Received: 18 May 2021 | Revised: 25 June 2021 | Accepted: 30 June 2021

DOI: 10.1002/wcms.1563

ADVANCED REVIEW

WIREs COMPUTATIONAL MOLECULAR SCIENCE WILEY

# Computational studies of protein–drug binding affinity changes upon mutations in the drug target

**Ran Friedman**

Department of Chemistry and Biomedical Sciences, Linnæus University, Kalmar, Sweden

**Correspondence**
Ran Friedman, Department of Chemistry and Biomedical Sciences, Linnæus University, Kalmar, SE-391 82, Sweden.
Email: ran.friedman@lnu.se

**Funding information**
The Swedish Cancer Society, Grant/Award Number: CAN 2018/362

**Edited by:** Peter R. Schreiner, Editor-in-Chief

## Abstract

Mutations that lead to drug resistance limit the efficacy of antibiotics, antiviral drugs, targeted cancer therapies, and other treatments. Accurately calculating protein–drug binding affinity changes upon mutations in the drug target is of high interest as this can yield a better understanding into how such mutations drive drug-resistance, especially when the mutation in question does not directly interfere with binding of the drug. The main aim of this article is to provide an up-to-date reference on the computational tools that are available for the calculation of Gibbs energy (free energy) changes upon mutation, their strengths, and limitations. The methods that are discussed include free energy calculations (free energy perturbation, thermodynamic integration, multistate Bennett acceptance ratio), analysis of molecular dynamics simulations (linear interaction energy, molecular mechanics [MM]/Poisson–Boltzmann solvated area, and MM/generalized Born solvated area), and methods that involve quantum mechanical calculations (including QM/MM). The possibility to use machine learning is also introduced. Given that the benefit of accurately calculating binding affinity changes upon mutation depends on comparing calculated values with experimental measurements, a brief survey on experimental methods and observables is provided. Examples of computational studies that go beyond calculating the Gibbs energy changes are given. Factors that need to be addressed by the computational chemist and potential pitfalls are discussed at length.

This article is categorized under:

Structure and Mechanism > Computational Biochemistry and Biophysics
Molecular and Statistical Mechanics > Free Energy Methods
Molecular and Statistical Mechanics > Molecular Interactions

**KEYWORDS**

density functional theory, free energy perturbation, molecular dynamics, protein–drug interactions, resistance mutations

## 1 | INTRODUCTION

Modern development of drugs to treat infections, which started about a century ago,[1] has changed the world as we knew it. Unfortunately, many strains of bacteria quickly become adapted to natural and semi-synthetic antibiotics,[2] and to synthetic antimicrobial agents. Although most infections can still be treated by small molecular drugs, multidrug resistance bacterial strains have become increasingly more common, and pose a serious threat to the health and well-being of humans and animals. Many other pathogens such as viruses, fungi, and unicellular parasites have become partially of fully resistant to the arsenal of drugs that are available. Targeted cancer therapies suffer from the same faith. These drugs are used with the aim of inhibiting the growth of tumors by targeting cell growth and proliferation mechanisms that are specific to the tumor; such drugs often work very well when the treatment begins but tumor cells gradually become resistant to therapy. Resistance to anticancer chemotherapy is also common. One of the most prevalent mechanisms by which viruses, cancer cells, and microorganisms become refractory to treatment is the development of mutations in the protein that is the molecular drug target.[3] A single substitution of an amino acid residue can be enough to drive drug resistance. Realizing that resistance is likely to emerge sooner or later, there is a need to develop treatment options that would do a better job in avoiding it. Combination therapies often perform better than single drugs in terms of efficacy and lower risk for resistance, but might also lead to additional side effects, and to increased risk for toxicity from several drugs and drug–drug interactions. Moreover, once combination therapies fail due to resistance, the availability of other therapies is limited. Thus, there is an urgent need to develop new therapies that are better suited to counteract resistance mutations.

The development of new drugs that overcome resistance mutations requires an understanding of the factors that contribute to resistance. In some cases, an examination of a crystal structure of the drug-bound molecular target is enough to envision how the mutation leads to resistance. However, in many cases, the situation is more complex as the mutated residue is not located at the vicinity of the drug, and it is unclear why the mutation is selected and leads to resistance. Fully accounting for the effects of mutations on drug resistance requires understanding of the target protein's dynamics and how it binds the drug. Simulation based methods have come to the point in which they are accurate enough to provide an insight into such processes and give clues about drug resistance.[4] In some cases, such simulations are enough to fully explain the changes in drug-binding thermodynamics. In others, a combination of data from structural, biophysical, and computational studies is required. The objective of this article is to provide a survey of the approaches that are used to this aim. After briefly introducing resistance mutations from an evolutionary standpoint, binding affinities are discussed in detail (describing both the experimental and simulation methods that are used to estimate these). Accurately estimating the binding affinity changes upon mutation would greatly increase our ability to design better drugs. Unfortunately, this is in many cases not straightforward and current limitations, as well as stumbling blocks are discussed at length in this article, after surveying the available computational methods. Finally, some case studies where simulations were used to provide a better understanding of drug resistance are presented. For completeness, additional factors that may lead to resistance are briefly discussed before the Conclusions section.

