Received: 28 October 2022 | Revised: 20 December 2022 | Accepted: 7 January 2023

DOI: 10.1002/jcc.27077

SOFTWARE NOTE

Journal of COMPUTATIONAL CHEMISTRY WILEY

NE-RDFE: A protocol and toolkit for computing relative dissociation free energies with GROMACS between dissimilar molecules using bidirectional nonequilibrium dual topology schemes

Marina Macchiagodena | Marco Pagliai | Piero Procacci

Dipartimento di Chimica "Ugo Schiff", Università degli Studi di Firenze, Sesto Fiorentino, Italy

Correspondence
Piero Procacci, Dipartimento di Chimica "Ugo Schiff", Universit`a degli Studi di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy.
Email: piero.procacci@unifi.it

Funding information
European Union - NextGenerationEU - CUP, Grant/Award Number: B83C22002830001

Abstract

We describe a step-by-step protocol and toolkit for the computation of the relative dissociation free energy (RDFE) with the GROMACS molecular dynamics package, based on a novel bidirectional nonequilibrium alchemical approach. The proposed methodology does not require any intervention on the code and allows computing with good accuracy the RDFE between small molecules with arbitrary differences in volume, charge, and chemical topology. The procedure is illustrated for the challenging SAMPL9 batch of host–guest pairs. The article is supplemented by a detailed online tutorial, available at https://procacci.github.io/vdssb_gromacs/NE-RDFE and by a public Zenodo repository available at https://zenodo.org/record/6982932.

KEYWORDS
alchemical calculations, BAR, GROMACS, HPC, Jarzynski, molecular dynamics, non-equilibrium, relative free-energy calculations, SAMPL challenges

1 | INTRODUCTION

The determination of accurate relative dissociation free energies (RDFEs) using computational techniques based on molecular dynamics (MD) simulations is becoming a major industrial asset[1] in the hit-to-lead optimization for drug discovery. According to the consensus approach in MD-based calculations, the RDFE is computed via a thermodynamic cycle where the ligand is transmuted into a strictly congeneric compound[2] in the bound state and in bulk solvent. The RDFE is given by the difference between the transmutation free energies in the two legs of the cycle. The latter are obtained by setting up the so-called alchemical stratification,[3] that is, a discrete series of nonphysical thermodynamic states where the system is characterized by an alchemical parameter λ defining a chimeric ligand with intermediate bonded and nonbonded potential functions gradually connecting the two end-state congeneric ligands. The free energy between the alchemical λ-states is computed via the so-called Bennett acceptance ratio (BAR)[4] in the context of free-energy perturbation theory (FEP).[5]

Sampling issues and dependence on the starting conditions are known to critically impair the accuracy and reliability of RDFE predictions.[6–9] In state-of-the-art FEP implementations, to overcome such issues and boost the transitions between conformational states, the calculation is performed using a Hamiltonian Replica Exchange Method (HREM)[10–13] allowing the periodic exchanges between λ-configurations (λ-hopping) with a probability regulated by a Metropolis-like criterion. At the same time, the potential function involving the transmuting ligand and the binding site (the so-called "hot-region"[14]) is also down-scaled (i.e., heated) along the replica progression, with a "temperature" reaching the climax at the center of the stratification and being normal at the true thermodynamic end-states.[14] Such alchemical λ-hopping scheme with the hot-zone is

known with the acronym FEP+ and has been recently implemented in the commercial suite Desmond[15] by the s/w company Schrodinger.

FEP+, while using simultaneously the alchemical and the hot-zone scaling factors in the generalized ensemble, is nonetheless implemented as a one-dimensional HREM with no scaling at the end-states.[16] As such, the methodology may suffer from limitations such a sluggish convergence[17,18] or poor enhanced sampling capabilities in challenging systems.[16,19,20] Also, transmutations are usually limited to the so-called R-group perturbations, where a single substituent or atom is transformed into another substituent or atom. FEP-based RDFE generally do not involve scaffold-hopping or ring breaking[21,22] and transmutations between congeneric molecules differing by more than one substituent are generally carried out through the so-called perturbation graphs, using uninteresting connecting intermediate ligands.[23,24] Recently, modifications of the common FEP-based single topology approach with dummy atoms have been proposed to cope with ring breaking and scaffold hopping in RBFE calculations.[25–27] However, these techniques are quite complex and system-dependent. Hence, unlike the proposed methodology, they are not easily amenable to be routinely implemented in hit-to-lead in silico projects. Limits of these FEP-based methodologies for scaffold hopping were amply discussed in [28] and in the recent authoritative reviews on RBFE and ABFE (see, e.g., [29, 30]).

Recently, we have devised a bidirectional nonequilibrium (NE) alchemical technique affording the calculation of the RDFE between non-congeneric chemical compounds differing in volume, chemical connectivity and net charge (NE-RDFE).[31] The method relies on a preliminary HREM sampling *of the end-states* where the fully coupled ligand, say A, coexists with the decoupled partner, say (B), with the definition of the hot-region involving both ligands and the host binding site. The dual topology (DT) end-states are then connected by a series of fast NE independent alchemical trajectories where the ligand is decoupled and the ghost partner is re-coupled, conducted in both the forward and reverse direction (i.e. A(B) → (A)B and B(A) → (B)A). The RDFE is recovered from the forward and reverse work distribution exploiting the Crooks theorem[32] and the Bennett acceptance ratio.[4,33] The dissipation (defined as the difference between the mean NE work and the underlying free energy, and affecting accuracy and precision in NE-RDFE calculations) is significantly tamed as the ghost molecule (or most of it) is let grow (i.e., its interaction with the environment is alchemically switched on) in the cavity that is occupied by decoupling partner, using a weak harmonic restraint between the centers of mass of the two ligands, with no effects on the resulting RDFE as rigorously proved in Reference 31.

