Journal of Computer-Aided Molecular Design (2005) 19: 765–770
DOI 10.1007/s10822-005-9021-3

Can free energy calculations be fast and accurate at the same time? Binding of low-affinity, non-peptide inhibitors to the SH2 domain of the src protein

Christophe Chipot[a,*], Xavier Rozanska[a,b] & Surjit B. Dixit[a,c]
[a]*Equipe de dynamique des assemblages membranaires, UMR CNRS/UHP 7565, Institut nancéien de chimie moléculaire, Université Henri Poincaré, BP 239, 54506, Vandoeuvre-lès-Nancy cedex, France;* [b]*Institut für Chemie, Humboldt-Universität zu Berlin, Berlin, Germany;* [c]*Wesleyan University, Middletown, CT, 06459, USA*

Received 15 July 2005; accepted 11 October 2005

Key words: free-energy calculations, molecular dynamics simulations, rational drug design

Summary

The usefulness of free-energy calculations in non-academic environments, in general, and in the pharmaceutical industry, in particular, is a long-time debated issue, often considered from the angle of cost/performance criteria. In the context of the rational drug design of low-affinity, non-peptide inhibitors to the SH2 domain of the ^{pp60}src tyrosine kinase, the continuing difficulties encountered in an attempt to obtain accurate free-energy estimates are addressed. free-energy calculations can provide a convincing answer, assuming that two key-requirements are fulfilled: (i) thorough sampling of the configurational space is necessary to minimize the statistical error, hence raising the question: to which extent can we sacrifice the computational effort, yet without jeopardizing the precision of the free-energy calculation? (ii) the sensitivity of binding free-energies to the parameters utilized imposes an appropriate parametrization of the potential energy function, especially for non-peptide molecules that are usually poorly described by multipurpose macromolecular force fields. Employing the free-energy perturbation method, accurate ranking, within ± 0.7 kcal/mol, is obtained in the case of four non-peptide mimes of a sequence recognized by the ^{pp60}src SH2 domain.

Introduction

One of the grand challenges of free-energy calculations is to assume in the near future a predictive role in ranking ligands of potential pharmaceutical interest, agonist or antagonist, according to their relative affinity towards a given protein. The usefulness of such numerical simulations outside a pure academic environment can be appraised by answering the following question: Can free-energy calculations provide a convincing answer significantly faster than experiments are carried out in an industrial setting? Early encouraging results had triggered much excitement in the community, opening new vistas for *de novo*, rational drug design, but where do we stand today?

Beyond the fundamental need of appropriate sampling to yield converged ensemble averages – today, easily achievable in some favorable cases by means of inexpensive clusters of personal computers – the necessity of well-parameterized potential energy functions, suitable for non-peptide ligands, rapidly turns out to constitute a critical bottleneck for the routine use of free-energy calculations in the pharmaceutical industry. Closely related to the parametrization of the force field, setting up free-energy calculations – i.e. defining the alternate

*To whom correspondence should be addressed. E-mail: Christophe.Chipot@edam.uhp-nancy.fr

topology of the mutated moieties and the initial set of coordinates – is sufficiently time-consuming to be incompatible with the high-throughput requirements of industrial environments. In essence, this is where the paradox of free-energy calculations lies: They have not yet come of age to be considered as black box routine jobs, but must evolve in that direction to become part of the arsenal of computational tools available to the pharmaceutical industry modeler.

Selection of potent ligand candidates in large data bases containing up to several millions of real or virtual molecules, employing a screening funnel that involves, as we move downwards, increasingly complex searching tools ranging from crude geometrical recognition to more sophisticated flexible molecular docking, offers new prospects for *de novo* drug design. In this pipeline of screening methods, free-energy calculations should evidently be positioned at the very end, i.e. at the level of the optimization loop aimed at a limited number of ligands, which also checks adsorption–metabolism/toxicology (ADME/TOX) properties. Ideally, as selection in the funnel proceeds, the computational effort should remain constant – *viz*. the amount of CPU time necessary to perform free-energy calculations on a few candidates is equivalent to that involved in the rough geometrical recognition over a number of molecules five to six orders of magnitude larger.