## 2 | HOW DO RESISTANCE MUTATIONS EMERGE?

To understand how resistance mutations emerge we should consider the principles that guide the evolution of species. The *species* here can be a microorganism, a virus, or a tumor, whose evolutionary development matches that of an organism in many senses.[5] Mutations in the DNA happen all the time, and only rarely lead to any benefit to the organism by increasing its *fitness* (capacity to produce offspring). When the pathogen (or tumor) is treated by drugs, there is an extreme modification of the pathogen's *fitness landscape* (Figure 1), which is a multidimensional landscape that matches genotype to fitness. Suddenly, the drug's effects make other evolutionary considerations (speed of growth, utilization of resources, etc.) much smaller in importance relative to drug resistance. Consequently, any mutation that makes the pathogen resistant is to be selected. The organism cannot decide which genes to modify, its survival now depends on adaptation in any way possible. Clearly, mutations in the molecular drug target are one of the most common mechanisms of resistance, especially against monotherapy with a specific drug.

Given that the new mutations are selected based on the need to overcome the effect of the drug, such mutations might slow down the proliferation of the pathogen. In rapidly evolving species (viruses and microorganisms[7,8]), and even in cancers[9] additional mutations may emerge to compensate for this. These are called “compensatory mutations.” Fortunately, this process cannot continue without limits, as too many mutations in the molecular targets either cannot

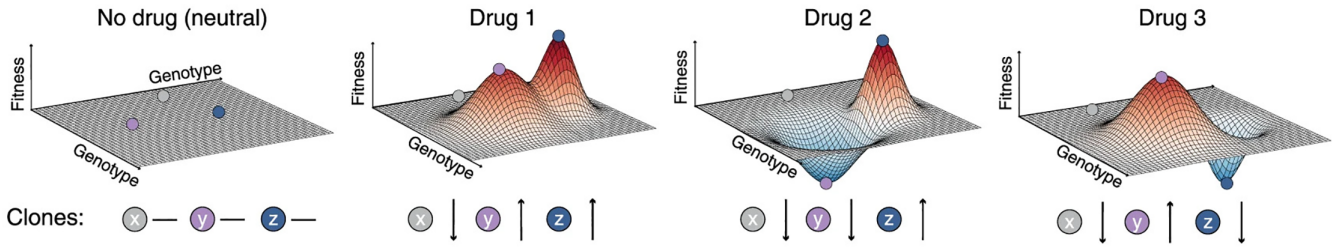


**FIGURE 1** The fitness landscape is dramatically influenced by drug treatment. Three clones (x, y, z) are equally fit in the absence of drugs. Drug 1 favors clones y and z, drug 2 favors clone z, and drug 3 favors clone y. Adapted from Reference 6, CC-BY 4.0 license

be tolerated, or limit the fitness of the pathogen.[10,11] This supports the notion that drugs that are more robust towards resistance are needed despite past experience with antibiotics and anti-parasitic drugs that were abandoned owing to gradual lack of efficacy. To develop such drugs there is a need to understand in details how resistance mutations affect drug binding.

## 3 | BINDING AFFINITIES

The *binding affinity* describes the strength of the interaction between (usually two) molecules that bind together. When discussing protein–drug interactions, the binding affinity (often abbreviated as just *affinity*) is a measurement of how well the drug binds to the protein. Resistance mutations often decrease the binding affinity.

The formal definition of affinity refers to *reaction affinity*, as described in the Appendix. As it is possible to calculate the reaction affinity for a protein–drug interaction, we are more often interested in equilibrium conditions, where the reaction is characterized by the association constant $K_a$ (sometimes referred to as the binding constant, $K_b$) or the (more commonly used) dissociation constant $K_d$ (see Section 3.1 for definitions). $K_d$ is given in concentration units; the lower it is, the higher the affinity.