The NE-RDFE methodology was implemented in the ORAC code[34] and successfully applied to the SAMPL9[35] host–guest systems.[31] In this note, we describe in detail how this inherently massively parallel technique can be straightforwardly adapted to the GROMACS code[36] with no intervention on the source code. Due to the lack of any modification on the GROMACS source code, there are technical differences in the ORAC and GROMACS implementation of NE-RDFE involving in particular the λ alchemical protocol. These differences and their impact on the calculation will be highlighted in the following sections.

In the accompanying Zenodo archive at https://zenodo.org/record/6982932, we provide all necessary files and ancillary scripts to run NE-RDFE calculations with GROMACS for the case the SAMPL9 host–guest systems involving the WP6 (carboxy-pillar[6]arene) cavitand with ammonium/diammonium cationic guests. While the data refer to the SAMPL9 host–guest systems, the Zenodo setup and ancillary scripts[37] can be applied to any RDFE hit-to-lead project by following the guidelines in the GitHub tutorial https://procacci.github.io/vdssb_gromacs/NE-RDFE.

2 | THEORETICAL BACKGROUND

The theory of bidirectional NE-RDFE is described in full detail in Reference 31. Here, we provide a brief summary of how the method works. NE-RDFE relies on a preliminary high-quality sampling using Replica Exchange Solute Tempering (REST)[12,14] of the complex (the *bound state*) and of the ligand in bulk solvent (the *unbound state*), as well as an enhanced sampling of the ligand in the gas-phase for all guests. The initial end-states configurations of the DT system can then be prepared by combining gas-phase configurations of the ghost ligand (B) with those of the fully coupled ligand A in the bound and unbound state, taking care that the distance between the centers of mass of the two ligands, with arbitrary mutual orientation, follows a Gaussian distribution with a variance of RT/k, where k is the force constant of the center of mass COM-COM harmonic restraint. These A(B) combined configurations in the bound or unbound state could have been generated by a simulation with the ghost (B) tethered to A via the COM-COM harmonic potential. Also note that k of the order of a few kcal/molÅ^2 is sufficient to set the distance between the ligands COM of order of the fraction of the Angstrom at standard temperature. While the free energy of the transmutation is independent on k,[31] the dissipation and hence accuracy and precision in a real calculation (few hundreds of NE trajectories) are not. An excessively loose restraint may let the initially ghost ligand grow far from the binding site with a high probability, whereas a too tight COM-COM restraint can hinder the readjustment of the growing ligand in the binding pocket.

Once the initial DT A(B) states have been prepared, the NE transitions A(B) → (A)B are controlled in GROMACS via the topology A and B files with linear interpolation between the initial A(B) and arrival, (A)B, potential function of the ligands. The harmonic restraint between the A and B ligands must be constantly enforced during the whole NE trajectories. Soft-core regularization[38,39] for both electrostatic and Lennard–Jones interactions are needed at the end-states $\lambda = 0$ and $\lambda = 1$ to avoid catastrophic singularities while numerically integrating the equations of motion. In a given thermodynamic state (bound or unbound), the NE transitions are conducted in both the forward, A(B) → (A)B, and reverse, B(A) → (B)A, direction, performed with inverted time schedule, that is, with the same time duration. So, to compute an

