Virtual Double-System Single-Box for Absolute Dissociation Free Energy Calculations in GROMACS

Marina Macchiagodena,[*,†] Maurice Karrenbrock,[†,‡] Marco Pagliai,[†] and Piero Procacci[*,†]

†*Dipartimento di Chimica "Ugo Schiff", Università degli Studi di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy*

‡*Present address: Institut des Sciences Pharmaceutiques De Suisse Occidentale-Université de Genève, CMU – Rue Michel Servet 1, CH-1211 Genève 4, Switzerland*

E-mail: marina.macchiagodena@unifi.it; piero.procacci@unifi.it

Abstract

We describe a step-by-step protocol for the computation of absolute dissociation free energy with GROMACS code and PLUMED library, which exploits a combination of advanced sampling techniques and nonequilibrium alchemical methodologies. The computational protocol has been automated through an open source python middleware (HPC_Drug) which allows to set up the GROMACS/PLUMED input files for execution on high performing computing facilities. The proposed protocol, by exploiting its inherent parallelism and the power of the GROMACS code on graphical processing units, has the potential to afford accurate and precise estimates of the dissociation constants in drug-receptor systems described at the atomistic level. The procedure has been applied to the calculation of the absolute dissociation free energy of PF-07321332, an oral antiviral proposed by Pfizer, with the main protease ($3CL^{pro}$) of SARS-CoV-2.

1 Introduction

In the last decades various computational methodologies, based on molecular dynamics (MD) simulations with explicit solvent, have been devised to improve (beyond the traditional docking approach) the calculation of absolute dissociation free energy in drug receptor systems. In this context, the so-called alchemical approach[1] emerged recently as one of the best-automated and most widely used MD-based free energy methods. The alchemical protocol relies on a stratification of non physical intermediate states gradually connecting the ligand-environment potential energy function of the target end-states,[2] recovering the corresponding free energy change by a sum of free energy contributions evaluated along the stratification using free energy perturbation (FEP).[3] Alchemical FEP, due to well know sampling issues especially at low ligand-environment coupling,[4] is generally applied in industrial applications to the calculation of *relative* binding free energies (RBFE),[5–8] whereby a ligand is *transmuted* into a strictly congeneric compound.

As recently noted,[8,9] a reliable estimate of the *absolute* dissociation free energy (ADFE) via MD is a key requirement for virtual screening campaigns in industrial and academic settings in order to evaluate distant hits that are not easily amenable for RBFE calculations. Here we present the fine-tuning of nonequilibrium (NE) alchemical transformations for ADFE calculations using the program GROMACS[10] on high performing computing (HPC) platforms. The approach is called NEW-vDSSB. The acronym NEW[2] stands for Non-Equilibrium Work. The work histograms, obtained by a swarm of rapid and independent NE alchemical trajectories connecting the target thermodynamic end-states, are related to the corresponding free energy via the Jarzynski[11] and Crooks[12] theorems. vDSSB stands for *virtual* double-system-single-box, a technique recently developed by some of us[13–15] as a variant of the so-called double-system-single-box (DSSB) method.[16,17]

NEW-vDSSB has been first implemented in the ORAC molecular dynamics package[18] and has been successfully applied, for example, to the calculation of absolute binding free energy (ABFE) of ligands of the SARS-CoV-2 main protease.[13] In this note, we describe

an automated protocol, managed by a python middleware (HPC_Drug), for implementing the NEW-vDSSB method on heterogeneous (GPU/CPU) HPC facilities using the GROMACS popular molecular dynamics package[10,19,20] and the PLUMED library,[21,22] requiring no intervention on the GROMACS source code.

The paper is organized as follows: we first provide some basic information on the fast switching alchemical technique employed for ADFE calculations. We then describe the computational steps to use NEW-vDSSB methodology with GROMACS. For each step, we make available the corresponding input files locally produced by the automated HPC_Drug python middleware for execution on the remote HPC facility. In the last section, as a real-world example, we have calculated on the CINECA[23] M100 CPU-GPU heterogeneous platform the ADFE of PF-07321332 against the SARS-CoV-2 main protease ($3CL^{pro}$). PF-07321332 is a $3CL^{pro}$ inhibitor proposed by Pfizer and exhibiting nanomolar affinity and capable of suppressing virus replication in human cells at sub-micromolar concentrations.[24]