### 3.1 | Definitions

#### 3.1.1 | Association, dissociation, and inhibition constants

Given a reaction between a protein, $P$ and any ligand $L$ such as a drug molecule:

$$P + L \rightleftharpoons PL. \tag{1}$$

The association constant is given by the concentration of the complex divided by the product of the concentrations of the reactants:

$$K_a = \frac{[PL]}{[P][L]}, \tag{2}$$

whereas the dissociation constant is the reciprocal of $K_a$, defined as:

$$K_d = \frac{[P][L]}{[PL]} = \frac{k_{-1}}{k_1}. \tag{3}$$

Here, $k_{-1}$ and $k_1$ are the rate constants for ligand dissociation and association, respectively. An alternative formalism is to use $k_{off}$ instead of $k_{-1}$ and $k_{on}$ instead of $k_1$.

When the protein in question is an enzyme and the drug is an enzyme inhibitor, the inhibition constant $K_i$ is often used which has the same meaning of $K_d$ when the inhibition is competitive. The two constants thus often represent the same reaction.

### 3.1.2 | Inhibitory concentration

The inhibitory concentration, or IC is the concentration of an inhibitor that is required to achieve a certain percent of inhibition of a given system. The most commonly used term is $IC_{50}$. When the concentration of an inhibitor is equal to $IC_{50}$, the reaction or process is inhibited by 50%. Other IC values are sometimes also observed in the literature such as $IC_{90}$, which is by definition higher than $IC_{50}$ since this is the concentration that is required to achieve 90% inhibition. $IC_{50}$ values may be measured for enzyme inhibition in vitro, but the system can also be different. As an example, in studies involving anticancer drugs, $IC_{50}$ values are often measured for cell growth. More generally, such values can represent any biochemical assay, where the inhibition of a biological activity is measured, for example, by the equilibrium concentration of a downstream reaction product.

For the typical case of a competitive binding reaction to an enzyme that follows Michaelis–Menten kinetics, $IC_{50}$ is related to $K_i$ as:

$$IC_{50} = K_i\left(1+\frac{[S]}{K_M}\right), \tag{4}$$

$K_M$ is the Michaelis constant of the enzyme. As this equation shows, $IC_{50}$ depends on the concentration of the substrate. Furthermore, when the mode of inhibition is different, so is the relationship between $IC_{50}$ and $K_i$. For further discussion about this relationship see Reference 12.

### 3.1.3 | Effective concentration

In analogy to $IC_{50}$, $EC_{50}$ is the concentration of an agonist that is required to achieve 50% of the maximal response, and other values such as $EC_{90}$ are sometimes used. Although the International Union of Pharmacology defines $EC_{50}$ for the effect of an agonist,[13] it is often used in the literature to represent the potency of a drug in a biochemical assay, irrespective if the drug is an agonist or antagonist.

### 3.1.4 | Gibbs energy of the binding reaction

For protein–drug binding under constant temperature and pressure, the Gibbs energy of the binding reaction $\Delta G_b$ is given by:

$$\Delta G_b = -RT\ \ln\ K_a = RT\ \ln\ K_d, \tag{5}$$

where $R$ is the gas constant and $T$ is the absolute temperature. Representative values are given in Table 1. Importantly, a mere change of 1.4 kcal/mol in $\Delta G_b$ is enough to modify the dissociation constant 10-fold.

The terms "free binding energy" or "binding free energy" as often found in the literature normally refer to $\Delta G_b$.

### 3.1.5 | Estimation of changes in the Gibbs energy upon resistance mutations

Clinically, resistance mutations limit the efficacy of drugs, so that the drug is no longer useful. We are thus often interested in estimating the Gibbs energy change upon mutation:

$$\Delta\Delta G_{b(S\rightarrow R)} = \Delta G_{b(R)} - \Delta G_{b(S)}. \tag{6}$$

**TABLE 1** Representative values of $\Delta G_b$ for the formation of a protein–drug complex as a function of $K_d$

| $K_d$ | 1 pM | 1 nM | 10 nM | 100 nM | 1 μM |
|---|---|---|---|---|---|
| $\Delta G_b$ (kcal/mol) | −16.3 | −12.3 | −10.9 | −9.5 | −8.2 |

Here, $S$ and $R$ refer to the drug sensitive and drug resistant variants, respectively. $\Delta\Delta G_{b(S\rightarrow R)}$ can be estimated from measurements of the dissociation constants for the binding reaction of the sensitive and resistant strains as

$$\Delta\Delta G_{b(S\rightarrow R)} = RT\ \ln\frac{K_{d(R)}}{K_{d(S)}}. \tag{7}$$

When not available, $K_d$ values might be replaced by $IC_{50}$ or $EC_{50}$. Note, however, that such modifications always imply the assumption that the changes to the protein structure and dynamics are minimal and do not affect its activity, and that the experiments were run under the same conditions (temperature, pressure, reagent concentrations, etc.).