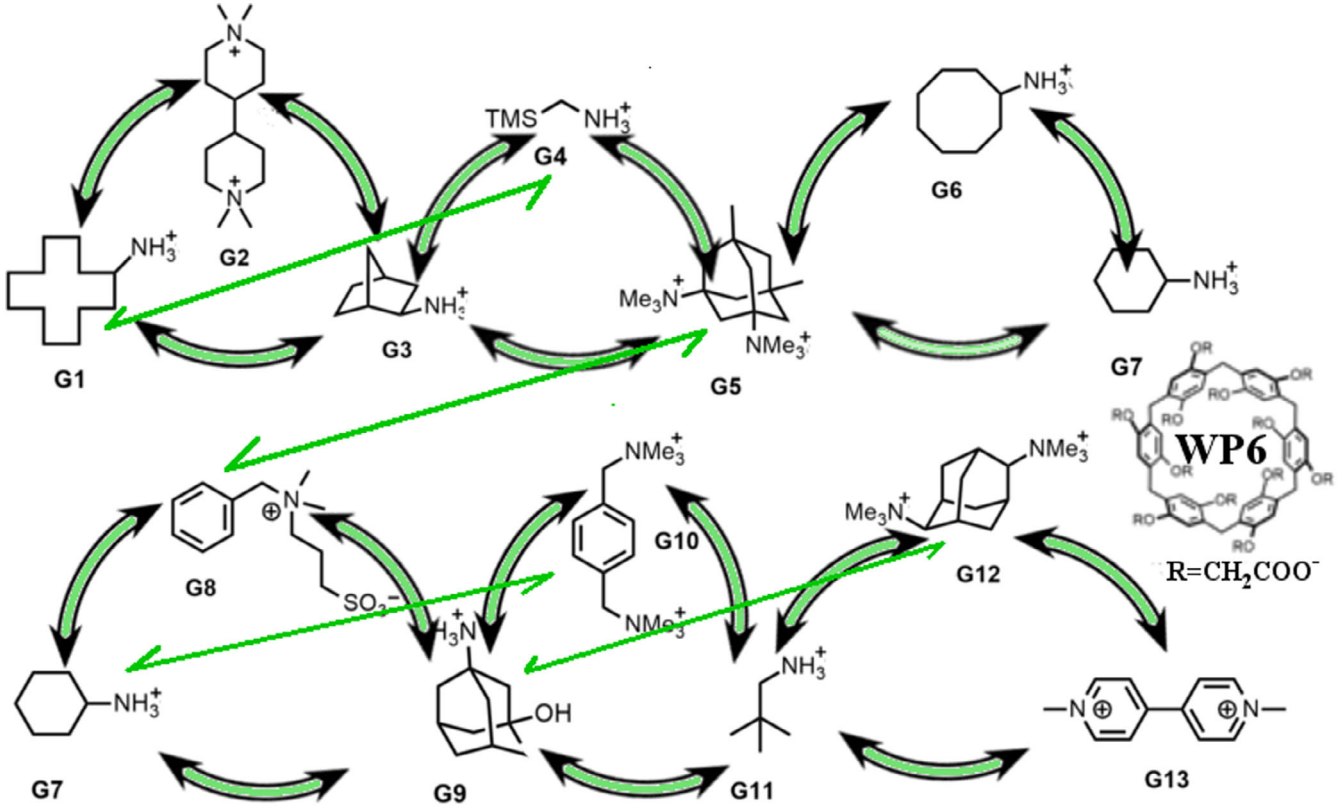

FIGURE 1 Guests (G1, G2… GN) and host (WP6) of the SAMPL9 challenge. The green arrows (curved or straight) connect the pairs for which we computed the RBFEs. In particular, the curved arrows connect two compounds belonging to the same cycle; the straight arrows connect two compounds on two different cycles.

RDFE, one must run four NE transitions, in the forward and reverse sense and in the bound and unbound states.

The forward, W_f, and reverse, W_r, alchemical work of the transmutation in the n A(B) → (A)B, and B(A) → (B)A NE transitions, is evaluated by numerical integration of the $\partial H/\partial\lambda$ derivatives, tabulated during the transitions at regular time intervals. These work values are used in the BAR equation to compute the unique root ΔG given by the crossing point of the forward $P_f(W)$ and the mirror-symmetric of the reverse distribution $P_r(-W)$:

$$\sum_{i=1}^{n}\frac{1}{1+e^{\beta(W_i(f)-\Delta G)}}-\sum_{i=1}^{n}\frac{1}{1+e^{\beta(W_i(r)+\Delta G)}}=0. \quad (1)$$

Equation (1) is accurate and precise,[33] so long that the two distributions $P_f(W)$ and $P_r(-W)$ overlap significantly.[40] The error on ΔG for finite sampling is assessed by bootstrap with resampling on the collection of n forward and reverse values or can be analytically determined by taking the derivative of Equation (1) with respect to ΔG affording the variance-related Fischer information.[33] The RDFE referring to the A(B) → (A)B transmutation is given by the difference between the corresponding transmutation free energies in the bound state and in bulk, ΔG_b, ΔG_u, each evaluated via Equation (1). As the unbound and bound state work values are statistically independent random variables by design, the uncertainty on the RDFE estimate is obtained by summing in quadrature the errors for ΔG_b and ΔG_u.

3 | RBFE CALCULATION WORKFLOW

In this section, we describe the step-by-step procedure for computing NE-RDFE with GROMACS. The procedure is also described in the companion GitHub tutorial.[41] All necessary data and files for reproducing the results on the SAMPL9 host-guest systems (reported further below in the section "Results" and in the Supporting Information (SI)) can be downloaded from the Zenodo public repository.[37]

3.1 | Preparation of the starting states for the DT NE transitions

Force field specification on the SAMPL9 host–guest systems (see Figure 1) can be found in References 31, 35 and in the Zenodo archive[37] (directory lib). We will assume that high-quality sampling of the end-state equilibrium simulations of the bound and unbound states with the guests at full coupling are available. We shall also assume that an enhanced sampling of the gas-phase configurations for all guests is also available. The former and the latter can be straightforwardly obtained with GROMACS patched with the PLUMED library[42,43] using Hamiltonian Replica exchange with solute tempering (ST-HREM). In References 44, 45 the reader can find excellent tutorials for implementing ST-HREM simulations with GROMACS. These configurations, obtained via ST-HREM for the

bound, unbound, and gaseous state of the SAMPL9 WP6 guests can be downloaded as PDB files from the Zenodo archive.[37]

Using these configurations we can produce the starting state for any arbitrary transmutation to be performed in both senses using NE-MD. Say that we want to make NE transitions to determine the RDFE for the transmutation of G2 into to G1. As shown in Figure 1, this transmutation involves a net charge change and a significant volume change between two dissimilar chemical compounds. We first combine, as outlined in the previous section, the equilibrium configurations of the bound state of G2 with those of the gas-phase of G1, thus obtaining the starting (equilibrium) states (pdb or gro files) for the G2(G1) DT system in the bound state. We repeat this process for the unbound state.

We then construct the top file for the DT system G2(G1) in the bound and unbound state. This files should contain the atomic type's specification for the host and the two ligands G2 and (G1). To this end, we may use the information in the G1, G2, and WP6 itp files produced with standard tools for the parameterization of ligands such as LigParGen,[46] PrimaDORAC,[47] or Antechamber.[48] The atomic types for the ghost species (G1) must be preceded by the string DUM_ and their atomic Lennard–Jones (LJ) parameters σ and ϵ should be set to zero. To keep the intramolecular LJ atom-atom interactions of (G1) during the NE transitions, we include in the top file all nonbonded atom-atom interactions between the dummy atomic types. Below we sketch out the structure of the topol.top for the G2(G1) DT system in the bound state. The parameters for G1, G2, and WP6 have been generated with PrimaDORAC.[47]

```
; This is the topol.top file for a NE transition in the bound state
[defaults]
; nbfunc        comb-rule      gen-pairs      fudgeLJ      fudgeQQ
1               2              yes            0.5          0.8333
[ atomtypes ]
;NAME           AT.NUM         MASS           CHARGE       PTYPE      SIGMA        EPSILON
c3              6              12.010         0.0000       A          0.3398       0.4510
c               6              12.010         0.0000       A          0.3315       0.4134
ca              6              12.010         0.0000       A          0.3315       0.4134
;... these are the NAME SIGMA and EPS for the atoms of the ghost
DUM_c3          6              12.010         0.0000       A          0.0000       0.0000
DUM_c           6              12.010         0.0000       A          0.0000       0.0000
DUM_ca          6              12.010         0.0000       A          0.0000       0.0000

[nonbond_params]
; i j           func           sigma          epsilon
DUM_c3          DUM_c3         1              0.3398       0.4510
DUM_c3          DUM_c          1              0.3357       0.4318
DUM_c3          DUM_ca         1              0.3357       0.4318
...

#include "ghost.itp"
#include "coupl.itp"
#include "wp6.itp"
#include "opc3.itp"

[ system ]
; Name
G1(G2)

[molecules]
; Compound      #mols
GHS               1
CPL               1
WP6               1
OPC            1634
```

