.7	−5.8	−5.3	−5.8	−18.9	−32.9	−43.0
25	0.006	−11.2	15.6	−47.7	−32.1	17.8	6.4	−4.8	−5.6	−5.2	−5.6	−19.5	−31.3	−42.6
26	0.005	−11.3	5.4	−46.8	−41.4	22.9	11.7	−0.2	−5.5	−5.1	−5.6	−23.5	−35.2	−47.1
27	0.0031	−11.6	12.9	−53.4	−40.5	25.3	11.0	−5.1	−5.9	−5.4	−5.9	−20.6	−35.4	−51.5
28	0.001	−12.3	27.3	−41.0	−13.6	−9.3	−18.4	−20.2	−4.7	−4.4	−4.7	−27.4	−36.7	−38.5

Values have been estimated after energy minimization of the complexes in explicit water solvent. Calculated free energies of binding $\Delta G'_{bind}$ are obtained by summing ΔE_{MM}, ΔG_{psolv} and ΔG_{npsolv} contributions.

[a] Experimental free energies of binding (kcal/mol) according to $\Delta G = -RT\ln(1/IC_{50})$.

[b] Gas-phase interaction energies $\Delta E_{MM} = \Delta E_{el} + \Delta E_{vdW}$.

[c] Polar solvation free energies according to Delphi, Amber PB and GB.

[d] Nonpolar solvation free energies estimated with Molsurf according to $G_{npsolv} = \gamma SASA + b$, where $\gamma = 0.0072$ kcal mol^{-1} Å^{-2} and $b = 0$ kcal mol^{-1} for use in amber PBSA and GBSA.

[e] Nonpolar solvation energies with Molsurf and $\gamma = 0.00542$ kcal mol^{-1} Å^{-2} and $b = 0.92$ kcal mol^{-1} for use in Delphi.

[f] Nonpolar solvation free energies with LCPO and $\gamma = 0.0072$ kcal mol^{-1} Å^{-2} and $b = 0$ kcal mol^{-1}.

[g] Estimated free energies of binding $\Delta G'_{bind} = \Delta E_{MM} + \Delta G_{psolv} + \Delta G_{npsolv}$ where ΔG_{psolv} was estimated with Delphi, Amber PBSA and GBSA.

Table 2. Average CPU times (min) on a dual core 2.4 GHz AMD Opteron processor 280 with 8 GB memory[a]

Structure generation	Min	Free energy evaluation	Min
Energy minimization with explicit water	11.60	Delphi	4.01
Energy minimization with GBSA	77.23	Amber PBSA	1.1
Energy minimization with distance-dielectric constant $\varepsilon = 4r$	2.70	Amber GBSA	0.13
Energy minimization/MD/re-minimization with distance-dielectric constant $\varepsilon = 4r$	6.86		

[a] CPU time needed to set up input files, amber topologies and update coordinates in the automated procedure is 0.05 min.

is unfavorable to binding, because the negatively charged inhibitors of our training set are strongly penalized by the presence of a net charge of −5 on ALR2. To test the dependence of the free energy results on the net charge of the protein, free energy predictions were repeated on ALR2 complexes neutralized with counterions. In the neutral complexes, ΔE_{el} becomes, on average, 83 ± 2 kcal/mol more negative, while ΔE_{vdW} values remain almost unaffected, varying only 0.6 ± 0.7 kcal/mol (data not shown). Therefore, electrostatic interactions should be regarded as important, in agreement with the observed requirement of an 'anion' binding hole in ALR2 for efficient binding of inhibitors.[43–48]

Polar solvation free energies (ΔG_{psolv}) included in Table 1 have been calculated with Delphi, Amber PB and GB, using Amber ff03 charges for the protein and AM1-BCC charges for the ligand in all three methods (i.e., the same charge sets used for generating the structures). In general, the observed variation of ΔG_{psolv} values reflects variations in ΔE_{el}. Interestingly, when the complexes are neutralized with counterions, ΔG_{psolv} values become 81 ± 2 kcal/mol higher in both PB and GB calculations (data not shown). Therefore, the higher electrostatic interaction (ΔE_{el}) observed on the neutralized complexes is compensated by an almost equivalent desolvation penalty, leading to $\Delta G'_{bind}$ values that are very similar to those calculated on the charged protein (on average, a 1.5 kcal/mol difference in $\Delta G'_{bind}$ was observed). ΔG_{psolv} values calculated with Delphi are, on average, 10 and 21 kcal/mol higher than values calculated with Amber PB and GB, respectively. However, the polar solvation free energies were significantly correlated in the series of compounds, squared regression coefficients being 0.97 between Delphi and Amber PB, 0.73 between Delphi and Amber GB, and 0.79 between Amber PB and GB. Therefore, while absolute values of ΔG_{psolv} clearly differ among the three methods, their variation along the series is significantly correlated, especially in the case of Delphi and Amber PB which both solve the Poisson–Boltzmann equation. While this finding was previously observed analysing large collections of structures sampled during MD, we show for the first time that the same holds for energy minimized structures, and for a considerably larger number of compounds in the training set compared to previous investigations. Considering that Delphi, which is often regarded as one of the most accurate methods for estimating polar solvation energies, is significantly more cpu-consuming than Amber PB and GB (Table 2), our results suggest that Amber PB ($r^2 = 0.97$), and to a lesser extent GB ($r^2 = 0.73$), can be useful alternatives. This is an important aspect of the validation of our procedure, given that a good balance between calculation accuracy and speed is crucial when large databases of molecules have to be processed in a virtual screening application.

The nonpolar solvation free energy contributions (ΔG_{npsolv}) have been computed using the Molsurf and LCPO methods, according to the equation $G_{npsolv} = \gamma SASA + b$, where SASA is the solvent accessible surface area. Since Delphi uses Parse radii and was parameterized with slightly different γ and b values compared to Amber PB and GB, Table 1 includes the ΔG_{npsolv} values to be used together with Delphi polar solvation energies. With the same γ and b parameters, Molsurf and LCPO give almost identical nonpolar solvation free energies (Table 1). Therefore, Molsurf was used for subsequent analyses. Not surprisingly, ΔG_{npsolv} values are always negative (i.e., favorable to binding) and highly correlated with ΔE_{vdW} ($r^2 = 0.92$). Besides polar interactions at the 'anion' hole of ALR2 established between the negatively charged portion of the inhibitors and Tyr48, His110, Trp111 and the oxidized cofactor $NADP^+$, hydrophobic interactions at a secondary hydrophobic pocket lined by Trp111, Leu300, and Phe121 also contribute favorably to activity.[44–49]

