An Accurate Free Energy Method for Solvation of Organic Compounds and Binding to Proteins

Omer Tayfuroglu, Muslum Yildiz, Lee-Wright Pearson and Abdulkadir Kocak[1]

Department of Chemistry, Gebze Technical University, 41400, Kocaeli/Turkey

KEYWORDS: Binding free energy, Covid-19, drug design, molecular dynamics

ABSTRACT. Here, we introduce a new strategy to estimate free energies using single end-state molecular dynamics simulation trajectories. The method is adopted from ANI-1ccx neural network potentials (Machine Learning) for the Atomic Simulation Environment (ASE) and predicts the single point energies at the accuracy of CCSD(T)/CBS level for the entire configurational space that is sampled by Molecular Dynamics (MD) simulations. Our preliminary results show that the method can be as accurate as Bennet-Acceptance-Ration (BAR) with much reduced computational cost. Not only does it enable to calculate solvation free energies of small organic compounds, but it is also possible to predict absolute and relative binding free energies in ligand-protein complex systems. Rapid calculation also enables to screen small organic molecules from databases as potent inhibitors to any drug targets.

[1] Corresponding to: Abdulkadir Kocak, Email: kocak@gtu.edu.tr, Phone: +902626053083

1. Introduction

The novel coronavirus, SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) is a

causative agent of Coronavirus Disease 2019 (COVID-19).[1] COVID-19 is now a global pandemic with unbearable economic burden and puts life of millions at risk around the world. The life cycle of the virus as many other viruses has an essential proteolytic auto-processing step.[2] [3-4] The 3CL-protease takes an important role in freeing individual functional proteins from single-chain polyprotein that has been translated by the host cell translation machinery.[5] Interfering to this step has been shown to have capability of inhibiting virus replication.[6] Therefore, 3CL-protease is one of the valid and most attractive drug targets for Covid-19 treatment.

Developing a drug molecule from scratch for curing a disease is quite expensive and time consuming. Instead, repurposing an FDA approved drug is a much faster and less expensive strategy since the time is very critical parameter in controlling such invasive viruses. Even testing all FDA approved drugs experimentally one by one may cause intolerable waste of the time. To help to ease the problem, computational calculations can be very handy. There have already been several computational studies about predicting an efficient FDA approved drugs that can be repurposed for COVID-19.[7-17]

In searching for drug candidates as inhibitors, there are two core questions that can be addressed: (a) what is the binding mode? (i.e. where does it bind from the enzyme?) (b) what is the binding affinity? (binding free energy).[18] Inhibition of an enzyme can be involved in various mechanisms. For instance, an inhibitor can bind to the enzyme from the active site in the absence of substrate (competitive inhibition) forming an additional or it can bind in the presence of the

substrate (i.e. enzyme-substrate complexed) from the allosteric site (uncompetitive inhibition). Alternatively, it can reduce the activity of the enzyme by binding from allosteric site of the enzyme either free or complex (non-competitive inhibition).[19]

In all these mechanisms the equilibrium constant, K_s belonging to the reaction between the enzyme and substrate is shifted towards the reactants by existing inhibitor in the media. Kinetic studies to understand the activity loss of the enzyme on substrate by the inhibitor report values such as inhibition equilibrium constant, K_i or half maximal inhibitory concentration, IC50. Both these experimental parameters are proportional (they are equal in special cases). Thermodynamic equilibrium constant K_i is related to the standard binding Gibbs free energy of the system by:

$\Delta G^o = -RT\ln(K_i)$

The ΔG^o can also be calculated from the thermodynamic potentials. Thus, it is possible to compare experimental ΔG^o determined from K_i and theoretical ΔG^o predicted by thermodynamic potentials using computational methods. However, logarithmic relation between the K_i and ΔG^o complicates the reliability of the calculations. An error of 3-4 kcal is acceptable for the most computations, but this makes thousands of times less efficient inhibitor. Therefore, most computational methods fail to estimate the experimental binding affinity trend (relative binding free energies) among the inhibitors. Reproducing the absolute standard free energies of binding is even more difficult for the computational methods. Moreover, a trade-off between the computational accuracy and speed must be made.

However, almost all of the studies screen databases listing the FDA approved drugs at the molecular docking level, which is mostly based on geometrical alignment of a ligand into a binding pocket, namely the active site of 3C-like protease ($3CL^{pro}$). Despite the speed of docking

approach, they are very coarse methods and almost never predict the correct experimental binding affinity trend among the inhibitors.[7-17]

More sophisticated methods to calculate the potential binding free energy of inhibitor candidate to the protein ranges from post molecular dynamics simulations such as Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA)[20-23] to perturbation methods such as Bennett acceptance ratio (BAR),[24-31] the latter being much more accurate.

Although calculation of MMPBSA energies is quite fast as it requires only one MD simulation for each Protein-Ligand (PL) complex system and thus applicable to screen many drug candidates, it lacks of explicit water definition and uses the energies at the Molecular Mechanics (MM) level, very low level when compared to QM level definitions.

Perturbation methods like BAR requires several intermediate lambda (λ-) states in which the ligand can be decoupled, annihilated, or pulled. It uses the explicit water definition and its prediction for solvation free energies is quite successful. However, it still uses MM energy terms, and requires non-physical intermediate states. There have been several attempts to correct the MM energy in the BAR calculation by re-weighting the end-states with QM-MM definitions. In these post-BAR methods such as NBB, BAR+TP, only part of the system could be defined at the QM level.[26, 29, 32-36] It could bring very little improvements in to the solvation free energies estimated by BAR.

