



Assessing the stability of free-energy perturbation calculations by performing variations in the method

Francesco Manzoni¹ · Ulf Ryde¹

Received: 20 December 2017 / Accepted: 6 March 2018 / Published online: 13 March 2018
© The Author(s) 2018

Abstract

We have calculated relative binding affinities for eight tetrafluorophenyl-triazole-thiogalactoside inhibitors of galectin-3 with the alchemical free-energy perturbation approach. We obtain a mean absolute deviation from experimental estimates of only 2–3 kJ/mol and a correlation coefficient (R^2) of 0.5–0.8 for seven relative affinities spanning a range of up to 11 kJ/mol. We also studied the effect of using different methods to calculate the charges of the inhibitor and different sizes of the perturbed group (the atoms that are described by soft-core potentials and are allowed to have differing coordinates). However, the various approaches gave rather similar results and it is not possible to point out one approach as consistently and significantly better than the others. Instead, we suggest that such small and reasonable variations in the computational method can be used to check how stable the calculated results are and to obtain a more accurate estimate of the uncertainty than if performing only one calculation with a single computational setup.

Keywords Free-energy perturbation · Galectin-3 · Independent simulations · Ligand-binding affinity · RESP charges · AM1-BCC charges

Introduction

Predicting the affinity of a small molecule to a biomacromolecule is one of the greatest challenges in computational chemistry [1, 2]. If such binding affinities could be accurately predicted for arbitrary drug candidates, significant parts of drug development could be done by computers. Consequently, many methods have been developed for this aim [1], e.g. docking [3], MM/PBSA (molecular mechanics combined with Poisson–Boltzmann and solvent-accessible surface-area solvation) [4, 5] and linear interaction energy methods [6]. However, the most accurate results are typically

obtained with free-energy perturbation (FEP) and other free-energy simulation methods [7, 8]. These are based on strict statistical mechanics theory and should in principle be limited only by the accuracy of the energy function employed and the sampling of the phase space.

Recent large-scale tests have shown that FEP calculations can often provide relative binding free energies with a mean absolute deviation (MAD) from experimental estimates of 4–6 kJ/mol [9–12]. However, the performance varies with the protein target and the type of ligands; in some cases, errors of over 20 kJ/mol are encountered. Likewise, the performance in prospective blind test (in which the experimental affinities are not known when the calculations are run), the performance is often worse, with root-mean-square deviations (RMSD) of 6–18 kJ/mol [13, 14].

It has repeatedly been pointed out that the results of FEP calculations, as well as other binding-affinity calculations, strongly depend on the starting conditions [15–18]. Therefore, the uncertainty of a single calculation, even if it is quite long, gives a too optimistic estimate of the precision of the results. To obtain a valid error estimate, instead a number of statistically independent simulations should be run [15–22]. Such independent calculations are typically generated by employing different random-number seeds when setting

This article is dedicated to the memory of Francesco Manzoni, who died from a heart attack on March 12th at the age of 27.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10822-018-0110-5>) contains supplementary material, which is available to authorized users.

✉ Ulf Ryde
Ulf.Ryde@teokem.lu.se

¹ Theoretical Chemistry, Department of Chemistry, Chemical Centre, Lund University, P. O. Box 124, 221 00 Lund, Sweden

up the starting velocities in the molecular dynamics (MD) simulations (these velocities are not known and therefore arbitrary, besides being taken from a Maxwell–Boltzmann distribution at the studied temperature). We have shown that the sampling can be further enhanced by employing other more or less arbitrary choices made during the setup of the simulations [20]. In particular, the solvation of the macromolecule in the simulation box is arbitrary, typically performed by overlaying an equilibrated box of solvent molecules and removing those that overlap with the solute. Therefore, equivalent independent simulations can be generated either by using different equilibrated solvent boxes or by simply rotating or translating the solute before the solvation [20]. This approach (together with different starting velocities), called solvent-induced independent trajectories, has been used in several of our studies [23–25].

However, there are many additional choices that the user needs to make during the setup of the simulation, e.g. selection of alternative conformation of disordered residues, the protonation of all residues with acid or base constants in the physiological range (in particular His residues), rotation of many groups, especially Asn, Gln, Ser, Thr and crystal-water molecules [20, 26]. These choices are of another type, because they are not completely arbitrary. Instead, there are typically better or worse choices and, in many cases, a single choice is correct. However, it may be very time-consuming to find the optimum choice in every case. Therefore, different users and software employ various heuristic approaches to make this assignment in a reasonable time, which naturally gives different solutions [27–31]. We have studied how these choices affect the results of the simulations and have shown that unless the affected site is close to the binding site of the ligand, the effect is minimal and therefore instead can be employed to enhance the sampling of the simulations [20].