## 3.2 | Commonly used experimental methods for binding affinities

Experimental values are often used by computational chemists as a reference, and the ability to reproduce measured values within a small margin might be seen as evidence that the computational approach is accurate. However, to correctly estimate the credibility of computed values, it is necessary to understand the nature of the measurements. A short summary of the common approaches to estimate binding affinities is provided below. For additional methods, see References 14,15.

### 3.2.1 | Radioligand binding assays

Radioligand binding assays are simple and useful. A radioactive ligand ($^3H$ and $^{125}I$ are used most often) is synthesized and incubated with the (usually membrane-bound) receptor. Binding is measured by scintillation after washing excess ligand. Radioligand assays are accurate, but the associated costs and toxicity limit their usability. In general, titration of the ligand makes it straightforward to calculate $EC_{50}$ from the maximal effect (signal). When keeping the concentration of either the protein or the drug much higher than the other it is possible to estimate $K_d$.[16]

### 3.2.2 | Fluorescence spectroscopy

Assays that employ fluorescence spectroscopy rely on a difference in fluorescence intensity upon binding of a ligand. Trp residues in proteins are highly useful in this sense since they possess an intrinsic fluorescence that depends on the environment of the residue. Alternatively, an extrinsic probe can be used. Different spectroscopical methods can be utilized for the measurement. Fluorescence assays are popular and useful but limited by internal fluorescence, fluorescence quenching (upon binding of exogenous probes), availability of suitable probes, potential interference between the probe and the binding reaction, and other factors.

### 3.2.3 | Calorimetry

Calorimetric assays utilize heat transfer generated upon the reaction and thus the protein and ligand can be used directly, that is, there is no need for additional probes. Isothermal titration calorimetry measurements yield not only $\Delta G_b$ but also its components (enthalpy and entropy changes). Such data are very useful, not the least to a theoretician!

Differential scanning calorimetry (DSC) can also be used to estimate protein–ligand binding.[17] In this case, the protein is heated and the phase change temperature $T_m$, heat capacity as a function of the temperature, and enthalpy of unfolding are estimated. The same measurement is thereafter repeated in the presence of a ligand, enabling the estimation of $K_d$.

### 3.2.4 | Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) combines fluorescence spectroscopy and $T_m$ changes. In such assays, changes in the signal are recorded as a function of temperature to estimate $T_m$ in the absence and presence of a ligand, from

which $K_d$ is estimated subject to considerations about the binding reaction. DSF can be used for high-throughput screening, but the use of complementary methods might be needed for higher accuracy.[18]

#### 3.2.5 | Surface plasmon resonance

In surface plasmon resonance (SPR), one of the reactants is immobilized on a surface while the other flows through a sensor. The signal is proportional to the mass that is gained through the reaction, yielding rate constants $k_{-1}$ and $k_1$, from which $K_d$ is calculated.

#### 3.2.6 | Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) techniques are based on measurements of matter subject to magnetic radiation. The radiosignal that is detected gives information on the chemical environment. Generally speaking, signals that are recorded for a protein change as a ligand is bound (and vice versa), allowing the estimation of $K_d$ by titration. Depending on the assay, additional information, such as structural changes, might also be extracted.

#### 3.2.7 | Complications that arise when estimating $\Delta\Delta G_{b(S\rightarrow R)}$ rather than $K_d$

The procedure to estimate $\Delta\Delta G_{b(S\rightarrow R)}$ is in principle straightforward, by using Equation (6). However, care should be taken when relying on indirect measurements. When thermal denaturation is used (DSC, DSF), there can be differences with respect to the aggregation of the unfolded states of the sensitive and resistant variants. Similarly, protein aggregation might affect the binding recorded by SPR measurements. Additionally, mutations might affect the binding of external probes used in fluorimetric assays. Overall, while it is sometimes easier to calculate $\Delta\Delta G_{b(S\rightarrow R)}$ rather than $\Delta G_b$, the required measurements are often more complicated and the computational chemist should be aware of any potential limitations.

### 3.3 | Experimental methods to study enzyme inhibition

Enzymes are popular targets in drug discovery[19] not the least as targets for inhibition in cancer,[20] antiviral therapy,[21] and antibacterial therapy,[22,23] where resistance mutations are clinically important. For a review on enzymes in drug discovery, see Reference 24.