In addition to BAR and post-BAR methods, others such as Free Energy Perturbation (FEP), Thermodynamic Integration (TI),[25-26, 31, 37-41] Multistate BAR (MBAR)[42-44] are also commonly used and proven to be sufficiently accurate in estimating solvation free energies along with relative and absolute binding free energies. However, all these methods require very high computational cost. Screening all the drug candidates using these methods may not be very

practical. For a typical free energy of binding, these methods require at least 30 λ-states for the decoupling of L from PL complex and another that many for decoupling L from water (i.e. solvation free energy of ligand), which increases the computational cost by 60-folds.

Machine Learning (ML) techniques have attracted a great interest for the past decade due to their successful algorithms to scientific questions including but not limited to chemical reactions,[45-46] potential energy surfaces,[47-50] forces,[51-53] atomization energies,[54-55] and protein-ligand complex scorings.[56]. One of the most promising aspects of the ML techniques is that the trained models can be applied to new systems (transferrable).[57] Recently several models using active learning such as ANI-1,[58] ANI-1x[59] and ANI-1cxx[60] have been trained to calculate DFT and CCSD(T) energies of small organic compounds containing C, H, O and N atoms in non-equilibrium conformations. It has been shown to reproduce the energies at the accuracy of the CCSD(T)/CBS level with billions of time faster than the actual QM calculations.[57]

Here, we introduce a new strategy to estimate free energies of solvation of small organic compounds and binding to proteins in explicit solvent using single end-state MD simulations. The method is adopted from ANI-1ccx neural network potentials (Machine Learning) for the Atomic Simulation Environment (ASE) and predicts the single point energies of the entire system at the accuracy of the CCSD(T)/CBS level. Our results show the method can be sufficiently accurate to measure solvation/binding free energies and fast making it applicable to screen FDA approved drugs.

Theory

Linear Response Approximation

The insertion of the ligand to an environment of solvent (solvation free energy) or receptor (binding free energy) can be defined by a coupling parameter, λ.[61] At each λ-state, the potential energy of the system would be:

$$U(\lambda) = U_0 + \lambda.V$$

where V is the interaction energy between the ligand and environment. U_0 is the reference potential energy, which includes all energy terms except for the ligand-environment interaction. The partition function is;

$$Z(\lambda) = e^{-F(\lambda)/kT} = \int e^{-\lambda V/kT} e^{-U_0/kT}$$

where F is the free energy depending on λ-state. Re-organizing the above equation,

$$F(\lambda) = -\frac{1}{\beta} ln[Z(\lambda)] \text{ where } \beta = \frac{1}{kT}$$

Thus,

$$\frac{\partial F}{\partial \lambda} = \langle V \rangle_\lambda$$

The equation above is only valid when the ligand-environment interaction is proportional to coupling term λ. This results in;

$$\Delta F = \alpha . \langle V \rangle_1$$

In the limiting case, this equation yields as;

$$\frac{\Delta F}{\langle V \rangle_1} = 1/2 \text{ if } \langle V \rangle_0 \ll \frac{\partial^2 F}{\partial \lambda^2} .^{61}$$

The equation implies that the free energy change is nearly half-the potential between the ligand-environment interaction in the complex configurational ensemble.[20] So for the de-coupling process;

$$\Delta F = \frac{1}{2}\langle V \rangle_0$$

Here, we replaced the ML estimated CCSD(T)/CBS energies with MM energies at the λ=0 (i.e. coupled/bound) configurations. For each MD frame, three different energies were calculated $\Delta U_{complex}$, the ligand is complexed with environment (solvent or protein), ΔU_{lig}, bare ligand that is extracted from the MD frame, ΔU_{env}, the environment that is also extracted from the same MD frame which included the parts of either solvent (in the case of solvation free energy) or protein + solvent (in the case of binding free energy). And combining these three enrgy terms by $\Delta U = \Delta U_{complex} - (\Delta U_{lig} + \Delta U_{env})$. Finally, the averaged ensemble for the free energy was calculated by the standard methods introduced above using only one-endstate simulations. Interestingly, we found that experimental free energy (of solvation or binding) is approximately half of our calculated free energy. One should note that the energies calculated are of single point energies at the ML estimated CCSD(T)/CBS level in the clasical MD configurations. It does not include the Zero Point Vibrational Energy (ZPVE), entropy and PV terms.

.

Methods

System Preparation

For the binding studies, the crystal structures of $3CL^{pro}$ of SARS-CoV-2 of which the IC50 values are known were retrieved from Protein Data Bank (PDB). PDB IDs: 2GZ7, 4YOJ, 4MDS, 2ALV, 4WY3. Monomer structures were aligned and H-atoms were added using PyMol. The pKa values were calculated using Propka 3.1. The ligands were scanned for conformers. For the solvation free energies, we calculated molecules that have extensively available data in literature[34] in order for a comparison. Using Gaussian 16 software, the ligands were first optimized at the B3LYP/6-31G* and ESP charges belonging to the conformer with the minimum

energy for each ligand were generated at the HF/6-31G* level. Using Antechamber, RESP charges and GAFF force field atom types were generated. Acypype was used to convert Amber type input files to Gromacs.

MD Simulations

The molecular dynamics simulations were carried out using Gromacs 2018 software package [62] with all-atom model of Amber ff99SB-ILDN force field [63] implemented in Gromacs.

For the binding studies, the protein-ligand complex (~4700 atoms) was placed in the center of a dodecahedron box. For the solvation studies, the ligand was placed in the center of a cubic box. Each system was solvated in the TIP3P model type water [64] with a cell margin distance of 10 Å for each dimension. The protein-ligand systems with ~70,000 atoms were neutralized and salted by 0.15 M NaCl. The neutral ligand only systems were not salted. Classical harmonic motions of all bonds were constrained to their equilibrium values with LINCS algorithm.