Perhaps the key to the generalized and routine use of free-energy calculations for molecular systems of pharmacological relevance is the demonstration that this methodology can be applied fruitfully to problems that clearly go beyond the typical scope of an academic environment. Collaborative projects with the pharmaceutical industry provide such a framework. In the context of the search for therapeutic agents targeted at osteoporosis and other bone-related diseases, free-energy calculations have been applied to complexes formed by the multi-domain protein ^{pp60}src tyrosine kinase associated to non-peptide inhibitors. ^{pp60}src tyrosine kinase is involved in signal transduction pathways and is implicated in osteoclast-mediated bone resorption [1]. Of particular interest, its src homology domain 2 (SH2), a common recognition motif of highly conserved protein sequence, binds preferentially phosphotyrosine (pY)-containing peptides. In most circumstances, the latter adopt an extended conformation mimicking a two-pronged plug that interacts at two distinct anchoring sites of the protein – i.e. the hydrophilic phosphotyrosine pocket and the hydrophobic pocket – separated by a flat surface [2, 3]. For instance, the prototypical tetrapeptide pYEEI – a sequence found on the platelet-derived growth factor (PDGF) receptor upon activation, appears to recognize the ^{pp60}src SH2 domain with an appropriate specificity.

In the following section, the methodological background and the computational details of the free-energy calculations are presented. Next, the relative protein–ligand affinities for a series of inhibitors of the ^{pp60}src SH2 domain are reported, discussing the binding mode in these complexes. This contribution closes with an introspection on the usefulness of free-energy calculations in the pharmaceutical industry, and a glimpse into their prospective role in a screening funnel-like strategy.

Methodology

The free-energy changes corresponding to the transformation of reference inhibitor **I1** into alternate inhibitors **I2**, **I3** and **I4** were obtained from the thermodynamic cycle shown in Figure 1. These "alchemical transformations" were performed using the dual-topology paradigm [4], wherein the initial state and the final state are defined by means of distinct, non-interacting topologies. The free-energy perturbation (FEP) approach [5] was employed to scale the interactions of the transformed moieties with their environment by means of a linear parameter, λ:

$$\Delta A_{\lambda\rightarrow\lambda+\delta\lambda} = -\frac{1}{\beta}\ln\langle\exp\{-\beta[\mathscr{V}(\mathbf{x},\lambda+\delta\lambda)-\mathscr{V}(\mathbf{x};\lambda)]\}\rangle_{\lambda} \quad (1)$$

$\beta = 1/k_{\mathrm{B}}T$, where k_{B} is the Boltzmann constant, T is the temperature of the N-particle system, and $\langle\cdots\rangle_{\lambda}$ denotes an ensemble average over configurations representative of the reference state defined by λ. $\mathscr{V}(\mathbf{x};\lambda)$ is the potential energy function, that depends upon the set of Cartesian coordinates, $\{\mathbf{x}\}$, and λ. For each transformation shown in Figure 1, 200 intermediate states of uneven widths, $\delta\lambda$, were simulated for 2 ps, following 2 ps of