In this paper, we examine another way to generate independent simulations. The user has to make additional choices when setting up the simulations, e.g. the force field, in particular the atomic partial charges, as well as the details of the FEP calculations. In the AMBER force fields [32], the partial charges are normally obtained by a restrained fit to the electrostatic potential (RESP) [33] around the molecule, calculated with quantum mechanical (QM) methods, typically at the Hartree–Fock/6–31G* level of theory. However, a cheaper method has also been developed, based on semi-empirical calculations and bond charge corrections (AM1-BCC) [34]. The two sets of charges have been parametrised to give similar results [35, 36]. In addition, it is not fully specified how the molecule of interest should be treated before the charges are calculated. Often, it is optimised to give bonds and angles compatible with the QM method, but for elongated polar molecules, the molecule may curl up to make as many internal hydrogen bonds as possible. In this

paper, we compare results obtained with RESP and AM1-BCC charges and optimised with either with Hartree–Fock or AM1. Moreover, we compare two selections of the perturbed group, viz. either only the atoms that differ between the two ligands or the whole chemical group involved (a tetrafluorophenyl ring). The perturbed group affects both which atoms are assigned soft-core potentials and which atoms are allowed to have differing coordinates [37]. We first show that the six different FEP schemes employed give nearly identical results, without any clear trends indicating that any of the approaches is better than the others, and therefore can be combined to provide enhanced sampling by variations in the computational method.

As the test case, we employ the binding of eight substituted tetrafluorophenyl-triazole-thiogalactosides to the carbohydrate-recognition domain of galectin-3 [38]. Galectin-3 is a mammalian β -galactoside binding protein involved in glycoprotein trafficking, signalling, cell adhesion, angiogenesis, macrophage activation and apoptosis [39–43]. It has been implicated in inflammation, immunity, cancer development and metastasis [44]. Previous FEP studies of this protein have given quite large errors in the calculated relative binding affinities [45], but in this study, all calculations give excellent results, with MADs of 2–3 kJ/mol.

Computational methods

Molecular dynamics simulations

The molecular dynamics (MD) and FEP calculations were based on the X-ray crystal structure of a related compound bound to galectin-3C (PDB id 5E89) [46]. The various ligands were built by manually modifying this ligand. All crystal-water molecules were kept in the simulations. Each galectin-3 complex was solvated in an octahedral box of water molecules extending at least 10 Å from the solute using the tleap module of the Amber software [32], so that 4965–5593 water molecules were included in the simulations. The simulations were set up in the same way as in our previous studies of galectin-3 [26, 31, 47]: All Glu and Asp residues were assumed to be negatively charged and all Lys and Arg residues positively charged, whereas the other residues were neutral. His-158 and 217 were protonated on the ND1 atoms, whereas and the other two His residues were protonated on the NE2 atom, in accordance with NMR measurements and previous extensive test calculations with MD [31]. This gave a net charge of +4 for the protein. No counter ions were used in the simulations. The protein was described by the Amber 99SB force field [48], water molecules with the TIP3P force field [49], whereas the ligands were treated with the general Amber force field [50].

Three different sets of charges were employed for the ligands: They were obtained either with the RESP method [33], using electrostatic potentials calculated with the Hartree–Fock/6–31G* approach or with the cheaper AM1–BCC approach [34, 35]. In the former case, geometries were first optimised at the Hartree–Fock/6–31G level of theory, giving rise to the RH charge set. In the latter case, we either used the same geometries (BH charges) or geometries optimised with the semiempirical AM1 method (BA charges) [51]. The various charges are shown in Table S2 in the supplementary material. The BH and BA charges are quite similar with a mean absolute deviation (MAD) of 0.007 e, whereas the RH charges are more different with a MAD of 0.11 e to the other two charge sets.

The MD simulations were performed with the sander module of Amber 14 [32]. The temperature was kept constant at 300 K using a Langevin thermostat [52] with a collision frequency of 2.0 ps^{-1} , and the pressure was kept constant at 1 atm using a weak-coupling isotropic algorithm [53] with a relaxation time of 1 ps. Particle-mesh Ewald summation [54] with a fourth-order B spline interpolation and a tolerance of 10^{-5} was used to calculate electrostatic energies and forces. The cut-off for the Lennard–Jones interactions was set to 8 \AA and the non-bonded pair list was updated every 50 fs. The SHAKE algorithm [55] was used to constrain bonds involving hydrogen atoms so that a 2 fs time step could be used.

Free-energy perturbations

We have calculated the relative free energy of the eight inhibitors for galectin-3 in Fig. 1. They are all 3-(4-(2,3,5,6-tetrafluorophenyl)-1,2,3-triazol-1-yl)-thiogalactosides with a varying substituent in the para-position on the tetrafluorophenyl group. They will simply be named after this varying substituent, F, OH, OMe, OEt, NH_2 , NHMe, NMe_2 , or pyrrolidine (Pyr). These inhibitors were connected by seven transformations, keeping the chemical differences as small as possible: $\text{OH} \rightarrow \text{F}$, $\text{OMe} \rightarrow \text{OH}$, $\text{OEt} \rightarrow \text{OMe}$, $\text{NHMe} \rightarrow \text{OMe}$, $\text{NMe}_2 \rightarrow \text{NHMe}$, $\text{NMe}_2 \rightarrow \text{NH}_2$ and $\text{Pyr} \rightarrow \text{F}$, as is shown in Fig. 1b.

Relative binding free energies ($\Delta\Delta G_{\text{bind}}$) were calculated for these transformations with FEP using a thermodynamic cycle that involves the conversion of one ligand (L_1) to the other (L_2) both in the protein binding site and in solution [25, 56]. The free energies of the transformations were calculated using the multi-state Bennett acceptance ratio (MBAR) approach, calculated with the PYMBAR software [57]. We employed the single transformation-approach [25, 58] in which both electrostatic and van der Waals interactions are modified in the same step. The calculations employed soft-core versions of both the van der Waals and Coulomb potentials [59, 60].