GHS, CPL, WP6, and OPC are the (residue) names, in the starting G2(G1) pdb or gro files, of the ghost (G1) and the fully coupled ligand G2, the host and the solvent, respectively. The topol.top for the G2(G1) system for the unbound state has the same structure simply missing the specifications referring to the host. The structure of the itp files in the #include statements is standard except for the [atoms] directive of the G1 and G2 itp files. The itp file of the initially fully coupled G2, coupl.itp, has lines of the type:

```
[atoms]
;                                                    INITIAL STATE (lambda=0)          FINAL STATE (lambda=1)
;     TYPE      RESID     RESNAME     PDB-NAME     IGRP      CHRGE       MASS        TYPE       charge      mass
1     c3        1         CPL         C01          1         -0.1033     12.0100     DUM_c3     0.0000      12.0100
2     hc        1         CPL         H01          1          0.0854      1.0080     DUM_hc     0.0000       1.0080
3     hc        1         CPL         H02          1          0.0854      1.0080     DUM_hc     0.0000       1.0080
...
```

while that of the ghost (G1) has [atoms] lines of the type:

```
[ atoms ]
;                                                    INITIAL STATE (lambda=0)          FINAL STATE (lambda=1)
;          TYPE      RESID     RESNAME     PDB-NAME     IGRP      CHRGE       MASS        TYPE       charge      mass
1          DUM_c3    1         CPL         C01          1         0.0000      12.0100     c3         -0.0794     12.0100
2          DUM_hc    1         CPL         H01          1         0.0000       1.0080     hc          0.0380      1.0080
3          DUM_hc    1         CPL         H02          1         0.0000       1.0080     hc          0.0380      1.0080
...
```

This complex step has been fully automatized with the script make_rbfe_dir.bash provided in the bin directory of the Zenodo repository.[37] For example, for preparing the GROMACS top and itp files for the transmutation of G2 into G1 in the bound state, the command is launched from the main directory NE-RDFE (containing all data for the SAMPL9 systems) as.

```
$ make_rbfe_dir.bash g01 g02 b
```

This command combines the (WP6-bound) configurations of G2 with the gas-phase configuration of G1 generating the starting configurations for the G2(G1) → (G2)G1 transition and the associated top and itp files. All these files are found in a newly created sub-directory named b-g01-g02. In particular, the starting equilibrium PDB files, containing the ghost and the coupled ligand, will be located in a sub-directory of b-g01-g02 named b-g01-PDBS.

To produce the corresponding directory for the unbound state, it suffices to relaunch the command as.

```
$ make_rbfe_dir.bash g01 g02 u
```

All files for performing the G2(G1) → (G2)G1 transition in the unbound state will be generated in the sub-directory of the main NE-RDFE directory named u-g01-g02. The reverse transitions G1(G2) → (G1)G2 for the bound and unbound states can be prepared with the commands.

```
$ make_rbfe_dir.bash g02 g01 b

$ make_rbfe_dir.bash g02 g01 u
```

3.2 | Generation of the multiple directories from the initial A(B) sampling and submission of the parallel job for the swarm of the NE transitions on the HPC system

Once the starting configuration and top/itp files have been generated in the b-g01-g02, b-g02-g01 and u-g01-g02, u-g02-g01 directories, in each of these four directories we have to create as many sub-directories as starting configurations G2(G1) or G1(G2) in the bound

and unbound states and produce corresponding tpr data files using the GROMACS gmx grompp command.

```
gmx grompp   -f  filename.mdp  -c filename.pdb -p../topol.top
    -maxwarn 100 -o topol.tpr
```

The mdp main input files for the bound and unbound states are identical except possibly for the number of steps (usually higher in the bound state). The simulation setup provided in the mdp Zenodo example is standard, with PME[49] for electrostatics, constraints on H-bonds, Parrinello-Rahman[50] for constant pressure and the Bussi velocity scaling[51] for constant temperature. Temperature and pressure are set to 298 K and 0.1 MPa, respectively. The mdp files contain the specification for the free energy module. The initial λ value should be set to 0 in all cases, with a positive delta-lambda equal to 1/nsteps. For the SAMPL9 systems, the duration of NE alchemical transitions specified in the mdp files is of 1 ns and 0.5 ns for the bound and unbound state, respectively. In the mdp we must also impose the harmonic restraint between the two guests via the pull command. In GROMACS, in topology-controlled transmutations, the coupling-intramol = no directive has no effect on the intramolecular electrostatic interaction that are regulated by the λ-coupling. Unlike in ORAC, where the decoupled state in both legs of the cycle corresponds to the gas-phase ligand, in the GROMACS NE-RDFE implementation the decoupled state is a hybrid state with full LJ intramolecular interactions and zero charges. Also, at variance with ORAC where growing and annihilating species do not feel each other, in GROMACS the two ligands are subject to λ-dependent intermolecular interactions while transmuting. The λ path connecting the end-state in an NE-RDFE calculation is hence significantly different in the ORAC and GROMACS implementations. The final RDFE, a state function, should be unaffected by the differences in the λ path as the end-states in the alchemical thermodynamic cycle are identical in ORAC and GROMACS.

