Revised manuscript.

The final version was published in:

Protein-ligand Interaction Energy-based Entropy Calculations: Fundamental Challenges For Flexible Systems

Gergely Kohut[a], Adam Liwo[b], Szilvia Bősze[c], Tamás Beke-Somfai[a,d*] , Sergey A. Samsonov[b*]

[a]Institute of Materials and Environmental Chemistry, Research Centre for Natural Sciences, Hungarian Academy of Sciences H-1117 Budapest, Magyar tudósok körútja 2, Hungary

[b]Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland

[c]MTA-ELTE Research Group of Peptide Chemistry, Hung. Acad. Sci., Eotvos LorandUniversity, Budapest 112, P.O. Box 32, H-1518 Budapest, Hungary

[d]Department of Chemical and Biological Engineering, Physical Chemistry, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

AUTHOR INFORMATION

Corresponding Authors

Tamás Beke-Somfai: beke-somfai.tamas@ttk.mta.hu

Sergey A. Samsonov: sergey.samsonov@ug.edu.pl

ABSTRACT: Entropy calculations represent one of the most challenging steps in obtaining the binding free energy in biomolecular systems. A novel computationally effective approach (IE) was recently proposed to calculate the entropy based on the computation of protein-ligand interaction energy directly from molecular dynamics (MD) simulations. We present a study focused on the application of this method to flexible molecular systems and compare its performance with well-established normal mode (NM) and quasiharmonic (QH) entropy calculation approaches. Our results raise substantial concerns on the general applicability of IE in terms of reproducibility, reasonable absolute values of the entropy and agreement with NM and QM approaches. IE shows significant variation in the computed entropy values depending on the MD frames chosen for calculations. These deviations render reproducibility of IE calculations to be far from sufficient. We conclude that IE is recommended to be used after substantial modifications with respect to its sampling methodology.

Introduction

Understanding the principles of intermolecular recognition between diverse biomacromolecules is a key in establishing molecular basis of all biochemically relevant processes in the cell. Both experimental and theoretical approaches are being constantly developed to 40be able to answer more effectively the fundamental question of how several biomolecules form a complex and so carry out their biological function[1]. Proper characterization of the structure-function relationship in biomacromolecular systems is a critical step for novel drug discovery and therefore represents one of the central topics in the field of biomedical research[2,3]. Although experimental data on structures of molecular complexes can provide an essential insight into the function of

2

corresponding systems, computational approaches are of high value. They not only support, complement and explain experimental findings but could also provide the detailed analysis of the data, understanding the contribution of separate energy terms and their connection to the participating molecules, which is mostly unapproachable from experiments. Moreover, they assist the rational guiding of further experiments leading to the minimization of the time and resources invested[4]. However, a major sensitive feature of computational methods that always needs careful evaluation is the reliability of the produced results, especially in terms of their comparison with the corresponding experimental data. The latter is key for increasing trust and widespread use of the available methods, highly aiding understanding of a plethora of molecular processes.

Among others, calculating binding affinity, a crucial property of complex systems, represents a subject of many diverse computational approaches[5]. However, despite successes in free energy calculations[6] by using molecular dynamics (MD)-based methods as rigorous free energy perturbation and thermodynamic integration methods[7], Molecular Mechanics Poisson Boltzmann Surface Area (MM-PBSA)[8,9] or linear interaction energy[10] and Monte Carlo simulation-based methodologies[11], accurate estimation of the binding affinity of a molecular system still represents a serious challenge[12]. One of the most commonly applied MD-based free energy calculation method, MM-PBSA and its modified version MM-GBSA (Generalized Born Approximation for MM-PBSA)[13], which implements an implicit solvent model, yields reliable enthalpy values[14,15,16]. Nevertheless, its bottleneck is the calculation of the entropic component, which can potentially increase the overall uncertainty in the calculated binding energies dramatically[17,18]. A popular

method to calculate entropy, which is widely used with the MM-PBSA approach, is the normal-mode (NM) analysis of harmonic frequencies obtained from minimized snapshots of MD simulations[9] and estimates the conformational entropy of a molecular system *in vacuo*. This method, being computationally expensive, has a solid capability to yield potentially useful results in reproducing experimental binding affinities[19]. An alternative method is quasiharmonic approximation (QH), which is essentially faster but in general less accurate and has a number of limitations[20]. QH involves calculations of the covariance matrix of atomic coordinates from MD simulations, interprets the covariances resulting from a harmonic energy surface, while neglecting anharmonic energy surface, and computes QH force constants, which are then used to compute the entropy[21]. A QH approach yields the conformational entropy of a molecule in an "effective" quadratic potential, implicitly including effects of solvent, which renders its results to be different from results from the NM approach. Moreover, the conformational entropy calculated by both above-mentioned methods represent only a part of the total entropy change in a binding process. In comparison to the above-mentioned methods, a relatively novel approach to calculate the entropy based on the computation of protein-ligand interaction energy reported by Duan et al. can use the already generated data from MD simulations without any additional computational cost[22]. The method, which will be related to as IE in this paper, calculates the entropy as following. The gas-phase component of the binding free energy, which is obtained directly from the MD simulation is derived:

