SAMPL6 Octanol-Water Partition Coefficients from Alchemical Free Energy Calculations with MBIS Atomic Charges.

Maximiliano Riquelme · Esteban Vöhringer-Martinez*

Abstract In molecular modeling the description of the interactions between molecules forms the basis for a correct prediction of macroscopic observables. Here, we derive atomic charges from the implicitly polarized electron density of eleven molecules in the SAMPL6 challenge using the Hirshfeld-I and Minimal Basis Set Iterative Stockholder(MBIS) partitioning method. These atomic charges combined with other parameters in the GAFF force field and different water/octanol models were then used in alchemical free energy calculations to obtain hydration and solvation free energies, which after correction for the polarization cost, result in the blind prediction of the partition coefficient. From the tested partitioning methods and water models the S-MBIS atomic charges with the TIP3P water model presented the smallest deviation from the experiment. Conformational dependence of the free energies and the energetic cost associated with the polarization of the electron density are discussed.

Maximiliano Riquelme
Instituto de Ciencias Químicas
Facultad de Ciencias
Universidad Austral de Chile
Valdivia
Chile

Esteban Vöhringer-Martinez
Departamento de Físico-Química
Facultad de Ciencias Químicas
Universidad de Concepción
Concepción
Chile
Tel.: +56-41-2204986
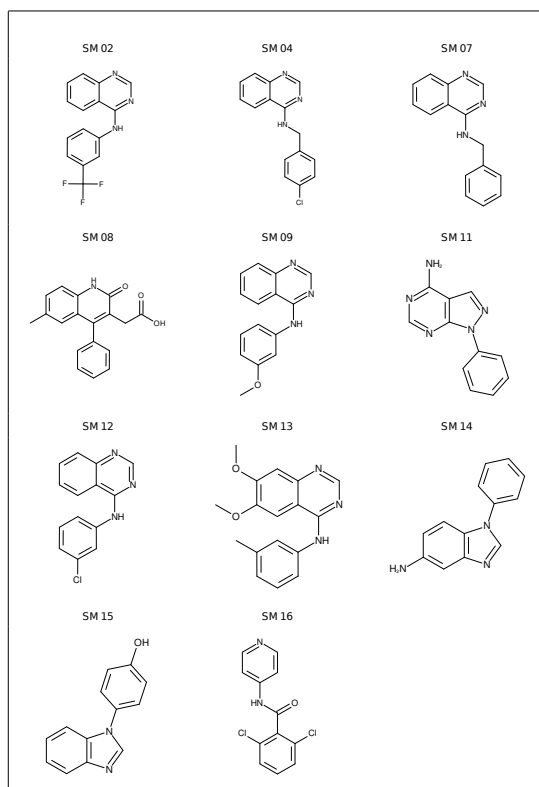Fax: +56-41-2204986
E-mail: evohringer@udec.cl

Fig. 1 Set of molecules proposed by the SAMPL6 challenge.

1 Introduction

In the past eleven years SAMPL challenges have included blind prediction of a variety of different properties such as hydration free energy [1], binding affinity of host-guest systems [2,3], distribution coefficients [4,5] and calculations of pka [6]. They have made an important contribution to the development of new methods and computational tools [7] and increased the accuracy in the prediction of each of these properties. The interest in obtaining more efficient and accurate methodologies to predict these properties lies mainly in their pharmaceutical, biochemical and environmental relevance. Indirectly, all these properties are also related to the prediction of binding free energies between a ligand and a protein serving as model systems for the validation of different methods.

The challenge of SAMPL6 part II consisted in the determination of octanol water partition coefficients of 11 molecules (see Fig. 1) that are similar to fragments of protein kinase inhibitors and are a subset of molecules that were part of the pka SAMPL6 challenge[6]. Predicting the logarithm of the partition coefficient is challenging both experimentally[8] and computationally, and in

general there are not many reports in which different computational methodologies are tested blindly (without knowledge of the experimental results) and then compared to each other.

Together with QSPR approaches implicit solvation models in combination with electronic structure methods have been widely used and parameterized to reproduce partition coefficients with great success as the winner of this challenge using COSMOtherm and the second using a QSPR method reveal. From a molecular modeling perspective partition coefficients are important experimental observables used to validate force fields for small drug like molecules. The increased computational power in combination with established free energy calculation methods make their prediction mainly dependent on the description of the force field since sampling of the relevant phase space is achieved in the simulation time.