The generation of the tpr for the swarm of the NE transitions has been automated by the Zenodo-provided script maketpr.bash. By launching this command in each of the four DT directories b-g01-g02, b-g02-g01 and u-g01-g02, u-g02-g01, for each initial DT state prepared in Step 1, a sub-directory with a corresponding tpr file is generated using the appropriate gmx grompp command.

3.3 | Running the NE alchemical simulations on a HPC system and production of the dhdl files

Once the initial DT states have been fully prepared in the form of GROMACS tpr files using the maketpr.bash, we are now ready to run the four parallel jobs on the HPC that will afford the calculation of the RDFE for the G2(G1) → (G2)G1 process. This step is system-dependent. On the M100/CINECA HCP system,[52] the workload manager is SLURM.[53] In the Zenodo archive,[37] we provide an application script (make_submit_slr.bash) for generating in the four DT directories a SLURM submission script for the M100 heterogeneous platform. Each M100 node has 4 NVIDIA Volta V100 GPUs and 128 IBM-Power9 CPU cores. The submission script is hence designed to use $n/4$ nodes in total, running n MPI processes, each corresponding to an NE trajectory using 1 GPU and 32 OpenMP threads. An example of this script is given below for the DT directory b-g01-g02.

```
#!/bin/bash
#SBATCH --job-name b-g01-g02
#SBATCH -N1 --ntasks-per-node=4
#SBATCH --cpus-per-task=32
#SBATCH --ntasks-per-socket=2
#SBATCH --time=1:00:00
#SBATCH --gres=gpu:4
#SBATCH --nodes=50
#SBATCH -A IscrB_NE-RBFE
#SBATCH --partition=m100_usr_prod
#SBATCH --qos=m100_qos_bprod

#load gromacs module (CINECA M100)
module load profile/lifesc
module load autoload gromacs/2021.4

mdir='ls -d b[0-9][0-9][0-9][0-9]'

mpirun -n PROC gmx_mpi mdrun -notunepme -v -multidir
  $mdir -s\
topol.tpr -ntomp 32 -pin on
```

GROMACS, with the multidir option, runs in this case 200 alchemical simulations using the tpr files stored in the 200 directories (created using maketpr.bash) matching the string b[0–9][0–9][0–9][0–9]. The G2(G1) system size for the bound state includes ≈ 5 K atoms and the parallel job on M100 is completed in the less than 20 wall clock minutes. The notunepme option is used to avoid that the direct lattice contribution for the finite size correction[54] in an alchemical process involving a net charge change (as in the G2 to G1 transmutation) is modulated by a λ-dependent η during the transition, with η being the Ewald convergence parameter. Note that the Direct-Lattice Finite Size Correction (DLFSC) in GROMACS is evaluated during the transition via the function ewald_charge_correction in the ewald.cpp module and not a posteriori as in ORAC. So in GROMACS the DLFSC contribution to the free energy is computed for each NE trajectory and bears a volume dependency (due to e.g., the volume disparity of the G1 and G2

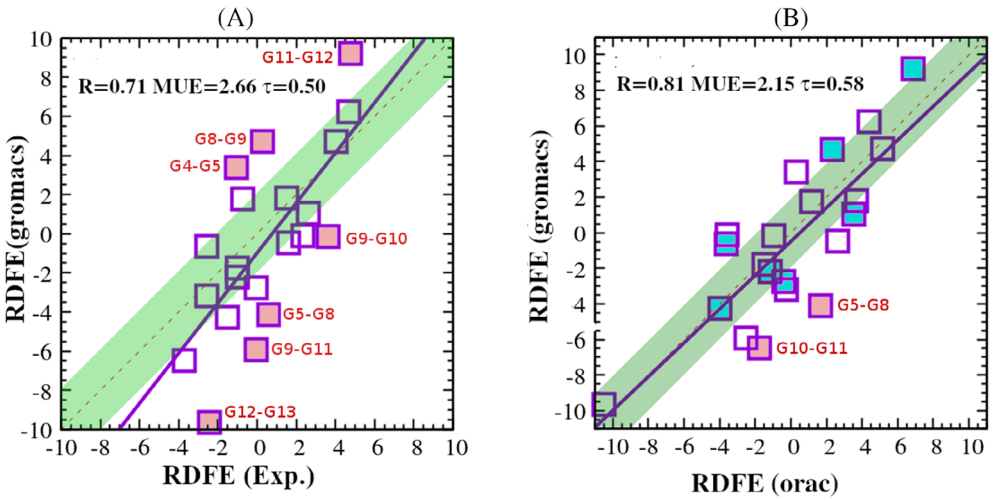

FIGURE 2 a) correlation diagram between experimental and GROMACS-computed RDFEs (units are in kcal/mol). Outliers (≥3 kcal/mol) b) are marked in red. b) Correlation diagram between ORAC-computed and GROMACS-computed RDFEs (units are in kcal/mol). Outliers (≥3 kcal/mol) are marked in red. The direct (one step) ORAC-computed transmutations are marked in blue color.

TABLE 1 Correlation metrics

	R	*a*	*b*	MUE	τ	ρ_c	MSE
Exp-GROMACS	0.71	1.28	−0.94	2.66	0.50	0.58	0.82
Exp-ORAC	0.68	1.04	−0.37	1.99	0.56	0.61	0.35
ORAC-GROMACS	0.81	0.95	−0.47	2.15	0.58	0.79	0.47

Note: *R is* Pearson's coefficient; *a is the* best-fitting line slope; *b* is the best-fitting line intercept; MUE is the mean unsigned error; τ is the Kendall rank coefficient; ρ_c is the Lin concordance coefficient; MSE is the mean-signed error. Units of *b*, MSE, and MUE are in kcal/mol.