$$\Delta G_{gas} = -KT \ln \frac{\int dq_w \, dq_p \, dq_l \, e^{-\beta(E_p + E_l + E_{pl}^{int} + E_w + E_{pw}^{int} + E_{lw}^{int})}}{\int dq_w \, dq_p \, dq_l \, e^{-\beta(E_p + E_l + E_w + E_{pw}^{int} + E_{lw}^{int})}}$$

$$= -KT \ln \left[\frac{1}{\langle e^{\beta E_{pl}^{int}} \rangle} \right] = KT \ln [\langle e^{\beta E_{pl}^{int}} \rangle]$$

$$= KT \ln [e^{\beta \langle E_{pl}^{int} \rangle} \langle e^{\beta(E_{pl}^{int} - \langle E_{pl}^{int} \rangle)} \rangle]$$

$$= \langle E_{pl}^{int} \rangle + KT \ln \langle e^{\beta \Delta E_{pl}^{int}} \rangle$$

$$= \langle E_{pl}^{int} \rangle - T \Delta S \qquad\qquad (1)$$

E_p, E_l, E_w are protein, ligand and solvent internal energies, respectively; E_{pl}^{int}, E_{pl}^{int}, W_{lw}^{int} are

protein−ligand, protein−solvent and ligand−solvent interaction energies, respectively; $\langle E_{pl}^{int} \rangle$ is

the protein−ligand interaction energy averaged in the ensemble, and

$\Delta E_{pl}^{int} = E_{pl}^{int} - \langle E_{pl}^{int} \rangle$ (2) is the difference between protein−ligand interaction energy and its

average value. Hence, final equation of interaction entropy is defined by the protein-ligand

interaction energy and is expressed by the following equation:

$$-T \Delta S = KT \ln \langle e^{\beta E_{pl}^{int}} \rangle \qquad (3)$$

The relevant ensemble averages can be calculated by averaging over MD trajectory:

$$\langle E_{pl}^{int} \rangle = \frac{1}{T} \int_0^T E_{pl}^{int}(t) \, dt = \frac{1}{N} \sum_{i=1}^{N} E_{pl}^{int}(t_i) \qquad (4)$$

The averaging procedure can be further formulated as: $\langle e^{\beta \Delta E_{pl}^{int}} \rangle_{arithmetic} = \frac{1}{N} \sum_{i=1}^{N} e^{\beta E_{pl}^{int}(t_i)} \qquad (5)$

Authors claim that the method is essentially superior to the NM approach demonstrating it to be computationally efficient and numerically reliable[22]. Since the method was described in 2016, it has been widely applied and cited in the literature as original papers[23-31] and reviews[32, 33].

The conceptual break-through arising from the theory behind IE method is that the entropy could be defined in a highly efficient and reliable way directly from the analysis of MD simulations without any entropy-specific post-processing calculations. The motivation to carry out an evaluative case study focusing on IE applicability[22] was grounded on the following observation that has drawn our attention: in practice, according to the described original protocols, the calculated entropic component of the free energy of binding is determined by a single spike (the highest positive value) of the protein-ligand interaction energies obtained in a course of the MD simulation in comparison to its average value. This could potentially suggest that if such a spike does not occur during the analyzed MD simulation or, on contrary, if a bigger spike occurs, the calculated entropy value obtained should be essentially different. Taking into account the length of the analyzed MD simulation from the original procedure, such calculations could be insufficient to compensate the effect of all the possible spikes in values of the interaction energy between the receptor and the ligand and so lead to low reproducibility of the calculated entropy values. In order to deal with the mentioned challenge in the absence of sufficient thermal equilibration of the system, it is possible to impose an energy cut-off for such an energy spike in order to smooth its impact on the calculated entropy values[34]. In our work, we rigorously analyzed and pointed out the limitations of IE in terms of its reliability and reproducibility for several case study systems with the focus on flexible molecular systems.

Methods

Structures. We used the experimental structure of urokinase-type plasminogen activator in complex with 4-(aminomethyl) benzoic acid (PDB ID: 3KGP, 2.35 Å) that was previously used in the dataset in the original study on IE method of Duan et al.[22]. The structures for the complex between CM15 peptide and suramin were obtained by the molecular docking procedure as described further.

