

A Parallel Multi-Step Approach to Concerted Alchemical Free Energy Calculations

Ramesh Kumar C, Research Scholar, Department of Chemistry, Radha Govind University, Ramgarh, Jharkhand.

Dr. Shiv Brat Singh, Assistant Professor, Department of Chemistry, Radha Govind University, Ramgarh, Jharkhand

Abstract - The reliability of AFE computations decreases for transformations involving more than 5 heavy atoms. This study examines the efficacy of performing individual transformations of running charge, Van der Waals, and bond parameters with variable allocation per step, as opposed to the standard practice of transforming all parameters in a single step in most AFE workflows. The MultiStep protocol is more advantageous for the bound leg as compared to the one-step ("SoftCore") approach, whereas the free leg does not exhibit any such benefits. Based on Cresset's additional research, it has been determined that the Softcore methodology and the MultiStep technique offer comparable benefits. This study emphasizes the advantages of analyzing a First Episode Psychosis (FEP) approach and comparing it with an alternative strategy.

Keywords: Multi-Step, Alchemical, Free Energy, FEP.

1. Introduction

The computation of relative binding free energy (RBFE) is a valuable technique for aiding in the optimization of ligands during the initial phases of drug development. This holds true for both the hit-to-lead and lead-optimization stages.⁹⁵ Despite the resilience of the technique and

its broad range of academic implementations, there remain several technological limitations that constrain practitioners from freely exploring chemical spaces without constraints. Although a few commercial tools are available, the full potential of the technique has yet to be realized. The RBFE technique involves the conversion of one ligand to another, wherein the decoupling parameter is designated as the specified parameter. The technique involves partitioning the transformation process into discrete bins, wherein the parameters undergo iterative modifications in a bin-wise fashion. Each bin is characterized by a set of perturbed parameters that are progressively altered. In ideal circumstances, specifically when reliable RBFE transformations are available, the accurate statistical estimation of the alteration in free energy between states can be achieved through a high phase space overlap between said states. Consequently, this results in a reliable estimation of the relative free energy between the two ligands, namely the endstates with values of 0.0 and 1.0. Thermodynamic Integration (TI^{271,272}) and the Multistate Bennett Acceptance Ratio (MBAR) are widely utilized statistical estimators. Regrettably, there exist situations where it becomes imperative to conduct perturbations that exhibit a low likelihood of attaining a suitable phase space overlap. An instance of this phenomenon could be the execution

of significant alterations (comprising over five substantial atoms) within a chemically diverse congeneric series. In general, significant modifications are tackled by augmenting sampling through the utilization of a greater number of windows. The efficacy of this approach is attributed to the subdivision of charge, Van der Waals (VdW), and bond parameter transformations into smaller, successive modifications across adjacent windows. Conversely, a disadvantage of this approach is its requirement for a substantial augmentation of computational resources.

Given that RBE perturbations involve alterations to multiple parameters, such as partial charges, Van der Waals forces, and bond parameters, there exist various approaches to address these concerns during simulation runs. Simultaneously modifying all parameters in a concentrated phase is a frequently employed approach and is therefore commonly advocated as the preferred method. Numerous examples exist where alternate strategies are employed, wherein not all parameters are modified simultaneously but rather progressively. Prior analyses have been conducted to assess the distinctions between concerted and multistep strategies. Currently, there exists a scarcity of scholarly inquiry in this field that juxtaposes diverse techniques during the execution of bound-phase RBE calculations. Furthermore, to the utmost of our comprehension, research of this nature that employs MBAR as the estimator of free energy or encompasses n steps or more has not been previously documented.

The preliminary results of this investigation suggest that a multistep protocol may be more effective than a concerted treatment for the bound leg, while no discernible benefit of the multistep protocol was observed for the free leg. These findings are qualitative in nature. Further research was conducted by

Cresset, which revealed that the concerted approach did not offer any noticeable disadvantages compared to the multistep strategy. This paper presents significant data that underscore the benefits of analyzing an RBE technique and conducting a comparative assessment of the two approaches. During the course of this study, additional observations were gathered, resulting in the optimization of the settings for Flare V6.