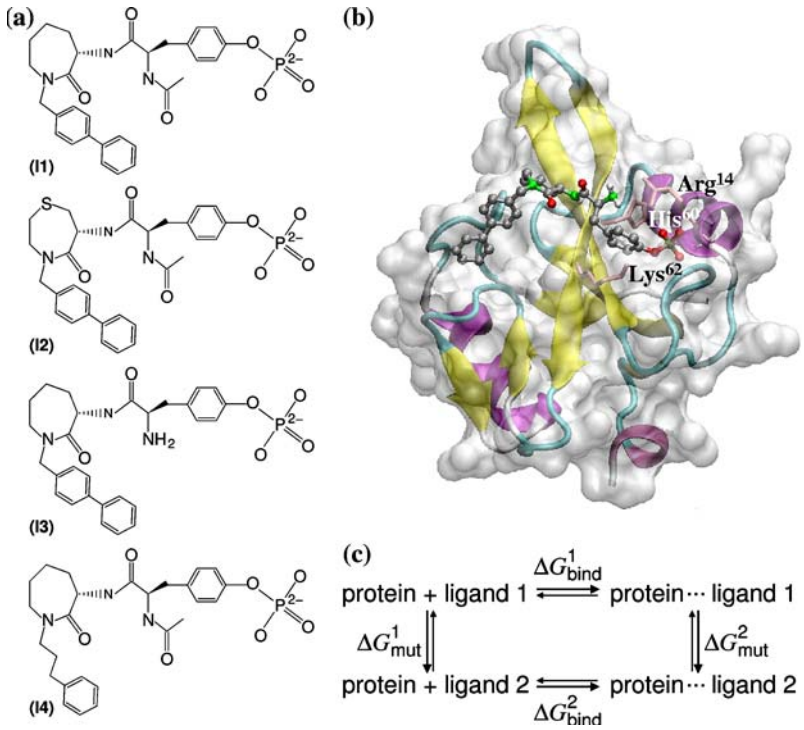

Figure 1. Non-peptide inhibitors of the SH2 domain of the ^{pp60}src tyrosine kinase (a). Complex formed by the ^{pp60}src SH2 domain and inhibitor **I3** bound to the surface of the latter. Note the diphenyl moiety interacting with the hydrophobic pocket, while the pY motif is buried in the hydrophilic pocket (b). Thermodynamic cycle utilized to estimate relative protein–ligand binding free energies [9]. Horizontal transformations are determined experimentally and generally are not amenable to statistical simulations. Vertical transformations correspond to "alchemical transformations" of the ligand in the free state (left) and in the bound state (right), so that $\Delta G^2_{\mathrm{bind}} - \Delta G^1_{\mathrm{bind}} = \Delta G^2_{\mathrm{mut}} - \Delta G^1_{\mathrm{mut}}$ (c).

equilibration. This corresponds to a total simulation length of 800 ps for each transformation of Figure 1, i.e. 1.6 ns to complete the thermodynamic cycle. Assuming that each free-energy difference computed at a given λ-state constitutes an independent observable, the error was determined using a first-order expansion of the free-energy [6]. Alternatively, the error could be estimated based on the hysteresis between the direct and the reverse transformations. This route, often chosen in free-energy calculations, is, however, arguable for essentially two reasons. From a theoretical standpoint, it is far from clear whether the forward and the backward mutations share the same convergence properties [7] – in fact, the well-known example of particle insertion [8] constitutes a prototypical case, wherein the insertion and deletion transformations do no converge towards the same result. From a non-academic perspective, free-energy calculations are sufficiently demanding to brush aside the idea of appraising the error by means of two separate simulations.

The ligands and the protein–ligand complexes were simulated in aqueous solution, under periodic boundary conditions, using the TIP3P water model [10]. In all, 6856 water molecules solvated each protein–ligand assembly in a cubic box of initial size $60.3 \times 60.3 \times 60.3$ Å^3. An appropriate number of sodium and chlorine counter-ions were included to ensure the electric neutrality of the system. The structures of the protein–ligand complexes were determined by X-ray crystallography [2, 3]. The protein and the ligands were described using the all-atom CHARMM force field. Missing terms of the force field, in particular torsional barriers, were determined quantum-mechanically at the MP2/6-31G* level of theory. The point charges borne by the ligands were derived from the electrostatic potential, including a reaction field contribution [11].

