This LiveCoMS document is maintained online on GitHub at https://github.com/michellab/alchemical-best-practices

To provide feedback, suggestions, or help improve it, please visit the GitHub repository and participate via the issue tracker.

Best Practices for Alchemical Free Energy Calculations : v0.1

John D. Chodera[1*], Antonia S. J. S. Mey[2*], Julien Michel[2*], David L. Mobley[3*], Conor Parks[4*], Julia E. Rice[5*], Michael Shirts[6*]

[1]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York NY 10065; [2]Departments of Pharmaceutical Sciences and Chemistry, University of California, Irvine; [3]EaStCHEM School of Chemistry, David Brewster road, Joseph Black Building, The King's Buildings, Edinburgh, EH9 3FJ, UK; [4]Department of Chemistry, University of California, San Diego CA 92093; [5]IBM Almaden Research Center, San Jose, CA 95120; [6]Chemical and Biological Engineering, University of Colorado Boulder, Boulder, CO 80309

Abstract Alchemical free energy calculations can be a useful tool for predicting free energy differences associated with the transfer of small molecules from one environment to another. The hallmark of these methods is the use of modified potential energy functions to represent *alchemical* intermediate states that cannot exist in chemistry; by analyzing simulation data collected from a series of bridging alchemical thermodynamic states, transfer free energies (or differences in transfer free energies) can be computed with orders of magnitude less simulation time than observing the process spontaneously. While these methods are highly flexible, care must be taken in avoiding common pitfalls to ensure that computed free energy differences can be robust and reproducible for the chosen forcefield, and that appropriate corrections are included to permit comparison with experimental data. In this paper, we review current best practices for several popular application domains of alchemical free energy calculations, including relative and absolute small molecule binding free energy calculations to biomolecular targets.

JDC: What about biopolymer mutations?

This version is dated December 3, 2017

***For correspondence:**

john.chodera@choderalab.org (JDC); antonia.mey@ed.ac.uk (ASJSM); dmobley@mobleylab.org (DLM); info@julienmichel.net (JM); cdparks@ucsd.edu (CP); jrice@ibm.com (JER); michael.shirts@colorado.edu (MS)

Todo list

1 Introduction

ASJSM: @Volunteer write Intro

(

Alchemical free energy calculations have become a mature technology for computing various properties related to the transfer of chemical species from one environment to another. The domain of applicability for these calculations now involves such varied applications as the computation of protein-ligand binding free energies [**?**], ligand selectivities [**?**], partition and distribution coefficients between different liquid phases (such as octanol-water partition coefficients [**?**]), loss of affinity due to resistance mutations [**?**], and changes in protein thermostability due to engineered mutations [**?**].

The defining characteristic of alchemical free energy calculations is the use of a series of modified potential functions $U(x; \lambda)$ in which an alchemical parameter λ modulates interactions in a manner that cannot occur in real chemical systems.

One or more simulations are used to collect data from a multitude of alchemical states to compute a free energy difference between a chemical state (λ_0) and another chemical or alchemical reference state (λ_1),

$$\Delta f \equiv f(\lambda_1) - f(\lambda_0) = -\ln \frac{Z(\lambda_1)}{Z(\lambda_0)} \quad (1)$$

where the dimensionless free energy $f(\lambda) \equiv \beta F(\lambda)$ is given in terms of partition functions $Z(\lambda)$,

$$Z(\lambda) = \int dx\, e^{-u(x;\lambda)}. \quad (2)$$

Here, the inverse temperature $\beta \equiv (k_B T)^{-1}$ where k_B is the Boltzmann constant, and the *reduced potential* $u(x;\lambda)$ [**?**] is generally given by a trace over thermodynamic parameters with their conjugate dynamical variables,

$$u(x;\lambda) \equiv \beta \left[U(x;\lambda) + p\,V(x) + \sum_{i=1}^{N} \mu_i N_i(x) + \cdots \right] \quad (3)$$

where the collection of thermodynamic and alchemical parameters $\theta \equiv \{\beta, \lambda, p, \mu, \ldots\}$ defines a *thermodynamic state*. The physical transformation of interest may require several free energy differences Δf to be computed in order to produce an estimate of the overall desired quantity.

2 Prerequisites

Readers are expected to have a good working knowledge of both biomolecular simulations and the theory underlying alchemical free energy calculations. Good reviews of alchemical free energy calculation theory and practice can be found at [CITE REVIEWS].