#### 3.3.1 | Enzymatic assays

Enzyme inhibition can be studied directly, by following on the product of the reaction and drawing a response curve from which $IC_{50}$ is extracted. Spectroscopy, scintillation (using radioactive atoms that are traced), and many other methods such as SPR and NMR (see the previous subsection) can be used.

#### 3.3.2 | Inhibition of cell growth

When cells depend on a certain enzyme for their growth, it is often easier to estimate the effect of an enzyme inhibitor on cellular growth. In a typical assay, starting from a given density of cells, the cells are left with increasing concentrations of the inhibitor for a certain period of time (e.g., 48 h) and a growth curve is drawn for each concentration, from which $IC_{50}$ is estimated. Repeated measurements can be made with cells that carry mutant proteins to estimate their sensitivity. Care should be taken, however, not to over-interpret the results when compared to calculated binding energies. The effect of the drug on cells likely depends on other factors than enzyme inhibition, such as the expression of proteins other than the drug target, permeability, efflux, and more. Still, it is possible in some cases to generate cells

that are rather homogeneous except for a given mutation. For example, Ba/F3 cells are particularly useful in kinase drug discovery, and as they can be made to depend on a specific kinase for their growth; isolation of mutants can thereafter be performed.[25]

$IC_{50}$ values measured for cell-growth experiments may represent the clinical situation better than those measured at the enzyme level. One must however note that these depend on the cell-line in question; both the actual values and the difference between sensitive and mutant strains may differ as a function of the cell line. On the other hand, issues such as drug uptake by the cells, efflux, and drug metabolism are not expected to vary much as long as it is only the drug target that is modified (as is the case with Ba/F3 cells for example).

### 3.3.3 | Antibacterial compounds: minimum inhibitory concentrations

In antibacterial drug development, the *minimum inhibitory concentration* (MIC) is often used to assess the efficacy of compounds against bacteria.[26] MIC is defined as the lowest concentration needed to inhibit the visible growth of a microorganism after an incubation for a given period of time (18–20 h is typical[26]). $MIC_{50}$ and $MIC_{90}$ are related terms and represent the concentrations of the compound which are needed in order to inhibit the growth of 50% and 90% of multiple isolates, respectively.[27] Fold changes in MIC or related values indicate resistance, and can be mapped to specific mutations.[28] However, care should be taken to examine whether additional genetic changes can be ruled out.

## 4 | ESTIMATION OF THE GIBBS BINDING ENERGY CHANGE UPON MUTATION, $\Delta\Delta G_{b(S\rightarrow R)}$, BY THEORETICAL METHODS

The estimation of protein–drug affinities by computational methods remains a challenge for computational chemists.[29,30] It is desired to achieve "chemical accuracy," that is, to be within 1.0 kcal/mol of the experiment which is about 1.5 kT (kT is the thermal energy, and is equal to 0.6 kcal/mol at room temperature). When calculating binding Gibbs energy changes upon mutations, $\Delta\Delta G_{b(S\rightarrow R)}$, it is important to be able to yield values that are meaningful in the clinic, where a fivefold increase in $K_d$ (or sometimes less[31] can lead to drug resistance. This makes the requirement for accuracy even stronger. On the other hand, $\Delta\Delta G_{b(S\rightarrow R)}$ values are sometimes easier to calculate, and usually yield more accurate values than absolute Gibbs binding energies.

Some theoretical methods to calculate $\Delta\Delta G_{b(S\rightarrow R)}$ values are presented in this section. In general, the methods require considering the two variants (sensitive and resistant) separately. Often, $\Delta\Delta G_{b(S\rightarrow R)}$ is estimated from the separate binding energies (Equation (6)). The individual $\Delta G_b$ values need not necessarily be so highly accurate as systematic errors may cancel when the two values are compared.

### 4.1 | Molecular mechanics based approaches

Molecular mechanics (MM) based approaches rely on extensive sampling of the system in question. Energies are estimated using a potential energy function and a set of parameters collectively termed *force field* (FF). There are many free, academic, and proprietary software that can be used for such calculations. Normally, parameters are already available for proteins, lipids, nucleic acids, salt ions and water but not for drugs. Modeling of the drugs can be facilitated by use of specialized tools or webservers developed to this aim (such as ATB,[32] CHARMM-GUI,[33,34] easyAmber,[35] LipParGen,[36] and SwissParam,[37] see Table 2). As such servers do not always correctly recognize the molecular structure, it is advisable to run a short simulation of the drug in solution to make sure that the energy is conserved, that aromatic and hetero-aromatic rings are planar (or do not deviate much from planarity) and that the structure is not otherwise distorted.