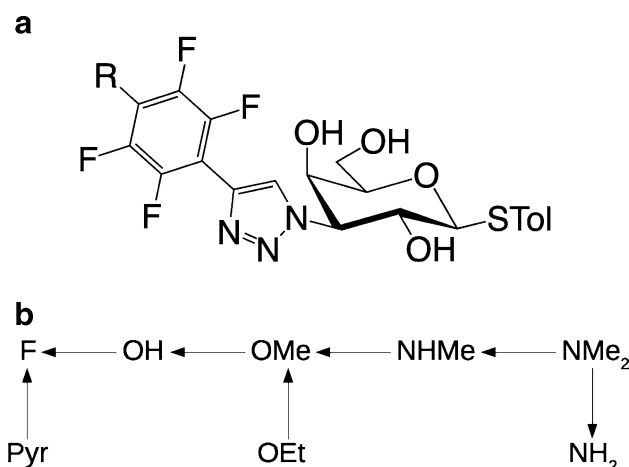


Fig. 1 **a** The eight studied galectin-3 inhibitors with $R = \text{F}$, OH , OMe , OEt , NH_2 , NHMe , NMe_2 , or pyrrolidine. STol is *S*-para-toluene. **b** The seven transformations studied by FEP

To improve the convergence of the free-energy difference, the transformation $L_1 \rightarrow L_2$ was divided into several small steps, involving intermediate states, defined by the potential energy $V(\lambda) = (1 - \lambda)V_0 + \lambda V_1$, where V_0 and V_1 are the potential energies of the L_1 and L_2 states, respectively. Thirteen λ values were used (0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 and 1). In all calculations, we used a single-topology scheme and the pmemd module of the Amber software with two sets of coordinates for the atoms that differ between L_1 and L_2 [37].

In the FEP calculations, we included in the perturbed group either only atoms directly involved in the perturbation (SP, small perturbed group), i.e. those in the para substituent, or all atoms of the terminal substituted tetrafluorophenyl group (LP, large perturbed group). In the present implementation of FEP [37], the perturbed group not only defines what atoms have a varying force field (including the soft-core Lennard–Jones and Coulomb potential), but also the atoms that are allowed to have different coordinates in the simulations.

The FEP simulations were performed in the following way: The system at each λ value was minimized for 500 cycles, with all atoms except water molecules and hydrogen atoms restrained to their start position with a force constant of 418 kJ/mol/\AA^2 . This was followed by a 20 ps constant-pressure simulation, using the same constraints, and a 1 ns constant-pressure simulation without any restraints. Finally, a 5 ns constant-pressure production run was performed, during which coordinates and energies were sampled every 10 ps. In total, six sets of calculations were performed, which will be denoted by the charges employed for the ligands (RH, BA or BH) and the size of the perturbed group (SP or LP), e.g. BA/LP.

Error estimates, quality and overlap measures

All reported uncertainties are standard errors of the mean (standard deviations divided by the square root of the number of samples). The uncertainty of the MBAR free energies calculated at each λ value was estimated by bootstrapping using the PYMBAR software [57] and the total uncertainty was obtained by error propagation.

The performance of the free-energy estimates was quantified by the mean absolute deviation (MAD), the mean signed deviation (MSD), the root-mean-square deviation (RMSD), the maximum error (Max), the correlation coefficient (R^2) and Kendall's rank correlation coefficient (τ_r) compared to experimental data [38]. The latter was calculated only for the transformations that were explicitly studied, not for all combinations that can be formed from these transformations. Moreover, it was also evaluated considering only differences (both experimental and calculated) that are statistically significant at the 90% level ($\tau_{r,90}$) [37]. Note that R^2 depends on the direction of the perturbation (i.e. whether $L_1 \rightarrow L_2$ or $L_2 \rightarrow L_1$ is considered, which is arbitrary). This was solved by considering both directions (both forward and backward) for all perturbations when these two measures were calculated. The standard deviation of the quality measures was obtained by a simple simulation approach [17]. For each transformation, 1000 Gaussian-distributed random numbers were generated with the mean and standard deviation equal to the MBAR and experimental results for that transformation. Then, the quality measures were calculated for each of these 1000 sets of simulated results and the standard error over the 1000 sets is reported as the uncertainty.

For all λ values of all perturbations, we have monitored five overlap measures, to ensure that the overlap of the studied distributions is satisfactory, viz. the Bhattacharyya coefficient Ω [61], the Wu and Kofke overlap measures of the energy probability distributions (K_{AB}) [62] and their bias metrics (Π) [62], the weight of the maximum term in the exponential average (w_{\max}) [63] and the difference of the forward and backward exponential average estimate ($\Delta\Delta G_{EA}$) [10]. In all calculations, $\Pi > 0.4$ and the other overlap measures were also in the safe range [10, 64], except in a single case for the OMe \rightarrow OH perturbation with BH/LP ($w_{\max} = 0.7$). Therefore, we conclude that the overlap is satisfactory for the simulations.

Result and discussion

We have studied the binding affinity of the eight substituted tetrafluorophenyl-triazole-thiogalactoside inhibitors of galectin-3, shown in Fig. 1a. Relative binding free energies were calculated for seven pairs of ligands, as is illustrated in Fig. 1b. The affinities were calculated by FEP with the MBAR approach. They are compared to experimental affinities obtained by competitive fluorescence polarization measurements (Table 1) [38, 65, 66].