3 Scope and Goals

- Preparation (focused on aspects unique to alchemical free energies), execution, and analysis of:
  - Transfer free energies (hydration free energies, partition coefficients, etc.)
  - Binding free energies
- Not in scope:
  - Advanced ligand binding topics such as:
  - Covalent inhibition
  - Association which is not 1:1
  - Endpoint free energy methods
  - PMF binding free energy methods
  - Force field choice: Should be a separate document (review)

Goals:

- Checklist and background for new practitioners of relative and absolute alchemical free energy calculations: What should you pay attention to in setting up and running calculations in common codes and why.
- Provide guidance to authors as to what should be reported about their protocols in a Methods section, either conforming to standard or reporting where they did not conform and why. Useful to reviewers as well.

4 Checklist

ASJSM: @Volunteer This needs to be revised and expanded, some initial thoughts were just thrown in.

An attempt at identifying most important checklist items.

STEP 0 – KNOW WHAT YOU WANT TO SIMULATE

What are the first questions that need addressing before setting up a molecular dynamics simulation Extensive explanation for the checklist questions can be found in section 5.

- ☐ Can I get the required accuracy with the simulation I want to carry out
- ☐
- ☐ And finally

SIMULATION PREPARATION

How do I get started setting up an alchemical free energy calculation Extensive explanation for the checklist questions can be found in section 6.

- ☐ Have I followed the Best practices for biomolecular simulation set up?
- ☐ In a relative simulation, will I run into problems with clashing geometries in the ligand transformation or crystal waters?

ABSOLUTE SIMULATIONS

What are the main things I need to consider for an absolute alchemical free energy calculation? Extensive explanation for the checklist questions can be found in section 7.

- ☐ Topology
- ☐ Restraints
- ☐ Standard state handling

RELATIVE SIMULATIONS

What are the main things I need to consider for an relative alchemical free energy calculation? Extensive explanation for the checklist questions can be found in section 7.

- ☐ First thing
- ☐ Also remember
- ☐ And finally

ANALYSIS

This is all about analysis of the simulation Extensive explanation for the checklist questions can be found in section **??**.

- ☐ Are my simulations converged enough?
- ☐ Am I using the right analysis techniques?

5 Step 0 – What can be expected from alchemical simulations?

- What level of accuracy you can expect?
- What timescales and how many transformations can you address given available computational resources?
- Can you even hope to tackle the problem you are attempting?

6 Step 1 – Simulation prerequesites

1. Generate geometry of initial state: Reference biomolecular simulation preparation best practices
2. Relative: Generate geometry of final state (Mey)
    - e.g. ligand ideally shouldn't clash with receptor, etc.; satisfy constraints that might be imposed by protocol such as overlapping atoms, etc.
    - reference biomolecular preparation setup practices for placing ligand into binding site, but elaborate on constraints that must be considered for relative free energy calculations

7 Step 2 – Simulation protocol selection

1. Relative: Select a topology and produce an atom mapping for transformation (Mobley)
    - Watching out for constraints to bonds for hydrogen: these cannot be allowed to change without including Jacobian terms
    - Share atoms between initial and final ligands if possible, otherwise colocalize pair of compounds and exclude their interactions with one another
    - Ring breaking/forming: Special are is needed; cite references
    - Tools: enlist Hannes Loeffler (developer of FESetup)?
    - Clarify terminology: Dual-topology, single-topology, etc.
2. Absolute: Identify **restraints and standard state handling**
    - Practical use of Boresch restraints/other forms of restraints. How to choose atoms involved. etc. (e.g. see Heinzelmann/Gilson BRD4 work; Chodera also has stuff to add)
    - Levi Naden: Problems with analytical approximation to standard-state correction for Boresch restraints
    - Harmonic and flat-bottom restraints
    - Clarify terminology: Double decoupling, etc. See Feature Box below.
3. Identify whether **net charge is changing** and how this will be handled: Issues that still need to be resolved: Charge corrections vs alchemically modify counterions
4. Select an **alchemical pathway**
    (a) Choice of alchemical Hamiltonian
        - Softcore potentials are always recommended, but might not ALWAYS be necessary, e.g.:
        - Pure changes in parameters of atoms that don't insert/delete atoms (turn into dummies/from dummies)
        - If roughly isosteric (e.g. lambda dynamics work from C. Brooks)

    (b) Choice of discrete alchemical protocol (Shirts, Mey, Chodera)
        - Many options: Adaptive scheme, Chebyshev polynomials, linear spacing, "choose your next lambda from data at this lambda", optimal thermodynamic length approaches (separately: Shirts, Sivak, Huafeng Xu).
        - Levi Naden had paper with lambda protocol which worked for all cases – methane solvation, host-guest (including disappearing host)

    (c) Relative Three-stage protocol (discharge unique initial atoms, transform LJ, charge unique final atoms) vs softcore electrostatics/LJ