Finally, Table 1 reports the estimated free energies of binding, expressed as the sum of the interaction energies in vacuo (ΔE_{MM}), and the polar (ΔG_{psolv}) and nonpolar (ΔG_{npsolv}) solvation free energies calculated with the different methods discussed above. Figure 1a–c reports the regression plots of the experimental versus computed ΔG of binding according to Delphi, Amber PBSA and GBSA. As expected, none of the methods could reproduce the absolute ΔG_{expt} values quantitatively. Long MD simulations and the inclusion of entropy contributions are required to reach this goal. However, the computed $\Delta G'_{bind}$ and the experimental ΔG_{expt} values were significantly correlated (Delphi $r^2 = 0.84$, $s = 0.66$, $F = 141$; Amber PBSA $r^2 = 0.81$, $s = 0.72$, $F = 114$; Amber GBSA $r^2 = 0.66$, $s = 0.98$, $F = 51$, Fig. 1a–c). Notably, this is the first report in which such good correlations have been obtained with single structures and a large set of inhibitors. In reproducing free energies of binding, Delphi proved to perform only slightly better than Amber PBSA. The GBSA approach, considered a good approximation of the more accurate but also more computationally expensive PBSA method, gave a slightly worse, but still significant relationship. The poorer performance of GBSA was also observed in previous investigations.[18]

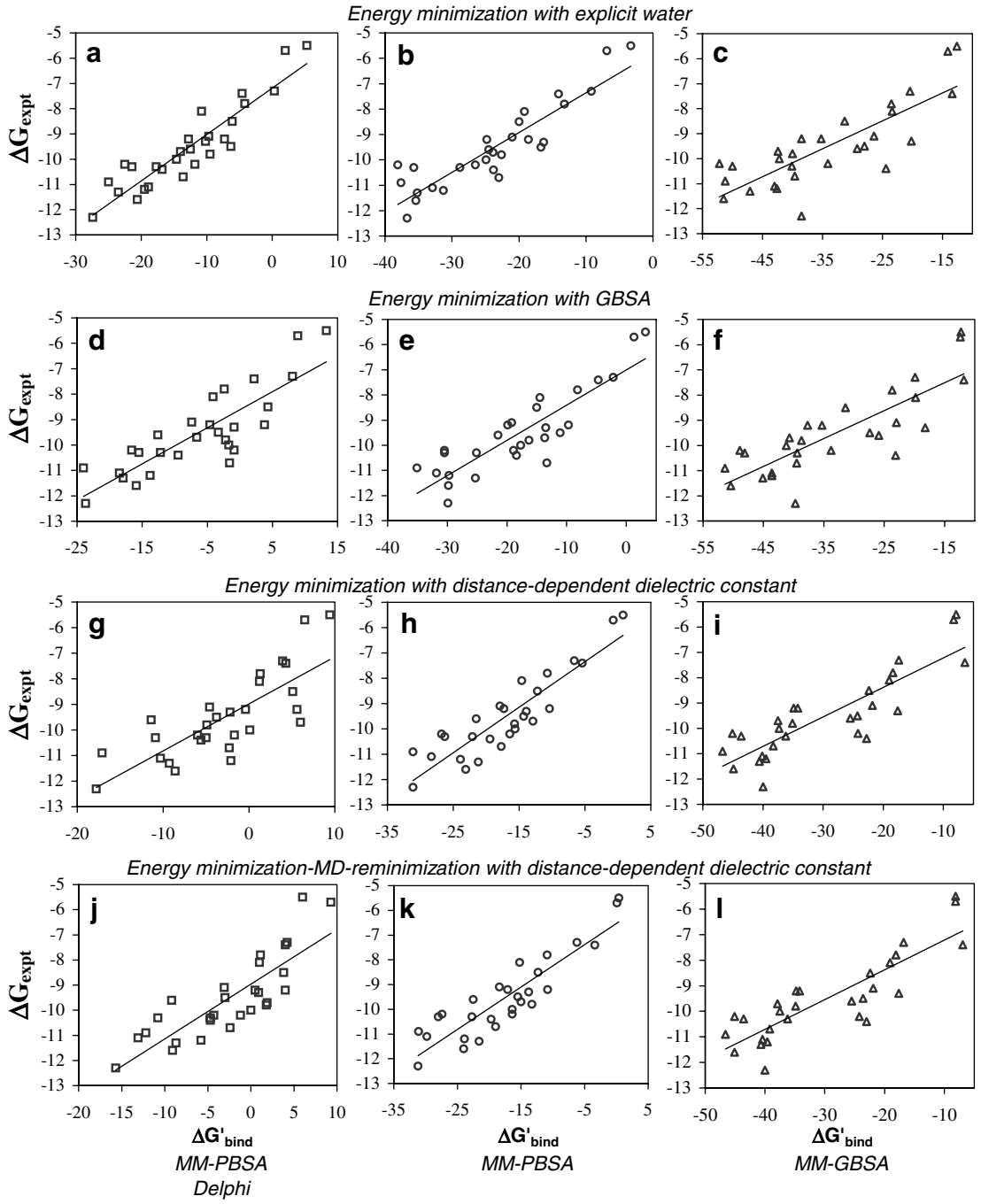


Figure 1. Experimental versus estimated free energies of binding (kcal/mol) of the 28 ARL2 inhibitors. Free energies were obtained after minimization with explicit water (a–c), minimization with GBSA (d–f), minimization with distance-dependent dielectric constant $\varepsilon = 4r$ (g–i), and a multistep minimization–MD–re-minimization protocol with distance-dependent dielectric constant (j–l). For each of these, solvation free energies were evaluated with Delphi (plots a, d, g, j), Amber PBSA (plots b, e, h, k), and Amber GBSA (plots c, f, i, l).

2.2. Free energies of binding using structures minimized with GBSA implicit solvation model

Extensive validations were then performed using structures of ALR2-inhibitor complexes minimized with an implicit solvent model. First, we tested whether the various methods were able to predict differences in binding free energies when structures were energy minimized with GBSA. Table 3 reports the free energies obtained after GBSA minimization, and Figure 1d–f shows the regression plots between experimental and estimated free energies of binding obtained using these structures. For convenience, only ΔE_{MM}, ΔG_{solv} (the sum of ΔG_{psolv} and ΔG_{npsolv}), and $\Delta G'_{bind}$ values have been reported in Table 3. Significant relationships between computed $\Delta G'_{bind}$ and experimental ΔG_{expt} values were still observed, especially when free energies were evaluated with Amber PBSA (Delphi $r^2 = 0.71$, $s = 0.91$, $F = 63$; Amber PBSA $r^2 = 0.77$, $s = 0.81$, $F = 85$; Amber GBSA $r^2 = 0.67$, $s = 0.96$, $F = 53$, Fig. 1d–f). Compared to sim-