2 Background

Our protocol for ADFE calculation is based on the vDSSB approach. In a nutshell, vDSSB is a NE variant of the alchemical technique whereby the ADFE is computed in a thermodynamic cycle as the difference of the ligand solvation free energy in the bulk solvent and in the solvated complex. vDSSB relies on the *enhanced sampling* of the unbound (decoupled-ligand in bulk) and bound (coupled protein-bound ligand) end-states, followed by the NE step, consisting in the production of two swarms of NE independent trajectories where the ligand is gradually recoupled (unbound leg) and decoupled (bound leg) via a driven alchemical parameter λ, producing two samples of n_u growth and n_b annihilation work values. The n_b and n_u work values are combined as two independent random variables obtaining $n_b \times n_u$ work values, referring to the process of bringing the ligand *from* the bound state *to* the bulk solvent. The convolution of the two work histograms allows to

increase significantly the resolution of resulting work histogram with benefit for accuracy and precision of the ADFE estimate. At variance with DSSB, in vDSSB different protocol for the bound and the unbound state can be used without violation of the Crooks[12] and Jarzynski[11] theorems, hence allowing to choose the optimal box size and time protocol according to the physical dimension of the solute and the nature of the environment. As such, vDSSB is an inherently parallel algorithm, affording very efficient performances on massively parallel facilities. The free energy and the corresponding confidence interval are recovered, in a fast post-processing stage, by constructing the convolution of the recoupling (unbound leg) and decoupling (bound leg) work histograms to which the Jarzynski theorem is applied, and by correcting for a standard state dependent term related to the translational binding site volume. For further details on vDSSB theory and technicalities, we refer to Refs.[13,15,25] vDSSB, as any other MD-based technology for ADFE calculations, requires expertise and know-how which may deter the average end-user. In the following we describe a step-by-step protocol where the most complex tasks in the vDSSB workflow have been automated by an *ad-hoc* devised python middleware.

3 ADFE calculation workflow

The vDSSB procedure for the HPC calculation of ADFE's (schematically described in Figure 1) is made up by 6 consecutive steps, starting from the 3D structure of receptor and ligand. In the following, we provide an overview of the vDSSB steps to arrive at the ADFE estimate. A companion step-by-step technical tutorial for the calculation of the ADFE in the PF-07321332-3CL^{pro} system is provided in the Supporting Information (SI).

3.1 Step 1: Docking (local)

Assumining that the experimental co-crystal structure of the complex structure is not available, we generate the initial bound state pose by way a preliminary docking calculation of