Six different sets of FEP calculations were performed to see how the results changed with variations in the computational method. First, three different sets of charges were employed for the ligands: They were obtained either with the RESP method, based on Hartree–Fock/6–31G* calculations, or with the cheaper AM1-BCC approach. In the former case, geometries were first optimised at the Hartree–Fock/6–31G

Table 1 Calculated relative binding free energies (kJ/mol), obtained with three different sets of charges for the ligands (RH, BA and BH) and two perturbed groups (SP or LP)

	RH/SP	RH/LP	BA/SP	BA/LP	BH/SP	BH/LP	Consensus	Exp.
OMe \rightarrow OH	0.2 \pm 0.4	1.0 \pm 0.7	-0.8 \pm 0.4	1.6 \pm 0.6	-0.6 \pm 0.4	0.9 \pm 0.6	0.4 \pm 0.4	0.6 \pm 0.3
NHMe \rightarrow OMe	-0.8 \pm 0.3	-3.3 \pm 0.6	-6.0 \pm 0.4	-4.8 \pm 0.6	-6.7 \pm 0.4	-6.4 \pm 0.6	-4.7 \pm 0.9	0.0 \pm 0.3
NMe ₂ \rightarrow NHMe	-5.8 \pm 0.5	-4.4 \pm 0.7	-1.6 \pm 0.5	-1.1 \pm 0.6	-3.2 \pm 0.5	-2.2 \pm 0.7	-3.0 \pm 0.7	-2.0 \pm 0.2
NMe ₂ \rightarrow NH ₂	-1.7 \pm 0.5	-2.9 \pm 0.7	-3.9 \pm 0.5	-3.6 \pm 0.7	-2.9 \pm 0.5	-5.1 \pm 0.7	-3.3 \pm 0.5	-3.2 \pm 0.2
OEt \rightarrow OMe	-2.8 \pm 0.4	2.7 \pm 0.7	-4.2 \pm 0.4	-1.4 \pm 0.6	-3.3 \pm 0.4	-2.7 \pm 0.6	-1.9 \pm 1.0	-4.0 \pm 0.4
Pyr \rightarrow F	-10.4 \pm 0.6	-7.5 \pm 0.8	-10.4 \pm 0.6	-9.1 \pm 0.7	-9.0 \pm 0.6	-8.3 \pm 0.7	-9.1 \pm 0.5	-11.2
OH \rightarrow F	-0.4 \pm 0.2	-2.1 \pm 0.6	1.7 \pm 0.2	-1.0 \pm 0.5	1.0 \pm 0.2	-0.4 \pm 0.6	-0.2 \pm 0.6	-4.8 \pm 0.2
MAD	1.8 \pm 0.2	2.8 \pm 0.3	2.3 \pm 0.2	2.2 \pm 0.3	2.6 \pm 0.2	2.5 \pm 0.3	2.1 \pm 0.3	
RMSD	2.3 \pm 0.2	3.4 \pm 0.3	3.4 \pm 0.2	2.7 \pm 0.3	3.5 \pm 0.2	3.3 \pm 0.3	2.7 \pm 0.3	
MSD	0.4 \pm 0.2	1.1 \pm 0.3	-0.1 \pm 0.3	0.7 \pm 0.3	0.0 \pm 0.3	0.0 \pm 0.3	0.4 \pm 0.3	
Max	4.3 \pm 0.3	6.6 \pm 0.8	6.4 \pm 0.3	4.8 \pm 0.6	6.7 \pm 0.4	6.4 \pm 0.7	4.7 \pm 0.7	
R^2	0.79 \pm 0.03	0.54 \pm 0.07	0.61 \pm 0.04	0.71 \pm 0.06	0.55 \pm 0.04	0.60 \pm 0.06	0.71 \pm 0.06	
τ_r	1.00 \pm 0.16	0.67 \pm 0.10	0.33 \pm 0.08	1.00 \pm 0.13	0.33 \pm 0.10	1.00 \pm 0.18	1.00 \pm 0.22	
τ_{r90}	1.00 \pm 0.04	0.60 \pm 0.02	0.33 \pm 0.08	1.00 \pm 0.13	0.60 \pm 0.04	1.00 \pm 0.00	1.00 \pm 0.08	

Experimental relative affinities are given in the last column [38]

level of theory (RH charge set). In the latter case, we either used the same geometries (BH) or geometries optimised with the semiempirical AM1 method (BA). Moreover, in the FEP calculations, we included in the perturbed group either only atoms directly involved in the perturbation (SP), i.e. those in the para substituent or all atoms of the terminal substituted tetrafluorophenyl group (LP).

The results ($\Delta\Delta G_{\text{bind}}$) of all calculations are shown in Table 1. This table also contains seven quality estimates, viz. the mean absolute deviation (MAD), the root-mean-square deviation (RMSD), the mean signed deviation (MSD), the maximum error (Max), the correlation coefficient (R^2), Kendall's rank correlation coefficient including only the relative energies considered (τ_r), as well as the same correlation coefficient calculated only for those experimental and calculated energies that are statistically significant at the 90% level (τ_{r90}).