Ligands parameterization for docking and MD simulations. Both 4-(aminomethyl) benzoic acid and suramin (Scheme 1) were parameterized in antechamber package [35] of the AMBER 16 MD programs suite[36] using GAFF force field parameters[37] with AM1-BCC charge module[38].

Docking with Autodock 3. In the docking calculations, CM15 was used as a receptor and suramin as a ligand. Since CM15 is a disordered peptide, its 21 conformations were obtained by several diverse MD approaches carried out in AMBER 16[36], GROMACS[39], ECEPP[40] and coarse-grained package UNRES[41]. The RMSD of the backbone atoms for obtained CM15 structures was 5.6±1.4 Å, which provided a high structural variety of the peptide conformations. For blind docking simulations for each of the peptide conformations, Autodock 3 (AD3)[42] was used with the protocol optimized for the glycosaminoglycan ligands[43], which similarly to suramin are linear and contain numerous sulfate and sulfonate groups. Flexible ligand docking was carried out within a box with a grid step of 0.375 Å. The Lamarckian genetic algorithm was applied with the following parameters: initial population size of 300, 10^5 generations, $9995 \cdot 10^5$ energy evaluations, 10^3 independent runs. All torsional angles besides C-S for sulfonates were flexible during docking. In overall, this corresponded to 10 torsional angles degrees of freedom for the

suramin molecule. Fifty top docking solutions were clustered by the DBSCAN algorithm[44] with a neighbourhood search radius of 2 Å and minimal number of points of 2. One representative from each cluster corresponding to the highest AD3 score was further chosen for MD-based analysis.

Molecular dynamics and MM-GBSA calculations. MD simulations were run for protein-ligand complexes using the following protocol. The complexes were solvated in a periodic TIP3P truncated octahedron water box with a minimal distance of 15 Å between complex atoms and the box, and charge neutralizing counter-ions were added. The ff14SB and GAFF force field parameters were used for protein and a small molecule, respectively. Prior to MD simulations, two energy-minimization steps were carried out: first, steepest descent (1500 cycles) and conjugate gradient (100 cycles) with harmonic force restraints of 10 kcal/(mol·Å) applied on protein and small molecule atoms; then, steepest descent (3000 cycles) and conjugate gradient (3000 cycles) without any restraints. This was followed by the process of system heating from 0 to 300 K for 10 ps, and a MD equilibration at 300 K and 10^6 Pa in NTP for 100 ps. Following the equilibration procedure, productive MD runs of 5 ns + 1 ns (see Entropy calculations section) and 100 ns were carried out for 3KGP and CM15-suramin complexes, respectively. An 8 Å cut-off for non-bonded interactions and Particle Mesh Ewald method for long-range electrostatic interactions were applied. The simulations were carried out in NTP with Langevin temperature coupling with collision frequency parameter $\gamma = 1$ ps^{-1} and integration step of 2 fs with the SHAKE algorithm applied to the bonds containing hydrogen atoms. MD trajectories were recorded every 10 ps. Energetic post-processing of the trajectories was done using MM-GBSA with igb = 2. MD simulations for each complex structure was repeated three times.

Entropy calculations. Entropy calculations were carried out using three different methods: interaction entropy (IE), normal mode (NM) and quasiharmonic approximation (QH). In case of the 3KGP complex, it was initially simulated for 5 ns and then the last 1 ns was repeated three times with different random seeds to be consistent with Duan's protocol[22]. During the last 1 ns the coordinates were saved every 0.1 ps resulting in 10000 frames. IE calculation and NM analysis were performed on each 10th frame covering altogether 1000 frames, while QH was carried out for 10000 frames due to its higher pretension of more frequent sampling[45]. For the CM15-suramin complex the entropic contribution was determined from the last 10 ns of the corresponding trajectory using 1000 frames for each method.

We compared the data obtained by IE to the results obtained by NM and QH for one of the systems reported in the study of Duan et al.[23], containing a protein and a ligand 4-(aminomethyl) benzoic acid (Scheme 1). The second system we used contained a potential drug-carrier CM15, which is a disordered 15 aa peptide (KWKLFKKIGAVLKVL-amide), and a ligand suramin, which can induce the folding of CM15[46].