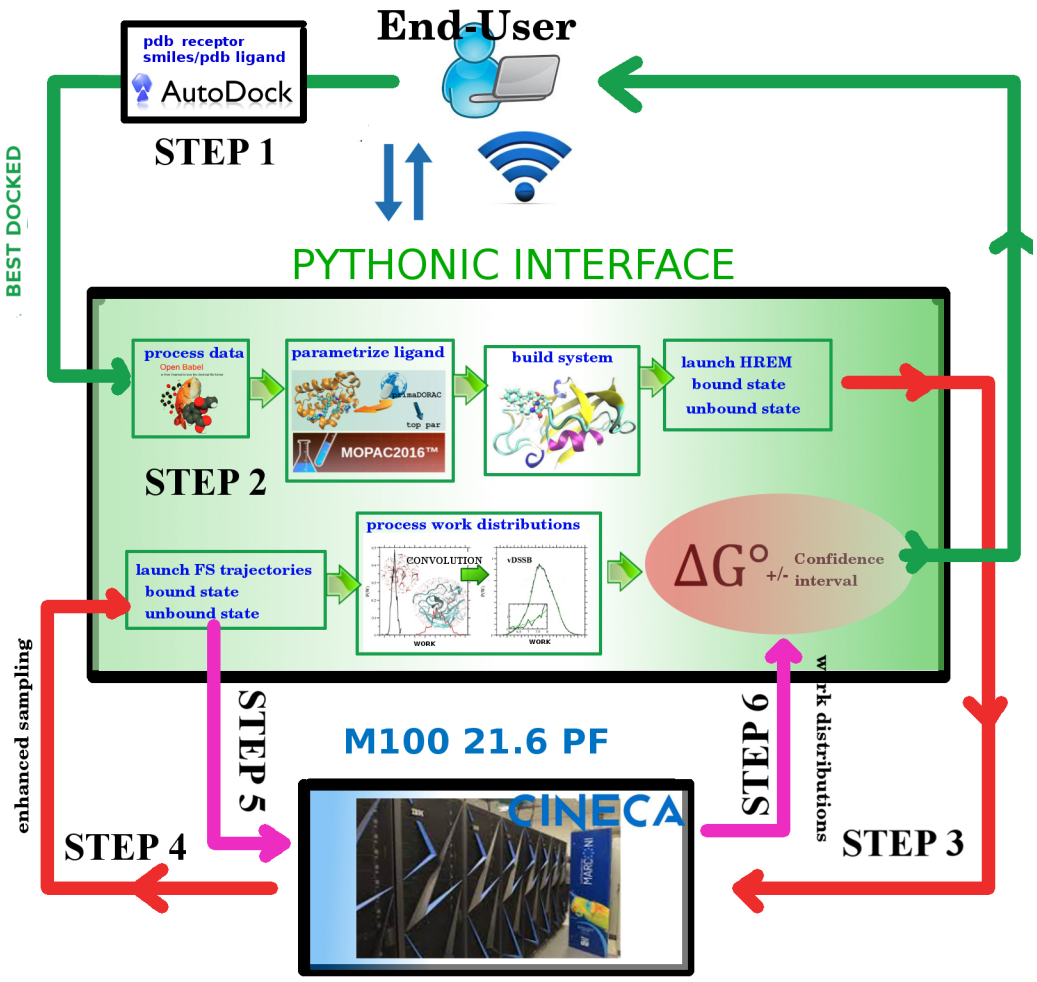

Figure 1: HPC_Drug/vDSSB layout for ADFE calculations: Dark green lines defines end-user minimal tasks. Light green arrows define tasks and data-flow within the python HPC_Drug automaton; red (HREM) and magenta lines (FSDAM) represent connections and data-flow between the python library and the underlying HPC.

the ligand-receptor system using Autodock4.[26] Starting from the receptor pdbcode and the ligand SMILES or pdb structure, the docking step can be completed in few minutes for a typical target of 5K atoms on a local workstation using the application script `docking.bash` for unix platforms providing a convenient interface for Autodock4. The script is included in SI with details on docking parameters. The lowest energy docked conformation is the "best-docked" bound state conformation for following vDSSB steps (used as complex with receptor, complex_ligand.pdb).

3.2 Step 2: HPC_Drug for HREM set-up (local)

Starting from the "best-docked" conformation generated in Step 1 (complex_ligand.pdb), we produce the GROMACS input files to perform Hamiltonian Replica Exchange (HREM) MD simulation for HPC execution of the target bound and unbound states. vDSSB HREM for the bound state relies on a definition of the "hot zone" comprising the ligand and nearby residues where the enhanced sampling is implemented. The hot zone for the unbound state is the ligand itself. These complex steps, which are central in the vDSSB approach, are assisted by the HPC_drug middleware,[27] an effective tool for the guided submission of vDSSB job on a HPC system.

HPC_drug is a python application distributed under the AGPL that can be cloned from the `github` repository `https://github.com/MauriceKarrenbrock/HPC_Drug`. All installation instructions are detailed at the GitHub site. HPC_drug reads a user-defined input file with minimal information (receptor-ligand pdb (complex_ligand.pdb) file, force field, water model, pH etc). For complex operations (e.g. replica exchange setup and NE input files productions) HPC_Drug has a solid default behaviour, which can be changed by the end-user in the main HPC_Drug input file, if necessary. Once started, HPC_Drug performs, *with no further user intervention*, a series of operations consisting in i) the generation of the GROMACS topology files using PrimaDORAC[28] as the default tool for the ligand force field parametrization; ii) a preliminary minimization of the complex with the user-selected force field; iii) the NPT equilibration in standard conditions of the resulting solvated receptor-ligand structure (using as default the TIP3P[29] model for water or a user-selected model) in a optimal MD box; iv) the production of the input files (ready for use) to perform HREM simulations for bound and unbound ligand state on the HPC platforms. The GROMACS HREM input files automatically define the hot-zone for the bound state based on ligand-residue distance criteria.[30] The hot-zone is implemented generating multiple topology files along the replica progression via PLUMED,[21,22] a community-wide open source library for enhanced sampling techniques and free energy methods. PLUMED is used by HPC_Drug

as a standalone tool. For HPC execution, GROMACS needs to be patched with PLUMED. Most of HPC systems for scientific applications provides modules for executing jobs with GROMACS/PLUMED.

HPC_Drug produces the `mdp` GROMACS main input file for HREM execution of the bound state with a series of "pulling" directives whereby the ligand is tethered to the receptor via a weak harmonic potential acting on the centers of mass (COM) of the partners. These directives generate output files, during HREM simulation step, containing information about the COM-COM distance distributions, allowing to calculate the binding site volume correction for the dissociation free energy.[13,15]

The final output of HPC_Drug (produced in few minutes on a local workstation for a typical system with 50K atoms) consists in two directories, one for the bound state and the other for the gas-phase unbound state, where all necessary data for HREM execution on a HPC (including examples of batch submission scripts) are stored. As previously stated some advanced options (like the number of HREM replicates, number of replica and HREM simulation time) can be changed in the input file. An example of HPC_Drug input file is reported in SI.