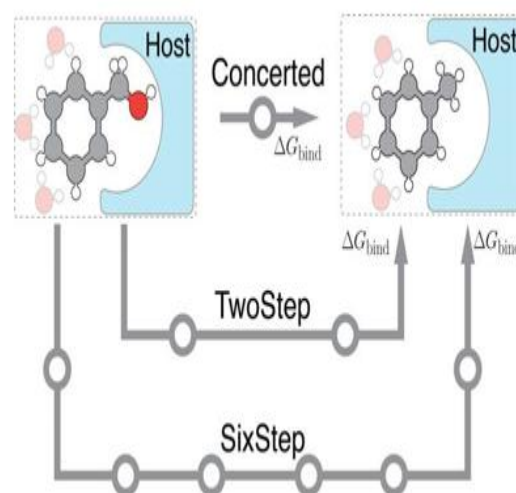


Figure 1: Diagrammatic representation of the concerted ($nsteps = 1$) method now in use and the alternative methods ($nsteps > 1$) provided here. Toluene-benzyl alcohol binding free energy was calculated as an example. Each 'step' is shown as a grey circle.

2. Literature Review

The study utilized the SOMD model, which was implemented in BioSimSpace version 2020.1.0=py37h1de35cc 97.241, for all simulations conducted. The duration of each simulation in any given phase may vary depending on the analysis being conducted. However, it can be assumed that each simulation was carried out for a period of 4 nanoseconds, unless explicitly specified otherwise. The force fields employed for the treatment of the waters were GAFF2, ff14SB, and TIP3P.

A. In the context of a molecular transformation utilizing RBE

A differentiation is made between hard atoms, which are constituent to the maximum common substructure (MCS), and soft atoms, which are located outside of the MCS and may be converted into or from a dummy atom. This disparity pertains to atoms that constitute the MCS or the maximum common substructure. The aforementioned proposition posits that the ligand endpoints exhibit a degree of commonality in their hard atoms, while the soft atoms are exclusively present in one of the ligand endpoints. This is a logical conclusion upon further contemplation.

The present study proposes N approaches, denoted as $nstep$, which correspond to the various methods for executing the protocol's total number of steps.

The MBAR, which stands for Multistate Bennett Acceptance Ratio, is a modern statistical estimator utilized for determining free energies. It considers all feasible states and is explained in a more concise manner on alchemy.org/wiki/Multistate_Bennett_Acceptance_Ratio (122,124).

The overlap matrix is a graphical representation generated by the MBAR methodology that illustrates the extent of phase space overlap among the states or windows utilized by MBAR. The aforementioned diagram is commonly denoted as an overlap matrix in academic literature.⁹⁵

The study examined the impact of perturbations on different systems, namely vacuum, free, and bound. Specifically, the perturbation was simulated using the ligand in a vacuum, the ligand in a 3 nm³ cubic water box, or the ligand in a protein in a 10 nm³ cubic water box. This pertains to the state of the ligand, whether it was in a vacuum, unbound, or bound.

B. In the context of this study, a multi-stage approach was devised to facilitate the development of a method.

The multistep method has been implemented on the BioSimSpace simulation platform, building upon the pre-existing Sire/OpenMM-MD (SOMD) code architecture.^{216,241} The inspiration for the development of the multistep approach was derived from BioSimSpace, as opposed to SOMD. The multi-stage approach comprises six sequential steps, as follows:

To facilitate the transformation of soft atoms into dummies, Coulomb terms should be adjusted to a value of 0.0. This modification applies specifically to soft atoms undergoing discharge.

To eliminate the softness of atoms undergoing a transformation into fake atoms, the LJ/Van der Waals terms can be perturbed to a value of 0.0.

The modification of the hard property involves the transfer of all Coulomb and LJ/VdW expressions for rigid atoms from λ_0 to λ_1 .

Induce perturbations in all bond parameters (angles, dihedrals) to modify the bonds.

To induce softness in atoms undergoing a transition from dummy, it is suggested to modify the LJ/VdW terms to λ_1 .

To modify the Coulombic terms for soft atoms transitioning from a dummy state, it is necessary to adjust the λ_1 value.

The initial conditions for each subsequent phase are established based on the parameters utilized to derive the outcomes of the preceding phase. In the second step of the process, atoms with low electronegativity have already undergone discharge at a wavelength of 0. Multistep approaches involve the adjustment of various categories of parameters, with each

step representing a distinct adjustment. For instance, in the context of a perturbation that entails the conversion of a methyl group into a hydrogen atom, the adjustment process would involve the following steps: firstly, the partial charges would be adjusted to 0 across step 1; secondly, the Van der Waals (VdW) parameters would be set to 0 across step 2; and finally, the partial charges would be adjusted in the ligand scaffold across step 3 to account for the absence of the methyl group. The terms that are bonded have been updated to the current status. In the present context, it is unnecessary to execute steps 5 and 6. The determination of the G value for each n can be accomplished through the utilization of established techniques for estimating free energy.