Scheme 1. Chemical structures of 4-(aminomethyl) benzoic acid (left) and suramin (right).
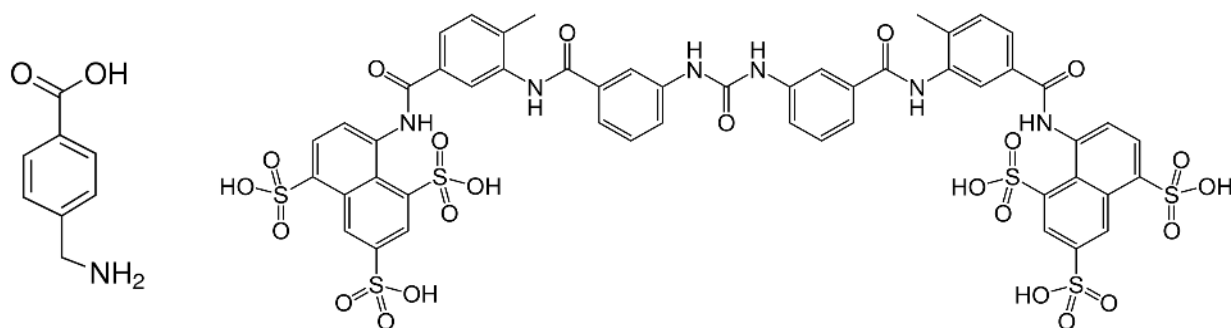
Taking into account the disordered structure of CM15, entropic contribution to its binding affinity could be particularly decisive, and the system is very suitable to address potential discrepancies in entropy calculations arising from its flexible nature.

Results and Discussion

First, we carried out three repetitive MD simulations of the urokinase-type plasminogen activator in complex with 4-(aminomethyl) benzoic acid taken from the dataset of Duan. The obtained results clearly show that the values of the entropy are strongly dependent on the frame corresponding to the highest spike of the interaction energy and are not significantly changed after this event due to the exponential function in the equation 3 (Figure 1).
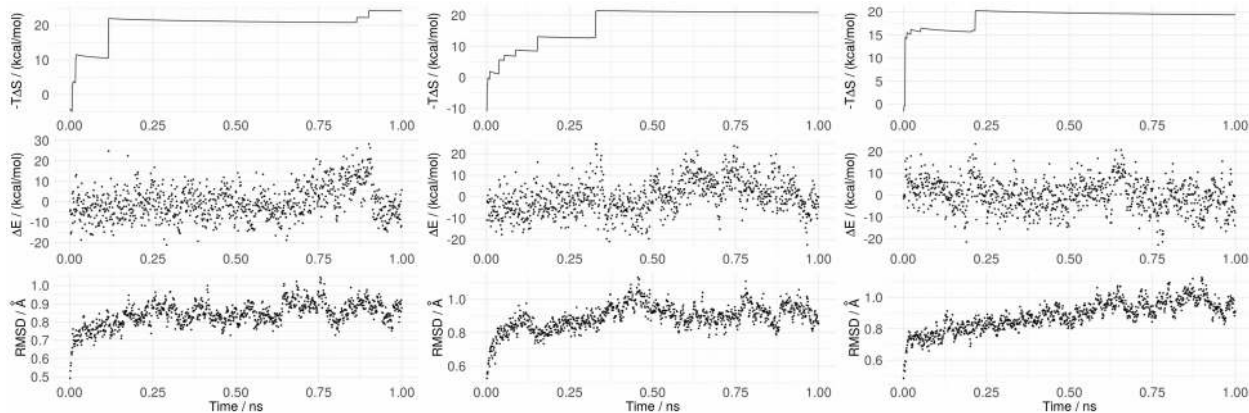

Figure 1. Performance of IE method in three independent MD simulations starting from the same initial configuration. Top panel: calculated entropy. Middle panel: protein-ligand interaction energy. Bottom panel: RMSD (Å) of the ligand with the reference to itself at the start of the MD simulation.

This means that in case the simulation is prolonged, and there is a potentially higher change of the interaction energy, the value of the calculated entropy would be essentially increased. Therefore, such calculations could yield consistent results only in case a very long and probably practically unfeasible MD simulation is carried out so that the system is well converged in terms of the changes of interaction energy. Taking into account that the analyzed 1 ns is a rather short MD simulation, such expectations are not likely to be realized. The averaging procedure (5) was used to calculate the entropy in the system yielding 24.2, 20.9, 19.4 kcal/mol, respectively. These values are very close to the ones values obtained in the study of Duan et al.[22], while the observed differences are expected due be to the use of a different parameterization procedure of the ligand in our study (see Supporting Materials: Methods section). Comparison of these results with the values of the entropy obtained by NM and QH reveal essential differences in the performance of the methods in terms of the absolute values obtained. Apart from this, the values obtained in three different simulations are essentially less deviating from each other in case of NM and QH than of IE (Table 1). Based on these findings, we conclude the reproducibility of IE is less robust.