    (d) Absolute
        - Select a **common alchemically-eliminated end state**
        - Decoupled vs annihilated for electrostatics and LJ
        - Sequential electrostatics and LJ versus simultaneous (recommend sequential)

    (e) Concerns: Part of AMBER still can't run at endpoints (lambda = 0 or 1); SANDER cannot but PMEMD can.
5. Determine whether you need to **handle multiple binding modes**: (Mobley)
    - In absolute: Confine-and-release, BLUES
    - In relative (ugh!!)
6. Determine **stopping conditions** Uncertainty-directed stopping criteria can ensure target uncertainty is achieved
7. Select which **data should be saved and with which frequency**
    - What data to save: dU/dlambda, Delta E's between neighbor for BAR, between further for MBAR, ...
    - BAR captures most of info with well-optimized lambda protocol, but MBAR when perhaps not, except when there are way too many lambda values.

- Recommend against solely relying on TI when possible
- Recommend cross-comparing methods (TI (spline, trapezoid, etc.), BAR, MBAR) as diagnosis of trouble

8 Step 3 – Overview of available analysis techniques

1. Detecting boundary between equilibrated and production regions (Chodera: http://dx.doi.org/10.1021/acs.jctc.5b00784)
2. Decorrelating samples for analysis
   - Subsample different lambdas based on correlation times
   - Ensure all simulations at least 50x correlation time
3. Examining output data for common problems with discussions of what exactly to plot or look at; examples of typical curves for dV/dlambda and free energy versus lambda, for example
   - Make sure ligand doesn't tumble out of binding site (Mey has observed this)
   - Significant discrepancies between different free energy estimators (TI, BAR, MBAR)
   - Poor replica mixing (for replica-exchange)
   - Correlation time as a function of lambda as it would be expected to be a smooth
   - Dependence on initial conformation
   - Torsional analysis: Is it stuck in specific states? Only very rarely transitions?
   - More "usual suspects"
4. Estimators for free energies
   - MBAR recommended if all energy differences are available
   - BAR just as good for highly optimized lambda values
   - TI should be roughly concordant, but quadrature error hard to quantify
   - Other variants useful in special circumstances (e.g. Z. Tan stochastic version)
5. Computing and reporting uncertainties on free energies correlated bootstrap v. timeseries analysis
6. Other considerations for many transformations Cycle closure error

9 Terminology and abbreviations

- Feature Box covering major technical terms and abbreviations
- Examples:
  - EXP, BAR, MBAR
  - Double decoupling, single-topology, dual-topology, hybrid-topology, coupled-topology
  - FEP (free energy peturbation), alchemical, AFE (alchemical free energy)

10 Available software – a summary

- Commercial:
  - FEP+
- Free or low-cost for academics / commercial for industry:
  - CHARMM / DOMDEC / CHARMM-OPENMM
  - TIES and AMBER FEW? (Peter Coveney)
  - AMBER / PMEMD
- Free (libre) open source:
  - SIRE
  - YANK
  - gromacs
  - pmx for mutations
- Setup tools
  - FESetup: AMBER, gromacs, Sire
  - Lomap/Lomap2 : Relative alchemical transformation graph planning
- Analysis tools:
  - Free Energy Workflows: Sire-specific free energy map analysis using weighted path averages https://github.com/michellab/freenrgworkflows
  - Alchemlyb: Multipackage free energy analysis https://github.com/alchemistry/alchemlyb
  - pymbar: MBAR implementation, but have to roll your own analysis wrapper https://github.com/choderalab/pymbar

11 Online resources

- http://www.ks.uiuc.edu/Training/Workshop/Urbana_2010A/lectures/TCBG-2010.pdf
- Basic Ingredients of Free Energy Calculations: A Review (DOI:10.1002/jcc.21450)
- Good Practices in Free-Energy Calculations (DOI:10.1021/jp102971x)
- Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges (doi:10.1016/j.sbi.2011.01.011)
- Alchemistry wiki: http://www.alchemistry.org/wiki/Best_Practices

Author Contributions

(Explain the contributions of the different authors here)

For a more detailed description of author contributions, see the GitHub issue tracking and changelog at https://github.com/michellab/alchemical-best-practices.

Other Contributions

(Explain the contributions of any non-author contributors here) For a more detailed description of contributions from the community and others, see the GitHub issue tracking and changelog at https://github.com/michellab/alchemical-best-practices.

Potentially Conflicting Interests

Declare any potentially conflicting interests here, whether or not they pose an actual conflict in your view.

Funding Information

FMS acknowledges the support of NSF grant CHE-1111111.