Table 3. Free energies (kcal/mol) according to Delphi, Amber PBSA and GBSA calculations performed on structures generated with minimization with GBSA, a distance-dependent dielectric function, and minimization–MD–re-minimization with distance-dependent dielectric function

n°	ΔG_{expt}	ΔE_{MM}	ΔG_{solv} Delphi	ΔG_{solv} PBSA	ΔG_{solv} GBSA	$\Delta G'_{bind}$ Delphi	$\Delta G'_{bind}$ PBSA	$\Delta G'_{bind}$ GBSA	n°	ΔG_{expt}	ΔE_{MM}	ΔG_{solv} Delphi	ΔG_{solv} PBSA	ΔG_{solv} GBSA	$\Delta G'_{bind}$ Delphi	$\Delta G'_{bind}$ PBSA	$\Delta G'_{bind}$ GBSA
Minimization with GBSA																	
1	−5.5	−2.7	15.9	5.8	−9.7	13.3	3.2	−12.3	**15**	−9.9	−23.8	21.6	7.5	−14.9	−2.2	−16.3	−38.7
2	−5.7	−4.1	13.1	5.4	−8.2	8.9	1.3	−12.4	**16**	−10.0	−26.7	25.0	9.0	−14.5	−1.7	−17.7	−41.2
3	−7.3	−8.0	16.1	5.8	−11.9	8.1	−2.2	−19.9	**17**	−10.2	−29.7	28.8	10.8	−4.1	−0.9	−18.9	−33.8
4	−7.4	−2.8	5.0	−1.9	−9.0	2.2	−4.7	−11.8	**18**	−10.2	−34.7	18.2	4.2	−14.1	−16.6	−30.5	−48.9
5	−7.8	−9.0	6.6	0.8	−14.6	−2.4	−8.2	−23.7	**19**	−10.3	−17.8	5.7	−7.3	−21.6	−12.2	−25.1	−39.4
6	−8.1	−5.7	1.6	−8.9	−14.1	−4.1	−14.5	−19.8	**20**	−10.3	−34.4	18.9	3.9	−13.7	−15.5	−30.5	−48.1
7	−8.5	−29.7	34.0	14.7	−1.7	4.3	−15.0	−31.4	**21**	−10.4	−8.0	−1.5	−10.4	−15.1	−9.5	−18.4	−23.1
8	−9.1	−12.8	5.4	−6.4	−10.2	−7.4	−19.2	−23.0	**22**	−10.7	−15.1	13.5	1.8	−24.4	−1.6	−13.3	−39.5
9	−9.2	−7.9	11.7	−1.8	−27.4	3.8	−9.7	−35.3	**23**	−10.9	−35.1	11.1	0.0	−16.3	−24.0	−35.1	−51.4
10	−9.2	−16.9	12.3	−3.0	−20.9	−4.6	−19.9	−37.7	**24**	−11.1	−28.6	10.1	−3.2	−15.0	−18.5	−31.8	−43.6
11	−9.3	−4.6	3.8	−8.8	−13.6	−0.9	−13.5	−18.2	**25**	−11.2	−29.4	15.6	−0.3	−14.2	−13.8	−29.7	−43.6
12	−9.5	−42.5	43.7	31.4	15.1	−3.3	−11.1	−27.4	**26**	−11.3	−30.6	12.8	5.3	−14.5	−17.9	−25.3	−45.1
13	−9.6	−9.6	−2.9	−11.8	−16.3	−12.6	−21.5	−25.9	**27**	−11.6	−34.7	18.8	5.0	−15.7	−15.9	−29.8	−50.4
14	−9.7	−15.0	8.3	1.3	−25.7	−6.6	−13.7	−40.7	**28**	−12.3	−11.5	−12.2	−18.4	−28.2	−23.7	−29.9	−39.7
Minimization with distance-dependent dielectric constant 4r																	
1	−5.5	−1.4	10.9	2.2	−6.5	9.4	0.8	−7.9	**15**	−9.9	−18.5	13.6	2.8	−16.6	−4.9	−15.7	−35.1
2	−5.7	−4.1	10.6	3.4	−4.2	6.5	−0.7	−8.3	**16**	−10.0	−21.9	22.0	6.3	−15.5	0.1	−15.6	−37.3
3	−7.3	−9.4	13.3	2.8	−8.0	3.9	−6.6	−17.4	**17**	−10.2	−19.1	17.5	2.8	−5.1	−1.7	−16.4	−24.2
4	−7.4	−3.1	7.4	−2.2	−3.3	4.3	−5.4	−6.4	**18**	−10.2	−30.3	24.2	3.5	−14.8	−6.0	−26.8	−45.1
5	−7.8	−5.5	6.8	−5.3	−12.9	1.3	−10.7	−18.4	**19**	−10.3	−16.2	11.2	−5.9	−20.0	−5.0	−22.1	−36.2
6	−8.1	−7.9	9.0	−6.7	−11.1	1.2	−14.6	−19.0	**20**	−10.3	−29.8	18.9	3.5	−13.8	−10.9	−26.3	−43.6
7	−8.5	−20.2	25.3	7.9	−2.2	5.1	−12.2	−22.4	**21**	−10.4	−10.1	4.5	−9.4	−12.7	−5.6	−19.4	−22.8
8	−9.1	−13.7	9.1	−4.2	−8.1	−4.6	−17.9	−21.8	**22**	−10.7	−17.0	14.7	−0.7	−21.4	−2.3	−17.7	−38.3
9	−9.2	−14.5	20.0	4.1	−20.4	5.6	−10.4	−34.9	**23**	−10.9	−29.4	12.4	−1.6	−17.3	−17.1	−31.1	−46.7
10	−9.2	−18.1	17.7	0.8	−16.2	−0.4	−17.3	−34.2	**24**	−11.1	−23.0	12.7	−5.3	−17.1	−10.3	−28.3	−40.1
11	−9.3	−6.5	4.3	−7.3	−11.1	−2.2	−13.9	−17.6	**25**	−11.2	−25.5	23.4	1.5	−14.1	−2.1	−23.9	−39.5
12	−9.5	−42.2	38.4	27.9	17.9	−3.8	−14.3	−24.3	**26**	−11.3	−22.0	12.7	0.8	−18.6	−9.3	−21.2	−40.6
13	−9.6	−11.9	0.5	−9.5	−13.6	−11.4	−21.5	−25.5	**27**	−11.6	−28.9	20.3	5.8	−16.0	−8.6	−23.1	−44.9
14	−9.7	−14.2	20.2	1.3	−23.3	6.0	−12.9	−37.5	**28**	−12.3	−13.5	−4.3	−17.6	−26.4	−17.8	−31.1	−40.0
Minimization/MD/re-minimization with distance-dependent dielectric constant 4r																	
1	−5.5	2.1	3.9	−1.7	−10.1	6.0	0.4	−8.1	**15**	−9.9	−18.9	20.7	5.6	−16.0	1.8	−13.3	−34.9
2	−5.7	−1.4	10.7	1.5	−6.7	9.3	0.1	−8.1	**16**	−10.0	−23.1	23.1	6.7	−14.4	0.0	−16.4	−37.5
3	−7.3	−9.1	13.2	2.9	−7.8	4.2	−6.2	−16.8	**17**	−10.2	−18.9	17.7	2.5	−5.2	−1.2	−16.4	−24.2
4	−7.4	−2.6	6.6	−0.7	−4.3	4.0	−3.4	−6.9	**18**	−10.2	−30.3	26.1	3.0	−14.8	−4.3	−27.4	−45.1
5	−7.8	−5.4	6.5	−5.6	−12.7	1.1	−10.9	−18.1	**19**	−10.3	−16.2	11.6	−6.5	−20.0	−4.7	−22.7	−36.2
6	−8.1	−8.4	9.4	−6.7	−10.7	1.0	−15.2	−19.1	**20**	−10.3	−30.0	19.2	2.0	−13.6	−10.8	−28.0	−43.6
7	−8.5	−20.4	24.1	8.1	−2.0	3.8	−12.3	−22.4	**21**	−10.4	−10.6	5.9	−9.2	−12.4	−4.7	−19.7	−23.0
8	−9.1	−14.1	10.9	−4.3	−7.8	−3.1	−18.4	−21.9	**22**	−10.7	−18.7	16.4	−0.3	−20.4	−2.4	−19.0	−39.2
9	−9.2	−15.3	19.2	4.5	−19.5	4.0	−10.8	−34.8	**23**	−10.9	−29.2	17.0	−1.8	−17.3	−12.2	−31.1	−46.6
10	−9.2	−18.1	18.5	0.9	−16.1	0.5	−17.1	−34.2	**24**	−11.1	−23.6	10.6	−6.2	−16.7	−13.1	−29.8	−40.4
11	−9.3	−6.8	7.6	−7.1	−10.8	0.9	−13.8	−17.6	**25**	−11.2	−26.2	20.4	2.3	−13.5	−5.8	−23.9	−39.6
12	−9.5	−43.1	40.1	27.6	19.6	−3.0	−15.5	−23.6	**26**	−11.3	−22.4	13.7	0.8	−18.3	−8.7	−21.6	−40.7
13	−9.6	−12.4	3.2	−10.1	−13.1	−9.2	−22.5	−25.5	**27**	−11.6	−29.4	20.3	5.4	−15.7	−9.1	−24.0	−45.1
14	−9.7	−14.9	16.8	−0.1	−23.0	1.9	−15.0	−37.9	**28**	−12.3	−13.3	−2.5	−18.0	−26.8	−15.7	−31.2	−40.0