When considering the more flexible CM15-suramin complex, the entropic contribution to the binding could be more representative and have a stronger impact on the full free energy of the system due to the disordered nature of the peptide. Namely, for the estimation of the binding affinities in similar complexes, entropy calculations are of special significance because MD simulations are often used as a post-processing step after docking to score the obtained docking solutions and to choose a more plausible prediction[47]. Comparison of the absolute values

1

obtained from MD simulations shows significant differences in the values calculated by the three methods for this system (Table 1). The discrepancies between NM, QH and IE are essentially higher than in the urokinase system described above. This is reflected by the higher spikes of the interaction energies obtained here (Supporting Information: Figures S1, S2, S3), where the contribution of the electrostatic interactions is decisive. However, the distributions of the entropy values arising from altogether 69 MD simulations are much narrower for NM and QH, than for the IE method (Figure 2). In addition, the correlation between the values obtained by NM and QH is higher than their correlations with IE (Table 2). These observations underline the fact that according to the principle of the IE approach, the entropic term is purely determined by the interactions between the complex counterparts averaged over the equilibrated MD trajectory of their complex and neglects potential changes of their configurations occurring upon complex formation. Therefore, the application of IE for a case system from the dataset of Duan et al.[22] yielded far more meaningful results than in case of CM15-suramin complexes, where significant conformational changes in both complex counterparts are observed upon binding. In addition, a very high interaction energy variances observed for different CM15-suramin docked structures in the MD analysis lead to the high differences in the IE-derived entropy values for this system. This suggests that IE methodology is not intended to yield meaningful data when applied for ranking of the diverse and *a priori* not sufficiently equilibrated structures obtained by a molecular docking approach for highly flexible systems.

Table 1. The entropy calculations obtained by three methods: IE, NM and QH. The data for urokinase – benzoic acid complex (PDB ID: 3KGP) and several representatives of CM15-suramin complexes are shown, which were obtained from three repeated simulations[a]

| name | $<E^{int}>$ | ΔG_{sol} | ΔH | -TΔS | | | ΔG_{bind} | | |
				IE	NM	QH	IE	NM	QH
3kgp_1	-106.1	79.1	-27.0	24.2	18.2	17.9	-2.8	-8.8	-9.1
3kgp_2	-105.0	79.0	-26.1	20.9	17.7	17.1	-5.2	-8.3	-9.0
3kgp_3	-112.9	84.8	-28.1	19.4	17.3	18.7	-8.7	-10.8	-9.4
CM-S_22_1	-971.3	936.9	-34.4	241.4	34.9	33.3	207.0	0.5	-1.1
CM-S_22_2	-1191.0	1131.5	-59.5	236.6	43.3	38.5	177.1	-16.2	-21.1
CM-S_22_3	-1237.8	1174.1	-63.7	156.8	40.2	39.8	93.1	-23.5	-23.9
CM-S_29_1	-1337.7	1283.3	-54.4	150.2	40.9	37.9	95.8	-13.5	-16.5
CM-S_29_2	-1132.5	1085.7	-46.8	134.9	39.8	37.1	88.1	-7.0	-9.6
CM-S_29_3	-1116.7	1055.6	-61.1	104.5	38.4	40.4	43.4	-22.7	-20.7
CM-S_41_1	-1275.3	1219.7	-55.6	143.4	45.3	38.1	87.8	-10.3	-17.5
CM-S_41_2	-1387.7	1326.0	-61.6	186.0	43.7	40.3	124.4	-18.0	-21.3
CM-S_41_3	-1219.6	1164.6	-55.0	107.9	42.5	39.1	52.9	-12.4	-15.9
CM-S_22_RES1	-1083.6	1039.2	-44.4	62.9	42.9	23.1	18.5	-1.5	-21.3
CM-S_22_RES2	-1049.7	1004.8	-44.9	76.1	46.2	22.6	31.2	1.3	-22.3
CM-S_22_RES3	-1084.6	1038.5	-46.1	90.5	40.2	23.1	44.4	-5.9	-23.0
CM-S_29_ES1	-1044.9	998.0	-46.9	185.3	42.0	23.2	138.4	-4.9	-23.7
CM-S_29_RES2	-1051.4	1004.6	-46.8	165.7	41.8	23.4	118.9	-5.0	-23.4
CM-S_29_RES3	-1002.0	956.0	-46.0	130.2	40.1	25.8	84.2	-5.9	-20.2
CM-S_41_RES1	-1232.1	1193.2	-38.9	70.6	42.2	22.8	31.7	3.3	-16.1
CM-S_41_RES2	-1226.1	1187.5	-38.6	54.2	42.3	22.9	15.6	3.7	-15.7
CM-S_41_RES3	-1228.6	1190.0	-38.6	56.8	42.2	23.0	18.2	3.6	-15.6
ASP_ARG_1	-22.3	21.8	-0.5	11.1	0.2	-28.6	10.6	-0.3	-29.1
ASP_ARG_2	-18.3	17.9	-0.4	7.3	0.01	-28.6	6.9	-0.4	-29.0
ASP_ARG_3	-15.2	15.0	-0.2	4.4	-0.7	-28.9	4.2	-0.9	-29.1
ASP_ARG_RES1	-32.0	31.3	-0.7	14.4	2.2	-24.6	13.7	1.5	-25.3
ASP_ARG_RES2	-45.8	44.0	-1.8	27.3	2.9	-25.1	25.5	1.1	-26.9
ASP_ARG_RES3	-32.5	32.1	-0.4	13.2	3.9	-25.0	12.8	3.5	-25.4