One of the most commonly used methods to validate force field performance are molecular dynamics simulations with an explicit solvent description as reported recently by Bannan *etal.*[9]. In this work the generalized AMBER force field (GAFF) and corrected dielectric GAFF (GAFF-DC) combined with a new autonomous tool for the creation of the input files (Solvation Toolkit) resulted in logP values with an root mean square error (RMSE) of 1.2 logarithm units compared to the experimental values. Other approaches combine molecular dynamics and solvation-free energy calculations with an implicit Generalized-Born / Solvent Accessible (GB/SA) approach for a group of 11,993 molecules with an RSME of 1.14 log units[10]. As in this challenge, better agreement with experiment has been obtained with electronic structure calculations and implicit solvation models as SMD, SM8, SM12 and the COSMO variants obtaining a mean absolute error (MAE) of approximately 0.6 log units for a set of 34 organic molecules and 55 fluorinated alcohols and carbohydrates[11]. Finally, there is a wide variety of empirical methods based on atoms and fragments such as KLOGP [12], ALOGP [13], XLOGP [14,15], which consist of regression models or neural networks that have been trained to reproduce the logP using a large set of experimental data.

Methods that use implicit solvents in general have good results partially because they were parameterized with solvation- and transfer - free energies of neutral solutes in water and different organic solvents [16,17]. The statistical models based on fragments are very fast to estimate the logP compared to the other two methods, but they have some drawbacks since they tend to overestimate the lipophilicity of large molecules and do not cover the entire chemical space, which creates greater uncertainty in the confidence of the results [10].

Molecular dynamics using explicit solvents account for the dynamic nature of the solvent molecules and specific hydrogen bond interactions, but at an increased computational cost. Crucial for the correct prediction of the associated free energies is the correct description of the interactions in the system by the different force fields. In previous challenges hydration free energies were found to depend significantly on the employed charge model[18]. More recently, we have shown that the electrostatic interactions described by polarized

Hirshfeld-I (HI) and Minimal Basis Set Iterative Stockholder (MBIS) atomic charges result in good agreement of hydration free energies of the FreeSolv database and partition coefficients of methylated DNA bases[19, 20].

The ability to reproduce the experimental logP by force field based molecular dynamics simulations can be conditioned to a large number of variables of the simulation protocol such as the initial conformation used in each solvent, the water and octanol solvent model, the simulation time (especially for octanol) and the description of the solute-solvent interactions. In polar solvents, the electrostatic interactions described by atomic charges dominate over other types as dispersion for example. Here, we address the capacity of two different methods to derive atomic charges from the polarized molecular electron density employing the theory of atoms in molecules[21]: the S-HI method (Hirshfeld-Iterative atomic charges using the implicit solvent SMD in the calculation of the electron density by electronic structure methods) and the atomic charges S-MBIS (using the alternative MBIS partitioning method) to calculate logP values for the 11 molecules of the SAMPL6 challenge with free energy calculations using explicit solvents. In addition, we also tested the influence of the water model, the conformation used in each solvent, and the total simulation time.

2 Methods

Based on the provided smiles strings we created conformers with RDKit and optimized their structure with the MMFF94s force field keeping only those conformers presenting a root mean square deviations (RMSD) of the heavy atom positions larger than 0.5 Åcompared to the most stable one. The obtained geometries were then optimized with the PM7 semiempirical ab-initio method with the MOPAC 2016 software. For SM02 also tautomers were studied were the hydrogen atom of the secondary amine group was moved to the closest nitrogen atom on the aromatic ring. This tautomer was more stable at the PM7 level in vacuum, but not in the DFT calculation mentioned below.

Once the conformations obtained by the previous procedure for each molecule were selected, each of the structures was optimized using the ORCA 4.0.0.2 [22] program at the BLYP theory level and the def2-TZVP base set. This was done in vacuo and using the implicit solvent SMD for water and octanol. Besides the test case of SM13 only the conformer with the lowest free energy in each solvent was used as starting structure for the free energy calculations.