Energy minimization was carried out to a maximum 100 $kJ.mol^{-1}.nm^{-1}$ force using Verlet cutoff scheme. For both long range electrostatic and Van der Waals interactions, a cutoff length of 12 Å was used with the Particle Mesh Ewald method (PME) (6^{th} order interpolation).[65] The neighbor list update frequency was set to 1 ps^{-1}. As with our earlier studies,[66-67] a stepwise energy minimization and equilibration schemes were used. Each minimization step consisted of a 5000 cycle of Steepest Descent and a subsequent 5000 cycles of l-bfgs integrators.

After minimization, each system was equilibrated within three steps using Langevin Dynamics. The first step consisted of a 1 ns of NVT ensemble. Experimental studies reported the solvation free energies at 298.15 K and binding free energies at 310 K. We have adjusted our simulations accordingly. The next steps consisted of NPT ensembles in which the systems were equilibrated to the 1 atm pressure by Berendsen for 200 ps and followed by Parrinello-

Rahman isotropic pressure coupling for 1 ns to a reference pressure of 1 atm. When systems were reached to the equilibrium, an MD simulation of ~10 ns were carried out.

Free Energy Calculations

For the BAR calculations, 10 equal decoupling steps (i.e. decoupling the ligand from the environment) of each coulombic and van der waals interactions, respectively (total 21 λ-windows) were used. For the BAR calculations of protein-ligand complex, a separate decoupling of the ligand from ligand+water system with the same parameters were performed and subtracted from the decoupling of ligand from protein+ligand+water complex system.[67]

MM-PBSA calculations were performed using g_mmpbsa script by Kumari et al.[21, 68]. For the calculations of protein-ligand complexes, we used the l=0 simulations of protein+ligand+water complexes. Only protein+ligand complex part from the trajectories were extracted and default parameters of implicit water with ε=80 and solute with ε=4 were used. Non-linear Poisson-Boltzmann equation with SASA model were used.

ML-QM calculations were performed using our custom build python scripts. For the calculations of solvation free energies, we used λ=0 simulations of ligand+water complexes whereas for the binding free energy calculations, we used λ=0 simulations of protein+ligand+water complexes.

$$\Delta G_{solv} = \langle G_{L(aq)} - G_{L(g)} - G_{(aq)} \rangle_{L(aq)}$$

$$\Delta G_{bind} = \langle G_{PL(aq)} - G_{L(g)} - G_{P(aq)} \rangle_{PL(aq)}$$

Results and discussion

Solvation Free Energies

In order to test the ML-QM method, we run 10-ns MD simulations for the molecules whose solvation free energies have been extensively studied. We first built a system in which the ligand

is replaced in the center of box and the box is solvated with just single water molecule. The ligands studied are given in Figure 1.

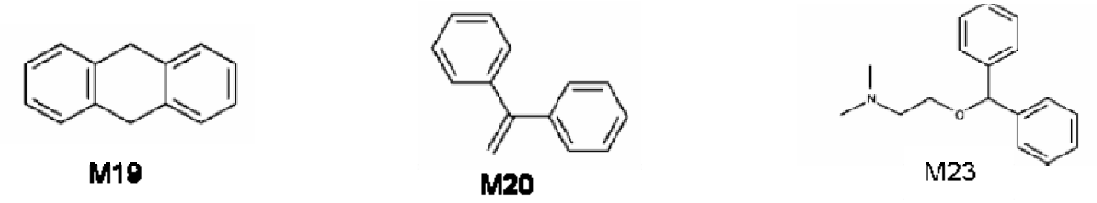


Figure 1 Molecules studied in this work for their solvation free energies

All the sampled configurations by the classical MD simulations were calculated for the single point energies using ML-QM method. Although the individual frames have very large energy fluctuations, the average energy of the system goes smoothly (Figure X)

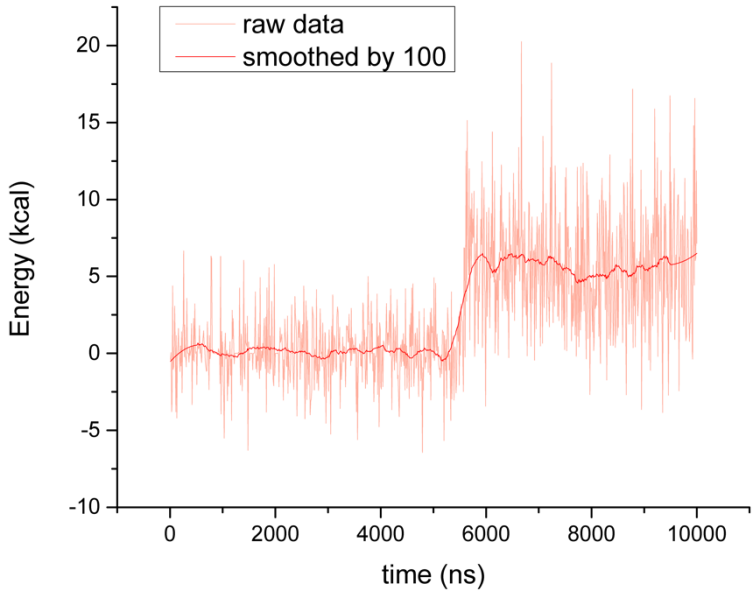


Figure 2 The ML-QM interaction energy calculated from the trajectory of MD simulation of M19 with 100 water molecules.

Since the degree of freedom for the water molecule is much more in the case of unbound state in which water molecule is too far from the ligand and thus not interacting with the ligand, the MD could sample only unbound state. Therefore, ML-QM energy for the binding was almost 0.