3.3 Step 3: HREM simulations (HPC)

Upon uploading on the remote HPC platform the two HPC_Drug generated HREM directories, the HREM simulations can straightforwardly executed on the HPC system. In the SI we included the batch script used to run the bound state HREM simulation on the heterogeneous Marconi100 HPC platform (CINECA), equipped with 4 Nvidia VOLTA GPU per node. In particular we ran 6 batteries (replicates) each of 24 replica on 36 nodes (144 Nvidia VOLTA GPU). For the unbound state we perform the HREM for a ligand molecule in gas phase; we ran 4 batteries each of 8 replica on 8 nodes (32 Nvidia VOLTA GPU).

3.4 Step 4: Selection of the (enhanced sampled) configurations

This post-processing operation is conveniently done on the HPC system. A preliminary analysis of the "pulling" output files in target state replicas might be needed to verify that the ligand did not leave the binding site despite the weak ligand-receptor tethering restraint. No less than 200 configurations files are extracted using the GROMACS tools from the target state HREM trajectories `xtc` files for the bound and the unbound state. Simple `bash` scripts to harvest configurations at regular intervals are provided in the SI. For the unbound state, the gas-phase HREM configurations of the ligand (`gro` files) are combined with a well equilibrated box of water molecules.

3.5 Step 5: Fast Switching Alchemical Simulations (HPC)

3.5.1 Bound state

The ligand decoupling in the bound state, starting from the HREM-generated `pdb` configurations, is straightforwardly implemented via the `multidir` GROMACS option whereby the independent alchemical trajectories are originated from the HREM-generated `pdb` configurations in dedicated directories. In the HPC submission script for Marconi100 (see the SI), each *independent* trajectory on the MPI layer is assigned to a GPU, in a very effective hybrid parallelization scheme. The resulting embarrassingly parallel run engages as many GPUs (MPI processes) as HREM generated `pdb` files. According to a consensus protocol,[2] the alchemical decoupling, for each independent trajectory, is done in two consecutive annihilation step and hence using two GROMACS `mdp` files (provided in the SI). In the first and second steps we switch off the electrostatic ligand-environment interactions, and the Lennard-Jones (LJ) potential, respectively, using a soft-core regularization.[31] Separate estimates of the electrostatic and LJ contributions to the dissociation free energy are useful to gain hints on how to modify the molecule to increase the affinity. Typical annihilation times depends on the ligand size. For a typical drug size ligand (300:400 MW), we use 0.375 ns for

the switching off the atomic charges and 0.750 ns for turning off the LJ ligand-environment interactions. Intramolecular interactions are always on during the alchemical process. An example of the batch script for submitting the fast-switching parallel job for the bound state is provided in the SI.

3.5.2 Unbound state

The fast switching growth run is technically similar to the bound state run. In each fast-growth trajectory (MPI process) we switch on the LJ and electrostatic potential in sequence in a total time (for a typical ligand) of 0.520 ns. Further details can be found in the SI.

3.6 Step 6: Calculation of dissociation free energy

The main output in the preceding Step 5 are the `dhdl.xvg` files, stored in the directories specified in the `multidir` option. In these files, the instantaneous values of the derivative of the Hamiltonian with respect to λ at each time step are recorded. For each alchemical trajectory, the work can be computed by numerical integration. This operation is done for the two contributions, electrostatic and LJ, and for the bound and unbound transformations. The resulting n_b and nu total work values, referring to the bound and unbound work histograms $P_b(W)$, $P_u(W)$, can be combined yielding $n_b \times n_u$ work data as $W_{ij} = W_i^b + W_j^u$ in the convolution $(P_b * P_u)(W) = \int dw\ P_b(W)P_u(W-w)$, affording the Jarzynski estimates of the free energy cost of transferring the ligand from the bound state to the bulk (dissociation free energy):

$$\Delta G_{\mathrm{vDSSB}} = -\beta^{-1} \ln \left(\frac{1}{n_b n_u} \sum_i \sum_j^{n_u * n_b} e^{-\beta W_{ij}} \right) \tag{1}$$

The confidence interval is evaluated by bootstrapping with resampling the bound and unbound work data *before* the combination of work values.

The free energy estimate can also be computed assuming that the convolution $(P_b *$

$P_u)(W)$ can be represented as a Gaussian mixture using:

$$\Delta G_{\mathrm{EM}} = -RT \ln \left[\sum_{i}^{c} w_i n(W, \mu_i, \sigma_i^2) \right] \tag{2}$$

where c is the number of components in the mixture, $n(W, \mu_i, \sigma_i^2)$ is the normal distribution of the i-th component with mean and variance μ_i, σ_i^2 and w_i is the i-th weight such that $\sum_i^c w_i = 1$. The parameters in Eq. 2 are determined from the $n_b \times n_u$ work data using the Expectation-Maximization (EM) algorithm.[32,33] Eq. 2 is a direct consequence of the Crooks theorem, providing and *unbiased* alternative estimate of the dissociation free energy if the distribution is the result of a combination of normal distributions. While ΔG_{EM} does not suffer of the inherent[34] bias of the Jarzynski estimate, Eq. 2 can be less precise as the EM fitting process can be ill-defined and unstable if the mixture is made by overlapping components with disparate variances. A discussion about Jarzynski and EM estimates and their differences can be found in Ref.[13,15]

All these post-processing calculations can be performed directly on the HPC using a simple script file (works.bash) detailed in the SI.

