

This is the author's final, peer-reviewed manuscript as accepted for publication (AAM). The version presented here may differ from the published version, or version of record, available through the publisher's website. This version does not track changes, errata, or withdrawals on the publisher's site.

FESetup: automatic setup for alchemical free energy simulations

Hannes H Loeffler, Julien Michel
and Christopher Woods

Published version information

Citation:

Loeffler HH, Michel J & Woods C. "FESetup: automatic setup for alchemical free energy simulations." *Journal of Chemical Information and Modelling*, vol. 55, no. 12 (2015): 2485-2490.

[doi:10.1021/acs.jcim.5b00368](https://doi.org/10.1021/acs.jcim.5b00368)

This version is made available in accordance with publisher policies. Please cite only the published version using the reference above.

This item was retrieved from **ePubs**, the Open Access archive of the Science and Technology Facilities Council, UK. Please contact epubs@stfc.ac.uk or go to <http://epubs.stfc.ac.uk/> for further information and policies.

Application Note

FESetup: Automating Setup for Alchemical Free Energy Simulations

Hannes H Loeffler, Julien Michel, and Christopher J. Woods

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.5b00368 • Publication Date (Web): 06 Nov 2015

Downloaded from <http://pubs.acs.org> on November 12, 2015

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.

FESetup: Automating Setup for Alchemical Free Energy Simulations

Hannes H Loeffler^{1,*,#}, Julien Michel^{2,#}, Christopher Woods^{3,#}

¹Scientific Computing Department, STFC Daresbury, Keckwick Lane, Warrington WA4 4AD,

UK; E-mail: Hannes.Loeffler@stfc.ac.uk

²EaStCHEM School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9

3JJ, UK

³BrisSynBio, University of Bristol, 8–10 Berkeley Square, Bristol BS8 1HH, UK

ABSTRACT: FESetup is a new pipeline tool which can be used flexibly within larger workflows. The tool aims to support fast and easy setup of alchemical free energy simulations for molecular simulation packages such as AMBER, GROMACS, Sire or NAMD. Post-processing methods like MM-PBSA and LIE can be set up as well. Ligands are automatically parametrized with AM1-BCC and atom mappings for a single topology description are computed with a maximum common substructure search (MCSS) algorithm. An abstract

1
2
3 molecular dynamics (MD) engine can be used for equilibration prior to free energy setup or
4
5 standalone. Currently, all modern AMBER force fields are supported. Ease of use, robustness of
6
7 the code and automation where it is feasible are the main development goals. The project
8
9 follows an open development model and we welcome contributions.
10
11
12
13
14
15
16
17
18
19
20

21 Introduction

22
23 The setup and input preparation for simulation has become an increasingly demanding task for
24
25 individual users. While many traditional work-flows are still being managed manually, today's
26
27 hardware capabilities allow the computation of vast amounts of data. For instance, a system of
28
29 ca. tens of thousands of atoms can easily be simulated by classical molecular dynamics for at
30
31 least tens of nanoseconds per day. Such small MD simulations would typically run most
32
33 efficiently on just a handful of CPU cores while modern hardware may have many thousands of
34
35 cores available. GPU and other accelerators have increased in popularity in recent times and
36
37 have tremendously enhanced computing power too especially for small scale installations. Thus,
38
39 there is a trend for carrying out a multitude of simulations in parallel allowing for large scale
40
41 comparative studies. It is also important to improve on reproducibility of simulation protocols as
42
43 manual setups are often poorly documented.
44
45
46
47
48

49 This necessarily means that numerous input files and control data needs to be created to allow
50
51 the running of large numbers of simulations. However, creating the necessary input files for a
52
53 simulation can be a laborious and time-consuming process. This was not a problem when
54
55 simulation run times vastly exceeded manual setup times. However, we are now at a stage where
56
57
58
59
60

1
2
3 manual setup is becoming a bottleneck, particularly when running large numbers of simulations.
4
5
6 It follows there is a growing need for automated setup tools that can minimize user time. The
7
8 key here is to automate work-flows for simulation setup to the maximum extent reasonable. Of
9
10 course, not every step will be easily amenable to automation for various reasons including “hard”
11
12 problems in science such as computing missing structural data from insufficient information and
13
14 limited development of present day algorithms. However, the goal must still be to automate
15
16 what is possible yet at the same time accept that automatic procedures may not always be
17
18 successful. Robust protocols are needed to minimize failure and, importantly, detect and handle
19
20 or report failures.
21
22
23

24
25 Relative alchemical free energy simulation¹ is one example where popular simulation packages
26
27 still offer limited support for setup. This process is very tedious, time-consuming and can easily
28
29 lead to errors as the user may have to edit dozens of files or more and reorder hundreds of lines
30
31 of input by hand. Alchemical free energy simulations certainly require considerably more
32
33 computing resources than e.g. docking methods so this may be one reason for limited uptake
34
35 within the simulation community. However, also the aforementioned obstacles met during setup
36
37 may deter potential users. Therefore, alchemical simulation setup is an interesting target for
38
39 automated simulation setup, especially considering its potential role in drug design and lead
40
41 optimization^{2a}.
42
43
44