We prepared other independent simulations in which we increased the number water molecules (nH_2O) in the box. Each time the average ML-QM energies were almost 0. This has been the case until nH_2O=100. Figure 3a shows the interaction energy between the ligand and water molecules calculated by ML-QM energy difference: $\Delta U_{ML-QM} = U_{ML-QM}^{lig-wat} - \left(U_{ML-QM}^{lig} + U_{ML-QM}^{wat}\right)$. Here, $U_{ML-QM}^{lig-wat}$ is the total energy of the system including ligand and water. U_{ML-QM}^{lig} and $U_{ML-QM}^{lig} + U_{ML-QM}^{wat}$ are the energies of free ligand and free water molecules, respectively, which have been calculated by extracting from the trajectory of complex structure. When we kept adding more water in the system and running other MD simulations, we found that starting from nH_2O=100, the system undergoes from the unbound state to bound state by a drastic energy jump after ~9.5 ns. The degree of freedom for the water molecules in the bound state must be limited by the apparent interaction with the ligand. The same jump occurred for all of the other MD runs with increased amount of water molecules up to 200 (Figure 3b) but in the earlier time. Further addition of water molecules in the system causes the system to stay in the bound state all the time (Figure 3c). It is very apparent that the unbound states and bound states have the same averaged ensembles (Figure 3d).

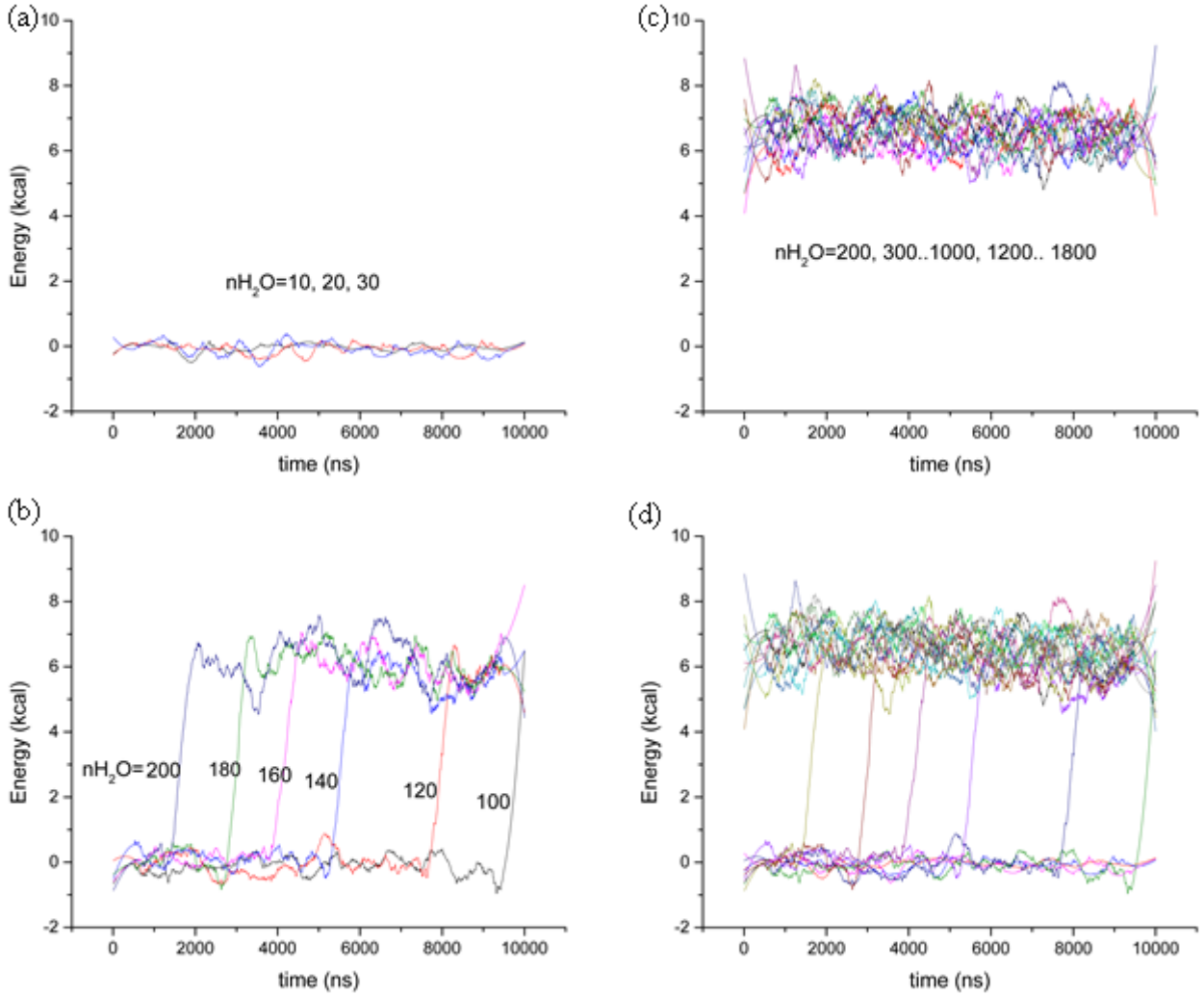
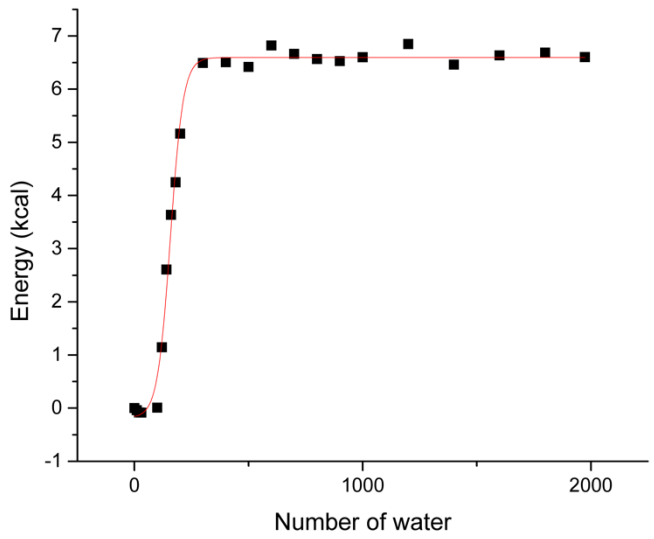