The differences between the $\Delta\Delta G_{\text{bind}}$ results obtained with the small and large perturbed groups (SP and LP) are up to 5 kJ/mol for RH and 2–3 kJ/mol for BA and BH, with MADs of 1–2 kJ/mol. The largest difference is for the OEt \rightarrow OMe perturbation for both the RH and BA charges. However, owing to the good precision of the simulations (0.2–0.8 kJ/mol), the differences are statistically significant for four (RH) or three of the calculations. Consequently, two (BH) to five (RH) of the quality estimates are also significantly different between the calculations with different perturbed groups. For the RH and BH charges, the SP calculations give the better results, whereas for the BA charges, the opposite is true. Therefore, it is hard to draw any firm conclusions from this variation. Apparently, there are at least two opposing effects for variations in the perturbed group. A larger perturbed group allows for a larger movement of atoms in the ligand, which may lead to improved results if the two groups actually have a different geometry. However, this larger variation in the coordinates also introduces more freedom in the system that may give rise to more random noise. The latter is reflected by a 0.2–0.3 kJ/mol higher uncertainty in all LP results, compared to the SP results. The difference is somewhat larger for RH than for the other two charge sets.

The BA and BH charge sets give very similar results, with differences in the individual $\Delta\Delta G_{\text{bind}}$ values of less than 1.6 kJ/mol (0.9–1.1 kJ/mol MAD differences). In fact, only one of the 14 values has a statistically significant difference at the 95% level. Likewise, only one of the quality estimates gives a significant difference. On the other hand, the RH charges are more different, giving rise to differences in the $\Delta\Delta G_{\text{bind}}$ results of up to 6 kJ/mol with MADs of 2 kJ/mol. Two to five of the individual results are significantly different for the four series. This leads to significant differences in nearly all of the quality estimates with the small perturbed group (SP), the RH charges always giving the better results.

However, with the large perturbed group (LP), the larger uncertainty in the individual results make the differences in the quality measures statistically significant only in three cases, involving τ_r and τ_{r90} . In this case, RH/LP always give the worse results.

Thus, again, it seems hard to draw any firm conclusions about the performance of the three charge sets—in fact RH/SP gives the best results among all calculations, whereas RH/LP gives the worse results, perhaps together with BH/SP. It seems that the RESP charges give strongly varying results, simply because they are larger in magnitude than the BCC charges. This prohibits any firm recommendation of any of the charge or perturbed-group methods. Instead, the observed variations in the results and performance of the six approaches have to be seen as a measurement of the sensitivity of the results to various quite reasonable variations in the theoretical method. Quite satisfying, the variation in the quality estimates is not extensive, e.g. 2–3 kJ/mol in the MAD and 0.5–0.8 in R^2 . Thus, a study like this can provide an estimate of the stability of the results, in addition to the precision, overlap measures and sampling estimates and therefore provide another approach to generate independent simulations [20] to better estimate the true uncertainty of the calculations.

The calculated $\Delta\Delta G_{\text{bind}}$ values are compared to the experimental results in Fig. 2. It can be seen that for five of the transformations, the range of the six approaches overlap with the ideal correlation, and for the remaining two, at least one of the results come within 1 kJ/mol of the experimental results. A natural choice is to test the consensus (i.e. average) results of the six methods with long simulations. These results are also included in Table 1 (with standard errors calculated from the variation of the results over the six methods). It can be seen that it comes close to the best results for all quality measures, e.g. $\text{MAD} = 2.1 \pm 0.3$, $R^2 = 0.71 \pm 0.06$ and $\tau_r = \tau_{r90} = 1.0 \pm 0.2$.

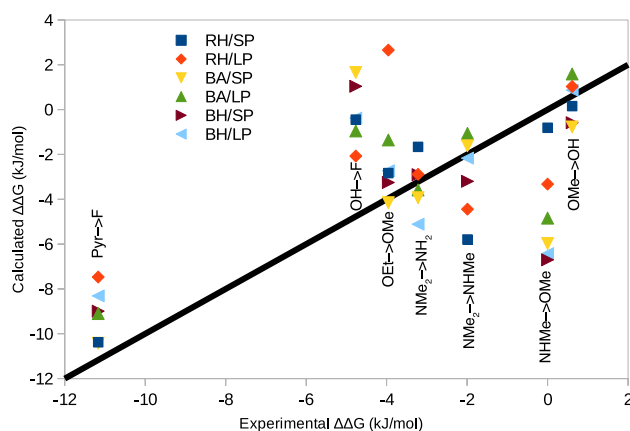


Fig. 2 Results of the FEP calculations, compared to the experimental data. The black line represents the perfect correlation

Taking all results together, it can be seen that $MAD = 1.8\text{--}3.1$ kJ/mol. This is a quite typical result of FEP for a well-behaving protein; previous large-scale studies have given MADs of 3–7 kJ/mol [9–12]. The RMSDs are slightly larger, 2.3–3.5 kJ/mol and the maximum error is 4–7 kJ/mol. Although these errors are small in energy terms, the MAD and maximum errors correspond to a factor of 2.1–3.5 and 6–15 in the binding constant, which are quite large from an experimental point of view. R^2 is 0.54–0.79, i.e. of an intermediate quality. However, R^2 is most affected by the Pyr \rightarrow F transformation, which gives the largest experimental difference and it is underestimated in all calculations; without this perturbation, R^2 is lower, 0.12–0.38. The reason for this is that all the other six relative affinities are less than 5 kJ/mol (in absolute terms), making it a formidable task to obtain an accurate correlation. On the other hand, Kendall's τ_r is perfect for half of the methods (RH/SP, BA/LP and BH/LP), showing that these calculations give the correct sign for all $\Delta\Delta G_{\text{bind}}$ estimates. The other three methods give one or two incorrect signs, but for BH/SP one of those is not statistically significant. This is arguably the most important quality measure, because during lead optimisation, the prime question is whether a new drug candidate will be better than the previous ones. Therefore, it seems that FEP may be a valuable approach to predict at least the sign of putative drug candidates of galectin-3.