Neutralizing ions (in case the solutes, i.e., protein and ligands, have a total charge different than zero together) and salt ions are normally added to the simulation box. The choice of ions[38] and their concentrations should be made in line with the experiments. Of note, there is a difference in affinity to proteins between $Na^+$ and $K^+$ ions,[39] which can also affect the organization of water.[40] Ion FF parameters[41] are also of importance, with multiply-charged ions posing a particular challenge.[42,43]

**TABLE 2** Tools that enable parametrization of ligands such as drugs for molecular mechanics calculations

| Tool | Webserver (Y/N) | Supported software | Supported ligand FF | Link |
|---|---|---|---|---|
| ATB | Y | Amber<br>GROMOS<br>GROMACS<br>LAMMPS | GROMOS 54A7 | https://atb.uq.edu.au |
| CHARMM-GUI | Y | Amber<br>CHARMM<br>Desmond<br>GENESIS<br>GROMACS<br>LAMMPS<br>NAMD<br>Open-MM | CGEN-FF | http://www.charmm-gui.org |
| easyAmber | N | Amber | GAFF | https://biokinet.belozersky.msu.ru/easyAmber |
| LigParGen | Y | BOSS<br>CHARMM<br>Desmond<br>GROMACS<br>LAMMPS<br>MCPRO<br>NAMD<br>Open-MM<br>TINKER | OPLS-AA | http://zarbi.chem.yale.edu/ligpargen |
| SwissParam | Y | CHARMM<br>GROMACS | MMFF | https://www.swissparam.ch |

### 4.1.1 | Estimation of dissociation constants directly from a simulation

A straightforward way to calculate $K_d$ is to run simulations that would sample enough binding and unbinding events[41,44] (Figure 2). If sampling of many such events is possible during a long molecular dynamics (MD) or Monte Carlo simulation, it is possible to approximate $K_d$ as

$$K_d = C\frac{1-\alpha}{\alpha}, \tag{8}$$

$C$ is the concentration of solutes (protein and drug) in the simulation box and $\alpha$ is the fraction of simulation time that the complex is present. The concentration of the solutes (proteins and drug) is usually considered with respect to the number of water molecules since the density of water is known. To consider the effect of mutations, $K_d$ is estimated for the sensitive strain first and thereafter for any desired mutants, allowing the estimation of fold resistance or of $\Delta\Delta G_{b(S\rightarrow R)}$ from Equations (5) and (6).

Although such calculation is simple in principle, sampling enough binding and unbinding events is often achievable only for fast reactions with relatively high $K_d$ (higher than $10^{-3}$ mol/dm$^3$), when atomistic simulations are used. It is also important to run multiple simulations (an ensemble) rather than a single one, even if very long, to ensure reproducible conclusions.[45] This is true in all cases when MD simulations are used, but is especially important here owing to the requirement for sampling of multiple events (which relies on a long trajectory).

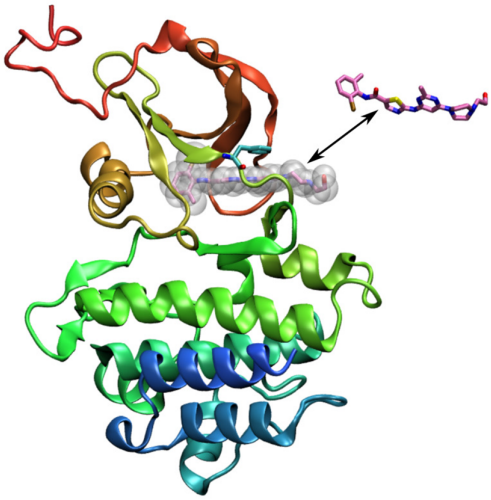

**FIGURE 2** Direct calculation of $K_d$ from MD simulations. Multiple binding and unbinding events must be evident. To study mutations, the simulations should be repeated for the wild type and mutant by modifying the residue that is mutated (cyan)

An alternative to simple calculations of $K_d$ from MD simulation is to use enhanced sampling methods, or coarse-grained (CG) simulations where better sampling is achieved but the accuracy may not be as high. CG simulations save computational time not only due to the use of fewer particles but also since the required time-step can be considerably longer (sometimes by more than an order of magnitude[46]), as the CG particles are heavier than the corresponding atoms and the frequency of vibrations is lower. Thorough parametrization of CG FF parameters makes such calculations increasingly more accurate.[47]