$$\Delta G_{\text{multistep}} = \sum_{n=1}^n \Delta G_{\lambda_n}. \quad (4.1)$$

The utilization of the Sire molecular framework for the management of perturbable molecules resulted in an immediate enhancement of the functionality of the Sire molecular object. A technique employed in Sire involves perturbable molecules that generate perturbation files elucidating the overall alterations in parameters, including charges, Van der Waals forces, and other relevant factors. The Sire molecule entity underwent modifications that enabled its toPertFile() method to accommodate a greater number of parameters. The aforementioned arguments delineate the particular stage within the multifaceted process at which the method is expected to generate a pertfile. An illustration of the construction of the pertfile atom words can be demonstrated through the use of pseudocode, which is presented below: For example, if the function toPertFile is called with the argument pert type="discharge soft", the resulting output will be as follows:

```

for atom in molecule do
  # retrieve native atom terms for lambda 0 and 1:
  LJ0 = atom.LJ0;
  LJ1 = atom.LJ1;
  charge0 = atom.charge0;
  charge1 = atom.charge1;
  # change charge1 for soft atoms, freeze all other terms to lambda 0:
  if atom == dummy then
    atom.LJ1 = LJ0;
    atom.charge0 = charge0;
    atom.charge1 = charge1;
  else
    atom.LJ1 = LJ0;
    atom.charge1 = charge0;
  end
  # now write updated atom terms to file:
  atom.LJ.write();
  atom.charges.write();
end

```

The first algorithm is a pseudocode representation of the atom terms written to a pertfile with a "discharge soft" pert type.

An further perk of editing pertfiles is that Sire can read its contents without any additional low-level reprogramming. Discharge mild and the five remaining phases are combined into one in this work's two-step technique.

3. Methodology

The bidirectional transformation principle advocates for the utilization of a four-step method, as opposed to a two-step approach, in achieving a perturbation that encompasses discharge, shrinkage, growth, and charging. The proposed approach entails an initial stage comprising the disintegration of atoms into inert entities, followed by four subsequent phases characterized by the reconstitution of said entities into their original atomic form. Nevertheless, the approach utilizing nsteps of four proves to be inadequate due to the emergence of supplementary intricacy resulting from partial charges and Van der Waals forces in rigid atoms, which are influenced by adjacent atoms, including pliant atoms. The insufficiency of the

nsteps equal to four method is further compounded by the increased complexity. Moreover, the alteration of soft atoms into dummy atoms and vice versa induces modifications in the bonding properties such as angles and dihedrals across the entire molecule, encompassing both soft and hard atoms. It is imperative to execute these two procedures autonomously during the interim period between the phases that pertain to the intricate atoms, in order to attain a comprehensive approach. The subsequent sections will demonstrate the results of our implementation of the multistep strategy. The methodology in question comprises six distinct procedures, namely discharging, shrinking, altering hard atoms, modifying bond terms, growing, and charging (as depicted in figure 4.4). For perturbations that entail only the shrinking or growth of atoms, it is possible to skip steps 1/2 or 5/6 (respectively) in this particular design. Omitting these steps is deemed acceptable as it is imperative that no parameters undergo modification during the aforementioned procedures.

Although a protocol encompassing each potential stage in a perturbation individually would require $nsteps = 6$ (as depicted in figure 4.4), the protocol to be employed henceforth will be ' $nsteps = 5$ '. This indicates that steps 3 and 4 have been consolidated into a single step, referred to as '3 flip'. The rationale behind this decision was to minimize computation time during testing. This decision was made based on the observation that steps 3 and 6 exhibited notably larger degrees of phase-space overlap in comparison to steps 1/2 and 4/5. Furthermore, this alternative was predicated on the observation that in the course of testing, these two procedures exhibited significantly greater degrees of phase-space overlap in comparison to steps 1/2 and 4/5. Furthermore, exclusively perturbations that consist of atoms that exhibit distinct growth or shrinkage are utilized. Redundant steps (1/2 during

growth and 4/5 during shrinkage) are removed prior to simulation. This measure guarantees that solely the most precise outcomes are attained. The multistep protocol entails the execution of three distinct steps for each perturbation, despite the fact that the protocol is designated as having a total of five steps, denoted as nsteps.