Finally, the standard dissociation free energy can be calculated by summing to $\Delta G_{\mathrm{vDSSB}}$ the volume correction and the finite-size correction, when a ligand is not electrically neutral.[35,36] The volume correction, evaluated from the variance of COM-COM distribution, is calculated as:

$$\Delta G_{\mathrm{vol}} = RT \ln(V_{\mathrm{site}}/V_0) \tag{3}$$

with $V_0 = 1661$ Å and $V_{\mathrm{site}} = \frac{4}{3}\pi(2\sigma)^3$, where σ^2 is the variance of the COM-COM distribution in the HREM stage obtained starting from pullx.xvg files. The finite-size correction, necessary only in case of charged ligands, is calculated applying the following equation:

$$\Delta G_{\mathrm{charge}} = -\frac{\pi}{2\alpha^2} \left\{ \frac{[Q_P^2 - (Q_P + Q_L)^2]}{V_{\mathrm{BOX}}^{(b)}} + \frac{Q_L^2}{V_{\mathrm{BOX}}^{(u)}} \right\} \tag{4}$$

where Q_P and Q_L are the net charge on the receptor and ligand, respectively, $V_{\mathrm{BOX}}^{(b/u)}$ are the MD box volumes of the bound and unbound states, and α is the Ewald convergence parameter. Summarizing, the vDSSB estimate of the standard dissociation energy is given by:

$$\Delta G_0 = \Delta G_{\mathrm{vDSSB}} + \Delta G_{\mathrm{vol}} + \Delta G_{\mathrm{charge}} \tag{5}$$

For technical further details on the post-processing stage for ADFE calculations, see the SI.

4 Application

During the Covid-19 pandemic, considerable efforts have been devoted to the identification an effective antiviral agent for SARS-CoV-2 via computational approaches.[37] Here we report on the calculation of the dissociation affinity of the PF-07321332 compound (structure in Figure 2a) against the main protease ($3CL^{pro}$) of the SARS-CoV-2 using the vDSSB technique.

Starting from the protein X-ray structure[38] and the PubChem[39] generated SMILES code of PF-07321332, we performed the docking calculation on the catalytic domain I+II of $3CL^{pro}$ obtaining as the best docking pose with a binding energy value of -9.07 kcal/mol the structure of the ligand-receptor complex reported in Figure 2b (**Step 1**).

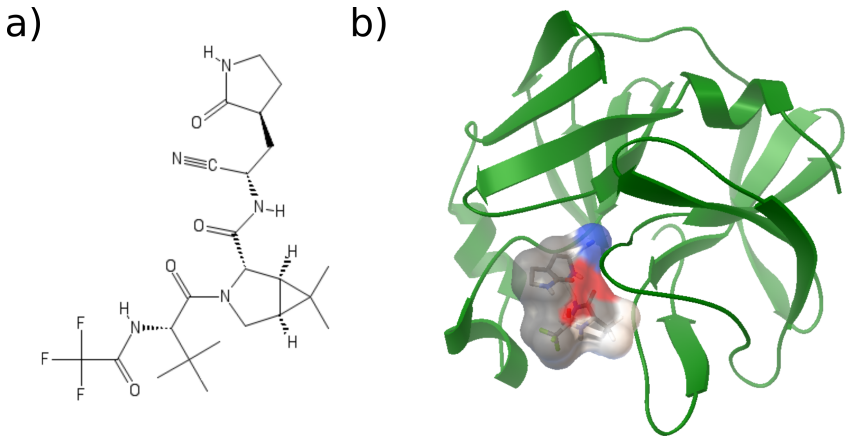

Figure 2: a) Structure of PF-07321332. b) Ligand-receptor complex with lowest docking binding free energy, " best-docked". The catalytic dyad (HIS41-CYS145) has been considered in the neutral state.[13] The surface shows the interactions between PF-07321332 and $3CL^{pro}$.

Using this starting configuration, the python application generates the two directories to perform HREM of bound and unbound state (**Step 2**).