ulations in water, structures minimized with GBSA yielded slightly lower regression coefficients when free energies of binding were evaluated with Delphi (0.71 vs 0.84) or Amber PBSA (0.77 vs 0.81), but almost identical when evaluated with Amber GBSA (0.67 vs 0.66). Notably, under the simulation conditions specified in the methods, GBSA minimization was, on average, 7 times slower than minimization in water (Table 2), and did not yield better results. While gas-phase interaction energies (ΔE_{MM}) are generally less negative but still significantly correlated with the ΔE_{MM} values calculated in water ($r^2 = 0.88$), solvation free energies are generally more negative but still correlated ($r^2 = 0.81$ for Delphi, 0.86 for Amber PBSA, and 0.91 for Amber GBSA). Therefore, a compensation between absolute values of interaction energies in vacuo and desolvation energies is generally observed comparing the two methods used to generate structures. Moreover, the finding that solvation free energies according to Amber GBSA highly correlate between structures minimized in explicit and implicit solvent (0.91) explains why the regression coefficients in $\Delta G'_{bind}/\Delta G_{expt}$ plots are very similar (0.67 vs 0.66). As a matter of fact, binding free energies estimated with GBSA proved to be less sensible to the method used to generate structures (Tables 1 and 3). However, Amber GBSA gave the worst agreement between experimental and calculated free energies of binding, at least in our test case. On the contrary, solvation free energies according to Delphi and Amber PBSA were slightly less correlated between water- and GBSA-minimized structures (0.81 in Delphi and 0.86 in PBSA), and a slightly reduced agreement between experimental and estimated binding free energies was observed (0.71 vs 0.84 for Delphi, and 0.77 vs 0.81 for Amber PBSA).

2.3. Free energies of binding using structures minimized with a distance-dependent dielectric constant

Free energy evaluations were then performed on structures minimized with a distance-dependent dielectric constant $\varepsilon = 4r$ (Table 3). These simulations were performed in order to investigate whether a fast minimization protocol of protein–ligand complexes could lead to free energy estimates in reasonable agreement with experiment. The best correlation between $\Delta G'_{bind}$ and ΔG_{expt} values (Fig. 1g–i) was observed with Amber PBSA ($r^2 = 0.80$, $s = 0.74$, $F = 107$). Importantly, such relationship is very similar to that obtained using explicit water ($r^2 = 0.81$, $s = 0.72$, $F = 114$), suggesting that free energy predictions with Amber PBSA work comparably well when structures are minimized with distance-dependent dielectric function. Notably, minimization with distance-dependent dielectric constant is, on average, 4 times faster than minimization in water (Table 2). The use of a larger cutoff for nonbonded interactions in minimization (24 Å or 36 Å instead of 12 Å) slowed down the calculation without improving the regression, $\Delta G'_{bind}$ differing, on average, of only 0.6 kcal/mol in PBSA and 0.2 kcal/mol in GBSA (data not shown). Binding free energies estimated with Delphi were significantly less correlated to experiment ($r^2 = 0.61$, $s = 1.0$, $F = 40$) compared to simulations in water ($r^2 = 0.84$, $s = 0.66$, $F = 141$) or with structures minimized in GBSA ($r^2 = 0.71$, $s = 0.91$, $F = 63$). The trend suggests that Delphi is more sensible to the method used to generate structures, and performs better when structures are minimized in explicit water solvent. As a matter of fact, Delphi was developed and parameterized for reproducing free energies of structures simulated in explicit solvent. Indeed, the correlation between solvation free energies calculated on structures minimized with a distance-dependent dielectric function and those evaluated in water-minimized structures is rather low in Delphi ($r^2 = 0.53$) compared to Amber PBSA ($r^2 = 0.71$). Regarding Amber GBSA, estimated and experimental free energies of binding were still correlated ($r^2 = 0.73$, $s = 0.88$, $F = 69$), but the plot is rather scattered and the statistical significance of the regression is lower than that obtained with Amber PBSA (Fig. 1i vs h). Again, free energies of binding evaluated with GBSA gave the poorest relationship with experiment.

2.4. Free energies of binding using structures generated with a multistep minimization-molecular dynamics protocol employing a distance-dependent dielectric constant

The last experiment regards the set up and validation of a multistep protocol in which the complexes were minimized with distance-dielectric function, MD was performed on the ligand only, and the structures were re-minimized after MD. Compared to the last test case, this protocol adds a molecular dynamics stage for the refinement of the orientation and conformation of the ligand, and was specifically devised for relieving the problem of incorrect ligand conformation and/or orientation that may result from automated docking tools. In fact, while all inhibitors in our training set have orientations and conformations validated by crystallography, typical applications in virtual screening should deal with the fact that docking tools in some cases fail to predict the 'correct' orientation or conformation of the ligands in the binding site. Our experience showed that MD, in a number of cases, can solve these problems provided that the orientation of the ligand assigned by docking is not too different (e.g., head to tail) from the correct one (personal communication).