45
46 Several attempts at automating the setup of free energy calculations have been reported
47
48 recently. The Free Energy Workflow (FEW) tool² is available for AMBER³ for the setup of
49
50 relative free energy simulations and the end-point methods MM-PB/GBSA⁴ and LIE⁵. Another
51
52 approach which supports AMBER 11 has been reported recently⁶. PMX is a program⁷ which
53
54 automates the setup of relative free energy simulations of side-chain mutations for GROMACS⁸,
55
56
57
58
59
60

1
2
3 while StaGE⁹ supports absolute hydration or binding free energy calculations with GROMACS.
4
5 LOMAP is another software project¹⁰ that does not directly provide simulation input data.
6
7 LOMAP reduces the graph of all possible ligand pairs (relative free energies) to a minimum set
8
9 based on a definition of similarity used to weight the graph's edges to solve the shortest path tree
10
11 problem. The binding affinity calculator¹¹ (BAC) is an automation tool for rapid computation
12
13 and analysis of ligand–receptor free energies using MM–PB/GBSA.
14
15

16
17 Here, we will discuss a new tool called FESetup. FESetup's advance over previous tools is
18
19 that it is designed to support alchemical free energy simulations in a range of different MD and
20
21 Monte Carlo (MC) packages. FESetup presents an abstraction of the setup process that is
22
23 independent of a given MD or MC code, and that is flexible enough to work within larger
24
25 workflows e.g. using docking software to provide receptor–drug structures. FESetup has been
26
27 built to be open source and flexible, providing a strong foundation to build setup workflows for
28
29 different free energy methods. Further goals are support for mutation of both non–covalently
30
31 and covalently bound moieties, maximal automation, robustness, the development of an API and
32
33 documentation of all outputs in a log file. In this Note, we will summarize current progress and
34
35 how FESetup can help the user to set up free energy calculations for codes like AMBER,
36
37 GROMACS, Sire¹² and also NAMD¹³. FESetup is free software (GPL) and is installed locally.
38
39 The project is developed in an open fashion where interested parties can contribute at all levels
40
41 including code contributions and interfacing to FESetup.
42
43
44
45
46
47
48
49

50 Methodology

51 Absolute vs Relative

52
53 There are principally two ways to compute alchemical free energies along a coupling
54
55 parameter λ ¹⁴: the *absolute* and the *relative* approach. *Absolute* free energy changes are obtained
56
57
58
59
60

1
2
3 by completely annihilating e.g. a ligand in solution¹⁵ and in the bound state (which yields an
4 absolute binding free energy), or by annihilating a ligand in different solvents (which yields an
5 absolute solvation or transfer free energy). The setup for this type of alchemical calculation in
6 modern packages like AMBER or GROMACS is very easy. Both programs allow the user to tag
7 a set of atoms for decoupling through specific keywords in the input file. Both also allow some
8 flexibility for separating the van-der-Waals free energy path from the electrostatic free energy
9 path. The user is *not* required to modify the topology directly as this is transparently done in-
10 code.
11
12
13
14
15
16
17
18
19
20
21
22

23 Single vs Dual-Topology

24 For *relative* free energy simulations there is a choice to be made between the single-topology
25 and the dual-topology approach^{16,17}. Codes like NAMD only allow the latter at the moment and
26 when this approach is applied to non-covalently bound molecules it suffers from the same
27 problem as the absolute approach. The final end states describe a “non-existing” molecule that
28 can freely drift through the receptor and this needs to be compensated for with adequate
29 restraints¹⁸.
30
31
32
33
34
35
36
37
38

39 Codes like AMBER or GROMACS implement hybrid approaches since they allow the
40 assignment of a single and a dual-topology region at the same time. The single-topology region
41 can serve as a convenient “anchor” to keep the ligands in place as the coordinates are shared and
42 only direct conversion of one atom to another is allowed to occur. The atoms within this region
43 are thus always present.
44
45
46
47
48
49
50