Figure 3. ML-QM interaction energy of M19 with different number of water molecules. Each data consists of the moving average of 100 points over 10,000 data set. (a) no interaction energy when the number of water molecules are much less than the configurational degree of freedom (b) the shift of the jump to an earlier time from unbound to bound state by addition of water molecules. (c) the system to stay in the bound state all the time when more than sufficient water molecules available. (d) the average energies of bound states to be all the same.

The average ML-QM energy of the entire MD simulation time can also be compared for each of the simulations with different number of water molecules. As in the case of simulations where

bound-unbound state jump occurred, the energy profile of the average energies fit into a sigmoidal plot in which the inflection point corresponds to the experimental solvation free energies. The Figure 4 shows the average energies with respect to number of waters in each simulations.



Model	Boltzmann	
Equation	y = A2 + (A1-A2)/(1 + exp((x-x0)/dx))	
Reduced Chi-Sqr	0.05	
Adj. R-Square	0.99	
	Value	Std Err.
A1	-0.18	0.12
A2	6.59	0.06
x0	157.34	2.46
dx	28.36	2.41
span	6.78	0.14
ΔG	**3.39**	**0.18**

Figure 4 The ML-QM interacting energy depending on the number of water. The square marks show the average interaction energy over the entire MD simulations for each M19-nH_2O complex (n=0, 10…, 30, 100, 120…200, 300…1000, 1200…1800, 1974)

It is apparent that the ML-QM interaction energy is constant for the simulations with more than sufficient waters exist. The empirically determined linearity constant is α=1/2, thus the free energy change is approximately half of the interaction energy between the ligand and environment.

Table 1 shows the experimental solvation free energies along with estimated by ML-QM and other techniques available in literature. ML-QM method significantly improve the results surpassing the conventional methods like BAR and MBAR.

Table 1 Experimental and calculated solvation free energies (kcal/mol). * values are from Ref.[34]

Mol. ID	Experiment	ML-QM	BAR	BAR*	NBB*	BAR+TP*	MBAR*
19	−3.8±0.1	-3.39	-3.63	−3.2	−3.0	−2.9	−3.2
20	−2.8±0.1	-3.35		−2.9	−1.1	−1.4	−2.9
23	−9.3±0.6	-9.35	-10.25	−3.8	−5.6	−5.6	−3.8

Absolute Binding Free Energies

Using the limiting case, we could successfully calculate the binding of available inhibitors to 3CLprotease of SARS-CoV-2, of which the IC50 values available on literature. We compared the success of the method with other two common computational methods.

Table 2 Experimental inhibition profiles of the 3CL protease enzymes SARS-CoV-2 compared with absolute binding free energies. Experimental data were calculated by –RTln(IC50) at 310 K.

PDB ID	Experimental	BAR	MMPBSA	ML-QM
2GZ7	-9.25	-9.78	-39.88	
4YOJ	-9.06	-8.32	-37.25	-7.70
4MDS	-7.37	-11.15	-38.24	-6.31
2ALV	-5.89	-21.03	-72.58	-4.42
4WY3	-5.13	-9.82	-56.63	-9.37

It should be noted that the absolute binding free energy calculations may not directly reproduce experimental values, since experimental values are of IC50 values rather than the pKi values. However, considering the experimental conditions are more or less similar, the energy profile

(i.e. the trend among the inhibitors) should at least be reproduced. Here, we see that ML-QM is the most accurate method to yield the correct trend among the compared methods (Figure 5).

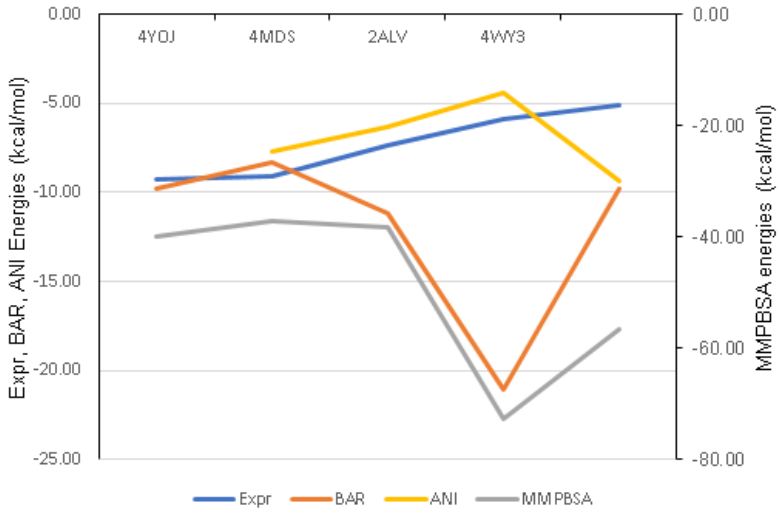


Figure 5 Calculated vs experimental binding profiles of inhibitors.

