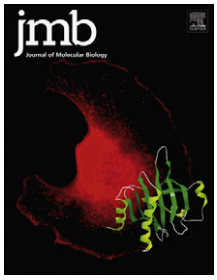

Free Energy Perturbation Calculations of the Thermodynamics of Protein Side-Chain Mutations

Thomas Steinbrecher[1], Robert Abel[2], Anthony Clark[2] and Richard Friesner[3]

*1 - **Schrödinger GmbH,** Dynamostr. 13, 68165 Mannheim, Germany*
*2 - **Schrödinger Inc.,** 120 West 45th Street, 17th Floor, New York, NY 10036, USA*
*3 - **Department of Chemistry,** Columbia University, 3000 Broadway, New York, NY 10027, USA*

***Correspondence to Thomas Steinbrecher and Richard Friesner:**. thomas.steinbrecher@schrodinger.com; raf8@columia.edu*
http://dx.doi.org/10.1016/j.jmb.2017.03.002
Edited by Anna Panchenko

Abstract

Protein side-chain mutation is fundamental both to natural evolutionary processes and to the engineering of protein therapeutics, which constitute an increasing fraction of important medications. Molecular simulation enables the prediction of the effects of mutation on properties such as binding affinity, secondary and tertiary structure, conformational dynamics, and thermal stability. A number of widely differing approaches have been applied to these predictions, including sequence-based algorithms, knowledge-based potential functions, and all-atom molecular mechanics calculations. Free energy perturbation theory, employing all-atom and explicit-solvent molecular dynamics simulations, is a rigorous physics-based approach for calculating thermodynamic effects of, for example, protein side-chain mutations. Over the past several years, we have initiated an investigation of the ability of our most recent free energy perturbation methodology to model the thermodynamics of protein mutation for two specific problems: protein–protein binding affinities and protein thermal stability. We highlight recent advances in the field and outline current and future challenges.

Introduction

Protein side-chain mutation is fundamental both to natural evolutionary processes and to the engineering of protein therapeutics, which constitute an increasing fraction of important medications [1–3]. Mutations enable proteins to adopt different shapes, catalyze new reactions, and engage in highly sophisticated forms of molecular recognition, involving both small-molecule ligands and other macromolecules. A wide range of natural events give rise to mutation in biological systems, including the effects of radiation and carcinogens, errors in reading tRNAs at the ribosome, and the generation of antigen-binding CDR loops of antibodies in the immune system. In the laboratory, techniques such as phage display [4–7] and directed evolution [8] enable huge numbers of monoclonal antibody variants to be screened for binding affinity to a targeted antigen.

Molecular simulation enables the prediction of the effects of mutation on properties such as binding affinity, secondary and tertiary structure, conformational dynamics, and thermal stability. Note that we use the term *mutation* throughout the text to indicate protein amino acid substitutions, that is, changes on the level of protein structure instead of DNA sequence. A number of widely differing approaches have been applied to these problems [9–11], including sequence-based algorithms, knowledge-based potential functions [12,13], and all-atom molecular mechanics calculations [14]. While some successes have been achieved in explaining the available experimental data, accurate and reliable predictive capability, for example, at the level needed to significantly impact a monoclonal antibody development project has not yet been demonstrated.

Method Overview

Free energy perturbation (FEP) theory, employing all-atom and explicit-solvent molecular dynamics

(MD) simulations, is a rigorous physics-based approach for calculating the thermodynamic effects of changes in chemical structures [15–17], such as, for example, protein side-chain mutations. FEP is one of a variety of free energy calculation techniques, which is based on using the famous Zwanzig formula [18] to compute the free energy difference between two states, by evaluating the potential energy differences averaged over a Boltzmann-weighted ensemble of their conformational space. In practice, the total change is typically broken down into multiple substeps by introducing intermediate states. The start and end state of the chemical change to be investigated are imagined to lie on a non-physical coordinate, commonly called λ, so that a change from $\lambda = 0$ to $\lambda = 1$ corresponds to a change from initial to final state, and an intermediate state at $\lambda = 0.5$ corresponds to a midway point between both. MD simulations are then conducted for each state to generate a representative conformational ensemble. A key feature of the approach is that free energy changes can be computed for arbitrary, even non-physical, changes in a system as long as the potential energy functions for both end states can be formulated. Thus, a residue on a protein can be "alchemically" transformed into a different residue during the course of the simulation, for example by removing an amino acid side chain in a protein wild-type structure and growing another side chain in its place, yielding the free energy difference between the original and mutant protein. Such transformation free energies are not interpretable on their own, but quantities such as protein–protein binding affinities can be computed via the use of a thermodynamic cycle setup so that all non-physical contributions cancel in the final result. In the case of a change in protein–protein binding affinity, the free energy differences for the unbound states, and bound complex, are determined separately and combined to generate an estimate of the experimentally observable binding free energy change. In principle, with a sufficiently accurate force field and convergence of the MD sampling (accessing all of the energetically important conformational states of the system), FEP should produce accurate free energy predictions.

FEP methods have historically been applied predominantly to the binding of small-molecule ligands to protein receptors. Early efforts [19], beginning in the 1980s, were hindered by inaccuracies in protein and ligand force fields, lack of sufficient computing power to achieve converged simulation results, and difficulties in setting up suitable initial conditions for the simulations. Nevertheless, progress was made over the past several decades as force fields improved and available computational power increased, and the potential of the technique became apparent [16,20–22]. FEP or similar calculations can be performed with the majority of molecular modeling suites in wide usage, but implementation details, simulation efficiency, and prediction accuracy can vary widely between software packages.

A number of years ago, we embarked on an effort to develop an FEP-based methodology for protein-ligand binding, which was suitable for robust application to structure-based drug design. Key ingredients included the implementation of FEP methods on graphics processing units, a force field with extensive coverage of the chemical space relevant for medicinal chemistry, enhanced simulation algorithms to prevent the system from becoming trapped in a local minimum, and automated tools for setup and assessment of convergence. Progress on this effort has been reported in a number of publications, and a summary of the latest results can be found in Refs. [22–28]. With a RMS error of less than 1 kcal/mol across thousands of test cases, and multiple examples demonstrating significant impact on structure-based drug discovery projects, the use of FEP to optimize small-molecule binding to proteins is becoming well established.

The use of FEP to predict the effects of protein mutations, particularly for protein engineering applications relevant to systems of biomedical interest, has been much less extensively investigated. A few early calculations of protein–protein binding affinities have been reported [29–32], but these publications generally considered small data sets and present results from sub-nanosecond simulations that are unlikely to be converged. More recent protein FEP studies using modern free energy calculation tools, such as in Ref. [33], are still rare. While the use of FEP to predict protein mutations does not require the development of an elaborated small-molecule ligand force field, the sampling challenges of converging protein mutation simulations may be greater than what is typically required in a small-molecule binding study, as mutations are more likely to induce significant conformational rearrangements.

FEP Calculation of Protein Mutation Thermodynamics

Over the past several years, we have initiated an investigation of the ability of our most recent FEP methodology to model the thermodynamics of protein mutation for two specific problems: protein–protein binding affinities and protein stability. These calculations employ a new protein force field, a component of OPLS3 [34], which has greatly superior performance as compared to earlier versions of OPLS (and is comparable to the best variants of the CHARMm [35,36] and AMBER [37–39] protein force fields). Publications describing this work in detail appear in the current issue of the Journal of Molecular Biology. Below, we provide a brief summary of the important results, followed by a discussion of interesting problems of biological function and protein engineering,