CrossMark

PERSPECTIVE

Collaborating to improve the use of free-energy and other quantitative methods in drug discovery

Bradley Sherborne[1] · Veerabahu Shanmugasundaram[2] · Alan C. Cheng[3] · Clara D. Christ[4] · Renee L. DesJarlais[5] · Jose S. Duca[6] · Richard A. Lewis[7] · Deborah A. Loughney[8] · Eric S. Manas[9] · Georgia B. McGaughey[10] · Catherine E. Peishoff[11] · Herman van Vlijmen[11]

Abstract In May and August, 2016, several pharmaceutical companies convened to discuss and compare experiences with Free Energy Perturbation (FEP). This unusual synchronization of interest was prompted by Schrödinger's FEP+ implementation and offered the opportunity to share fresh studies with FEP and enable broader discussions on the topic. This article summarizes key conclusions of the meetings, including a path forward of actions for this group to aid the accelerated evaluation, application and development of free energy and related quantitative, structure-based design methods.

Keywords FEP · Structure-based design · QSBD · Software · Review

Background

The promise of alchemical free energy (more commonly known as "free energy perturbation" or simply FEP) calculations for quantitative predictions of activity in rational drug design was first demonstrated in the late 80s [1]. In subsequent decades however, FEP failed to become a standard computational tool within the pharmaceutical industry due to lack of computational power, force field limitations and general usability challenges. Structure-based prediction of binding affinities remains an important, even aspirational benchmark for computational methods, although it is one of many needs for computation in the characterization and prioritization of candidate compounds in modern drug discovery. Last year, authors from Schrödinger, Nimbus and academic associates published a paper in JACS [2] entitled "Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field", describing the performance of the FEP+ product, reflecting investments and advancements at Schrödinger that have made FEP accessible to many more scientists. In May of this year, more than a dozen pharmaceutical companies met to compare

✉ Bradley Sherborne
brad.sherborne@merck.com

[1] Modeling and Informatics, Merck & Co., Inc., 2000 Galloping Hill Road, Kenilworth, NJ 07033, USA

[2] Discovery Sciences, Pfizer Worldwide Research & Development, 300 Eastern Point Road, Groton, CT 06340, USA

[3] Amgen Inc., 1120 Veterans Blvd., South San Francisco, CA 94080, USA

[4] Drug Discovery, BAYER Pharma AG, Muellerstrasse 178, 13353 Berlin, Germany

[5] Janssen Research & Development LLC, Welsh & McKean Roads, 41-4124, Spring House, PA 19477, USA

[6] Computer-Aided Drug Discovery, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, 100 Technology Square, Cambridge, MA 02139, USA

[7] Novartis Pharma Ag, 4002 Basel, Switzerland

[8] Bristol-Myers Squibb Company, PO Box 4000, Princeton, NJ 08540, USA

[9] Computational and Modeling Sciences, GlaxoSmithKline, 1250 South Collegeville Road, Collegeville, PA 19426, USA

[10] Modeling & Informatics, Vertex Pharmaceuticals Inc., 50 Northern Ave, Boston, MA 02210, USA

[11] Janssen Research & Development, Turnhoutseweg 30, 2340 Beerse, Belgium

experiences in evaluating FEP and other quantitative structure-based design methods on data from internal projects. With the usual rules of confidentiality in place, atomic, compound, or even target-level detail could not be readily exchanged among the participants. Despite these restrictions, the group quickly formed a preliminary path forward to leverage collective experiences with a goal of developing best practices, better defining the domain of applicability and building an environment and culture that encourages and accelerates the development of capabilities in this field.

The central role of the domain of applicability

Clearly multiple companies had applied free energy calculations to different datasets, yet all present had come to the same conclusion – in the best cases, the accuracy was in general agreement with the 0.8–1.1 kcal/mol Mean Unsigned Error (MUE) error in activities reported in the 2015 paper (a good outcome as reproducibility of any method is critical). Understanding where the method could be applied successfully, however, was the common challenge: it simply was not clear *a priori* for which targets and ligand modifications the method would be predictive and for which ones it would not.

There are some general situations that can cause inaccurate predictions, thus limiting the (still poorly understood) "domain of applicability"—e.g.

- incorrect system preparation through e.g. tautomer or protomer states
- protein disorder or plasticity
- low resolution crystal structures
- presence of metals
- multiple binding modes
- compounds of varying net charge
- changes in ring size
- differing number of water molecules in the binding site, or trapped water molecules during the simulation

Besides these situations, there are also seemingly "compatible" targets for which the companies found that FEP did not predict experimental data well. The issues with this latter set are not obvious and additional work is needed to identify them. Finally, even more subtle aspects to understanding the domain of applicability were experienced that need further investigation:

1. Selecting the right protein structure is important and, while a suitable compound test set can be used to select the "best" structure, this is not always possible in a discovery project.