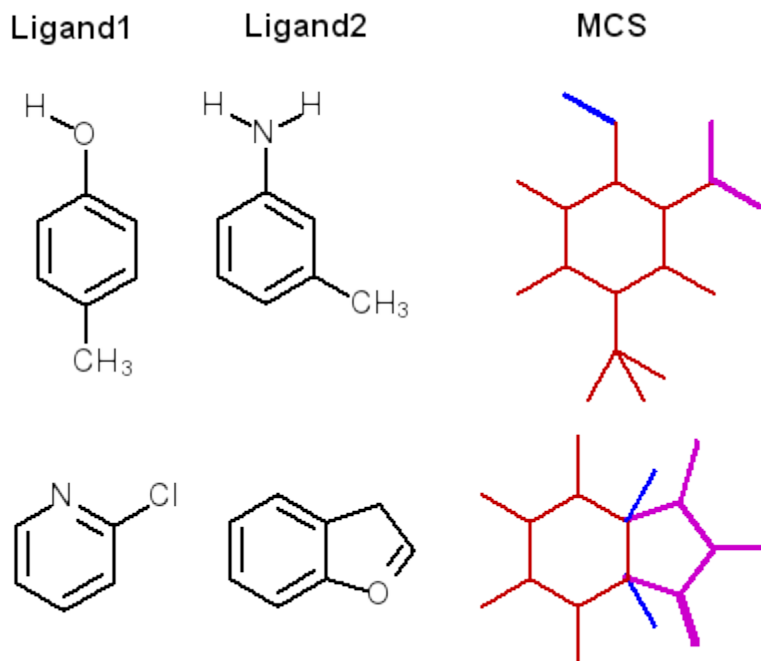
51 MCSS mapping

52 The single-topology approach requires a one-to-one mapping of equivalent atoms. While this
53 is generally a rather simple task for a human operator it requires some care when done
54
55
56
57
58
59
60

1
2
3 algorithmically. In cases where there is a well-defined anchor region as in protein backbone
4 atoms, the remaining atoms can be mapped through a simple distance criterion⁷. A general
5
6 solution is needed, however, in the case of arbitrary molecular structures.
7
8

9
10 Maximum common substructure search (MCSS) algorithms have been used in connection with
11 the definition of a similarity criterion to decide when two atoms or a bond match each other^{6,10}.

12
13 FESetup uses fmcs¹⁹ which is a connected MCSS algorithm²⁰. Our similarity criterion is that any
14 atom or any bond can match each other but rings must always match rings and rings can't be
15 broken²¹. Hydrogen atoms are explicitly included in the MCS. Our scheme essentially aims for
16 maximum overlap⁷, the idea being that a maximal single-topology description is the most
17 efficient protocol³⁹. This also implies that the number of vanishing or appearing atoms
18 (“dummy” atoms) is minimal (see Chart 1). In the definition of ref. 38 this implies an implicit
19 intermediate. The internal representation of molecules is package-independent and thus allows
20 generation of any arbitrary output format.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



26
27
28
29
30

Chart 1. Example for two ligand morph pairs. The MCS (red) is shown with explicit hydrogens on the right together with vanishing (blue) and appearing atoms (magenta).

31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

Two structures can principally have more than one common substructure²⁰. In our scheme, where we ignore atom and bond type identities, even more solutions are possible, e.g. the methanol to ethanol transformation has four unique solutions in terms of our definition of similarity. Different MCSs merely define a different pathway and the free energies obtained from a closed thermodynamic cycle should be independent of the pathway chosen to connect different molecules. However, some pathways may be more computationally efficient than others. This is especially so for ligands binding to a receptor, often involving a high degree of symmetry, where the binding mode may need to be preserved. Although a ligand can, in principal at least, adjust to λ dependent changes, sufficient sampling may be very hard to achieve in practice. The preservation of the binding mode is the responsibility of the user in FESetup.

54
55
56
57
58
59
60

Another issue is chirality because molecular graphs are, by definition, only two-dimensional. This means that chiral centers may be inverted because the MCSS algorithm simply follows the