2.1 Atomic charges

Atomic charges were obtained from the polarized electronic density of the most stable conformer of each of the 11 molecules proposed in the challenge at the BLYP/def2-TZVP level of theory using the SMD implicit solvent[17] for water and octanol. Two methods to partition the electronic density were used:

one based on the Hirshfeld-I [23] method and the other based on the Minimal Basis Iterative Stockholder method [24] using the Horton 2.0.0 program [25] as described in previous work [19,20]. After obtaining the charges, the chemically equivalent atoms by symmetry were averaged using the OpenEye tools (version 2017.2.1).

2.2 Free energy and partition coefficients

The free energy of hydration and solvation was obtained by means of al-chemical free energy calculations for each of the 11 molecules using standard protocols described in previous works [26,20] which allow to obtain free en-ergy values with very small uncertainty. The 11 molecules were solvated in approximately 1500 water molecules using the SPC/E [27] and TIP3P [28] water model for the calculation of hydration-free energy. For octanol, approxi-mately 140 molecules were added in a dodecahedron simulation box using the GROMACS simulation package 5.0.4 [29]. Then a short minimization was per-formed, and the system was equilibrated in a NVT and NPT ensemble using a time step of 2 fs in combination with stochastic dynamics [30] ($\tau = 2$ps) and the Parrinello-Rahman pressure coupling [31] ($\tau_p = 1$ps) algorithm using the compressibility of water or octanol. The electrostatic interactions were calcu-lated with the Particle-Mesh-Ewald method, [32] a cut-off radius of 1.2 nm, a PME-order of 6 and a spacing of 0.1 nm. The van der Waals interactions were scaled to zero via a switching function, which switches the potential to zero between 1.0 and 1.2 nm. The neighbor list was updated every 10 steps with the verlet cutoff-scheme implemented in GROMACS 5.0.4[29] and its cut-off was set to 1.2 nm. All bonds were constrained with the LINCS algorithm[33] of order 4 and the isotropic correction to the energy pressure due to missing van-der-Waals interactions was applied. [34]

After the equilibration of the system the free energy of hydration and solvation was calculated using an alchemical path using molecular dynamics simulations where first the electrostatic interactions of the solute with the sol-vent were turned off through a lambda parameter using the following lambda values [0.00, 0.50, 0.75, 1.00] and subsequently van-der-waals interactions were turned off with the following lambda values [00, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85 , 0.90, 0.95, 1.00] using soft core poten-tials with parameters $\sigma = 0.3$, $\alpha = 0.5$ and $p = 1$. For hydration, a total time of 5 ns was simulated for each lambda value and a time of 15 ns for solvation in octanol. The results of these simulations were analyzed using the alchemical analysis tool [35] with the MBAR method [36] to estimate the energy values free of hydration and solvation.

The logP was calculated from the values of free energy of hydration and solvation at 298 K by the following equation:

$$\log P_{\text{Octanol/H}_2\text{O}} = \frac{\Delta G_{\text{Hyd}} - \Delta G_{\text{solv}}}{RT \ln 10} \tag{1}$$

where R is the ideal gas constant and T the temperature.

2.3 LogP from implicit solvent SMD

Additionaly, we also calculated the logP values with electronic structure calculations using the implicit solvent SMD with the ORCA package 4.0.0.2 [22]. For this, the difference of free energy between the molecule in vacuum and using the implicit solvent SMD for water (free energy of hydration) was determined, and the difference of free energy for the molecule in vacuum and the solvent SMD octanol was also determined (solvation free energy), considering the standard state of 1 mol L^{-1}. Free energies were obtained under the rigid rotor and harmonic oscillator approximation.

3 Results

3.1 Conformational Analysis of Hydration and Solvation Free Energies
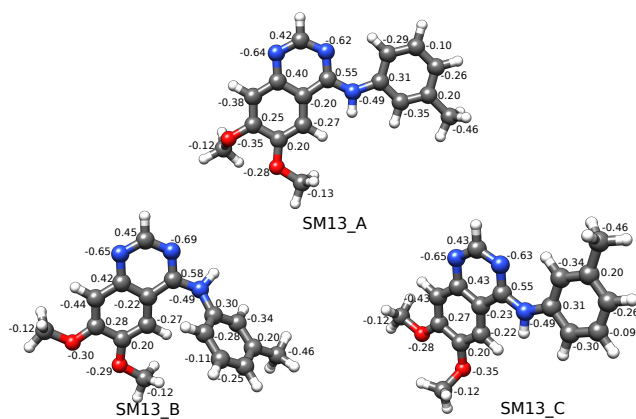

Fig. 2 The most stable conformers of SM13 at the BLYP/def2-TZVP level with the SMD model and the S-MBIS atomic charges of the non-hydrogen atoms in water.