### 4.1.2 | Free energy perturbation and thermodynamic integration

In free energy perturbation (FEP) and thermodynamic integration (TI) the system is allowed to transform between two states $A$ and $B$ which may differ in their atomic compositions or bonds (e.g., when considering mutations, Figure 3). The free energy profile is recorded along the transition and its integration yields the free energy that is required for the transformation. Such methods are thus suitable in particular for calculation of $\Delta\Delta G_{b(S\to R)}$ values in a straightforward manner. For such calculations, the drug is kept bound while the system is transformed from the sensitive variant $S$ to the resistant variant $R$ or vice versa. A similar calculation is carried out without the drug, to account for the free energy of change in the drug-free protein:

$$\Delta\Delta G_{b(S\to R)} = \Delta G^{complex}_{S\to R} - \Delta G^{protein}_{S\to R}. \tag{9}$$

$\Delta G^{complex}_{S\to R}$ is the Gibbs energy associated with mutating the protein in the complex (protein–drug) and $\Delta G^{protein}_{S\to R}$ is the Gibbs energy associated with mutating the protein in a simulation without the drug present.

Since we are interested in calculating the change between systems $S$ and $R$ that never sample the same energy basin (the mutations do not just happen spontaneously), there is a need to sample intermediate configurations in which the system is transformed in a non-natural (referred to as "alchemical") manner between the two states. This is done by using a coupling parameter, $\lambda$, such that $\lambda$ varies between 0 (representing the initial system, e.g., $S$) and 1 (representing the final system, e.g., $R$). By varying $\lambda$ in small steps, it is possible to calculate the free energy change as[48]:

$$\Delta F = -\frac{1}{\beta}\sum_{i=1}^{N}\left\langle e^{-\beta[H_{i+1}(\mathbf{p},\mathbf{q})]-[H_i(\mathbf{p},\mathbf{q})]}\right\rangle_i. \tag{10}$$

Here, $\beta = (kT)^{-1}$ where $k$ is the Boltzmann constant and $T$ is the temperature, $\langle\cdots\rangle$ represents an ensemble average, $H(\mathbf{p}, \mathbf{q};\lambda)$ is the Hamiltonian of the system which depends on the momenta $\mathbf{p}$ and coordinates $\mathbf{q}$ of the system and on the

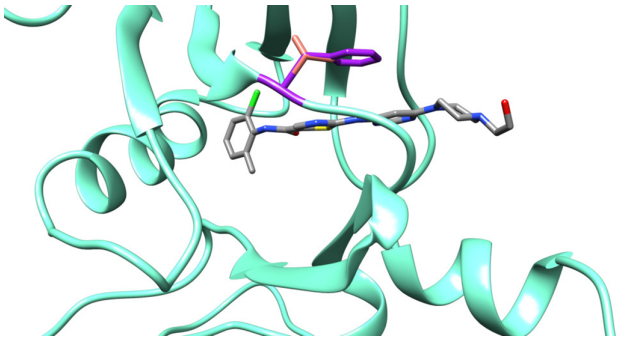

**FIGURE 3** Direct calculation of the binding energy change from simulations. With methods such as FEP and TI, it is possible to estimate the free energy change when gradually mutating a residue from its initial state to a final state. In the example shown, the structure of the Abl1 protein is presented where a single mutation, $Phe^{317} \rightarrow Val^{317}$ leads to drug resistance. In the simulations, the side chain of $Phe^{317}$ (violet) will be gradually mutated into $Val^{317}$ (pink). The simulations should be run in the presence and absence of the drug (Equation (9)). The drug (dasatinib) is shown here in light gray

simulated system which is given by the value of $\lambda$. $N$ is the number of steps in which $\lambda$ is modified, such that $H_1(\mathbf{p}, \mathbf{q}) = H(\mathbf{p}, \mathbf{q};0)$ and $H_N(\mathbf{p}, \mathbf{q}) = H(\mathbf{p}, \mathbf{q};1)$ with all other values of $i$ representing values of $\lambda$ such that $0 < \lambda_i < \lambda_{i+1}$. This method to calculate changes in the free energy (written here in general as $\Delta F$, but usually used with the NpT ensemble to yield the change in Gibbs energy) is called FEP.

Alternatively, it is possible to consider the change in $\lambda$ and calculate the free energy along the pathway as[49]:

$$\Delta F = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda. \tag{11}$$

This method is known as TI. There are different ways to calculate $\Delta\Delta G_{b(S\rightarrow R)}$ or indeed $\Delta F$ from TI,[48] including discrete TI (DTI), in which, similar to FEP, multiple equilibrium simulations are run, each for a discrete value of $\lambda$. The free energy is then estimated by integrating over $\left\langle \frac{dH}{d\lambda} \right\rangle_{\lambda_i}$. As clear from Equation (11), there is always a need to run simulations with multiple values of $\lambda$.