TABLE 2 Cycle closure conditions in the network of Figure 1. The BAR entries (values in kcal/mol) refer to the sum of the RDFEs of the cycle, each computed using BAR.

Cycle	BAR
g01→g02 g02→g03 g03→g01	−0.29 ± 0.99
g03→g04 g04→g05 g05→g03	−0.23 ± 0.95
g05→g06 g06→g07 g07→g05	−0.23 ± 1.00
g07→g08 g08→g09 g09→g07	0.71 ± 1.10
g09→g10 g10→g11 g11→g09	−0.72 ± 0.89
g11→g12 g12→g13 g13→g11	−0.33 ± 0.80
g05→g06 g06→g07 g07→g08 g08→g05	1.22 ± 1.21
g01→g02 g02→g03 g03→g04 g04→g01	0.00 ± 1.09

species) that is not included in the approximated ORAC DLFSC. As we have chosen here, as in Refs. [28, 31], a neutralization protocol based on the simple uniform background plasma in PME with no counter-ions, and due to the high charge of the WP6 host (−12 *e*), we may expect some non-negligible difference in the DLFSC contribution in ORAC and GROMACS.

3.4 | Calculation of the RBFE from the dhdl files using BAR and v-DSSB-BAR

Once the parallel jobs launched from each of the 4 DT directories (b-g01-g02, b-g02-g01 and u-g01-g02, u-g02-g01) are completed, we should have in each of the corresponding *n* sub-directories the dhdl.xvg files with the time record of the derivative of the potential energy with respect to the λ alchemical parameter. We now compute from the dhdl.xvg files by integration the alchemical work done in the forward (b-g01-g02, u-g01-g02) and reverse (b-g02-g01, u-g02-g01) NE trajectories so as to apply twice Equation 1 for the bound and unbound processes to recover the RDFE for the transmutation of G2 into G1 as $\Delta G = \Delta G_b - \Delta G_u$. The Zenodo-provided[37] script RDFE.bash does all these operations automatically, including the evaluation of the 95% confidence interval by bootstrapping with resampling. For usage of the RDFE.bash, we refer to the GitHub tutorial[41] and to the documentation provided in the Zenodo archive.[37]

In the SI, the full sequence of UNIX commands for computing the RDFE of the transmutation of G2 into G1 is reported in Scheme S1.

4 | RESULTS ON SAMPL9 HOST–GUEST SYSTEMS

In Figure 1, we show the host–guest systems of the SAMPL9 challenge. The arrows indicate the 22 transmutations considered.

Results for the RDFE are shown in Figure 2 where we report the correlation diagram between experimental[35,55] and GROMACS-computed RDFEs, and the correlation diagram between GROMACS-computed and ORAC-computed RDFEs. RDFEs with corresponding error bars are also tabulated in Tables S1 and S2 of the SI. In Table 1, we report the full mutual correlation metrics regarding experimental, GROMACS-computed and ORAC-computed RDFEs.

The agreement between the experimental and GROMACS-computed RDFEs is good and in line with expectations.[31] Outliers do not depend on volume change or on the chemical dissimilarity index (e.g., Tanimoto).

The deviations between experimental and GROMACS-computed show, nonetheless, a weak correlation with the net charge change, highlighting the critical role played by electrostatics in the SAMPL9 challenge.

In Table 2, we report the cycle-closure conditions (CCC) using the BAR-based estimates of the RDFEs (Table S1 of the SI) for the six 3-nodes and the two 4-nodes cycles of Figure 1. Errors have been summed in quadrature. As it can be seen the CCC is satisfied within error bars for all eight cycles, demonstrating the robustness of the NE-RDFE estimates.

The agreement between ORAC-computed[31] and GROMACS-computed NE-RDFE is good (R = 0.81) although somewhat below expectations, most notably for the MUE. The deviations for the one-step ORAC transmutations are similar for those RDFEs that were obtained by combining more than one ORAC-computed transitions. The λ-path used in Reference 31is markedly different from that used in GROMACS: (i) the duration time of the alchemical transformations are different (1 ns here vs. 0.72 ns in Reference 31 for the bound state, and 0.5 ns here versus 0.36 ns in Reference 31 for the unbound state); (ii) the soft-core regularization is different. GROMACS uses a Beutler soft-core,[38] while ORAC uses a shifted potential[31]; (iii) the interactions between the two transmuting species are not present in ORAC while they are enforced in GROMACS; (iv) in ORAC calculations, the intramolecular energy of the transmuting ligand was not coupled to the alchemical parameter; in GROMACS only the LJ part of the intramolecular energy was λ-independent, whereas the electrostatic part was annihilated or created.

None of such differences should affect the final RDFE in the alchemical thermodynamic cycle. Therefore, their impact on accuracy and precision might arise from differences in the dissipation (i.e., in the overlap of the forward and reverse work distribution). As it can be seen from Table S1 of the SI, the errors with GROMACS are in general smaller than those obtained with ORAC, indicating that the λ-path protocol adopted in GROMACS is less dissipative with larger overlap of the forward and reverse distributions.