To test the effect of using different conformations in the free energy calculations we selected the SM13 molecule because of its large number of rotable single bonds. We identified the three most stable conformers in water and octanol at the BLYP/def2-TZVP level by rotamer generation (RdKit) and geometry optimization. The three conformers differ mainly in the torsional angle between the two phenyl rings and the relative orientation of the methoxy groups (see Fig. 2). For each conformation free energies in each solvent were calculated using the respective S-MBIS atomic charges and the SPC/E water model. All three conformations present the same values within the errors (see Table 1), probably because the flexible character of the molecule and only small variations in the atomic charges of each conformer (see Fig. 2). But,

when we corrected the free energies with the polarization energy we observed significant differences between the conformers. For the most stable conformer which has the two phenyl rings aligned on one plane this energy is largest. This is explained by the conjugated π system built by the two planar phenyl rings leading to larger polarizability and its associated energy cost. The different electronic properties of the three conformers is also reflected in the dipole moment of the most stable conformer which is 3.5 D larger in water than in vacuum at this level of theory. The same trends are also observed with the hybrid functional B3LYP which is known to result in smaller dipole moments than BLYP. The electronic response and polarization of the solute is dependent on the dielectric properties of the solvent which results in smaller polarization energy in octanol for all conformers (see Table 1). The polarization energy corrections are in water (octanol) 2-3 kcal/mol (1 kcal/mol) larger for the most stable conformer SM13_A. This conformational dependent polarization energy correction of the logP values increases the value by almost 1 unit for the most stable conformer.

Molecule	$\Delta G_{\mathrm{hyd}}^{\mathrm{elec-vdw}}$	$\Delta G_{\mathrm{oct}}^{\mathrm{elec-vdw}}$	E_{hyd}^{pol}	E_{hyd}^{pol}	ΔG_{hyd}	ΔG_{oct}	logP
SM13_A	-12.18	-18.69	5.01	3.75	-7.17	-14.94	5.69
SM13_B	-12.44	-18.41	3.06	2.29	-9.38	-16.12	4.94
SM13_C	-12.56	-19.24	2.25	2.68	-10.31	-16.56	4.58

Table 1 Hydration and solvation free energies with and without correction by the polarization energy in kcal mol^{-1} for the three most stable coformers of molecule SM13.

As will be shown below, our method overestimates logP values and one possible contribution to the error of all molecules with a common substructure as SM13 (e.g. SM02 and SM09) could arise from the overestimated polarization cost of the most stable conformer which is not representative for the not-planar conformation observed in the MD simulations which possess a smaller polarization energy correction and would lead to smaller logP values.

These molecules might, therefore, present one case where the dynamics of the solute and the solvent are required to provide the correct partitioning coefficients, free energies and polarization energy corrections.

3.2 Electron density partitioning and water model dependence

Figure 3 shows the calculated logP values for the eleven molecules compared to the experimental references starting from the most stable conformation. There is a significant dependence on the method used to parition the electron density in atomic contributions providing the S-HI and S-MBIS atomic charges. The MBIS atomic charges outperform the ones obtained with the HI partitioning method, which is in agreement with our previous results on hydration free energies for the FreeSolv database. The poor performance of the HI method could be explained by the presence of N-heterocycles in the structure of the
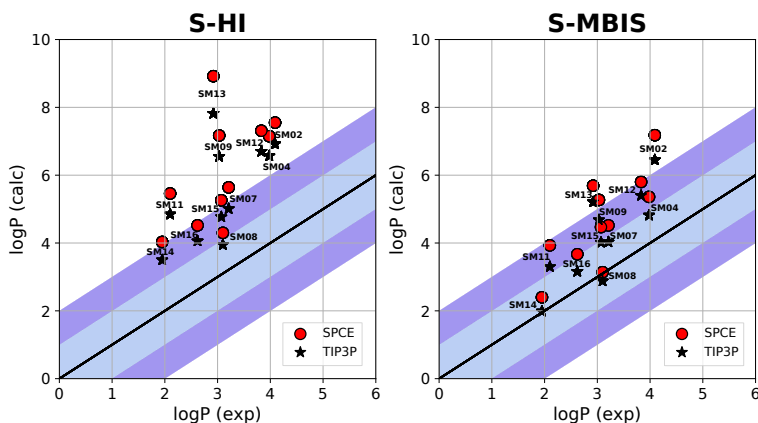
Fig. 3 Parity plots of logP values obtained with the S-MBIS and S-HI atomic charges and the SPC/E or TIP3P water model.