The two HPC_Drug generated directories are then transferred to the HPC system (CINECA M100). On the HPC, the bound HREM was carried out using 6 batteries and 24 replicas on 144 GPUs collecting about 142 ns on the target state in 24h, while for the unbound HREM we launched 4 batteries and 8 replicas on 32 GPUs, collecting 32 ns in about two wall clock hours (**Step 3**).

From the "pulling" directive output in the target states of the bound HREM directory, we computed the COM receptor-ligand distribution (Figure 3) to obtain the volume correction to the dissociation free energy. From the GROMACS `xtc` files in the bound HREM, we extracted 384 configurations of the target state sampled at regular intervals as starting points for the subsequent NE simulations corresponding to the fast annihilation of the ligand in the bound state. From the gas-phase HREM target state trajectories, we extracted 192 configurations of the ligand. These conformations were combined with a well equilibrated water box to provide the starting structures for the ligand growth NE process in the solvent (**Step 4**).

The fast switching stage produced 384 NE bound state trajectories lasting $\simeq 1.2$ ns and engaging as many GPUs in about two hours and 192 NE 0.520 ns unbound state trajectories on a many GPUs in about twenty minutes (**Step 5**).

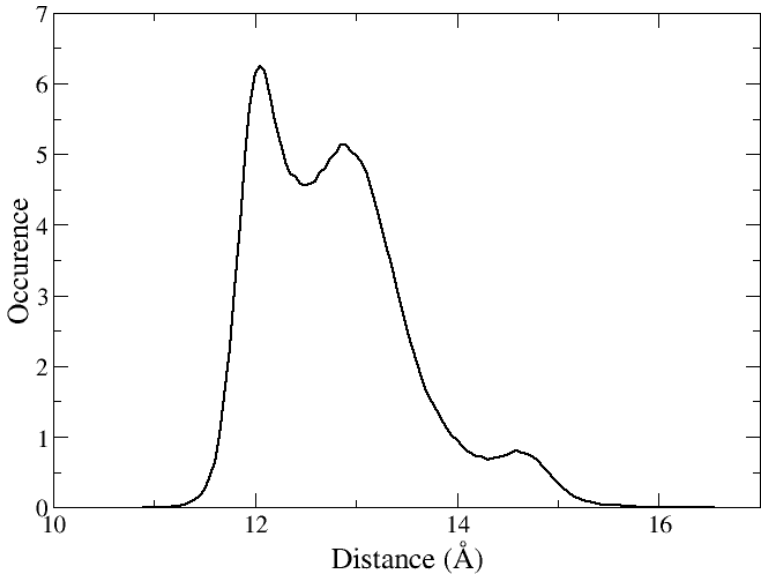

Figure 3: COM-COM distribution obtained from bound state HREM trajectories.

The `dhdl` files produced in the step 5 were finally post-processed on the HPC using the `work.bash` script obtaining the fast switching work values for bound (annihilation) and unbound (growth) states along with the estimates of the dissociation free energy $\Delta G_{\mathrm{vDSSB}}$ and ΔG_{EM} in Eq. 1 and Eq. 2, respectively (**Step 6**).

In Figure 4, we report the work distributions for bound and unbound state obtained and the convolution calculated by combining the bound and unbound work samples. The correction volume, obtained applying Eq. 3 and analyzing the Figure 3, is of -2.81 kcal/mol. The finite-charge correction is null since PF-07321332 is a neutral molecule. The final dissociation free energy estimate ($\Delta G_{\mathrm{vDSSB}}$, Eq. 1) is 5.8± 0.9 kcal/mol.