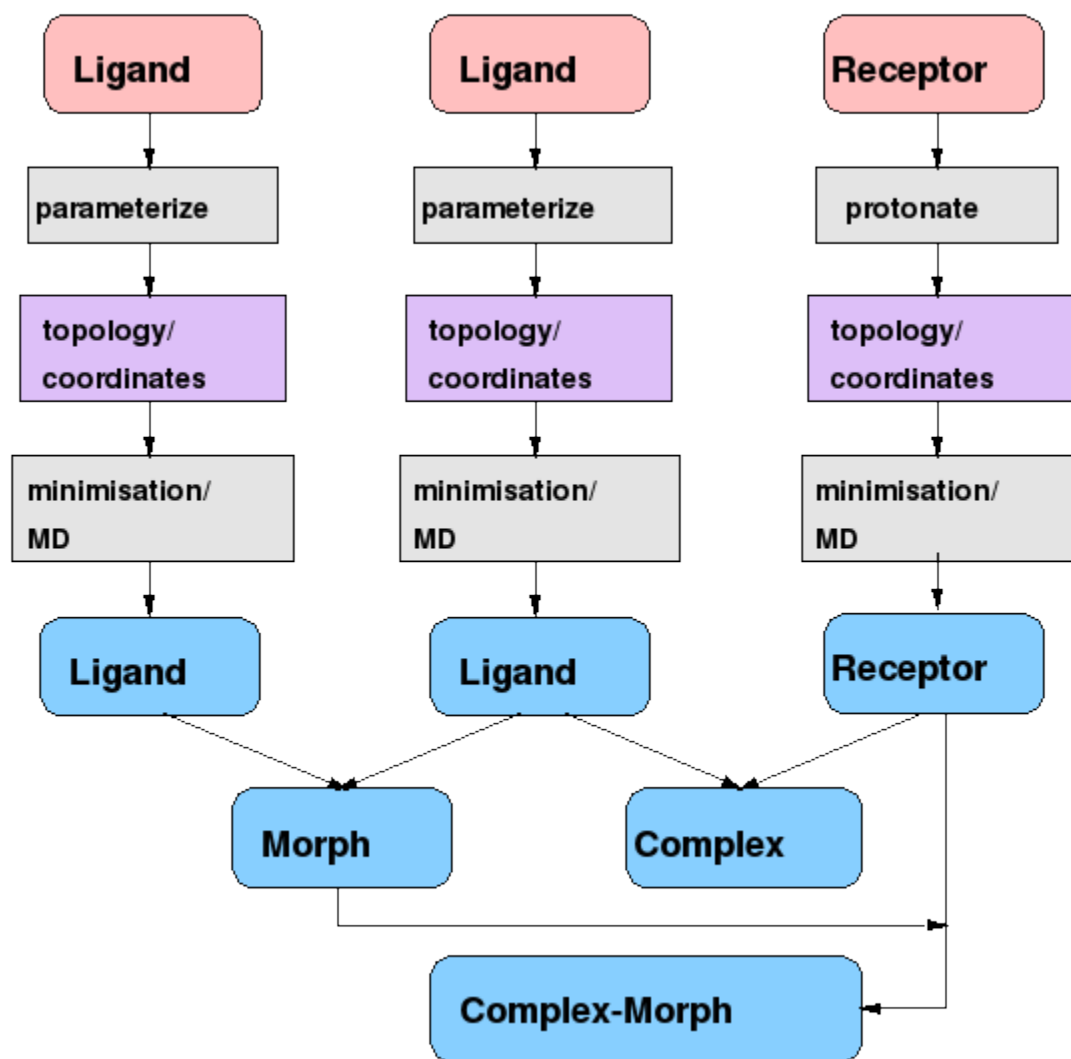
1
2
3 longest path through the graph. A solution has been proposed earlier⁶ but this removes potential
4 mappings in the R-groups and also the asymmetric carbon.
5
6

7
8 To handle all these complexities, FESetup allows the user to arbitrarily assign desired
9 mappings through a file mechanism. The user creates a file for each mutation containing pairs of
10 mapping indices for each corresponding atom pair. This is a very general solution that allows
11
12
13
14
15
16
17
18
19

Workflow

20 Scheme 1 depicts the principal work flow in FESetup to set up a relative free energy
21 calculation that morphs from ligand 1 to ligand 2 while bound to a receptor. Various third party
22 software and toolkits are recruited for setup. Openbabel²² is used to convert file formats, carry
23 out preliminary minimization and possibly create alternative conformations. The latter can be
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

useful to provide multiple starting structures and could potentially be used to create charges from multiple conformations. RDKit²³ is used to compute the MCS between a ligand pair (a “morph”) with the fmcs algorithm¹⁹. Sire is used to read AMBER topology files and provide data structures for force field parameters and allow for their manipulation. Furthermore, Sire can detect and select internal degrees of freedom for Monte Carlo (MC) simulations. The AmberTools antechamber and leap are used to create AMBER topology files including force field parameterization.



Scheme 1. The principal work flow in FESetup. Optional steps are in gray boxes. The pink boxes signify input structures and blue boxes ready simulation systems.

The work flows for ligand and receptor are independent from each other and arbitrary numbers of either can be run. The ligand is automatically parametrized with antechamber but users can supply parameters of their own. FESetup supports currently all modern AMBER force fields for biomolecules, lipids and carbohydrates. For the ligand the default force field is GAFF²⁴ but this can be overwritten, e.g. for a previous free energy study on carbohydrates²⁵ we have used the GLYCAM force field²⁶. By default atomic partial charges are derived with the AM1–BCC method²⁷.

1
2
3 The semi-empirical AM1-BCC charges are derived with sqm, the semi-empirical quantum
4 mechanics code used to compute the atomic partial charges, which is supplied with AmberTools.
5
6 The default convergence criteria are very tight, however. FESetup will reduce these criteria
7
8 when SCF convergence fails. As part of this, the MMFF94 force field may be used to slightly
9
10 modify the starting coordinates. This has been found to facilitate convergence in sqm in some
11
12 cases. The final charges are written by antechamber with 4 digit precision only which can lead
13
14 to rounding errors and a non-integer net charge. The code redistributes this charge to a higher
15
16 precision evenly among all atoms to match the total molecular charge.
17
18
19
20
21

22 Protonation, and more specifically the tautomeric, state of the ligand is the responsibility of the
23
24 user. A recent paper has pointed out the inherent problems of assigning these automatically with
25
26 chemoinformatic tools²⁸. Total charges are communicated through the input structure file.
27
28 Openbabel will read these from e.g. the PDB and SDF formats but *not* the popular MOL2
29
30 format. The latter format is typically used with the antechamber tool chain.
31
32
33