It should be noted that ML-QM and MMPBSA energies use single MD simulation trajectory, in which the ligand is fully interacting with the complex. The unbound states were created by removing the ligand from environment. Although the MMPBSA is as fast as ML-QM, the results are only meaningful in terms of relative binding energies. On the other hand, the BAR method uses 21 λ-states for the decoupling the ligand from the environment (protein and water) and another 21 λ -states for the solvation free energy of the ligand. The ML-QM surpasses both MMPBSA and BAR methods with increased speed and accuracy.

Summary and Conclusions

Here we applied the Machine Learning techniques to calculate the experimental solvation and binding energies with increased speed at the accuracy of CCSD(T)/CBS level. The results show that the predicted binding energies are half of the interaction energies, which is a limiting case of LIE and reproduce the experiments better that well known conventional methods like BAR and

MMPBSA. Since the method is fast and accurate, this method opens an area in the search of Covid-19 related inhibitors.

References

1. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* **2020,** *382* (8), 727-733.

2. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus Main Proteinase ($3CL^{pro}$) Structure: Basis for Design of Anti-SARS Drugs. *Science* **2003,** *300* (5626), 1763-1767.

3. Ziebuhr, J.; Snijder, E. J.; Gorbalenya, A. E. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *Journal of General Virology* **2000,** *81* (4), 853-879.

4. Marra, M. A.; Jones, S. J. M.; Astell, C. R.; Holt, R. A.; Brooks-Wilson, A.; Butterfield, Y. S. N.; Khattra, J.; Asano, J. K.; Barber, S. A.; Chan, S. Y.; Cloutier, A.; Coughlin, S. M.; Freeman, D.; Girn, N.; Griffith, O. L.; Leach, S. R.; Mayo, M.; McDonald, H.; Montgomery, S. B.; Pandoh, P. K.; Petrescu, A. S.; Robertson, A. G.; Schein, J. E.; Siddiqui, A.; Smailus, D. E.; Stott, J. M.; Yang, G. S.; Plummer, F.; Andonov, A.; Artsob, H.; Bastien, N.; Bernard, K.; Booth, T. F.; Bowness, D.; Czub, M.; Drebot, M.; Fernando, L.; Flick, R.; Garbutt, M.; Gray, M.; Grolla, A.; Jones, S.; Feldmann, H.; Meyers, A.; Kabani, A.; Li, Y.; Normand, S.; Stroher, U.; Tipples, G. A.; Tyler, S.; Vogrig, R.; Ward, D.; Watson, B.; Brunham, R. C.; Krajden, M.; Petric, M.; Skowronski, D. M.; Upton, C.; Roper, R. L. The Genome Sequence of the SARS-Associated Coronavirus. *Science* **2003,** *300* (5624), 1399-1404.

5. Hegyi, A.; Ziebuhr, J. Conservation of substrate specificities among coronavirus main proteases. *Journal of General Virology* **2002,** *83* (3), 595-599.

6. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020,** *368* (6489), 409-412.

7. Abdelli, I.; Hassani, F.; Brikci, S. B.; Ghalem, S. In silico study the inhibition of angiotensin converting enzyme 2 receptor of COVID-19 by Ammoides verticillata components harvested from Western Algeria. *Journal of Biomolecular Structure & Dynamics.*

8. Azeez, S. A.; Alhashim, Z. G.; Al Otaibi, W. M.; Alsuwat, H. S.; Ibrahim, A. M.; Almandil, N. B.; Borgio, J. F. State-of-the-art tools to identify druggable protein ligand of SARS-CoV-2. *Archives of Medical Science* **2020,** *16* (3), 497-507.

9. Belhassan, A.; Chtita, S.; Zaki, H.; Lakhlifi, T.; Bouachrine, M. Molecular docking analysis of N-substituted oseltamivir derivatives with the SARS-Cov-2 main protease. *Bioinformation* **2020,** *16* (5), 404-408.

10. Elmezayen, A. D.; Al-Obaidi, A.; Sahin, A. T.; Yelekci, K. Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *Journal of Biomolecular Structure & Dynamics*.

11. Enayatkhani, M.; Hasaniazad, M.; Faezi, S.; Guklani, H.; Davoodian, P.; Ahmadi, N.; Einakian, M. A.; Karmostaji, A.; Ahmadi, K. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *Journal of Biomolecular Structure & Dynamics*.

12. Hakmi, M.; Bouricha, E.; Kandoussi, I.; El Harti, J.; Ibrahimi, A. Repurposing of known anti-virals as potential inhibitors for SARS-CoV-2 main protease using molecular docking analysis. *Bioinformation* **2020,** *16* (4), 301-305.

13. Ibrahim, I. M.; Abdelmalek, D. H.; Elshahat, M. E.; Elfiky, A. A. COVID-19 spike-host cell receptor GRP78 binding site prediction. *Journal of Infection* **2020,** *80* (5), 554-562.

14. Kandeel, M.; Al-Nazawi, M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life Sciences* **2020,** *251*.

15. Reiner, Z.; Hatamipour, M.; Banach, M.; Pirro, M.; Al-Rasadi, K.; Jamialahmadi, T.; Radenkovic, D.; Montecucco, F.; Sahebkar, A. Statins and the COVID-19 main protease: in silico evidence on direct interaction. *Archives of Medical Science* **2020,** *16* (3), 490-496.

16. Sampangi-Ramaiah, M. H.; Vishwakarma, R.; Shaanker, R. U. Molecular docking analysis of selected natural products from plants for inhibition of SARS-CoV-2 main protease. *Current Science* **2020,** *118* (7), 1087-1092.