Charge Model	Average error	RMS	AUE	Kendall tau	Pearson R
S-HI SPC/E	3.03 ± 0.40	3.28 ± 0.44	3.04 ± 0.40	0.38 ± 0.24	0.58 ± 0.20
S-HI TIP3P	2.44 ± 0.32	2.67 ± 0.34	2.43 ± 0.32	0.42 ± 0.22	0.63 ± 0.17
S-MBIS SPC/E	1.59 ± 0.29	1.82 ± 0.26	1.59 ± 0.29	0.53 ± 0.25	0.79 ± 0.14
S-MBIS TIP3P	1.09 ± 0.23	1.35 ± 0.21	1.13 ± 0.22	0.53 ± 0.23	0.81 ± 0.12

Table 2 Statistical descriptors for each charge model combined with the two water models.

eleven molecules, which also presented large deviations in the hydration free energies. The MBIS paritioning method, which does not rely on the proatom electron density of the unstable anion, improved the hydration free energy in our previous study and also the logP values as evidenced in this study.

For both methods a water model dependence is observed although to a lesser extend. The TIP3P model results in better logP values than SPC/E although SPC/E is known to reproduce properties of liquid water more accurately. One possible explanation is that the GAFF vdW parameters are more consistent with the TIP3P water model, which is widely preferred for the simulations using AMBER and GAFF force fields. Since we did not alter these parameters when replacing the atomic charges in the GAFF force field this could explain the slightly better performance of this water model. But, the effect varies between the molecules and is not systematic, which suggest that an electron density based method to derive van der Waals parameters would be desirable to become independent on previously derived non-bonded parameters. Compared to other methods using molecular dynamics simulations and force fields as CGENF and GAFF participating in this challenge our results present a comparable RMSE when the S-MBIS atomic charges are combined with the TIP3P water model although some molecules present deviations larger than two logP units. Additionally, we also tested the effect of longer simulation times to calculate the octanol solvation free energy. Extend-
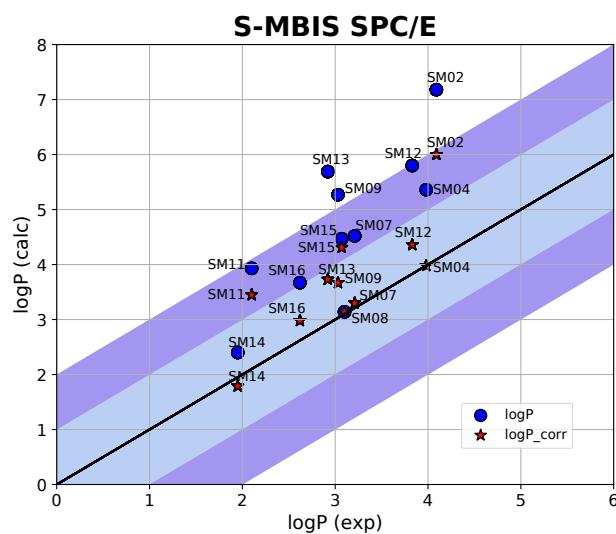
Fig. 4 Parity plot of logP values for the S-MBIS atomic charges with the SPCE water model and including the correction per functional group for the hydration free energies derived in ref [20].

ing the simulation time per lambda window from five to twenty nanoseconds did not change the free energy by more than 1 kcal/mol.

One possible explanation for the systematic overestimation of the logP values is the presence of significant amount of water in the octanol phase in the experimental measurements. Because we simulate a pure octanol phase our logP values could have been predicted systematically larger and explain the observed deviation.