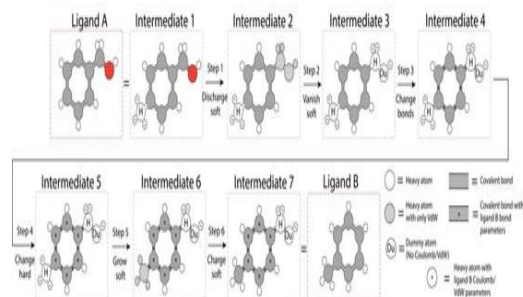


Figure 2: Multi-stage process schematic for ligand A to B. In this example, the top right loses methanol while the bottom left gains methyl. Section 3 provides detailed instructions. The following chemical process shows the ligand endpoints (red) and seven intermediates (grey; intermediate 1 = ligand A and intermediate 7 = ligand B). Hydrogens are not represented for clarity, although they should undergo the same disturbances as their heavier atom.

Confirmation of the efficacy of the multistage process using ethane and methanol

In order to determine whether or not the multistep protocol was appropriately developed, a validation study was carried out using a perturbation involving ethane and methanol. The analysis consisted of adequate sampling (17 windows with 4 ns in each window). This system was chosen above others due to the extensive amount of benchmarking work that had been done on it in a setting that was quite comparable.

The multistep methodology is shown correct by two crucial findings that are provided by this investigation. First of all, the multistep method (as well as the

concerted approach) predictions (6.35 kcal/mol) show a high degree of agreement with previously published relative free energy of hydration estimates (6.23, 6.22, 5.99, and 6.26 kcal/mol for AMBER, CHARMM, GROMACS, and SOMD, respectively).²²⁴ Since the research that was referred to was carried out five years before the work that was carried out in the present study, it is predicted that the 0.09 kcal/mol is due to incremental modifications in the SOMD codebase. This is because the work that was carried out in the current study was completed.

4. Results & Discussion

In light of the fact that this study is an examination of a multistep process in sequential order, we shall highlight different trial methodologies. After the discussion of the outcomes of an initial basic protocol ($nsteps = 2$, 4.3.1), which will be presented in the next subsections, the description of the final protocol ($nsteps = 5$, 4.3.2) will follow. In section 4.3.2, the notation $nsteps = 5$ will be replaced with the term 'multistep' to describe the protocol.

A. Method with $nSteps$ Equal to 2

The preliminary concept of dividing the standard technique into two phases (first shrinking soft atoms, then discharging soft atoms) was evaluated on a variety of eg5 inhibitors that were included in the Merck benchmarking set.¹²⁵ After making a selection of substantial perturbations (defined as having more than ten heavy atoms), simulations were run for a duration of 4 nanoseconds per angstrom window in both vacuum and solvated systems. The

MBAR overlap matrices for the $nsteps = 1$ protocol revealed poor overlap for the majority of the RBFs. On the other hand, the $nsteps = 2$ procedure indicated improvements in phase space overlap in situations where the transformations consisted of deleting (or decreasing) functional groups (see figure 4.3). The forward (or increasing) perturbation was the only one of these transformations that demonstrated poor overlap. This indicates that in order to build an atom from a dummy, first the water molecules need to be displaced by raising the VdW terms, and then the Coulomb terms need to be set (that is, charged). When reducing an atom to a dummy, on the other hand, the Coulomb terms should be switched off first (that is, discharged), followed by the Van der Waals terms. At this point, water molecules will be able to take the moiety's place. This is consistent with the methodologies discussed in previous studies.