All simulations have been carried out in the isothermal–isobaric ensemble, using the program NAMD [12, 13]. The temperature and the pressure were fixed at 300 K and 1 atm, respectively,

employing Langevin dynamics and the Langevin piston method [14]. Long-range electrostatic forces were taken into account by means of the particle mesh Ewald (PME) approach [15, 16].

The equations of motion were integrated with a 2-fs time step and all chemical bonds between heavy to hydrogen atoms were constrained to their equilibrium distance. The 1.6 ns simulation completing each thermodynamic cycle requires approximately 100 wall-clock hours on a cluster of eight AMD Athlon XP-1800 processors.

Results and discussion

From the onset, it may be observed that ligand **l1** adopts a binding mode very similar to that of pYEEI. The biphenyl moiety occupies the hydrophobic pocket entirely and the interaction of pY with the hydrophilic site is strong. The scaffold of the peptide forms steady van der Waals contacts with the surface of the protein. A persistent water molecule remains bridged between the carbonyl group of the lactam moiety and the amide -NH group of Lys^{62}. The relative binding free energies corresponding to the "alchemical transformations" of ligand **l1** are gathered in Table 1. Overall, the agreement between the computational and the experimental estimates is good, within ± 0.7 kcal/mol.

Replacement of one methylene group in the lactam scaffold by a sulfur atom – see Figure 1, reduces the binding affinity by about an order of magnitude. No glaring structural modification of the scaffold is observed upon this mutation. The two-pronged plug motif is well preserved. A closer look at the protein–ligand interface, however, reveals, as expected, a loss of van der Waals contacts, together with rather unfavorable electrostatic interactions where the point mutation occurred, resulting in a somewhat increased flexibility of **l2**.

Table 1. Point mutations of a series of non-peptide inhibitors of the SH2 domain of the ^{pp60}src kinase.

Transformation	Calculated $\Delta G^2_{mut} - \Delta G^1_{mut}$ (kcal/mol)	Experimental $\Delta G^2_{bind} - \Delta G^1_{bind}$ (kcal/mol)
l1 $\rightarrow$ **l2**	$+0.6 \pm 0.4$	$+1.3$
l1 $\rightarrow$ **l3**	$+2.9 \pm 0.4$	$+2.2$
l1 $\rightarrow$ **l4**	$+1.9 \pm 0.4$	$+1.8$

The FEP methodology was employed in association with the dual-topology paradigm. Experimental binding free energies were determined using isothermal titration micro-calorimetry at 25°C and pH 7, in a 50 mM hepes buffer.

Replacement of the amide group of **l1** by an amino group decreases the binding affinity more significantly. Here again, the modified ligand, **l3**, remains perfectly anchored in the hydrophobic and hydrophilic pockets upon mutation. Yet, the interaction formed by the –NAc moiety and Arg^{14} vanishes as the former is altered into $-NH_2$ thereby weakening the binding – i.e. approximately 40 times less than for **l1**.

The difference between ligands **l1** and **l4** is localized in their hydrophobic region. The first phenyl ring of the biphenyl moiety is replaced by two methylene groups. A glimpse into the complex formed by the ^{pp60}src SH2 domain and **l4** reveals that the second phenyl ring of **l1** and **l4** superimpose nicely. Positioning of the peptide scaffold is, however, modified, favoring the interaction of –NAc with His^{60} at the expense of Arg^{14}.

This structural change propagates all the way to the pY ring, which adopts an alternative conformation. Altogether, this corresponds to a radically different binding mode of the ligand. It is important to underline that the "alchemical transformation" of **l1** into **l4** is able to capture the structural modifications in the protein–ligand association, and, hence, provide a very precise estimate of the relative free energy. This may be ascribed to the rapidly relaxing degrees of freedom of the short non-peptide inhibitor, compatible with the time scale of the simulation.