34 After parametrization, leap is recruited to create topology and coordinate files for either
35
36 vacuum or a solvated simulation box with counter ions. Preset minimization and MD protocols
37
38 can be carried out to adjust the density and provide starting velocities via an abstraction interface
39
40 that supports a number of MD engines: AMBER, GROMACS, NAMD or DL_POLY²⁹.
41
42 Topologies and coordinates will be transparently converted to GROMACS and DL_POLY
43
44 formats (but alchemical free energy simulations are not yet supported for the latter). The direct
45
46 conversion has the advantage that the newest AMBER force fields are always immediately
47
48 available independent of a native port to a particular MD package.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The receptor can optionally be protonated with PROPKA3^{30,31} but otherwise the work flow is
4 the same as for the ligand. The receptor can optionally be aligned along the principal axes. Any
5 biomolecule which is supported by the AMBER force fields and leap may be used.
6
7

8
9
10 Two ligands can be combined into a morph pair. FESetup will determine the maximum
11 common substructure and use this to set up the mapped region with a set of common coordinates.
12 Any atoms not mapped in this way are described as softcore (dummy) atoms. The code will
13 create all necessary topology and coordinate files for the perturbed simulation for the MD
14 packages AMBER, GROMACS and Sire. NAMD is principally supported too because NAMD
15 can read AMBER files but it is dual-topology only (see discussion above).
16
17
18
19
20
21
22
23

24 The receptor can be combined with a complex to form a complex-morph. The solvated box is
25 created from the coordinates of the unperturbed simulation system. The only additional atoms
26 and their coordinates are those for the dummy atoms. These will be computed from the internal
27 coordinates of the other state with the existing atoms. This is not necessary for AMBER³²
28 because AMBER topologies can be created without explicit dummy atoms. The assumption is
29 made that all bonded terms involving softcore atoms will cancel in a thermodynamic cycle and
30 thus need not be explicitly computed. FESetup provides, however, setup with explicit dummy
31 atoms for AMBER too.
32
33
34
35
36
37
38
39
40
41
42

43 The force field parameters for dummy atoms have to be created except for AMBER which
44 handles this entirely internally. For GROMACS and Sire we follow a scheme of copying the
45 bond and angle force field terms from the respective other state i.e. the end state where the atoms
46 do exist. The end state of a bond or angle that involves dummy atoms has thus the same, non-
47 zero parameters as the original atom. This implies that these bonded parameters of softcore
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 atoms will be independent of the coupling parameter λ . Dihedrals or impropers involving dummy
4 atoms can be set to zero.
5
6
7
8
9

10 Input/Output

11
12 FESetup has been designed to be easy to use with little input from the user and with sensible
13 defaults but also at the same time to enable the user to overwrite decisions the software may
14 make. Input is handled through a shell script, called FESetup, which sets up the environment
15 and calls dGprep.py which is the actual Python code handling user input. The input file is in an
16 INI style format very close to the configuration file format as typically used on Microsoft
17 operating systems and is thus easy to understand and parse. The input file is divided into 4
18 sections: [global], [complex], [protein] (historical synonym for receptor) and [ligand] (historical
19 synonym for unparameterized molecule) sections.
20
21
22
23
24
25
26
27
28
29
30

31
32 FESetup will create all topology and template control files required for simulation. The input
33 files do not, however, prescribe a specific λ schedule. It is not clear *a priori* what λ path would
34 guarantee a smooth gradient (TI) or sufficient energetic overlap (FEP/BAR). This will depend on
35 the nature of the system and is still an open question.
36
37
38
39

40
41 The *Supporting Information* demonstrate typical input examples and also provide the results
42 for the relative hydration free energies of two test system. Further validation is presented
43 through single point energies of the end-states. Fully worked tutorials can be found on our web
44 page (see below).
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

API

With FESetup we also define an application programming interface (API). In fact, dGprep.py is an elaborate example of how to use the API. See the *Supporting Information* for details.

Online material

FESetup can be downloaded from <http://www.hecbiosim.ac.uk/fesetup>. The source code repository including a Wiki, tutorials, a manual page, a discussion board, bug tracker and feature tracker are hosted through CCP–Forge on <http://ccpforge.cse.rl.ac.uk/gf/project/ccpbiosim/>.

FESetup is developed in Python 2.7 and thus principally highly portable as are the third party dependencies. Binaries are available for Linux, MacOSX and Windows. FESetup is licensed as free software under the GPL2 license.

Applications

FESetup is already cited in a number of published studies and is used in several ongoing studies.