[a]The subscript after CM-S (staying for the complex CM15-suramin) corresponds to the number of the conformation of CM15 and the second subscript corresponds to the number of the simulation in three repeats. All values are in kcal/mol.

Figure 2. Density of probability for the entropy calculated by three methods: IE, NM and QH
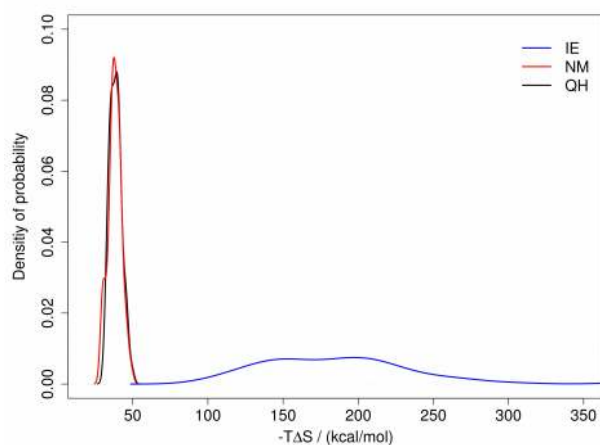
Table 2. Correlations between the entropy values obtained by three methods: IE, NM and QH. The data from altogether 69 MD simulations are shown.

	IE	NM	QH
IE	1	0.16	0.11
NM	0.16	1	0.34
QH	0.11	0.34	1

In case of the flexible system IE has always at least 60 kcal/mol deviation for all obtained entropy terms when compared to NM and QH (Table 1). This consequently leads to major deviations in the obtained free energy values. Unfortunately, the obtained results by IE suggest that complex formation is highly unfavorable. Therefore the simulation results provide an unfortunate misinterpretation of the binding preference even at a qualitative level.

Conclusions

To summarize our findings, we observe that: i) the occurrence of a frame with the highest spike of the interaction energy is indeed crucial and determining for the final value of the entropy in IE calculations; ii) therefore, it is dependent on the length of the MD simulation affecting the reproducibility of the method; iii) the events in the simulation occurring after this spike do not significantly affect the obtained entropy value; iv) IE method yields both absolute values as well as the distribution of the entropy essentially different from the ones in NM and QH. The latter two approaches provide rather similar entropy values even for a very flexible complex system. Despite the high potential of IE in obtaining entropy values computationally very effectively, the presented data suggest that there are strong limitations for its general applicability. It could be potentially reliable only for very well equilibrated complexes, which are fundamentally not accessible within the feasible range of MD simulations, or for those systems which according to the experiment have substantially low structural and thus intermolecular interaction energy changes. We suggest that IE should be applied after the rigorous calibration for a particular molecular system and used together with the NM approach as a general reference.

ASSOCIATED CONTENT

Supporting Information.

Figure S1, S2, S3. Performance of IE method in three independent MD simulations starting from the same initial configurations for CM15-suramin system: CM_S_22, CM_S_29, CM_S_41.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

REFERENCES

(1) Keskin, O; Gursoy, A; Ma, B; Nussinov, R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.* **2008**, *108,* 1225–1244.

(2) Thornton, J. M; Todd, A. E; Milburn, D; Borkakoti, N; Orengo, C. A. From structure to function: approaches and limitations. *Nat. Struct. Biol.* **2000,** *7,* l991–994.

(3) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulation and drug discovery. *BMC Biol.* **2000**, *9*, 71.

(4) Ramachandran, K. I.; Deepa, G; Krishnan Namboori, P.K. Computational chemistry and molecular modeling principles and applications*;* Springer: Berlin Heidelberg, Germany, 2008.

(5) Mobley, D. L.; Gilson, M. K. Predicting binding free energies: frontiers and fenchmarks. *Annu. Rev. Biophys*. **2017**, *46*, 531-558.