By and large, the present sampling strategy of 1.6 ns per transformation appears to be sufficient to rank the four ligands according to their affinity for the SH2 domain of ^{pp60}src. Beyond the numerical agreement between experimental and computed free-energy differences, this is confirmed by the very reasonable overlap of the density of states displayed in Figure 2. Poorly overlapping distributions is a source of statistical error that propagates and grows from $\lambda = 0$ to $\lambda = 1$. Deconvolution of the global imprecision of the calculation in terms of systematic and statistical errors is a daunting task. The first-order expansion of the free-energy [6], which provides an estimate of ± 0.4 kcal/mol, underestimates for both ligands **l2** and **l3** the imprecision of the computation. It

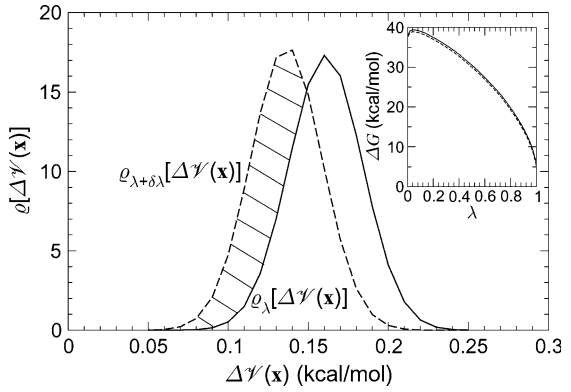

Figure 2. Typical density of states, $\varrho_{\lambda}[\Delta V(\mathbf{x})]$ and $\varrho_{\lambda+\delta\lambda}[\Delta V(\mathbf{x})]$, contiguous intermediate states (obtained from the direct and reverse transformations of **l1** into **l2** at $\lambda = 0.5$). Here, $\Delta V(\mathbf{x})$ stands for the difference in configurational energy within the window of interest. Under most circumstances, finite-length simulations do not allow the distributions to overlap perfectly, thereby, yielding an imprecision in the free-energy estimates. Inset: Change in the free-energy as a function of the linear parameter, λ, for the protein–ligand complex (solid line) and the free, unbound ligand (dashed line). The relative binding free-energy is a difference of usually large numbers.

would, therefore, seem from the latter that the potential energy function may not be optimal for handling this set of ligands, albeit further quantification of the systematic error is beyond the scope of the present work.

Conclusion

Undeniably, free-energy calculations constitute a tangible link between molecular statistical simulations and experiment, they are robust and reliable, and, in principle, they have acquired the status of being a standard tool in the arsenal of methods used routinely by the modeler. Why are these approaches yet only marginally employed outside an academic environment? One possible reason lies in the fortuitously promising results of pioneering free-energy calculations, shattered by 20 years of hindsight and improvements at both algorithmic and methodological levels [17]. Another, more likely rationale is rooted in the high-throughput requirements of the user – in particular the pharmaceutical industry, often incompatible with the time frame necessary for performing this type of numerical simulations.

The free-energy calculations reported here suggest that a reliable ranking of inhibitors to the ^{pp60}src SH2 domain can be established based on an affordable sampling strategy. Whereas the number of non-peptide ligands probed in the present study is very limited, it is envisioned that with an access to dedicated massively parallel architectures and grid computing, cost-effective and rigorous free-energy calculations may provide a fast and convincing answer, that can help rationalizing experimental observations, and, in some instances, play a predictive role in the development of new leads for a specific target. Granted that the computational effort of such simulations is incommensurable with that of a simple geometric recognition, the estimation of free-energy differences should be located at the end of a virtual screening pipeline.

Appropriate sampling of the configurational space is, however, only a necessary, although not sufficient condition to guarantee a reasonable accord with experiment. The precision, based on a single run, and the accuracy, obtained from different simulations, reflect the overall statistical error of free-energy calculations. Given a sampling strategy, free energies and their associated statistical error can be determined seamlessly, with minimal human involvement.