Table 3 includes the free energy results obtained after minimization/MD/re-minimization of the complexes with the distance-dependent dielectric function, and Figure 1j–l shows the regression plots. On average, the combined minimization–MD protocol was only 2.5 times slower than a simple minimization (Table 2). The absolute values of gas-phase interaction energies, solvation free energies, and estimated free energies of binding were very similar to those calculated after simple energy-minimization with distance-dependent dielectric function (Table 3). Free energies of binding estimated with Amber PBSA gave the best relationship with ΔG of binding ($r^2 = 0.80$, $s = 0.75$, $F = 105$, Fig. 1k), while relationships obtained with Delphi ($r^2 = 0.69$, $s = 0.94$, $F = 58$, Fig. 1j) and Amber GBSA ($r^2 = 0.73$, $s = 0.87$, $F = 71$, Fig. 1l) were less significant. Therefore, the results of the combined minimization/MD procedure were very similar to those obtained with simple minimization, at least in

our case in which the ligands already have correct initial orientations and conformations. However, we anticipate that the application of the intermediate MD refinement is crucial when docking tools fail to assign correct orientations and/or conformations of the ligand. In these cases, a simple minimization may not be enough to obtain correct geometries, as the energy barriers may be too high.

One last comment is worth for entropies of binding. In this study, the entropy contributions to binding have not been calculated because the calculation requires very long computing time (i.e., it is not applicable to virtual screening). Moreover, ΔS values obtained with normal mode analyses usually have large uncertainties.[18] Finally, it is still unclear how many snapshots of a protein–ligand complex are generally required to obtain converged entropy estimates. Surely, single structures like the ones analyzed here are not sufficient. Despite this, we have demonstrated that reasonable relationships between experimental and estimated free energies of binding can be achieved with fast minimization protocols and without the inclusion of entropy effects.

2.5. An automated procedure for refinement and rescore of docking results

An automated procedure for the refinement and rescore of virtual screening results was set up (Fig. 2). The procedure requires a pdb file containing the structure of the protein, and mol2 files containing the coordinates of the docked ligands. The coordinates of the docked ligand are merged with the protein to create the complex. Then, antechamber is used to create a topology file of the ligand in which atoms are described with gaff atom types and AM1-BCC charges. To make the procedure faster (charge calculation is time-consuming), atomic charges of the ligand are not computed during the procedure but read from the original mol2 file. This choice has the additional advantage that ligand charge calculations can be done only once, and used for any target protein. A separate script was written to calculate atomic partial charges of compounds in a database, using antechamber. Interestingly, the same set of AM1-BCC charges can also be used for automated ligand docking. Previous work showed that AM1-BCC charges perform well in automated docking.[50]

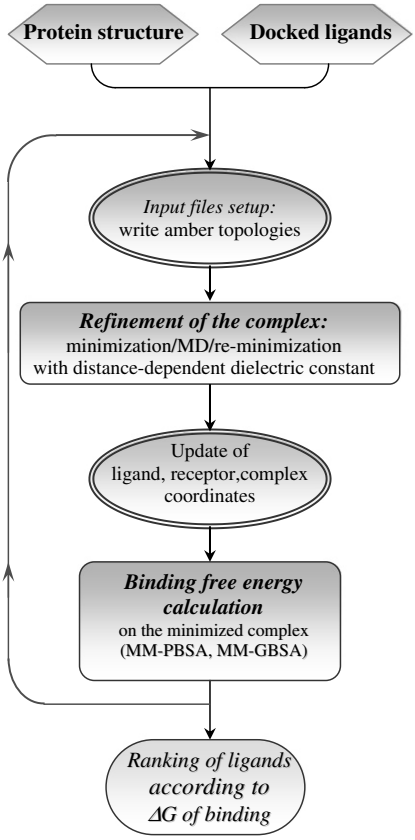


Figure 2. Flowchart of the automated procedure for the refinement of protein–ligand complexes and binding free energy evaluation.

Then, missing gaff force field parameters for the ligand are automatically assigned by parmcheck, and Amber topologies of ligand, receptor, and complex are written with leap (Amber 9). Minimization, MD, and final re-minimization of the complexes are performed with distance-dependent dielectric constant $\varepsilon = 4r$, using sander. For each of these steps, the procedure is highly 'flexible' in that it enables the user to set ad hoc refinement options (e.g., which residues are allowed to move during MD, the cutoff for nonbonded interactions, the number of cycles of minimization, etc.), depending on the application. After refinement of the complex, a pdb file is generated, the final coordinates of the ligand, receptor and complex are updated and used for binding free energy evaluation with Amber MM-PBSA and MM-GBSA, and the free energy results (ΔG_{MM}, ΔG_{solv} and $\Delta G'_{bind}$) are written to a file. When the procedure has finished analyzing one compound, the next ligand in the database is automatically prepared and processed (Fig. 2). In the procedure, big databases of compounds can be splitted in to several subsets, so that they can be run simultaneously on many CPUs. At the end of the screening, results are collected and ligands are ranked according to their estimated free energies of binding.

3. Conclusions

In this study, we have validated different simulation protocols for their ability in predicting free energies of binding of a set of 28 inhibitors of aldose reductase with known crystal structure. Compared to previously reported MM-PBSA applications,[12–19] this is, to our knowledge, the largest set of compounds used for predicting ligand affinities in silico with these methods. Different combinations of methods (Delphi, Amber PBSA, GBSA, Molsurf, LCPO), as well as different methods for generating the structures of the protein–ligand complexes (explicit water solvent, implicit solvent with GBSA minimization, or distance-dependent dielectric constant simulations), have been explored.

While MM-PBSA free energy predictions are usually performed on large collections of structures sampled

with MD,[12–14,16–18] we have shown that satisfactory relationships between experimental and computed free energies of binding can be obtained using single energy minimized structures. This finding was confirmed by another validation study performed on a set of DHFR inhibitors, in which free energy evaluations performed on large numbers of snapshots taken from 4 ns MD simulations in water were compared with those obtained in single minimized structures, and found to correlate similarly with the experimental ΔG of binding (manuscript in preparation). Therefore, our validation studies have important implications for virtual screening, where a good balance between calculation accuracy and speed is crucial when large databases of molecules have to be processed. These validations add to other recent studies in this direction, in which different simulation protocols are explored. In particular, Ryde et al.[18] suggest that if it is possible to identify and correct a few problematic cases with single minimized structures, this is undoubtedly the best choice as a compromise between accuracy and time consumption. Kuhn et al.[15] conclude that applying the MM-PBSA approach to a single, relaxed complex structure is an adequate and sometimes more accurate approach than the standard free energy averaging over molecular dynamics snapshots. Lyne et al.[19] found that MM-GBSA free energies of binding calculated after energy minimization of protein–ligand complexes have been successful in getting the correct relative rankings of a set of kinase inhibitors.