(6) Rodinger, T.; Pomès, R. Enhancing the accuracy, the efficiency and the scope of free energy simulations. *Curr. Opin. Struct. Biol.* **2005,** *15*, 164-170.

(7) Beveridge, D. L; Dicapua, F. M. Free-energy via molecular simulation - applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Bio*. **1989,** *18,* 431–492.

(8) Kollman, P. A; Massova, I; Reyes, C; Kuhn, B; Huo, S; Chong, L; Lee, M; Lee, T; Duan, Y; Wang, W. et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res*. **2000,** *33*, 889-897.

(9) Srinivasan, J; Cheatham, III, T. E; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J. Am. Chem. Soc*. **1998,** *120,* 9401-9409.

(10) Aqvist, J; Medina, C; Samuelsson, J. E. New method for predicting binding-affinity in computer-aided drug design. *Protein Eng*. **1994,** *7(3),* 385–391.

(11) Vitalis, A.; Pappu, R. V. Methods for Monte Carlo simulations of biomacromolecules. *Annu. Rep. Comput. Chem*. **2009**, *5*, 49–76.

(12) Hansen, N.; van Gunsteren, W. F. Practical aspects of free-energy calculations: a review. *J. Chem. Theory Comput*. **2014**, *10*, 2632-2647.

(13) Tsui, A.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers.* **2000-2001,** *56,* 275-91.

(14) Hout, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model*. **2011,** *51,* 69-82.

(15) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug. Discov*. **2015,** *10*, 449-461.

(16) Zeller, F.; Zacharias, M. Evaluation of generalized born model accuracy for absolute binding free energy calculations. *J. Phys. Chem. B*. **2014,** *118,* 7467-7474.

(17) Weis, A; Katebzadeh, K; Soderhjelm, P; Nilsson, I; Ryde, U. Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49,* 6596-6606.

(18) Homeyer, N.; Gohlke, H. Free energy calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area method. *Mol. Inf*. **2012**, *31*, 114–122.

(19) Genheden, S.; Kuhn, O.; Mikulskis, P.; Hoffmann, D.; Ryde, U. The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant. *J. Chem. Inf. Model*. **2012,** *52*, 2079-2088.

(20) Chang, C. E.; Chen, W.; Gilson, M. K. Evaluating the accuracy of the quasiharmonic approximation. *J. Chem. Theory. Comput.* **2005,** *1*, 1017-1028.

(21) Karplus, M.; Kushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromolecules.* **1981**, *14*, 325−332.

(22). Duan, L.; Liu, X.; Zhang, J. Z. H. Interaction entropy: a new paradigm for highly efficient and reliable computation of protein–ligand binding free energy. *J. Am. Chem. Soc*. **2016**, *138*, 5722-5728.

(23) Yan, Y.; Yang, M.; Ji. C. G.; Zhang, J. Z. H. Interaction entropy for computational alanine scanning. *J. Chem. Inf. Model*. **2017**, *57,* 1112-1122.

(24) Ben-Shalom, I. Y.; Pfeiffer-Marek, S.; Baringhaus K. H.; Gohlke H.; Efficient approximation of ligand rotational and translational entropy changes upon binding for use in MM-PBSA calculations. *J. Chem. Inf. Model.* **2017**, *57 (2)*, 170–189.

(25) Aldeghi, M.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Statistical analysis on the performance of molecular mechanics poisson–boltzmann surface area versus absolute binding free energy calculations: bromodomains as a case study. *J. Chem. Inf. Model.* **2017,** *57 (9)*, 2203–2221.

(26) Cebrián-Prats, A.; Rovira, T.; Saura, P.; González-Lafont, À.; Lluch, J. M. Inhibition of mammalian 15-Lipoxygenase by three Ebselen-like drugs. A QM/MM and MM/PBSA comparative study, *J. Phys. Chem*. *A*. **2017**, *121 (51)*, 9752-9763.

(27) Thai, N. Q.; Nguyen, H. L.; Linh, H. Q.; Liad, M. S.; Protocol for fast screening of multi-target drug candidates: application to Alzheimer's disease, *J. Mol. Graph. Model.* **2017***, 77*, 121-129.

(28) Balakumar, C.; Ramesh, M.; Tham, C. L.; Khathi, S. P.; Kozielski F.; Srinivasulu C.; Hampannavar, G. A.; Sayyad, N.; Soliman, M. E.; Karpoormath, R. Ligand- and structure-based *in silico* studies to identify kinesin spindle protein (KSP) inhibitors as potential anticancer agents. *J. Biomol. Struct. Dyn*. **2017,** in press, DOI: 10.1080/07391102.2017.1396255

(29) Cong, Y; Li; M. Feng, G; Li, Y; Wang, X; Duan, L. Trypsin-ligand binding affinities calculated using an effective interaction entropy method under polarized force field. *Sci. Rep.* **2017**, *7*, 17708.