The same, unfortunately, cannot be said for the design of simulations requiring the construction of either a single or a dual topology for the ligands investigated. Since most ligands of pharmacological relevance are small, organic, non-peptide molecules, adaptation of the macromolecular force field is often inevitable. Determination of point charges, optimization of geometries, evaluation of torsional energy barriers are examples of tasks that cannot be automatized in a straightforward fashion. Whereas this preamble to any free-energy calculation is a well-accepted, necessary step for academic environments, the human cost is generally too high to warrant its routine use in industrial settings. Coarser, obviously less accurate, yet widely utilized approaches, like molecular docking, will only be outstaged when the above issues are addressed in an effective way.

Acknowledgments

Jérôme Hénin and Bernard Maigret are gratefully acknowledged for fruitful and motivating discussions. The authors are indebted to Robert

Gilli (Université d'Aix-Marseille) for sharing the experimental micro-calorimetry data prior to publication. The Centre Informatique National de l'Enseignement Supérieur (CINES) and the centre de Calcul Réseaux et Visualisation Haute Performance (CRVHP) is acknowledged for generous provision of CPU time on their SGI Origin 3000 architectures. This work was funded in part by a Hoechst Marion Roussel sponsorship (GIP FR98 CHM 032).

References

1. Soriano, P., Montgomery, C., Geske, R. and Bradley, A., Cell, 64 (1991) 693.
2. Lange, G., Lesuisse, D., Deprez, P., Schoot, B., Loenze, P., Benard, D., Marquette, J.P., Broto, P., Sarubbi, E. and Mandine, E., J. Med. Chem., 45 (2002) 2915.
3. Lange, G., Lesuisse, D., Deprez, P., Schoot, B., Loenze, P., Benard, D., Marquette, J.P., Broto, P., Sarubbi, E. and Mandine, E., J. Med. Chem., 46 (2003) 5184.
4. Gao, J., Kuczera, K., Tidor, B. and Karplus, M., Science, 244 (1989) 1069.
5. Zwanzig, R.W., J. Chem. Phys., 22 (1954) 1420.
6. Hénin, J., Pohorille, A. and Chipot, C., J. Am. Chem. Soc., 127 (2005) 8478.
7. Mark, A.E., In Schleyer, P.v.R., Allinger, N.L., Clark, T., Gasteiger, J., Kollman, P.A., Schaefer, H.F. III and Schreiner, P.R. (Eds.), Encyclopedia of Computational Chemistry 2. Wiley and Sons, Chichester, 1998, pp. 1070–1083.
8. Widom, B., J. Chem. Phys., 39 (1963) 2808.
9. Tembe, B.L. and McCammon, J.A., Comp. Chem., 8 (1984) 281.
10. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L., J. Chem. Phys., 79 (1983) 926.
11. Chipot, C., J. Comput. Chem., 24 (2003) 409.
12. Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K. and Schulten, K., J. Comput. Phys., 151 (1999) 283.
13. Bhandarkar, M., Brunner, R., Chipot, C., Dalke, A., Dixit, S., Grayson, P., Gullingsrud, J., Gursoy, A., Humphrey,W., Hurwitz, D., Krawetz, N., Nelson, M., Phillips, J., Shinozaki, A., Zheng, G., Zhu, F., NAMD users guide, Version 2.5. Theoretical Biophysics Group, University of Illinois and Beckman Institute, 405 North Mathews, Urbana, Illinois 61801, September 2003.
14. Feller, S.E., Zhang, Y.H., Pastor, R.W. and Brooks, B.R., J. Chem. Phys., 103 (1995) 4613.
15. Darden, T.A., York, D.M. and Pedersen, L.G., J. Chem. Phys., 98 (1993) 10089.
16. Essman, U., Perera, L., Berkowitz, M., Darden, T., Lee, H. and Pedersen, L.G., J. Chem. Phys., 103 (1995) 8577.
17. Brandsdal, B.O., Österberg, F., Almlöf, M., Feierberg, I., Luzhkov, V.B. and Åqvist, J., Adv. Protein Chem., 66 (2003) 123.