3.3 Functional Group Corrections of Hydration Free Energies

In our previous study of hydration free energy we were able to assign corrections to the calculated values based on the functional group present in the 613 molecules. Here, we wanted to test if this correction would improve the obtained logP value, thereby identifying the error contribution from the hydration free energies. In Figure 4 we show that in all cases the inclusion of the correction improve the logP values suggesting that the prediction of the hydration free energy contributes considerably to the error and its improvement would also have an impact on the quality of the predicted logP values.
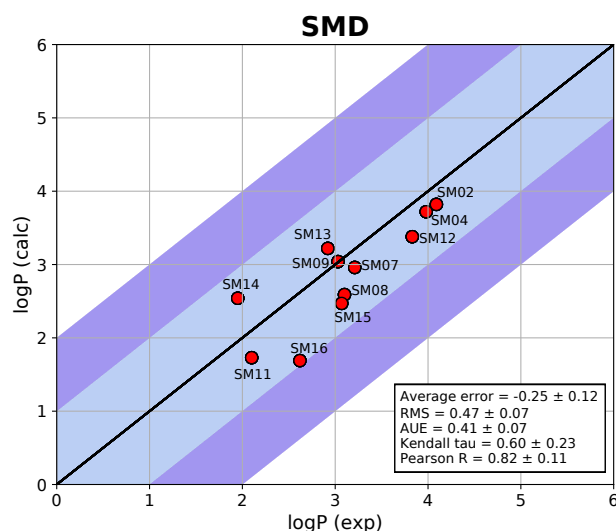
Fig. 5 Parity plots of LogP values calculated at the BLYP/def2-TZVP level of theory under the rigid-rotor harmonic oscillator approximation and the SMD solvation model and the experimental reference.

3.4 LogP from SMD Solvation Model

For the calculation of the atomic charges we had optimized the geometries of all molecules and calculated the vibration frequencies of all molecules with the SMD solvation model and the BLYP/def2-TZVP method. Based on these data we also calculated the hydration and octanol solvation free energies under the rigid rotor - harmonic oscillator approximation resulting in the logP values shown in the parity plot of Figure 5. The small RMSE is comparable to the best predicted values from the COSMOtherm. However, the good performance of the SMD solvation model has to be taken carefully because its parametrization was mainly based on data of octanol solvation free energies and partition coefficients. Therefore, the predictive property for other solvents might vary.

Our method does does not rely on experimental free energies and its only input is the polarized electron density which is obtained accurately from low computational cost DFT methods and mostly independent of the solvation model.

4 Conclusion

The results show that S-MBIS atomic charges derived from the polarized molecular electron densities of the eleven molecules combined with alchemical free energy calculations using explicit solvent (including polarization energy)

provide partition coefficients comparable to other small molecule force field. Considering that no parameters have to be adjusted in their derivation and their similar performance to other atomic charge derivation methods we think they provide a promising alternative in the derivation of the next generation small molecule force fields.

Supporting Information

Gromacs input files of all molecules and the calculated hydration and solvation free energies are provided as zip file or can be downloaded from
`https://github.com/evohringer/Supporting\_Information`