Conclusions

We have studied the binding of the eight substituted tetrafluorophenyl-triazole-thiogalactoside inhibitors in Fig. 1a to galectin-3 with FEP calculations. The results show that we can reproduce experimental relative binding affinities with a MAD of 2–3 kJ/mol, which is similar or better than FEP results obtained for other proteins [9–12]. This is quite satisfying, especially as we have had problems to reproduce experimental affinities with FEP before for this protein (7 kJ/mol error for a single perturbation obtained with similar methods) [45]. The correlation coefficient is also acceptable, 0.5–0.8, especially considering that six of the seven relative energies are below 5 kJ/mol. Moreover, most approaches give a perfect Kendall's τ .

We have also studied how the calculated binding affinities change with small variations in the computational method, employing either RESP or AM1-BCC charges for the ligands, optimising it with two different QM methods before the charge calculation and employing different sizes of the perturbed group in the FEP calculations (only the differing atoms or the whole substituted tetrafluorophenyl group). We show that the variations give only rather small differences in the calculated relative affinities and that none of the approaches is consistently and significantly

better than the others. Instead, we suggest that such small, but reasonable variations in the theoretical method could be employed to assess the stability of the results obtained.

Recently, it has been argued that the results of a single FEP calculation of binding energies is quite uncertain and that the precision estimated by standard methods underestimates the true uncertainty [22]. Instead, the true uncertainty should be estimated by performing a number of independent simulations, using different starting conditions. This has also repeatedly been suggested for other types of calculations based on MD simulations [15–21].

Alternatively, thermodynamic cycle-closure hysteresis has been suggested as a cheaper estimate of the uncertainty of the calculations [67, 68]. However, it is a rather blunt estimate, because it can give a good result by chance or by cancellation of errors. To be useful, all ligands should be involved in at least one cycle, requiring almost the same number of extra calculations as a set of independent simulations. Moreover, they normally require larger perturbations, because the original scheme (without cycles) is typically designed to make the perturbations as small as possible. Moreover, the cycle-closure hysteresis is independent of errors in the force field and therefore ignores method errors considered by the present approach.

Traditionally, independent simulations have been generated by using different random seeds for the starting velocities. We have employed other arbitrary choices in the setup, in particular the position of solvation water molecules, but also the selection of alternative conformations or the protonation of the residues in the protein [20, 23–25]. It is a natural extension to include also variations in the computational method. In analogy with the naming of the other approaches to obtain independent simulations [20], we suggest that this should be called methods-induced independent trajectories (MIIT). However, as for the selection of conformations and protonations, it needs to be checked that the different approaches do not give rise to results that are significantly different. Still, it is important in computational studies to ensure that the results are not sensitive to reasonable variations in the computational setup.