In our investigation, binding free energies estimated on protein–ligand complexes minimized with a distance-dependent dielectric constant $\varepsilon = 4r$ (the fastest method used to generate structures in this investigation) turned out to be significantly correlated with the experimental ΔG of binding. Interestingly, in the case of Amber PBSA, the correlation was very similar to that obtained using structures minimized in explicit solvent. Considering that Delphi, which is regarded as one of the most accurate methods for estimating solvation free energies, is significantly more cpu-consuming than Amber MM-PBSA and MM-GBSA, our results suggest that Amber PBSA and, to a lesser extent, GBSA, can be useful alternatives.

The results reported in this study are particularly encouraging considering that aldose reductase was a challenging test case for these applications. Firstly, the training set contained inhibitors that bind two distinct (closed and open) conformations of the enzyme, and we showed that the correlation with experimental binding free energies holds for both. Secondly, the complexes were generated starting from only one representative structure of the closed and one of the open conformation of ALR2, a choice that mimics typical docking screening applications in which a single or a limited number of different conformations of the same target are used. Finally, the enzyme binding site is formed by two adjacent pockets, one of which is highly polar and the other is hydrophobic. Since binding is driven by both electrostatic and hydrophobic forces, it is important that the balance between polar and nonpolar interactions is described appropriately and that both terms are calculated with sufficient accuracy.

A molecular dynamics stage for the refinement of the orientation and conformation of the ligands was introduced in the simulation, with the aim of refining geometries when automated docking tools fail to assign the correct conformation and/or orientation of the ligand. The combined minimization/MD protocol proposed here proved to give almost the same results of a simple minimization, in a case in which 28 ligands already have correct initial orientation and conformation. However, the application of the intermediate MD refinement becomes crucial when the geometries are not correct.

Finally, an automated procedure that integrates all the necessary steps required to generate structures and calculate free energies of binding of ligands in a database of compounds has been developed and tested. The results prospect that our procedure can be a useful tool for predicting relative binding free energies of structurally unrelated molecules in a given binding site, and holds promise as post-processing methods to more accurately score and rank hit compounds after docking screenings. The evaluation of the automated procedure in database screening is currently underway, and the results will be reported separately.

4. Experimental

4.1. ALR2 inhibitors

The training set of 28 ALR2 inhibitors used in this study is reported in Scheme 1. It has been selected according to three different criteria, that is, chemical diversity, significant variation in inhibitory activity, and availability of crystal structures. The set of inhibitors explores five orders of magnitude variation in activity,[25–34] corresponding to a ~7 kcal/mol difference in free energy of binding, and is therefore appropriate for establishing significant relationships between experimental and predicted free energies.[12] Secondly, the binding modes of these inhibitors were available from the crystal structures of the ALR2-inhibitor complexes, in some cases at ultra-high resolution (PDB codes 2INE, 2IQ0, 2IS7, 2INZ, 1AH0, 1X97, 1EL3, 1EKO, 1PWL, 1IEI, 1PWM, 1US0, 1T41, 1Z3N, 1MAR, 1Z8A).[27,31,33–43]

The training set is composed of three structurally unrelated classes of ALR2 inhibitors, with significant chemical diversity within each class (Scheme 1). Except for the sulfonyl-pyridazinone derivatives, that were discovered and reported more recently, the remaining molecules investigated in this study belong to the carboxylic acid and cyclic-imide classes of inhibitors, that constitute the two most important series of ARIs whose structure–activity relationships and binding features to ALR2 have been extensively investigated by crystallography and molecular modeling.[44–49] Carboxylic acid inhibitors bind at the substrate active site, the carboxylate functional group of the inhibitors being close to the nicotinamide C4 carbon of $NADP^+$ and giv-

ing hydrogen bonds with Tyr48, His110, and Trp111, three key residues for binding at the so-called 'anion hole' of ALR2. These residues and the positively charged nicotinamide ring of the oxidized cofactor form a positively charged binding site that recognizes and binds negatively charged inhibitors. Cyclic-imide inhibitors bind at the active site within hydrogen bonding distances to Tyr48, His110, and Trp111, and resemble the interactions described for carboxylates. Experimental evidences including ultra-high resolution crystallography confirmed that cyclic-imide inhibitors bind ALR2 in their anionic form arising from proton dissociation of the slightly acidic hydantoyn nitrogen. Sulfonylpyridazinone derivatives bind in a negatively charged state arising from deprotonation of the N2 nitrogen of the pyridazinone ring, recognized as an essential requisite for the inhibitor to assume the observed orientation.[42] Another interesting feature revealed by crystallography is the opening of a secondary hydrophobic pocket, lined by residues Trp111 and Leu300, which becomes accessible to inhibitors having suitable hydrophobic–aromatic substituents.[43–45]

According to these evidences, the 28 inhibitors were modeled in their deprotonated, negatively charged states, and partial atomic charges on atoms were calculated with the AM1-BCC[51] method implemented in the antechamber module of Amber 9.[22,52] Atom types and missing force-field parameters of the ligands were assigned based on the General Amber force field (gaff).[53]

4.2. Generation and refinement of the ALR2-inhibitor complexes with MM

The plenty of information available from ALR2 crystal structures suggests that two different conformations of ALR2, the so-called 'closed' and 'open' conformations, are usually explored depending on the type of inhibitor. Inhibitors having suitable hydrophobic–aromatic substituents generally induce the opening of the additional hydrophobic pocket, while those devoid of such substituents bind to the closed conformation.[43–45] Instead of generating the minimized ALR2-inhibitor complexes starting from the different crystal structures available, we chose to generate the complexes starting from one representative ALR2 structure with the open conformation and one representative with the closed conformation. This choice was motivated by our purpose to develop and test a procedure applicable to virtual screening, which is usually performed on a single or a limited number of protein conformations of the same target. It should be noted, however, that the various ALR2 crystal structures with the closed conformation are generally rather superimposable, and the same holds within structures in the open conformation. Therefore, the ALR2-inhibitor complexes of the ligands that bind to the open conformation (**6**, **8**–**11**, **13**–**16**, **18**–**28**) were generated starting from the crystal structure of human ALR2 in complex with inhibitor **23** (IDD552)[40] at 1.05 Å resolution (PDB entry lT41), and for those that bind to the closed conformation (**1**–**5**, **7**, **12**, **17**), the crystal structure of human ALR2 in complex with inhibitor **17** (fidarestat)[38] at 0.92 Å resolution (PDB entry 1PWM) was used. Each inhibitor was manually docked into the active site of 1T41 or 1PWM by superimposing its corresponding ALR2-inhibitor crystal structure to 1T41 or 1PWM, and the coordinates of the docked ligands were saved. For the inhibitors in Scheme 1 whose crystal structures in complex with ALR2 were not available, but that differed only for the introduction or replacement of small substituents, the initial orientation of the ligand was assumed to be the same of the closest member in the series with known crystal structure.