References

1. D.L. Mobley, K.L. Wymer, N.M. Lim, J.P. Guthrie, J. Comput. Aided Mol. Des. **28**(3), 135 (2014)
2. H.S. Muddana, M.K. Gilson, J. Comput. Aided Mol. Des. **26**(5), 517 (2012)
3. H.S. Muddana, A.T. Fenley, D.L. Mobley, M.K. Gilson, J. Comput. Aided Mol. Des. **28**(4), 305 (2014)
4. C.C. Bannan, K.H. Burley, M. Chiu, M.R. Shirts, M.K. Gilson, D.L. Mobley, J. Comput. Aided Mol. Des. **30**(11), 927 (2016)
5. A.S. Rustenburg, J. Dancer, B. Lin, J.A. Feng, D.F. Ortwine, D.L. Mobley, J.D. Chodera, J. Comput. Aided Mol. Des. **30**(11), 945 (2016)
6. C.C. Bannan, D.L. Mobley, A.G. Skillman, J. Comput. Aided Mol. Des. **32**(10), 1165 (2018)
7. P.V. Klimovich, D.L. Mobley, J. Comput. Aided Mol. Des. **29**(11), 1007 (2015)
8. M. Işık, D. Levorse, D.L. Mobley, T. Rhodes, J.D. Chodera, bioRxiv (2019)
9. C.C. Bannan, G. Calabr, D.Y. Kyu, D.L. Mobley, J. Chem. Theory Comput. **12**(8), 4015 (2016)
10. A. Daina, O. Michielin, V. Zoete, J. Chem. Inf. Model. **54**(12), 3284 (2014)
11. V. Kundi, J. Ho, J. Phys. Chem. B **123**(31), 6810 (2019)
12. G. Klopman, J.Y. Li, S. Wang, M. Dimayuga, J. Chem. Inf. Comput. Sci. **34**(4), 752 (1994)
13. A.K. Ghose, G.M. Crippen, J. Comput. Chem. **7**(4), 565 (1986)
14. R. Wang, Y. Fu, L. Lai, J. Chem. Inf. Comput. Sci. **37**(3), 615 (1997)
15. T. Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. Zhang, Y. Li, R. Wang, L. Lai, J. Chem. Inf. Model. **47**(6), 2140 (2007)
16. A.V. Marenich, C.J. Cramer, D.G. Truhlar, J. Chem. Theory Comput. **9**(1), 609 (2013)
17. A.V. Marenich, C.J. Cramer, D.G. Truhlar, J. Chem. Theory Comput. **5**(9), 2447 (2009)
18. H.S. Muddana, N.V. Sapra, A.T. Fenley, M.K. Gilson, J. Comput. Aided Mol. Des. **28**(3), 277 (2014)
19. A. Lara, M. Riquelme, E. Vhringer-Martinez, J. Comput. Chem. **39**(22), 1728 (2018)
20. M. Riquelme, A. Lara, D.L. Mobley, T. Verstraelen, A.R. Matamala, E. Vhringer-Martinez, J. Chem. Inf. Model. **58**(9), 1779 (2018)
21. F. Heidar-Zadeh, P.W. Ayers, T. Verstraelen, I. Vinogradov, E. Vhringer-Martinez, P. Bultinck, J. Phys. Chem. A **122**(17), 4219 (2018)
22. F. Neese, Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**(1), 73 (2011)
23. P. Bultinck, C. Van Alsenoy, P.W. Ayers, R. Carbó-Dorca, J. Chem. Phys. **126**(14), 144111 (2007)

24. T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, P.W. Ayers, J. Chem. Theory Comput. **12**(8), 3894 (2016)
25. T. Verstraelen, P. Tecmer, F. Heidar-Zadeh, K. Boguslawski, M. Chan, Y. Zhao, T.D. Kim, S. Vandenbrande, D. Yang, C.E. Gonzlez-Espinoza, S. Fias, P.A. Limacher, D. Berrocal, A. Malek, P.W. Ayers. Horton 2.0.0 (2015). Accessed: January 25, 2018
26. G. Duarte Ramos Matos, D.Y. Kyu, H.H. Loeffler, J.D. Chodera, M.R. Shirts, D.L. Mobley, J. Chem. Eng. Data **62**(5), 1559 (2017)
27. H. Berendsen, J.R. Grigera, T.P. Straatsma, J. Phys. Chem. **91**(24), 6269 (1987)
28. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M. Klein, J. Chem. Phys. **79**, 926 (1983)
29. M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, SoftwareX **1-2**, 19 (2015)
30. W.F. Van Gunsteren, H.J.C. Berendsen, Mol. Simul. **1**(3), 173 (1988)
31. M. Parrinello, A. Rahman, J. Appl. Phys. **52**(12), 7182 (1981)
32. U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, J. Chem. Phys. **103**(19), 8577 (1995)
33. B. Hess, H. Bekker, H. Berendsen, J. Fraaije, J. Comput. Chem. **18**(12), 1463 (1997)
34. M.R. Shirts, D.L. Mobley, J.D. Chodera, V.S. Pande, J. Phys. Chem. B **111**(45), 13052 (2007)
35. P.V. Klimovich, M.R. Shirts, D.L. Mobley, J. Comput. Aided Mol. Des. **29**(5), 397 (2015)
36. M.R. Shirts, J.D. Chodera, J. Chem. Phys. **129**(12), 124105 (2008)