2. In some cases it was seen that for the same protein one series was accurately predicted while another was not. Understanding what other features of the ligand and of ligand:protein interactions may affect performance of free energy methods will be key to avoid unexpected failure.

3. Occasionally predictions were found to be very inaccurate without any obvious cause. Further diagnostics are required to understand these cases which may relate to point 2 above.

Putting free energy methods in the drug discovery context

Is FEP ready for prime time as a core method in hit to lead and lead optimization chemistry? An accuracy of 0.6–1 kcal/mol can be regarded as necessary to guide medicinal chemistry optimization of ligands, as it would allow predictions of binding affinity to be correct to within roughly two- to fivefold. Higher accuracies are meaningless in most cases due to experimental error in the affinity assays. The previously mentioned MUE values of 0.8–1.1 kcal/mol indicate that for systems where the method works, FEP is suitable for ranking multiple ligand alternatives during medicinal chemistry optimization. The computational throughput of FEP is currently about 1 ligand idea per day per GPU. Depending on the number of new compound ideas per day per target in a project, clarifying which types of industrial drug discovery problems are compatible, tractable and of sufficient value is of great interest.

Science needs

There are multiple factors in the process between X-ray structure, ligands, and FEP prediction, such as choice of force field, ligand/protein preparation processes, protonation and tautomeric states, and sampling strategy, and the best approach to each of these aspects is something that still needs systematic research. (Coordinated, blinded prediction challenges offer the best opportunity to develop broad understanding and broadly applicable methods—see actions 6 & 7 below).

At a more detailed level, there are improvements to be made in force fields and conformational sampling to expand the range of applicability, and it is clear that access to relevant test sets is essential to improve both the computational details and parameters, as well as the workflow of setting up the protein–ligand system. (Generation and sharing of high quality datasets is necessary for the development of

methods and validation through blind challenges—see action 5 below.)

Finally, understanding the effort/benefit/performance landscape where free energy methods might be appropriate to use depends on good comparative data from different methods. (Common, practical guidelines for comparing quantitative methods, including the statistical means to quantify the results of particular ligand:protein studies are needed to accurately assess and compare performance—see actions 1–4 below.)

Committing to the future

The May meeting of this industry group gave a sense of common interest, purpose and excitement, and provided a good start to the difficult task of coordinating and openly communicating amongst dozens of pharmaceutical companies to help drive the science behind FEP. There is already good alignment with the NIH supported Drug Design Data Resource initiative (D3R) [3] that will ensure an open means to share data and hold blind predictions. Beyond that, the goodwill, open and honest communications between pharma, software vendors and academic researchers will be crucial in building upon the enthusiasm and urgency established by this collective group.

As a group, the following needs and commitments were outlined:

1. Recommend best practices to ensure that the accuracy of predictions is comparable between studies and in prospective use.
2. Recommend statistical methods for quantitating results (i.e. inclusion of 95% CI, RMSE, Kendal tau-b etc.).
3. Identify and recommend null models for use when comparing FEP to other binding affinity prediction methods that can be faster and higher-throughput, though competitive with more rigorous methods like FEP.
4. Identify factors that should be shared to better diagnose the likely success and domain of applicability of calculations (e.g. target class, resolution …).
5. Provide high quality datasets that can be used for benchmark purposes [4].
6. Assist in open, blind challenges organized through D3R [3] to maximize learnings and effectiveness.

7. Work with the D3R [3] and the wider academic community to encourage the development of a framework for community-wide challenges that maximizes interoperability and encourages crossover studies.

The industry group reconvened for a multi-day meeting in August to focus on the scientific lessons learnt so far and to fill in the details of the needs and commitments outlined in May. By bringing together our results and experiences, and by collaborating with software vendors and academia to improve the methods, this industry group is committed to guide and assist the wider community in establishing the right place for free energy and related quantitative structure-based design methods in the computational chemists' toolbox.

Acknowledgements

References

1. Bash PA, Singh UC, Brown FK, Langridge R, Kollman PA (1987) Science 235(4788):574–576
2. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, Murcko M, Fye L, Farid R, Lin T, Mobley DL, Jorgensen WL, Berne BJ, Freisner RA, Abel R (2015) J Am Chem Soc 137(7):2695–2703
3. D3R website at https://drugdesigndata.org/
4. See also:https://github.com/mobleylab/benchmarksets, http://doi.10.5281/zenodo.163963