In previous work³⁵ FESetup was used to prepare input files for multiple ligands bound to diverse proteins and in solution. Alchemical free energy calculations were not performed, but FESetup was very useful to automate the setup process and to ensure that a consistent setup protocol was applied throughout. The resulting input files were used to produce molecular dynamics simulation trajectories and these were processed by the software nautilus³³⁻³⁴ to compute enthalpies and entropies of water molecules in diverse protein binding sites using the grid cell theory method.

In another paper²⁵ the usefulness of the specialized carbohydrate force field GLYCAM and the general force field GAFF have been investigated in their application to a lectin complexed with

1
2
3 various mono-saccharide pyranoses. Alchemical free energy simulations have shown that
4
5 GLYCAM mostly outperforms GAFF. FESetup has been used to set up all free energy
6
7 simulations. The mapping feature proved to be very useful to preserve chirality and maximum
8
9 mapping at the same time.
10
11

12 13 14 Conclusions and Outlook

15 FESetup is a pipeline tool to make the setup of alchemical free energy simulations easier, faster
16
17 and less error-prone. The tool can flexibly be integrated into larger workflows taking in a wide
18
19 variety of structures. It creates simulation input for the MD packages AMBER, GROMACS,
20
21 Sire, and, to some extent, NAMD. A maximum common substructure algorithm is used to
22
23 automatically determine one-to-one mappings between start and end states. Ligands are
24
25 automatically and transparently parametrized to provide AM1-BCC charges.
26
27

28
29 An abstract MD engine can be used to equilibrate (minimize, heat, pressurize, restraint release)
30
31 the unperturbed simulation systems. These engines are available for AMBER, GROMACS,
32
33 NAMD and DL_POLY, and can be used independently from the alchemical setup code. In
34
35 principle, this can be developed into a general MD driver and combined with job submission
36
37 code to off-load compute-intensive tasks to a remote HPC system.
38
39

40
41 FESetup is free software released under the GPL2 license. The code is installed locally which
42
43 means that no confidential data has to be transferred over a network, is always available and the
44
45 source code can be inspected and modified. The third party software is freely available too. The
46
47 project follows an open development model accepting contributions of all kinds.
48
49

50
51 Future development will focus on expanding functionality to mutation of covalently bound
52
53 moieties like side-chains, and integration of absolute free energy setup. The goal here is to
54
55 allow arbitrary transformations just as with the current ligand support. Other popular MD
56
57
58
59
60

1
2
3 packages (support for the PERT module in CHARMM³⁶ is currently developed) and other
4
5 popular force fields like CHARMM or OPLS will be supported too. Alternative parametrization
6
7 schemes like RESP³⁷ can be integrated too as the work-flow is mostly automatic.
8
9

10 11 12 13 14 15 ASSOCIATED CONTENT

16
17
18 **Supporting Information.** Detailed input examples, examples of free energy changes computed
19
20 with different codes using FESetup generated input files. Table of single point energies for
21
22 several transformations setup for different simulation packages.
23
24

25 26 27 AUTHOR INFORMATION

28 29 **Corresponding Author**

30
31
32 *Hannes H Loeffler, *E-mail:* Hannes.Loeffler@stfc.ac.uk
33
34

35 36 **Author Contributions**

37
38 The manuscript was written through contributions of all authors. All authors have given approval
39
40 to the final version of the manuscript. ‡These authors contributed equally.
41
42

43 44 **Funding Sources**

45
46 HHL is supported through an EPSRC provided SLA, funding the core support of CCPBioSim.
47
48 CCPBioSim is the Collaborative Computational Project for Biomolecular Simulation funded by
49
50 EPSRC grants EP/J010588/1 and EP/M022609/1. JM is supported by a Royal Society
51
52 University Research Fellowship. CJW is grateful to the BBSRC (BB/K016601/1) and
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

BrisSynBio for funding. BrisSynBio is a BBSRC/EPSRC-funded Synthetic Biology Research Centre (BB/L01386X/1).

Notes

ACKNOWLEDGMENT

We are grateful to the continuing HPC support through the STFC Scientific Computing Department's SCARF cluster and the STFC Hartree Centre. CCP-Forge (EPSRC grant EP/L000342/1) is acknowledged for hosting the source code repository, the wiki and the tutorials. James Gebbie (HECBioSim) is acknowledged for designing and hosting the front-end page for FESetup and contributions towards the installer. Chin Yong and Ilian Todorov are acknowledged for helping with DL_FIELD and DL_POLY support. Jason Swails is acknowledged for helping with ParmEd.

ABBREVIATIONS

MD, Molecular Dynamics; MCSS, maximum common substructure search; MCS maximum common substructure; AM1, Austin Model 1 Hamiltonian; BCC, bond charge correction; SCF, self-consistent field; API, application programming interface; MC, Monte Carlo;

References

- (1) Michel, J. Current and emerging opportunities for molecular simulations in structure-based drug design. *Phys. Chem. Chem. Phys.* **2014**, *16*, 4465–4477.
- (2) Homeyer, N., Gohlke, H. FEW: A workflow tool for free energy calculations of ligand binding. *J. Comput. Chem.* **2013**, *34*, 965–973.
- (2a) Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, *42*, 724–733.