(30) Zhou, Y.; Liu, X.; Zhang, Y.; Peng, L; Zhang, J. Z. H. Residue-specific free energy analysis in ligand bindings to JAK2, *Mol. Phys.* **2018***, in press, DOI: 10.1080/00268976.2018.1442596

(31) Zou, Y.; Qian, Z.; Sun, Y.; Wei, G.; Zhang, Q. Orcein-related small molecule O4 destabilizes hIAPP protofibrils by interacting mostly with the amyloidogenic core region. *J. Phys. Chem. B.* **2017,** *121(39),* 9203-9212.

(32) Qui, L.; Yan, Y.; Sub, Z.; Song, J.; Zhang, J. Z. H. Interaction entropy for computational alanine scanning in protein–protein binding. *WIREs Comput. Mol. Sci.* **2018**, *8*, DOI: 10.1002/wcms.1342.

(33) Wang, C.; Greene, D.; Xiao L.; Qi, R.; Luo, R. Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci*., **2018**, *4*, 87, DOI: 10.3389/fmolb.2017.00087

(34) Sun, Z.; Yan Y. N.; Yang, M.; Zhang Y. Z. Interaction entropy for protein-protein binding. *J. Chem. Phys.* **2017**, *146*, 124124, DOI: https://doi.org/10.1063/1.4978893

(35) Wang, J.; Wang, W.; Kollman P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 2006, *25*, 247-260.

(36) Case, D., A.; Cerutti, D.S.; Cheatham, III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T.J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N. et al. AMBER 2017. 2017, University of California, San Francisco.

(37) Wang, R. M.; Wolf, J. W.; Caldwell, P. A.; Kollman, D. A.; Case, J. Development and testing of a general amber force field. *Comp. Chem*. 2004, *25*, 1157-1174.

(38) Jakalian, A.; Bush, B. L.; Jack, B. D.; Bayly, C. I.; Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comp. Chem.* 2000, *21*, 132-146.

(39) Abraham M. .J.; Murtola T.; Schulz R.; Páll S.; Smith J. C.; Hess B.; Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015, *1–2*, 19–25.

(40) Ripoll, D. R; Pottle, M. S.; Gibson, K. D.;Scheraga, H. A.;Liwo, A. Implementation of the ECEPP algorithm, the Monte Carlo minimization method, and the electrostatically driven Monte

Carlo method on the Kendall square research KSR1 computer. *J. Comput. Chem.*, 1995, *16,* 1153-1163.

(41) Liwo, A.; Czaplewski, A.; Oldziej, S.; Rojas, A. V.; Kazmierkiewicz, R.; Makowski, M.; Murarka, R. K.; Scheraga, H. A. Simulation of protein structure and dynamics with the coarse-grained UNRES force field. In: coarse-craining of condensed phase and biomolecular systems. *Taylor & Francis*, 2008, *8*, 107–122.

(42) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian algorithm an empirical binding free energy function. *J. Comput. Chem.* 1998, *19*, 1639-1662.
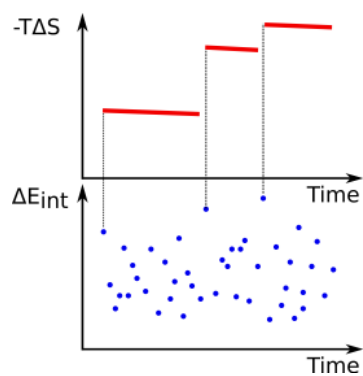
(43) Samsonov S. A.; Pisabarro, M. T. Computational analysis of interactions in structurally available protein-glycosaminoglycan complexes. *Glycobiology*, 2016, *26*, 850–861.

(44) Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *AAAI Press.* 1996, 226–231.

(45) Chang, C. E.; Chen, W.; Gilson, M. K. Evaluating the accuracy of the Quasiharmonic Approximation. *J. Chem. Theory. Comput.* 2005, *1*, 1017-1028.

(46) Zsila, F.; Bősze, S.; Horváti, K.; Szigyártó, I. C.; Beke-Somfai T. Drug and dye binding induced folding of the intrinsically disordered antimicrobial peptide CM15. *RSC Adv.*, **2017**, *7*, 41091-41097 .

(47) Wichapong, K.; Lawson, M.; Pianwanit, S.; Kokpol, S.; Sippl, W. Postprocessing of protein−ligand docking poses using linear response MM-PB/SA: Application to wee1 kinase inhibitors. W. *J. Chem. Inf. Model*. **2010**, *50*, 1574-1588.

TOC graphic