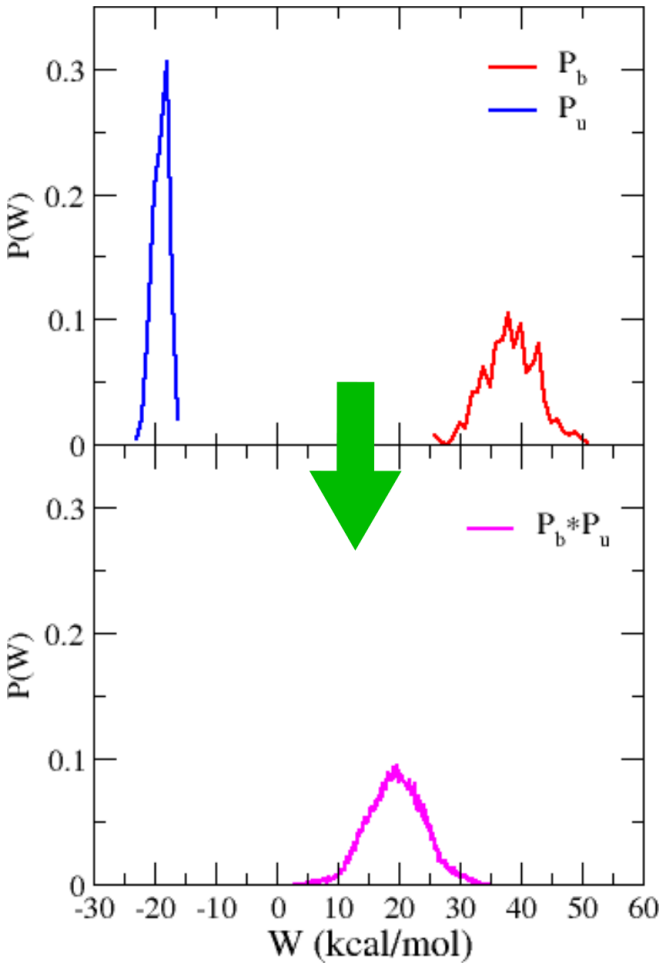

Figure 4: Upper panel: $P_u(W)$, $P_b(W)$ work histograms obtained from NE decoupling (bound) and recoupling (unbound) runs. Lower panel $(P_B * P_u)(W)$ convolution process.

5 Conclusions

We have described a step-by-step protocol for the calculations of the absolute dissociation free energy calculation using the GROMACS code and the PLUMED library, relying on the end-states enhanced sampling with HREM followed by the production of concurrent non equilibrium fast switching alchemical trajectories. The approach emulates a double-system-single-box ligand-receptor dissociation process by effectively combining the ligand annihilation work in the bound state with the ligand growth work in the bulk solvent. The methodology, specifically tailored for massively parallel architectures, allows to obtain a

reliable estimate of the ligand-receptor ADFE on GPU-accelerated HPC architectures in about one wall clock day. HPC submission of the complex HREM computational stage with GROMACS is greatly facilitated by the python middleware HPC_Drug, freely available at the GitHub site. A detailed tutorial is provided in the Supporting Information where the non-covalent dissociation affinity of the PF-07321332 inhibitor of the SARS-CoV-2 main protease is evaluated.

6 Data and Software Availability

HPC_Drug is publicly available at `https://github.com/MauriceKarrenbrock/HPC_Drug` under a AGPL licence. The simulations have been performed using GROMACS (https://www.gromacs.org/) and the PLUMED library (https://www.plumed.org/). Autodock4 (http://autodock.scripps.edu/) has been adopted for preliminary docking calculations. ORAC software (http://www1.chim.unifi.it/orac/) has been used for ligand force field parameters. All homemade scripts used to calculate PF-07321332-3CL^{pro} absolute dissociation free energy are provided in SI.zip file and described in Supporting Information (SI.pdf). All output files are reported on https://zenodo.org/record/5139374.

Supporting Information Available

Step-by-step technical tutorial for the calculation of the ADFE in the PF-07321332-3CL^{pro} system.

Acknowledgement

We acknowledge the Partnership for Advanced Computing in Europe (PRACE) for awarding us access to Marconi100 at CINECA consortium (Italy) (Pra20_CV28). We thank the CINECA staff for technical assistance on the Marconi100 HPC. We thank Giovanni Bussi

for assistance in the use of the PLUMED library with GROMACS. We thank Yuriy Khalak, Vytautas Gapsys, Matteo Aldeghi, Bert de Groot for their selfless collaborative attitude and precious assistance in performing nonequlibrium free energy calculations with GROMACS code. The authors thank MIUR-Italy ("Progetto Dipartimenti di Eccellenza 2018-2022" allocated to Department of Chemistry "Ugo Schiff").

References

(1) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden Thermodynamics of Mutant Proteins: a Molecular Dynamics Analysis. *Science* **1989**, *244*, 1069–1072.

(2) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem, B* **2010**, *114*, 10235–10253.

(3) Zwanzig, R. W. High-temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(4) Pal, R. K.; Gallicchio, E. Perturbation Potentials to Overcome Order/Disorder Transitions in Alchemical Binding Free Energy Calculations. *J. Chem. Phys.* **2019**, *151*, 124116.

(5) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.

(6) Cui, G.; Graves, A. P.; Manas, E. S. GRAM: A True Null Model for Relative Binding Affinity Predictions. *J. Chem. Inf. Model.* **2020**, *60*, 11–16.

(7) Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A. S. J. S.; Mackey, M.; Michel, J. Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *J. Chem. Inf. Model.* **2020**, *60*, 3120–3130.

(8) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W.

Rigorous Free Energy Simulations in Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60*, 4153–4169.

(9) Procacci, P. Methodological Uncertainties in Drug-Receptor Binding Free Energy Predictions Based on Classical Molecular Dynamics. *Curr. Opin. Struct. Biol.* **2021**, *67*, 127–134.