Hydrogens were added to the complexes using the internal coordinates of the Amber all-atom database. All Lys and Arg residues were positively charged and Glu and Asp residues negatively charged. All water molecules in the crystal structures were removed except for three conserved waters embedded into the protein, and interacting with His110, Thr19, Asn50, His83, and nearby backbone atoms. These water molecules were maintained for generating and minimizing the structures, but removed for free energy calculations. The parameters of the cofactor, $NADP^+$, were taken from our previous simulations.[47,48] All calculations in this study were performed with Amber 9, the ff03 force field[54] for the protein, and the gaff[53] force field for the ligands.

To refine the structures to be used for free energy analysis, energy minimizations of the entire protein–ligand complexes were performed in explicit or implicit water solvent models, using the sander module of Amber 9. Regarding simulations in explicit solvent, each complex prepared as described above was solvated in an octahedral box of TIP3P[55] water molecules extending 5 Å outside the protein on all sides, resulting in an average of about 4420 waters. The solvated complexes were minimized with 2000 steps of conjugate-gradient minimization without restraints, employing a residue-based cutoff of 8 Å. Two thousand steps of minimization proved to be sufficient to obtain converged free energy results in explicit and implicit water models; longer minimizations (e.g., 5000 steps) resulted in similar free energies of binding. The electrostatics were treated with the Particle-Mesh Ewald (PME)[56,57] method with a grid size of 72^3 Å, and a fourth order B-spline interpolation with a tolerance of 10^{-5}. Simulations in implicit solvent were performed with GBSA. Two thousand steps of conjugate-gradient minimization were performed with the Tsui and Case parameters (igb = 1, viz GB^{HCT} model in the original paper[58]), surface area was computed and included in the solvation term, and a cutoff of 18 Å for nonbonded interactions was used. In addition, minimization was also performed with a distance-dependent dielectric constant $\varepsilon = 4r$, with 2000 steps and a cutoff of 12 Å. Finally, a protocol consisting of 2000 steps of minimization on the entire complex, followed by 100 ps molecular dynamics in which only the ligand was allowed to move, plus another final minimization of the complex (2000 steps), was set up using distance-dependent dielectric conditions. Compared to energy minimization, this protocol adds a molecular dynamics stage for the refinement of the orientation and conformation of the ligand, and was specifically devised for relieving possibly incorrect assignment of ligand confor-

mation and/or orientation resulting from automated docking tools used in virtual screening studies. MD on the ligand was performed at 300 K, with SHAKE[59] turned on for bonds involving hydrogens, allowing a time-step of 2.0 fs.

After each energy minimization, visual inspection of the complexes was performed to make sure that the protein and the ligand remained close to conformation observed in the crystal structures.

4.3. MM-PBSA and MM-GBSA analyses

The MM-PBSA and MM-GBSA methods were used to calculate the binding free energies of the 28 inhibitors from their minimized ALR2 complexes.

The polar solvation free energies were calculated by solving the Poisson–Boltzmann (PB) equation either with Delphi[20,21] or the Amber PBSA module,[22] and with the generalized Born (GB) approach implemented in Amber 9.[60,61]

In Delphi, the grid spacing was set to 0.5 Å, and dielectric constants of 1 and 80 were used for the interior and exterior of the molecule, respectively. A cubic lattice was used, and the largest dimension of the cubic lattice was 80% filled by the longest dimension of the molecule. The dielectric boundary was defined using a 1.4 Å probe water on the atomic surface. The PB equation was solved using 1000 linear steps of finite difference. The Parse[62] set of radii were used for atoms of the molecule plus radii of 1.75, 1.948, and 2.22 Å, respectively, for F, Cl, and Br which are the radii for these halogens in the Amber ff03 force field. In Delphi calculations, we used the same atomic charges used to minimize the complexes, that is, the standard Amber charges (ff03)[54] for the protein and the AM1-BCC[51] charges for the cofactor and the ligands. In Amber PBSA, we used the Poisson–Boltzmann solver of Amber with the same parameters described for Delphi, with atomic cavity radii and atomic charges (ff03) taken from the Amber topology files. Finally, energy estimates with GBSA were made with the Tsui and Case parameters (igb = 1) and ff03 charges, with the same values of dielectric constants used for PB.

The hydrophobic contribution (ΔG_{npsolv}) to the solvation free energy was determined with solvent accessible surface area (SASA)[23] dependent methods using the equation $G_{\text{npsolv}} = \gamma\,\text{SASA} + b$. To compare the results, we have used both Molsurf[23] and LCPO[24]. Parameters were $\gamma = 0.00542$ kcal mol^{-1} Å^{-2} and $b = 0.92$ kcal mol^{-1} to be used in combination with Delphi polar solvation energies and Parse radii, and $\gamma = 0.0072$ kcal mol^{-1} Å^{-2} and $b = 0$ kcal mol^{-1} to be used in combination with Amber PB or GB polar solvation energies. These parameters are well documented in the literature.[12–19]

To investigate the dependence of the various free energy contributions on the electrostatics of binding, calculations in explicit water were repeated with complexes neutralized with counterions. At this purpose, the complexes were neutralized with counterions (Na^+), using the additions utility of leap (Amber 9). Then, the structures of the complexes were minimized in water and free energy predictions were performed using the same procedure described above.

The entropy estimates are usually obtained with normal mode analyses of harmonic frequencies calculated at the molecular mechanics level. However, in this study the entropy contributions to binding have not been calculated because the calculation requires very long computing time (i.e., it is not applicable to virtual screening) and the ΔS values usually have large error bars.[18]

References and notes

1. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.
2. Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.* **2001**, *308*, 377.
3. Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J. Comput. Chem.* **1992**, *13*, 505.
4. Lorber, D. M.; Shoichet, B. K. *Protein Sci.* **1998**, *7*, 938.
5. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006**, *49*, 5912.
6. Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.
7. Beveridge, D. L.; Dicapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
8. Reddy, M. R.; Erion, M. D. *Curr. Pharm. Des.* **2005**, *11*, 283.
9. Lybrand, T.; McCammon, J. A.; Wipff, G. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 833.
10. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Jr. *Acc. Chem. Res.* **2000**, *33*, 889.
11. Wang, J.; Morin, P.; Wang, W.; Kollman, P. *J. Am. Chem. Soc.* **2001**, *123*, 5221.
12. Pearlman, D. A. *J. Med. Chem.* **2005**, *48*, 7796.
13. Rafi, S. B.; Cui, G.; Song, K.; Cheng, X.; Tonge, P. J.; Simmerling, C. *J. Med. Chem.* **2006**, *49*, 4574.
14. Adekoya, O. A.; Willassen, N. P.; Sylte, I. *J. Struct. Biol.* **2006**, *153*, 129.
15. Kuhn, B.; Gerber, P.; Schultz-Gasch, T.; Stahl, M. *J. Med. Chem.* **2005**, *48*, 4040.
16. Steinbrecher, T.; Case, D. A.; Labahn, A. *J. Med. Chem.* **2006**, *49*, 1837.
17. Page, C. S.; Bates, P. A. *J. Comput. Chem.* **2006**, *27*, 1990.
18. Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. *J. Med. Chem.* **2006**, *49*, 6596.
19. Lyne, P. D.; Lamb, M. L.; Saeh, J. C. *J. Med. Chem.* **2006**, *49*, 4805.
20. Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507.
21. Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
22. Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *16*, 1668.
23. Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548.
24. Weiser, J.; Shenkin, P. S. W.; Clark Still, W. K. *J. Comput. Chem.* **1999**, *20*, 217.