17. Sudeep, H. V.; Gouthamchandra, K.; Shyamprasad, K. Molecular docking analysis of Withaferin A from Withania somnifera with the Glucose regulated protein 78 (GRP78) in comparison with the COVID-19 main protease. *Bioinformation* **2020,** *16* (5), 411-416.

18. Limongelli, V. Ligand binding free energy and kinetics calculation in 2020. *Wires Comput Mol Sci* **2020**.

19. Burlingham, B. T.; Widlanski, T. S. An intuitive look at the relationship of K-i and IC50: A more general use for the Dixon plot. *J Chem Educ* **2003,** *80* (2), 214-218.

20. Genheden, S.; Ryde, U. The Mm/Pbsa and Mm/Gbsa Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015,** *10*, 449.

21. Kumari, R.; Kumar, R.; Lynn, A. g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J Chem Inf Model* **2014,** *54* (7), 1951-1962.

22. Rifai, E. A.; van Dijk, M.; Vermeulen, N. P. E.; Yanuar, A.; Geerke, D. P. A Comparative Linear Interaction Energy and MM/PBSA Study on SIRT1-Ligand Binding Free Energy Calculation. *J Chem Inf Model* **2019,** *59* (9), 4018-4033.

23. Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys J* **2004,** *86* (1), 67-74.

24. Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976,** *22*, 245.

25. Bruckner, S.; Boresch, S. Efficiency of Alchemical Free Energy Simulations I: Practical Comparison of the Exponential Formula, Thermodynamic Integration and Bennett's Acceptance Ratio Method. *J. Comput. Chem.* **2011,** *32*, 1303.

26. de Ruiter, A.; Boresch, S.; Oostenbrink, C. Comparison of thermodynamic integration and Bennett's acceptance ratio for calculating relative protein-ligand binding free energies. *J. Comput. Chem.* **2013,** *34*, 1024.

27. Habeck, M. Bayesian Estimation of Free Energies From Equilibrium Simulations. *Phys Rev Lett* **2012,** *109* (10).

28. König, G.; Bruckner, S.; Boresch, S. Unorthodox Uses of Bennett's Acceptance Ratio Method. *J. Comput. Chem.* **2009,** *30*, 1712.

29. König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L. Multiscale Free Energy Simulations: An Efficient Method for Connecting Classical MD Simulations to QM or QM/MM Free Energies Using Non-Boltzmann Bennett Reweighting Schemes. *J Chem Theory Comput* **2014,** *10* (4), 1406-1419.

30. Luzhkov, V. B. On Relation between the Free-Energy Perturbation and Bennett's Acceptance Ratio Methods: Tracing the Influence of the Energy Gap. *J. Chem. Phys.* **2010,** *132*, 194104.

31. Shirts, M. R.; Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.* **2005,** *122*, 144107-1.

32. Cui, D.; Zhang, B. W.; Tan, Z. Q.; Levy, R. M. Ligand Binding Thermodynamic Cycles: Hysteresis, the Locally Weighted Histogram Analysis Method, and the Overlapping States Matrix. *J Chem Theory Comput* **2020,** *16* (1), 67-79.

33. Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O'Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; Swaney, D. L.; Tummino, T. A.; Huettenhain, R.; Kaake, R. M.; Richards, A. L.; Tutuncuoglu, B.; Foussard, H.; Batra, J.; Haas, K.; Modak, M.; Kim, M.; Haas, P.; Polacco, B. J.; Braberg, H.; Fabius, J. M.; Eckhardt, M.; Soucheray, M.; Bennett, M. J.; Cakir, M.; McGregor, M. J.; Li, Q.; Meyer, B.; Roesch, F.; Vallet, T.; Mac Kain, A.; Miorin, L.; Moreno, E.; Naing, Z. Z. C.; Zhou, Y.; Peng, S.; Shi, Y.; Zhang, Z.; Shen, W.; Kirby, I. T.; Melnyk, J. E.; Chorba, J. S.; Lou, K.; Dai, S. A.; Barrio-Hernandez, I.; Memon, D.; Hernandez-Armenta, C.; Lyu, J.; Mathy, C. J. P.; Perica, T.; Pilla, K. B.; Ganesan, S. J.; Saltzberg, D. J.; Rakesh, R.; Liu, X.; Rosenthal, S. B.; Calviello, L.; Venkataramanan, S.; Liboy-Lugo, J.; Lin, Y.; Huang, X.-P.; Liu, Y.; Wankowicz, S. A.; Bohn, M.; Safari, M.; Ugur, F. S.;

Koh, C.; Savar, N. S.; Tran, Q. D.; Shengjuler, D.; Fletcher, S. J.; O'Neal, M. C.; Cai, Y.; Chang, J. C. J.; Broadhurst, D. J.; Klippsten, S.; Sharp, P. P.; Wenzell, N. A.; Kuzuoglu, D.; Wang, H.-Y.; Trenker, R.; Young, J. M.; Cavero, D. A.; Hiatt, J.; Roth, T. L.; Rathore, U.; Subramanian, A.; Noack, J.; Hubert, M.; Stroud, R. M.; Frankel, A. D.; Rosenberg, O. S.; Verba, K. A.; Agard, D. A.; Ott, M.; Emerman, M.; Jura, N.; von Zastrow, M.; Verdin, E.; Ashworth, A.; Schwartz, O.; d'Enfert, C.; Mukherjee, S.; Jacobson, M.; Malik, H. S.; Fujimori, D. G.; Ideker, T.; Craik, C. S.; Floor, S. N.; Fraser, J. S.; Gross, J. D.; Sali, A.; Roth, B. L.; Ruggero, D.; Taunton, J.; Kortemme, T.; Beltrao, P.; Vignuzzi, M.; García-Sastre, A.; Shokat, K. M.; Shoichet, B. K.; Krogan, N. J. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**.