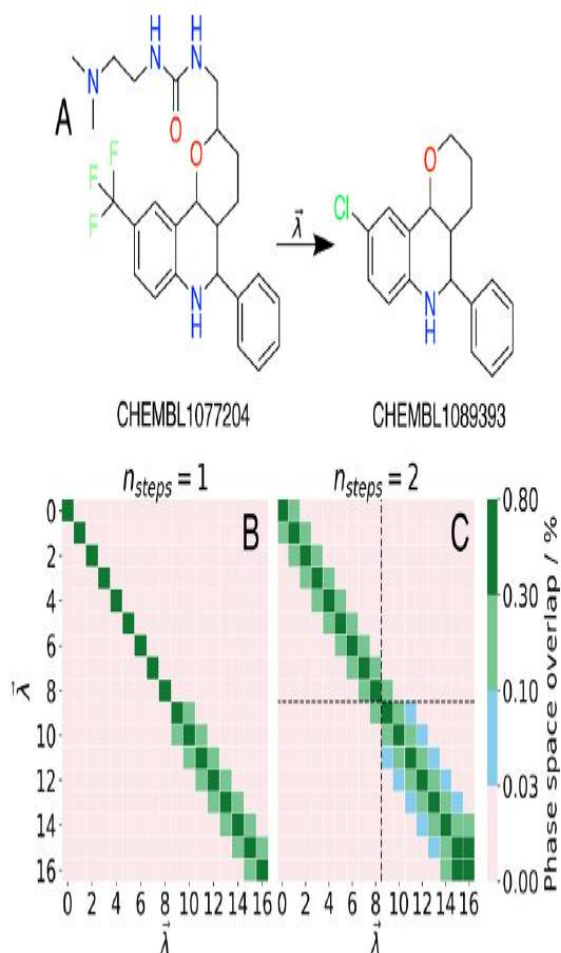


Figure 3: According to computed free energies in the free phase, the $n_{steps} = 2$ process has higher phase-space overlap than $n_{steps} = 1$. With $n = 17$, the free phase is simulated with 13 perturbed heavy atoms. The overlap matrix B for $n_{steps} = 1$ comprises blocks coloured by the fraction of phase space shared by the two windows. C: B without $n_{steps} = 2$. The dashed black lines indicate the ninth window, when partial charge changes become VdW adjustments. All hues except salmon need phase space overlap of 0.03 or more.

Conclusions

The present study outlines a technique that effectively partitions a coordinated approach utilizing RBF E into distinct phases. To the best of our knowledge, this study represents the initial attempt to conduct a direct comparison between the

two methods utilizing the identical RBF E code and undertaking a comprehensive analysis of the coordinated procedure. This research has yielded two significant findings, despite the lack of apparent benefits in utilizing the multistep approach over the concerted strategy. Initially, it appears that there exists a discrepancy in the way in which sulfonamide groups (and potentially other functional groups that share similarities with sulfonamide groups) are affected by the multistep methodology. Further investigation is required using updated versions of SOMD and alternative force fields to ascertain whether the observed discrepancy is attributable to algorithmic errors or systemic issues related to the expansion of these functional groups. The utilization of hysteresis as a reliability measure for perturbations involving the perturbation of a cyclical structure on the ligand scaffold indicates that the multistep technique outperforms the concerted approach. When contrasting the multistep approach with the concerted approach, this scenario arises. Given that cycle-growing disturbances are a common risk in RBF E, it would be advantageous to conduct a screening to identify potential solutions. This screening would be valuable as a novel approach to managing these disturbances could have significant implications for the field. However, a comprehensive analysis is imperative to fully assess the consistency of this impact.

References

- [1] S. Myers and A. Baker, Nature Biotechnology, 2001, 19, 727-730.
- [2] A. L. Nelson, E. Dhimolea and J. M. Reichert, Nature Reviews Drug Discovery, 2010, 9, 767-774.
- [3] I. Delany, R. Rappuoli and E. D. Gregorio, EMBO Molecular Medicine, 2014, 6, 708-720.
- [4] S. Morgan, P. Grootendorst, J. Lexchin, C. Cunningham and D. Greyson, Health Policy, 2011, 100, 4-17.

- [5] J. Hughes, S. Rees, S. Kalindjian and K. Philpott, *British Journal of Pharmacology*, 2011, 162, 1239-1249.
- [6] C. H. Wong, K. W. Siah and A. W. Lo, *Biostatistics*, 2018, 20, 273-286.
- [7] A. Bender and I. Cortes-Ciriano, *Drug Discovery Today*, 2021, 26, 511-524.
- [8] J. Ha, H. Park, J. Park and S. B. Park, *Cell Chemical Biology*, 2021, 28, 394-423.
- [9] S. Fox, S. Farr-Jones, L. Sopchak, A. Boggs, H. W. Nicely, R. Khoury and M. Biros, *SLAS Discovery*, 2006, 11, 864-869.
- [10] A. Bender and I. Cortes-Ciriano, *Drug Discovery Today*, 2021, 26, 1040-1052.
- [11] L. Zhao, H. L. Ciallella, L. M. Aleksunes and H. Zhu, *Drug Discovery Today*, 2020, 25, 1624-1638.
- [12] S. J. Y. Macalino, V. Gosu, S. Hong and S. Choi, *Archives of Pharmacal Research*, 2015, 38, 1686-1701.
- [13] G. Sliwoski, S. Kothiwale, J. Meiler and E. W. Lowe, *Pharmacological Reviews*, 2013, 66, 334-395.
- [14] Y. Hasin, M. Seldin and A. Lusic, *Genome Biology*, 2017, 18, year.
- [15] B. Dafniet, N. Cerisier, B. Boezio, A. Clary, P. Ducrot, T. Dorval, A. Gohier, D. Brown, K. Audouze and O. Taboureau, *Journal of Cheminformatics*, 2021, 13, year.