We finally should mention two features affecting the end-states that might be responsible for most of the observed deviations between ORAC-computed and GROMACS-computed RDFEs: (i) in GROMACS, the direct lattice PME correction for transmutations involving a change of net charge is included directly in the dhdl terms, thus bearing a volume dependence that is not accounted for in ORAC where the correction is evaluated a posteriori[31]; (ii) in our GROMACS calculation, there is a mismatch regarding the end-states and the corresponding starting configurations. Those corresponding to the decoupled ligand were prepared with ORAC using HREM on the molecule in the gas-phase with the full intramolecular potential. This is not the starting or final end-state in the GROMACS NE runs. The decoupled ligand is in an *hybrid* state, with the LJ intramolecular interaction fully on, and the atomic charges set to 0. Larger deviations between ORAC and GROMACS NE-RDFEs are hence expected for transmutations involving flexible ligands (e.g., g08) as in g05-g08 or involving a large charge change (again g05-g08). Large differences in volume of the ligands may also have an impact (as in g10-g11). NE-RDFEs computed with ORAC and GROMACS are reported in Table S1 of the SI.

5 | CONCLUSIONS

We have described a protocol for computing with the popular GROMACS code RDFEs between ligands differing in chemical connectivity, volume and net charge using a nonequilibrium DT alchemical approach. The methodology does not require any intervention on the source code. Results for the SAMPL9 host–guest systems are in good agreement with experimental data and strongly correlated with those obtained with a different code where the methodology was originally implemented (ORAC), despite significant differences in the details of the alchemical protocol.

At variance with standard FEP where convergence is required on each of the states of the λ stratification, the proposed method relies on the canonical (equilibrium) sampling of *the end-states only*. The latter can be reliably obtained using standard enhanced sampling techniques, which are particularly effective on the low-entropy[56] bound end-state and in the gas-phase decoupled ligand. The end states are connected by a few tens to few hundreds of fast (few ns or few hundreds of ps) nonequilibrium alchemical trajectory where one ligand is decoupled while the other is re-coupled, performed in both directions. The overlap between the forward and reverse work distribution, affecting the accuracy in BAR-based free energy estimates,[33] is tamed given that the initially decoupled ligand grows in a cavity that is devoid of solvent molecules or of host/receptor moieties, thanks to an harmonic restraints potential between the two ligands that has no effect on the final free energy. As we have seen for the challenging case of the SAMPL9 host–guest systems, these features of the NE-RDFE approach allows computing RDFEs that are not in the reach of current state-of-the-art FEP-based methods.

ACKNOWLEDGMENT

We acknowledge the Partnership for Advanced Computing in Europe (PRACE) for awarding us access to Marconi100 at CINECA consortium (Italy) (Pra20_CV28) and CINECA for allocation of CPU time (ISCRAB NE-RBFE). We thank the CINECA staff for technical assistance on the Marconi100 HPC. We thank Vytautas Gapsys for his selfless collaborative attitude and precious assistance in performing nonequlibrium free-energy calculations with the GROMACS code. The authors thank MIUR-Italy ("Progetto Dipartimenti di Eccellenza 2018-2022" allocated to Department of Chemistry "Ugo Schiff") and the National Recovery and Resilience Plan, Mission 4 Component 2 - Investment 1.4 - NATIONAL CENTER FOR HPC, BIG DATA AND QUANTUM COMPUTING - funded by the European Union -

NextGenerationEU - CUP B83C22002830001. Open Access Funding provided by Universita degli Studi di Firenze within the CRUI-CARE Agreement.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Zenodo at https://zenodo.org/record/6982932, reference number 6982932.