1
2
3 (3) Case, D. A.; Thomas E., III, C.; Darden, T.; Gohlke, H.; Luo, R.; Merz, Jr., K. M.;
4 Onufriev, A.; Simmerling, C.; Wang, B., Woods, R. J. The Amber biomolecular simulation
5 programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
6
7

8
9
10 (4) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A., Case, D. A. Continuum
11 Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J. Am.*
12 *Chem. Soc.* **1998**, *120*, 9401–9409.
13
14

15
16
17 (5) Gutiérrez-de-Terán, H., Åqvist, J. Linear Interaction Energy: Method and Applications in
18 Drug Design. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Springer New York,
19 2012; pp 305–323.
20
21
22

23
24 (6) Christ, C. D., Fox, T. Accuracy Assessment and Automation of Free Energy Calculations
25 for Drug Design. *J. Chem. Inf. Model.* **2014**, *54*, 108–120.
26
27

28
29 (7) Gapsys, V.; Michielssens, S.; Seeliger, D., de Groot, B. L. pmx: Automated protein
30 structure and topology generation for alchemical perturbations. *J. Comput. Chem.* **2015**, *36*, 348–
31 354.
32
33
34

35
36 (8) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E., Berendsen, H. J. C.
37 GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
38
39

40
41 (9) Lundborg, M., Lindahl, E. Automatic GROMACS Topology Generation and Comparisons
42 of Force Fields for Solvation Free Energy Calculations. *J. Phys. Chem. B* **2015**, *119*, 810–823.
43
44

45
46 (10) Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.,
47 Mobley, D. L. Lead optimization mapper: automating free energy calculations for lead
48 optimization. *J. Comp.-Aided Mol. Des.* **2013**, *27*, 755–770.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 (11) Sadiq, S. K.; Wright, D.; Watson, S. J.; Zasada, S. J.; Stoica, I., Coveney, P. V. Automated
4 Molecular Simulation Based Binding Affinity Calculator for Ligand-Bound HIV-1 Proteases. *J.*
5
6 *Chem. Theory Comput.* **2008**, *48*, 1909–1919.
7

8
9
10 (12) Woods, C. J., Michel, J. M. *Sire: An advanced, multiscale, molecular simulation*
11 *framework*, <http://siremol.org/> (accessed Sep 4, 2015).
12

13
14 (13) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.;
15 Skeel, R. D.; Kale, L., Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput.*
16 *Chem.* **2005**, *26*, 1781–1802.
17

18
19 (14) Hansen, N., van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A
20 Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632–2647.
21

22
23 (15) Mobley, D., Guthrie, J. FreeSolv: a database of experimental and calculated hydration
24 free energies, with input files. *J. Comp.-Aided Mol. Des.* **2014**, *28*, 711–720.
25

26
27 (16) Boresch, S., Karplus, M. The Role of Bonded Terms in Free Energy Simulations: 1.
28 Theoretical Analysis. *J. Phys. Chem. A* **1999**, *103*, 103–118.
29

30
31 (17) Michel, J., Essex, J. Prediction of protein--ligand binding affinity by free energy
32 simulations: assumptions, pitfalls and expectations. *J. Comp.-Aided Mol. Des.* **2010**, *24*, 639–
33 658.
34

35
36 (18) Gumbart, J. C.; Roux, B., Chipot, C. Standard Binding Free Energies from Computer
37 Simulations: What Is the Best Strategy?. *J. Chem. Theory Comput.* **2013**, *9*, 794–802.
38

39
40 (19) Dalke, A., Hastings, J. FMCS: a novel algorithm for the multiple MCS problem. *J.*
41 *Cheminform.* **2013**, *5*, O6.
42

43
44 (20) Raymond, J. W., Willett, P. Maximum common subgraph isomorphism algorithms for the
45 matching of chemical structures. *J. Comp.-Aided Mol. Des.* **2002**, *16*, 521–533.
46
47
48
49
50
51
52

- 1
2
3 (21) Liu, S.; Wang, L., Mobley, D. L. Is Ring Breaking Feasible in Relative Binding Free
4 Energy Calculations?. *J. Chem. Inf. Mod.* **2015**, *55*, 727–735.
5
6
7
8 (22) O'Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T., Hutchison, G. Open
9 Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33–46.
10
11
12 (23) Landrum, G. *RDKit: Open-source cheminformatics*, <http://www.rdkit.org> (accessed
13 Sep 4, 2015).
14
15
16
17 (24) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A., Case, D. A. Development and
18 testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
19
20
21
22 (25) Mishra, Sushil K; Calabro, Gaetano; Loeffler, Hannes H.; Michel, J.; Koča, J. Evaluation
23 of selected classical force fields for alchemical binding free energy calculations of protein-
24 carbohydrate complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3333–3345.
25
26
27
28 (26) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C.
29 R.; Foley, B. L., Woods, R. J. GLYCAM06: A generalizable biomolecular force field.
30 Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622–655.
31
32
33
34 (27) Jakalian, A.; Bush, B. L.; Jack, D. B., Bayly, C. I. Fast, efficient generation of high-
35 quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
36
37
38
39 (28) Sayle, R. So you think you understand tautomerism?. *J. Comp.-Aided Mol. Des.* **2010**, *24*,
40 485–496.
41
42
43
44 (29) Todorov, I. T.; Smith, W.; Trachenko, K., Dove, M. T. DL_POLY_3: new dimensions in
45 molecular dynamics simulations via massive parallelism. *J. Mater. Chem.* **2006**, *16*, 1911–1918.
46
47
48
49 (30) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M., Jensen, J. H. Improved Treatment
50 of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J.*
51 *Chem. Theory Comput.* **2011**, *7*, 2284–2295.
52
53
54
55
56
57
58
59
60