25. Yamaguchi, T.; Miura, K.; Usui, T.; Unno, R.; Matsumoto, Y.; Fukushima, M.; Mizuno, K.; Kondo, Y.; Baba, Y.; Furono, M. *Arzneim.-Forsch.* **1994**, *44*, 344.
26. Negoro, T.; Murata, M.; Ueda, S.; Fujitani, B.; Ono, Y.; Kuromiya, A.; Komiya, A.; Suzuki, K.; Matsumoto, J. *J. Med. Chem.* **1998**, *41*, 4118.
27. Rogniaux, H.; Van Dorsselaer, A.; Barth, P.; Biellmann, J. F.; Barbanton, J.; van Zandt, M.; Chevrier, B.; Howard, E.; Mitschler, A.; Potier, N.; Urzhumtseva, L.; Moras, D.; Podjarny, A. *J. Am. Soc. Mass. Spectrom.* **1999**, *10*, 635.
28. Van Zandt, M. C.; Sibley, E. O.; McCann, E. E.; Combs, K. J.; Flam, B.; Sawicki, D. R.; Sabetta, A.; Carrington, A.; Sredy, J.; Howard, E. *Bioorg. Med. Chem.* **2004**, *12*, 5661.
29. Mylari, B. L.; Armento, S. J.; Beebe, D. A.; Conn, E. L.; Coutcher, J. B.; Dina, M. S.; O'Gorman, M. T.; Linhares, M. C.; Martin, W. H.; Oates, P. J.; Tess, D. A.; Withbroe, G. J.; Zembrowski, W. J. *J. Med. Chem.* **2005**, *48*, 6326.
30. Mylari, B. L.; Larson, E. R.; Beyer, T. A.; Zembrowski, W. J.; Aldinger, C. E.; Dee, M. F.; Siegel, T. W.; Singleton, D. H. *J. Med. Chem.* **1991**, *34*, 108.
31. Van Zandt, M. C.; Jones, M. L.; Gunn, D. E.; Geraci, L. S.; Howard Jones, J.; Sawacki, D. R.; Sredy, J.; Jacot, J. L.; DiCioccio, A. T.; Petrova, T.; Mitschler, A.; Podjarny, A. D. *J. Med. Chem.* **2005**, *48*, 3141.
32. Barski, O. A.; Gabbay, K. H.; Grimshaw, C. E.; Bohren, K. M. *Biochemistry* **1995**, *34*, 11264.
33. Brownlee, J. M.; Carlson, E.; Milne, A. C.; Pape, E.; Harrison, D. H. *Bioorg. Chem.* **2006**, *34*, 424.
34. Kinoshita, T.; Miyake, H.; Fujii, T.; Takakura, S.; Goto, T. *Acta Crystallogr., Sect. D* **2002**, *58*, 622.
35. Urzhumtsev, A.; Tete-Favier, F.; Mitschler, A.; Barbanton, J.; Barth, P.; Urzhumtseva, L.; Biellmann, J. F.; Podjarny, A.; Moras, D. *Structure* **1997**, *5*, 601.
36. El-Kabbani, O. E.; Darmanin, C.; Oka, M.; Schulze-Briese, C.; Tomizaki, T.; Hazemann, I.; Mitschler, A.; Podjarni, A. *J. Med. Chem.* **2004**, *47*, 4530.
37. Calderone, V.; Chevrier, B.; Van Zandt, M.; Lamour, V.; Howard, E.; Poterszman, A.; Barth, P.; Mitschler, A.; Lu, J.; Dvornik, D. M.; Klebe, G.; Kraemer, O.; Moorman, A. R.; Moras, D.; Podjarny, A. *Acta Crystallogr., Sect. D* **2000**, *56*, 536.
38. El-Kabbani, O.; Darmanin, C.; Schneider, T. R.; Hazemann, I.; Ruiz, F.; Oka, M.; Joachimiak, A.; Schulze-Briese, C.; Tomizaki, T.; Mitschler, A.; Podjarny, A. *Proteins* **2004**, *55*, 805.
39. Howard, E. I.; Sanishvili, R.; Cachau, R. E.; Mitschler, A.; Chevrier, B.; Barth, P.; Lamour, V.; Van Zandt, M.; Sibley, E.; Bon, C.; Moras, D.; Schneider, T. R.; Joachimiak, A.; Podjarny, A. *Proteins: Struct. Funct. Genet.* **2004**, *55*, 792.
40. Ruiz, F.; Hazemann, I.; Mitschler, A.; Joachimiak, A.; Schneider, T.; Karplus, M.; Podjarny, A. *Acta Crystallogr., Sect. D* **2004**, *60*, 1347.
41. Wilson, D. K.; Tarle, I.; Petrash, J. M.; Quiocho, F. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9847.
42. Steuber, H.; Zentgraf, M.; Podjarni, A.; Heine, A.; Klebe, G. *J. Mol. Biol.* **2006**, *356*, 45.
43. El-Kabbani, O.; Ramsland, P.; Darmanin, C.; Chung, R. P. T.; Podjarny, A. *Proteins: Struct. Funct. Genet.* **2003**, *50*, 230.
44. Podjarny, A.; Cachau, R. E.; Schneider, T.; Van Zandt, T. M.; Joachimiak, A. *Cell. Mol. Life Sci.* **2004**, *61*, 763.
45. Costantino, L.; Rastelli, G.; Vianello, P.; Cignarella, G.; Barlocco, D. *Med. Res. Rev.* **1999**, *19*, 3.
46. Klebe, G.; Kramer, O.; Sotriffer, C. *Cell. Mol. Life Sci.* **2004**, *61*, 783.
47. Rastelli, G.; Vianello, P.; Barlocco, D.; Costantino, L.; Del Corso, A.; Mura, U. *Bioorg. Med. Chem. Lett.* **1997**, *14*, 1897.
48. Rastelli, G.; Costantino, L. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 641.
49. Costantino, L.; Rastelli, G.; Gamberini, M. C.; Giovannoni, M. P.; Dal Piaz, V.; Vianello, P.; Barlocco, D. *J. Med. Chem.* **1999**, *42*, 1894.
50. Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. *J. Mol. Biol.* **2002**, *322*, 339.
51. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132.
52. Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, 2006.
53. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollamn, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
54. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A. *J. Comput. Chem.* **2003**, *21*, 1999.
55. Åqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021.
56. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
57. Petersen, H. G. *J. Chem. Phys.* **1995**, *103*, 3668.
58. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.
59. Van Gusteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.
60. Feig, M.; Im, W.; Brooks, C. L., III *J. Chem. Phys.* **2004**, *120*, 903.
61. Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 265.
62. Sitkoff, D.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.