34. Jia, X. Y.; Wang, M. T.; Shao, Y. H.; Konig, G.; Brooks, B. R.; Zhang, J. Z. H.; Mei, Y. Calculations of Solvation Free Energy through Energy Reweighting from Molecular Mechanics to Quantum Mechanics. *J Chem Theory Comput* **2016,** *12* (2), 499-511.

35. Boresch, S.; Woodcock, H. L. Convergence of single-step free energy perturbation. *Molecular Physics* **2017,** *115* (9-12), 1200-1213.

36. König, G.; Bruckner, S.; Boresch, S. Absolute Hydration Free Energies of Blocked Amino Acids: Implications for Protein Solvation and Stability. *Biophys. J.* **2013,** *104*, 453.

37. Lawrenz, M.; Wereszczynski, J.; Ortiz-Sánchez, J. M.; Nichols, S. E.; McCammon, J. A. Thermodynamic integration to predict host-guest binding affinities. *J. Comput.-Aided Mol. Des.* **2012,** *26*, 569.

38. Steinbrecher, T.; Joung, I.; Case, D. A. Soft-core potentials in thermodynamic integration: Comparing one-and two-step transformations. *J. Comput. Chem.* **2011,** *32*, 3253.

39. Bruckner, S.; Boresch, S. Efficiency of Alchemical Free Energy Simulations II: Improvements for Thermodynamic Integration. *J. Comput. Chem.* **2011,** *32*, 1320.

40. Kästner, J.; Senn, H.; Thiel, S.; Otte, N.; Thiel, W. QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: Application to an enzymatic reaction. *J. Chem. Theory Comput.* **2006,** *2*, 452.

41. Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994,** *100*, 9025.

42. Gumbart, J. C.; Chipot, C.; Roux, B. Efficient Determination of Protein-protein Standard Binding Free Energies from First Principles. *J. Chem. Theory Comput.* **2013,** *9*, 3789.

43. Gumbart, J.; Chipot, C.; Schulten, K. Free-Energy Cost for Translocon-Assisted Insertion of Membrane Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108*, 3596.

44. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005,** *26*, 1781.

45. Klucznik, T.; Mikulak-Klucznik, B.; McCormack, M. P.; Lima, H.; Szymkuć, S.; Bhowmick, M.; Molga, K.; Zhou, Y.; Rickershauser, L.; Gajewska, E. P.; Toutchkine, A.; Dittwald, P.; Startek, M. P.; Kirkovits, G. J.; Roszak, R.; Adamski, A.; Sieredzińska, B.; Mrksich, M.; Trice, S. L. J.; Grzybowski, B. A. Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem* **2018,** *4* (3), 522-532.

46. Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **2018,** *360* (6385), 186-190.

47. Yao, K.; Herr, J. E.; Toth, David W.; McKintyre, R.; Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chemical Science* **2018,** *9* (8), 2261-2269.

48. Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci Adv* **2017,** *3* (5), e1603015-e1603015.

49. Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nature Communications* **2017,** *8* (1), 13890.

50. Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angewandte Chemie International Edition* **2017,** *56* (42), 12828-12840.

51. Li, Z.; Kermode, J. R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys Rev Lett* **2015,** *114* (9), 096405.

52. Glielmo, A.; Sollich, P.; De Vita, A. Accurate interatomic force fields via machine learning with covariant kernels. *Physical Review B* **2017,** *95* (21), 214302.

53. Kruglov, I.; Sergeev, O.; Yanilkin, A.; Oganov, A. R. Energy-free machine learning force field for aluminum. *Scientific Reports* **2017,** *7* (1), 8512.

54. Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys Rev Lett* **2012,** *108* (5), 058301.

55. Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J Chem Theory Comput* **2017,** *13* (11), 5255-5264.

56. Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein–Ligand Scoring with Convolutional Neural Networks. *J Chem Inf Model* **2017,** *57* (4), 942-957.

57. Justin, S. S.; Benjamin T., N.; Roman, Z.; Nicholas, L.; Christian, D.; Kipton, B.; Sergei, T.; Olexandr, I.; Adrian, R. *Outsmarting Quantum Chemistry Through Transfer Learning* 2018.

58. Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific Data* **2017,** *4* (1), 170193.

59. Smith, J. S.; Roitberg, A. E.; Isayev, O. Transforming Computational Drug Discovery with Machine Learning and AI. *ACS Med. Chem. Lett.* **2018,** *9*, 1065.

60. Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: Sampling chemical space with active learning. *The Journal of Chemical Physics* **2018,** *148* (24), 241733.

61. Su, Y.; Gallicchio, E.; Das, K.; Arnold, E.; Levy, R. M. Linear interaction energy (LIE) models for ligand binding in implicit solvent: Theory and application to the binding of NNRTIs to HIV-1 reverse transcriptase. *J Chem Theory Comput* **2007,** *3* (1), 256-277.

62. 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.

63. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010,** *78*, 1950.

64. Toukan, K.; Rahman, A. Molecular-dynamics study of atomic motions in water. *Physical Review B* **1985,** *31* (5), 2643-2648.

65. Darden, T. A.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993,** *98*, 10089.

66. Kocak, A.; Erol, I.; Yildiz, M.; Can, H. Computational insights into the protonation states of catalytic dyad in BACE1–acyl guanidine based inhibitor complex. *Journal of Molecular Graphics and Modelling* **2016,** *70*, 226-235.

67. Kocak, A.; Yildiz, M. Docking, molecular dynamics and free energy studies on aspartoacylase mutations involved in Canavan disease. *Journal of Molecular Graphics and Modelling* **2017,** *74*, 44-53.

68. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **2001,** *98* (18), 10037-41.