1
2
3 (31) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M., Jensen, J. H. PROPKA3:
4 Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem.*
5
6 *Theory Comput.* **2011**, *7*, 525–537.
7
8

9
10 (32) Kaus, J. W.; Pierce, L. T.; Walker, R. C., McCammon, J. A. Improving the Efficiency of
11
12 Free Energy Calculations in the Amber Molecular Dynamics Package. *J. Chem. Theory Comput.*
13
14 **2013**, *9*, 4131–4139.
15
16

17 (33) Michel, J.; Henchman, R. H.; Gerogiokas, G.; Southey, M. W. Y.; Mazanetz, M. P., Law,
18
19 R. J. Evaluation of Host--Guest Binding Thermodynamics of Model Cavities with Grid Cell
20
21 Theory. *J. Chem. Theory Comput.* **2014**, *10*, 4055–4068.
22
23

24 (34) Gerogiokas, G.; Calabro, G.; Henchman, R. H.; Southey, M. W. Y.; Law, R. J., Michel, J.
25
26 Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory. *J. Chem.*
27
28 *Theory Comput.* **2014**, *10*, 35–48.
29
30

31 (35) Gerogiokas, G.; Southey, M. W. Y.; Mazanetz, M. P.; Hefetz, A.; Bodkin, M.; Law, R. J.,
32
33 Michel, J. Correction: Evaluation of water displacement energetics in protein binding sites with
34
35 grid cell theory. *Phys. Chem. Chem. Phys.* **2015**, *17*, 16213–16213.
36
37

38 (36) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won,
39
40 Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig,
41
42 M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov,
43
44 V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.;
45
46 Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M., Karplus, M. CHARMM: The biomolecular
47
48 simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
49
50
51
52
53
54
55
56
57
58
59
60

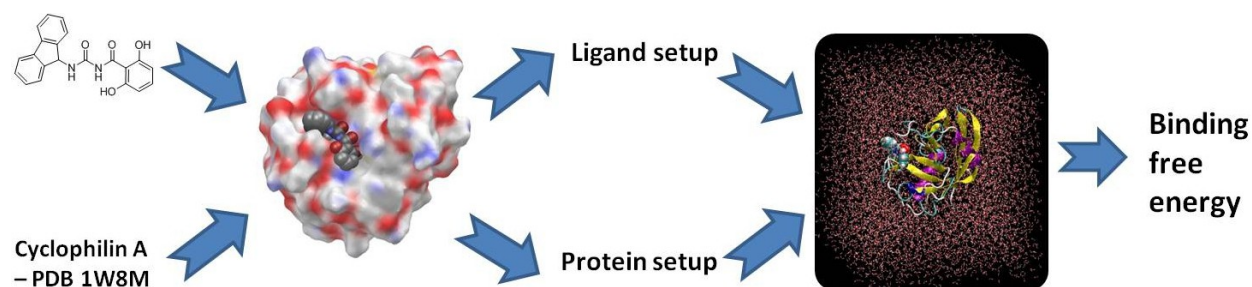
1
2
3 (37) Bayly, C. I.; Cieplak, P.; Cornell, W., Kollman, P. A. A well-behaved electrostatic
4 potential based method using charge restraints for deriving atomic charges: the RESP model. *J.*
5
6
7
8 *Phys. Chem.* **1993**, *97*, 10269–10280.

9
10 (38) Mobley, D. L.; Klimovich, Pavel V. Perspective: Alchemical free energy calculations for
11 drug discovery. *J. Chem. Phys.* **2012**, *137*, 230901.

12
13 (39) Michel, J; Verdonk M. L.; Essex J. W. Protein-ligand complexes: computation of the
14 relative binding free energy of different scaffolds and binding modes. *J. Chem. Theory and*
15
16
17
18
19
20 *Comp.* **2007**, *3*, 1645–1655.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Insert Table of Contents Graphic and Synopsis Here



A graphical workflow of FESetup. A protein from the PDB and a ligand are combined and solvated. Input files are created to carry out binding free energy simulations.