(10) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous Parallelization and Acceleration of Molecular Dynamics Simulations in GROMACS. *J. Chem. Phys.* **2020**, *153*, 134110.

(11) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.

(12) Crooks, G. E. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *J. Stat. Phys.* **1998**, *90*, 1481–1487.

(13) Macchiagodena, M.; Pagliai, M.; Karrenbrock, M.; Guarnieri, G.; Iannone, F.; Procacci, P. Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. *J. Chem. Theory Comput.* **2020**, *16*, 7160–7172.

(14) Procacci, P. A Remark on the Efficiency of the Double-System/Single-Box Nonequilibrium Approach in the SAMPL6 SAMPLing Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34*, 635–639.

(15) Macchiagodena, M.; Karrenbrock, M.; Pagliai, M.; Guarnieri, G.; Iannone, F.; Procacci, P. *Nonequilibrium Alchemical Simulations for the Development of Drugs Against Covid-19*; Springer US: New York, NY, pp 1–41.

(16) Gapsys, V.; Michielssens, S.; Peters, J.; de Groot, B.; Leonov, H. *Molecular Modeling*

of Protein; Humana Press, 2015; Chapter Calculation of Binding Free Energies, pp 173–209.

(17) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The SAMPL6 SAMPLing Challenge: Assessing the Reliability and Efficiency of Binding Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2020**, *34*, 601–633.

(18) Procacci, P. Hybrid MPI/OpenMP Implementation of the ORAC Molecular Dynamics Program for Generalized Ensemble and Fast Switching Alchemical Simulations. *J. Chem. Inf. Model.* **2016**, *56*, 1117–1121.

(19) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(20) 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.

(21) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comp. Phys. Commun.* **2014**, *185*, 604 – 613.

(22) Bussi, G. PLUMED. PDB, 2020; https://github.com/plumed/tuto-trieste-instructions (accessed 8 September 2021).

(23) Erbacci, G. Trends in HPC Architectures and Parallel Programmming, PRACE Winter School 2012: http://www.training.prace-ri.eu/uploads/tx_pracetmo/ParallelArchitectures-Erbacci.pdf (accessed 22 January 2016).

(24) Halford, B. Pfizer Unveils its Oral SARS-CoV-2 Inhibitor. *Chem. Eng. News* **2021**, *99*.

(25) Procacci, P.; Macchiagodena, M.; Pagliai, M.; Guarnieri, G.; Iannone, F. Interaction of Hydroxychloroquine with SARS-CoV2 Functional Proteins Using All-Atoms Non-Equilibrium Alchemical Simulations. *Chem. Commun.* **2020**, *56*, 8854–8856.

(26) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.

(27) Karrenbrock, M. HPC_Drug: a Python Application for Drug Development on High Performance Computing Platforms. M.Sc. thesis, Università degli Studi di Firenze, Firenze, Italy, 2020.

(28) Procacci, P. PrimaDORAC: A Free Web Interface for the Assignment of Partial Charges, Chemical Topology, and Bonded Parameters in Organic or Drug Molecules. *J. Chem. Inf. Model.* **2017**, *57*, 1240–1245.

(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(30) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431–9438.

(31) Beutler, T.; Mark, A.; van Schaik, R.; Gerber, P.; van Gunsteren, W. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222*, 5229–539.

(32) Dempster, A.; Laird, N.; Rubin, D. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. Royal Stat. Soc. B* **1977**, *39*, 1–38.

(33) Gupta, M. R.; Chen, Y. Theory and Use of the EM Algorithm. *Found. Trends Signal Process.* **2011**, *4*, 223–296.

(34) Gore, J.; Ritort, F.; Bustamante, C. Bias and Error in Estimates of Equilibrium Free-Energy Differences from Nonequilibrium Measurements. *Proc. Natl. Acad. Sci.* **2003**, *100*, 12564–12569.

(35) Darden, T.; Pearlman, D.; Pedersen, L. G. Ionic Charging Free Energies: Spherical Versus Periodic Boundary Conditions. *J. Chem. Phys.* **1998**, *109*, 10921–10935.

(36) Procacci, P.; Guarrasi, M.; Guarnieri, G. SAMPL6 Host–Guest Blind Predictions Using a Non Equilibrium Alchemical Approach. *J. Comput. Aided Mol. Des.* **2018**, *32*, 965–982.

(37) Chodera, J.; Lee, A. A.; London, N.; von Delft, F. Crowdsourcing Drug Discovery for Pandemics. *Nature Chemistry* **2020**,

(38) Liu, X.; Zhang, B.; Jin, Z.; Yang, H.; Rao, Z. The Crystal Structure of 2019-nCoV Main Protease in Complex with an Inhibitor N3. RSCB PDB, 2020; pdbode: 6LU7.

(39) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213.

Graphical TOC Entry