Acknowledgements This investigation has been supported by grants from the Swedish research council (project 2014-5540), from Knut and Alice Wallenberg Foundation (KAW 2013.0022), and from the European Spallation Source consortium. The computations were performed on computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc at Lund University and HPC2N at Umeå University.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chemie Int Ed* 41:2644–2676
- Jorgensen WL (2009) Efficient drug lead discovery and optimization. *Acc Chem Res* 42:724–733
- Kontoyianni M, Madhav P, Seibel ES (2008) Theoretical and practical considerations in virtual screening: a beaten field?. *Curr Med Chem* 15:107–116. <https://doi.org/10.2174/092986708783330566>
- Kollman PA, Massova I, Reyes CM et al (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897
- Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10:449–461. <https://doi.org/10.1517/17460441.2015.1032936>
- Åqvist J, Luzhkov VB, Brandsdal BO (2002) Ligand binding affinities from MD simulations. *Acc Chem Res* 35:358–365
- Wereszczynski J, McCammon JA (2012) Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Quart Rev Biophys* 45:1–25. <https://doi.org/10.1017/S0033583511000096>
- Hansen N, Van Gunsteren WF (2014) Practical aspects of free-energy calculations: a review. *J Chem Theory Comput* 10:2632–2647. <https://doi.org/10.1021/ct500161f>
- Christ C, Fox T (2013) Accuracy assessment and automation of free energy calculations for drug design. *J Chem Inf Model* 54:108–120. <https://doi.org/10.1021/ci4004199>
- Mikulskis P, Genheden S, Ryde U (2014) A large-scale test of free-energy simulation estimates of protein–ligand binding affinities. *J Chem Inf Model* 54:2794–2806. <https://doi.org/10.1021/ci5004027>
- Wang L, Wu Y, Deng Y et al (2015) Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc* 137:2695–2703. <https://doi.org/10.1021/ja512751q>
- Harder E, Damm W, Maple J et al (2016) OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J Chem Theory Comput* 12:281–296. <https://doi.org/10.1021/acs.jctc.5b00864>
- Gathiaka S, Liu S, Chiu M et al (2016) D3R grand challenge 2015: evaluation of protein–ligand pose and affinity predictions 1 introduction. *J Comput Mater Des* 30:651–668
- Gaieb Z, Liu S, Gathiaka S et al (2017) D3R grand challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J Comput Aided Mol Des*. <https://doi.org/10.1007/s10822-017-0088-4>
- Elofsson A, Nilsson L (1993) How consistent are molecular dynamics simulations? Comparing structure and dynamics in reduced and oxidized *Escherichia coli* thioredoxin. *J Mol Biol* 233:766–780
- Caves LS, Evanseck JD, Karplus M (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci* 7:649–666. <https://doi.org/10.1002/pro.5560070314>
- Genheden S, Ryde U (2010) How to obtain statistically converged MM/GBSA results. *J Comput Chem* 31:837–846. <https://doi.org/10.1002/jcc.21366>
- Lawrenz M, Baron R, McCammon JA (2009) Independent-trajectories thermodynamic-integration free-energy changes for biomolecular systems: determinants of H5N1 avian influenza virus neuraminidase inhibition by peramivir. *J Chem Theory Comput* 5:1106–1116. <https://doi.org/10.1021/ct800559d>
- Zagrovc B, Van Gunsteren WF (2007) Computational analysis of the mechanism and thermodynamics of inhibition of phosphodiesterase 5A by synthetic ligands. *J Chem Theory Comput* 3:301–311. <https://doi.org/10.1021/ct600322d>
- Genheden S, Ryde U (2011) A comparison of different initialization protocols to obtain statistically independent molecular dynamics simulations. *J Comput Chem* 32:187–195. <https://doi.org/10.1002/jcc.21546>
- Sadiq SK, Wright DW, Kenway OA, Coveney PV (2010) Accurate ensemble molecular dynamics binding free energy ranking of multidrug-resistance HIV-1 protease. *J Chem Inf Model* 50:890–905
- Bhati AP, Wan S, Wright DW, Coveney PV (2017) Rapid, accurate, precise, and reliable relative free energy prediction using ensemble based thermodynamic integration. *J Chem Theory Comput* 13:210–222. <https://doi.org/10.1021/acs.jctc.6b00979>
- Olsson MA, Söderhjelm P, Ryde U (2016) Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level. *J Comput Chem* 37:1589–1600. <https://doi.org/10.1002/jcc.24375>
- Genheden S, Mikulskis P, Hu L, Kongsted J, Söderhjelm P, Ryde U (2011) Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *J Am Chem Soc* 133:13081–13092. <https://doi.org/10.1021/ja202972m>
- Genheden S, Nilsson I, Ryde U (2010) Binding affinities of factor Xa inhibitors estimated by thermodynamic integration and MM/GBSA. *J Chem Inf Model* 51:947–958. <https://doi.org/10.1021/ci100458f>
- Genheden S, Diehl C, Akke M, Ryde U (2010) Starting-condition dependence of order parameters derived from molecular dynamics simulations. *J Chem Theory Comput* 6:2176–2190. <https://doi.org/10.1021/ct900696z>
- Forrest LR, Honig B (2005) An assessment of the accuracy of methods for predicting hydrogen positions in protein structures. *Proteins Struct Funct Genet* 61:296–309. <https://doi.org/10.1002/prot.20601>
- Hoof RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381:272–272
- Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pK_a values. *Proteins Struct Funct Genet* 61:704–721. <https://doi.org/10.1002/prot.20660>
- Gordon JC, Myers JB, Folta T et al (2005) H++: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic Acids Res* 33:368–371. <https://doi.org/10.1093/nar/gki464>
- Uranga J, Mikulskis P, Genheden S, Ryde U (2012) Can the protonation state of histidine residues be determined from molecular dynamics simulations? *Comput Theor Chem* 1000:75–84. <https://doi.org/10.1016/j.comptc.2012.09.025>
- Case DA, Berryman JT, Betz RM et al (2014) AMBER 14. University of California, San Francisco
- Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 97:10269–10280. <https://doi.org/10.1021/j100142a004>
- Jakalian A, Bush BL, Jack DB, Bayly CI (2000) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J Comput Chem* 21:132–146
- Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 23:1623–1641. <https://doi.org/10.1002/jcc.10128>
- Mikulskis P, Cioloboc D, Andrejić M et al (2014) Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid

- host-guest binding energies. *J Comput Aided Mol Des* 28:375–400. <https://doi.org/10.1007/s10822-014-9739-x>
37. Kaus JW, Pierce LT, Walker RC, Mccammon JA (2013) Improving the efficiency of free energy calculations in the amber molecular dynamics package. *J Chem Theory Comput* 9:4131–4139
 38. Peterson K (2018) Molecular basis for galectin-ligand interactions. In: Design, synthesis and analysis of ligands. PhD Thesis. Lund University. <http://portal.research.lu.se/ws/files/38414625/Thesis.pdf>
 39. Leffler H, Carlsson S, Hedlund M et al (2002) Introduction to galectins. *Glycoconj J* 19:433–440. <https://doi.org/10.1023/B:GLYC.0000014072.34840.04>
 40. MacKinnon AC, Farnworth SL, Hodgkinson PS et al (2008) Regulation of alternative macrophage activation by galectin-3. *J Immunol* 180:2650–2658. <https://doi.org/10.4049/jimmunol.180.4.2650>
 41. Delacour D, Koch A, Jacob R (2009) The role of galectins in protein trafficking. *Traffic* 10:1405–1413. <https://doi.org/10.1111/j.1600-0854.2009.00960.x>
 42. Liu FT, Rabinovich GA (2010) Galectins: regulators of acute and chronic inflammation. *Ann N Y Acad Sci* 1183:158–182. <https://doi.org/10.1111/j.1749-6632.2009.05131.x>
 43. Grigorian A, Demetriou M (2010) Manipulating cell surface glycoproteins by targeting *n*-glycan–galectin interactions. In: Fukuda M (ed) *Glycobiology*. Academic Press, Cambridge, pp 245–266
 44. Rabinovich GA, Liu F-T, Hirashima M, Anderson A (2007) An emerging role for galectins in tuning the immune response: lessons from experimental models of inflammatory disease, autoimmunity and cancer. *Scand J Immunol* 66:143–158. <https://doi.org/10.1111/j.1365-3083.2007.01986.x>
 45. Genheden S, Ryde U, Söderhjelm P (2015) Binding affinities by alchemical perturbation using QM/MM with a large QM system and polarizable MM model. *J Comput Chem* 36:2114–2124. <https://doi.org/10.1002/jcc.24048>
 46. Delaine T, Collins P, MacKinnon A et al (2016) Galectin-3-binding glycomimetics that strongly reduce bleomycin-induced lung fibrosis and modulate intracellular glycan recognition. *ChemBioChem* 17:1759–1770. <https://doi.org/10.1002/cbic.201600285>
 47. Genheden S, Akke M, Ryde U (2014) Conformational entropies and order parameters: convergence, reproducibility, and transferability. *J Chem Theory Comput* 10:432–438. <https://doi.org/10.1021/ct400747s>
 48. Hornak V, Abel R, Okur A et al (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct Funct Bioinf* 65:712–725. <https://doi.org/10.1002/prot.21123>
 49. Jorgensen WL, Chandrasekhar J, Madura JD et al (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935. <https://doi.org/10.1063/1.445869>
 50. Wang JM, Wolf RM, Caldwell JW et al (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174. <https://doi.org/10.1002/jcc.20035>
 51. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) A new general purpose quantum mechanical molecular model. *J Am Chem Soc* 107:3902–3909
 52. Wu X, Brooks BR (2003) Self-guided Langevin dynamics simulation method. *Chem Phys Lett* 381:512–518. <https://doi.org/10.1016/j.cplett.2003.10.013>
 53. Berendsen HJC, Postma JPM, van Gunsteren WF et al (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690
 54. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092
 55. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J Comput Phys* 23:327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5)
 56. Gilson MK, Given JA, Bush BL, Mccammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J* 72(3):1047–1069
 57. Shirts MR, Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* 129:124105. <https://doi.org/10.1063/1.2978177>
 58. Genheden S, Ryde U (2012) Improving the efficiency of protein-ligand binding free-energy calculations by system truncation. *J Chem Theory Comput* 8:1449–1458. <https://doi.org/10.1021/ct200853g>
 59. Steinbrecher T, Mobley DL, Case DA (2007) Nonlinear scaling schemes for Lennard–Jones interactions in free energy calculations. *J Chem Phys* 127:1–13. <https://doi.org/10.1063/1.2799191>
 60. Steinbrecher T, Joung I, Case DA (2011) Soft-core potentials in thermodynamic integration: comparing one- and two-step transformations. *J Comput Chem* 32:3253–3263. <https://doi.org/10.1002/jcc.21909>
 61. Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distributions. *Bull Cal Math Soc* 35:99–109
 62. Wu D, Kofke DA (2005) Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *J Chem Phys* 123:1–10. <https://doi.org/10.1063/1.1992483>
 63. Rod TH, Ryde U (2005) Quantum mechanical free energy barrier for an enzymatic reaction. *Phys Rev Lett* 94:138302. <https://doi.org/10.1103/PhysRevLett.94.138302>
 64. Olsson MA, Ryde U (2017) Comparison of QM/MM methods to obtain ligand-binding free energies. *J Chem Theory Comput* 13:2245–2253. <https://doi.org/10.1021/acs.jctc.6b01217>
 65. Sörme P, Kahl-Knutsson B, Huflejt M et al (2004) Fluorescence polarization as an analytical tool to evaluate galectin–ligand interactions. *Anal Biochem* 334:36–47. <https://doi.org/10.1016/j.ab.2004.06.042>
 66. Cumpstey I, Carlsson S, Leffler H, Nilsson UJ (2005) Synthesis of a phenyl thio-β-d-galactopyranoside library from 1,5-difluoro-2,4-dinitrobenzene: discovery of efficient and selective monosaccharide inhibitors of galectin-7. *Org Biomol Chem* 3:1922–1932. <https://doi.org/10.1039/B502354H>
 67. De Graaf C, Oostenbrink C, Keizers PHJ et al (2007) Free energies of binding of R- and S-propranolol to wild-type and F483A mutant cytochrome P450 2D6 from molecular dynamics simulations. *Eur Biophys J* 36:589–599. <https://doi.org/10.1007/s00249-006-0126-y>
 68. Liu S, Wu Y, Lin T et al (2013) Lead optimization mapper: automating free energy calculations for lead optimization. *J Comput Aided Mol Des* 27:755–770. <https://doi.org/10.1007/s10822-013-9678-y>