ORCID

Piero Procacci https://orcid.org/0000-0003-2667-3847

REFERENCES

[1] In-Silico Drug Discovery Market. **2022**; https://www.psmarketresearch.com/market-analysis/in-silico-drug-discovery-market, accessed June 12 2022.
[2] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, R. Abel, *J. Am. Chem. Soc.* **2015**, *137*, 2695.
[3] A. Pohorille, C. Jarzynski, C. Chipot, *J. Phys. Chem. B* **2010**, *114*, 10235.
[4] C. H. Bennett, *J. Comp. Phys.* **1976**, *22*, 245.
[5] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420.
[6] J. Chodera, D. Mobley, M. Shirts, R. Dixon, K. Branson, V. Pande, *Curr. Opin. Struct. Biol.* **2011**, *21*, 150.
[7] P. Procacci, *J. Mol. Graph. And Model.* **2017**, *71*, 233.
[8] P. Procacci, *Curr. Opin. Struct. Biol.* **2021**, *67*, 127.
[9] H. M. Baumann, V. Gapsys, B. L. de Groot, D. L. Mobley, *J. Phys. Chem. B* **2021**, *125*, 4241 PMID: 33905257.
[10] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141.
[11] P. Liu, B. Kim, R. A. Friesner, B. J. Berne, *Proc. Acad. Sci.* **2005**, *102*, 13749.
[12] S. Marsili, G. F. Signorini, R. Chelli, M. Marchi, P. Procacci, *J. Comput. Chem.* **2010**, *31*, 1106.
[13] L. Wang, R. A. Friesner, B. J. Berne, *J. Phys. Chem.* **2011**, *115*, 9431 PMID: 21714551.
[14] L. Wang, B. J. Berne, R. A. Friesner, *Proc. Nat. Acad. Sci.* **2012**, *109*, 1937.
[15] Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, 2006, November 11–17*; Association for Computing Machinery, New York, NY, United States **2006**.
[16] P. Procacci, *Molecules* **2022**, *27*, 4426.
[17] C. C. Bannan, K. H. Burley, M. Chiu, M. R. Shirts, M. K. Gilson, D. L. Mobley, *J. Comput. Aided Mol. Des.* **2016**, *30*, 927.
[18] D. Markthaler, M. Fleck, B. Stankiewicz, N. Hansen, *J. Chem. Theory Comput.* **2022**, *18*, 2569 PMID: 35298174.
[19] S. Wan, G. Tresadern, L. Perez-Benito, H. van Vlijmen, P. V. Coveney, *Adv. Theory Simul.* **2020**, *3*, 1900195.
[20] V. Gapsys, A. Yildirim, M. Aldeghi, Y. Khalak, D. van der Spoel, B. L. de Groot, *Commun. Chem.* **2021**, *4*, 61.
[21] Y.-S. Tung, M. S. Coumar, Y.-S. Wu, H.-Y. Shiao, J.-Y. Chang, J.-P. Liou, P. Shukla, C.-W. Chang, C.-Y. Chang, C.-C. Kuo, T.-K. Yeh, C.-Y. Lin, J.-S. Wu, S.-Y. Wu, C.-C. Liao, H.-P. Hsieh, *J. Med. Chem.* **2011**, *54*, 3076 PMID: 21434659.
[22] S. Liu, L. Wang, D. L. Mobley, *J. Chem. Inf. Model.* **2015**, *55*, 727 PMID: 25835054.
[23] R. Abel, L. Wang, E. D. Harder, B. J. Berne, R. A. Friesner, *Acc. Chem. Res.* **2017**, *50*, 1625 PMID: 28677954.
[24] L. F. Song, T.-S. Lee, C. Zhu, D. M. York, K. M. Merz, *J. Chem. Inf. Model.* **2019**, *59*, 3128 PMID: 31244091.
[25] L. Wang, Y. Deng, Y. Wu, B. Kim, D. N. LeBard, D. Wandschneider, M. Beachy, R. A. Friesner, R. Abel, *J. Chem. Theory Comput.* **2017**, *13*, 42 PMID: 27933808.
[26] J. Zou, Z. Li, S. Liu, C. Peng, D. Fang, X. Wan, Z. Lin, T.-S. Lee, D. P. Raleigh, M. Yang, C. Simmerling, *J. Chem. Theory Comput.* **2021**, *17*, 3710 PMID: 34029468.
[27] S. Azimi, S. Khuttan, J. Z. Wu, R. K. Pal, E. Gallicchio, *J. Chem. Inf. Model.* **2022**, *62*, 309.
[28] P. Procacci, G. Guarnieri, *J. Chem. Phys.* **2022**, *156*, 164104.
[29] Z. Cournia, B. K. Allen, T. Beuming, D. A. Pearlman, B. K. Radak, W. Sherman, *J. Chem. Inf. Model.* **2020**, *60*, 4153.
[30] Cournia, Z.; Chipot, C.; Roux, B.; York, D. M.; Sherman, W. *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Chapter 1, pp 1–38.
[31] P. Procacci, *J. Chem. Theory Comput.* **2022**, *18*, 4014.
[32] G. E. Crooks, *J. Stat. Phys.* **1998**, *90*, 1481.
[33] M. R. Shirts, E. Bair, G. Hooker, V. S. Pande, *Phys. Rev. Lett.* **2003**, *91*, 140601.
[34] P. Procacci, *J. Chem. Inf. Model.* **2016**, *56*, 1117.
[35] SAMPL9. https://github.com/samplchallenges/SAMPL9, accessed 13 January 2022.
[36] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1-2*, 19.
[37] Procacci, P. Relative binding free Energy calculations in GROMACS between dissimilar molecules using bidirectional nonequilibrium dual topology schemes. https://zenodo.org/record/6850702, accessed on 4 August 2022.
[38] T. Beutler, A. Mark, R. van Schaik, P. Gerber, W. van Gunsteren, *Chem. Phys. Lett.* **1994**, *222*, 5229.
[39] V. Gapsys, D. Seeliger, B. de Groot, *J. Chem. Teor. Comp.* **2012**, *8*, 2373.
[40] P. Procacci, *J. Chem. Phys.* **2019**, *151*, 144113.
[41] Tutorial for relative dissociation free energy calculations in GROMACS between dissimilar molecules using bidirectional nonequilibrium dual topology schemes. https://procacci.github.io/vdssb_gromacs/NE-RDFE, accessed on 21 July 2022.
[42] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comp. Phys. Commun.* **2014**, *185*, 604.
[43] M. Macchiagodena, M. Karrenbrock, M. Pagliai, P. Procacci, *J. Chem. Inf. Model.* **2021**, *61*, 5320.
[44] Using Hamiltonian replica exchange with GROMACS. https://www.plumed.org/doc-v2.6/user-doc/html/hrex.html, accessed: July 2022.
[45] Absolute Dissociation Free Energy calculations on HPCs: vDSSB tutorial for GROMACS users. https://procacci.github.io/vdssb_gromacs, accessed: July 2022.
[46] L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, W. L. Jorgensen, *Nucleic Acids Res.* **2017**, *45*, W331.
[47] P. Procacci, *J. Chem. Inf. Model.* **2017**, *57*, 1240.
[48] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graphics Modell.* **2006**, *25*, 247.
[49] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577.
[50] M. Parrinello, A. Rahman, *Phys. Rev. Lett.* **1980**, *45*, 1196.
[51] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.

[52] Consorzio Interuniversitario del Nord est Italiano Per il Calcolo Automatico (Interuniversity Consortium High Performance Systems) http://www.cineca.it (accessed: January 2018).
[53] Slurm Workload Manager Documentation https://slurm.schedmd.com (accessed: July 2022).
[54] T. Darden, D. Pearlman, L. G. Pedersen, *J. Chem. Phys.* **1998**, *109*, 10921.
[55] C.-L. Deng, M. Cheng, P. Y. Zavalij, L. Isaacs, *New J. Chem.* **2022**, *46*, 995.
[56] R. Pal, E. Gallicchio, *J. Chem. Phys.* **2019**, *151*, 124116.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: M. Macchiagodena, M. Pagliai, P. Procacci, *J. Comput. Chem.* **2023**, *44*(12), 1221. https://doi.org/10.1002/jcc.27077