In practice, the free energy is nowadays estimated from intermediates generated from alchemical simulations using methods such as Bennett acceptance ratio (BAR),[50] and its extension multistate Bennett acceptance ratio (MBAR).[51] BAR can be calculated directly in MD software packages such as CHARMM[52] and GROMACS.[53] MBAR is implemented in the PyMBAR[51] and FastMBAR[54] packages. Practical aspects of running such simulations are discussed in References 55,56.

TI and FEP are based on sound theory, and can in principle be highly accurate (subject to the accuracy of the FF, of course) for $\Delta\Delta G_{b(S\rightarrow R)}$ calculations. However, the computational requirements are often higher than non-alchemical methods such as MM/PBSA, MM/GBSA, and LIE (see Sections 4.1.3 and 4.1.4). As clear from Equations (10) and (11), accuracy will depend on the number of discrete steps used while varying $\lambda$. The length of the trajectory is also of importance, and both aspects will govern the speed by which the calculation is completed. Another aspect is the length of equilibrium simulations of the non-perturbed end states, where $\lambda = 0$ or 1.[57] Another approach, thermodynamic integration with enhanced sampling-protein mutations[58] uses an enhanced sampling protocol rather than unbiased MD simulations in TI.

There is an issue of concern when mutations do not conserve the overall charge of the protein. MD simulations normally apply periodic boundary conditions (PBCs, Figure 4a) to avoid artifacts arising in the boundary of the simulated system. In PBC, the simulation box is replicated in three dimensions, which mimics an infinite crystal. Forces between the atoms are however truncated at a certain distance that must be smaller than half the dimension of the simulation box. Non-bonded interactions (of which the Coulomb contributions are the most dominant) are calculated explicitly within the cutoff and approximated thereafter. When a charge is introduced in the system upon FEP or TI, simple calculation of $\Delta G_{(S\rightarrow R)}$ is no longer accurate (see Reference 59 for the contributing factors to this). The same is true when a charge is removed. Thus, when $\Delta G_{(S\rightarrow R)}$ involves a mutation from a charged amino acid to a neutral one (or vice versa), a simple calculation using Equations (10) and (11) is insufficient and corrections should be introduced. One

correction that has been suggested for such cases, involves Poisson–Boltzmann (PB) calculations on top of the simulations.[59] Another alternative is to change another charge simultaneously with the mutation, for example, by making a small ion neutral.[60]

When calculating $\Delta\Delta G_{b(S\to R)}$ (Equation (9)) from FEP or TI, the effect of charge modification might cancel out. Indeed, in a large-scale study, it was found that finite size effects were minimal for such calculations with several different systems and FFs.[57] Note that, in the aforementioned study the overall charge of the system was not kept zero upon mutation which is rather unusual in MD simulations, but can be dealt with to some extent.[61]

Setting up calculations of $\Delta\Delta G_{b(S\to R)}$ is tedious, but has been simplified by the use of tools such as pmx[62] and QresFEP.[63] The first tool is used in conjugation with the GROMACS simulation software.[53] The second is used with the Q package,[64] where calculations employ stochastic boundary conditions (SBCs, Figure 4b) rather than PBC. Charge-changing mutations can lead to artifacts even with SBC. In QresFEP, these can be dealt with by simultaneously modifying the charge of a small ion such as $Cl^-$.

### 4.1.3 | MM/PBSA and MM/GBSA

MM/PBSA (Poisson–Boltzmann solvated area) and the closely related MM/GBSA (generalized Born solvated area) are two popular methods that rely directly on equilibrium MD simulations to calculate ligand binding energies. The methods take advantage of the fact that the intermolecular and intramolecular interactions are anyway calculated in an MD simulation of a protein–ligand complex (or of other complexes, e.g., protein–peptide,[66] peptide–RNA,[67] etc.). Considering a protein and a ligand, the binding energy can be estimated as:

$$\Delta G_b = \langle G_{PL}\rangle - \langle G_P\rangle - \langle G_L\rangle. \tag{12}$$

Here, *PL*, *P*, and *L* refer to the complex, protein, and ligand, respectively. In most cases, a single simulation is run for each variant, from which $\Delta\Delta G_{b(S\to R)}$ values are estimated using Equation (9). To further reduce the cost of the calculations, it is possible to use single minimized structures (discussed in Reference 68), but MD simulations are run most often, either for better sampling or because further analysis than just calculating $\Delta G_b$ is of interest.

Estimation of the Gibbs energies from MD requires some approximations, as these are not directly available from the simulations. An approximation of the Gibbs energy for any of the systems can be achieved by energy decomposition:

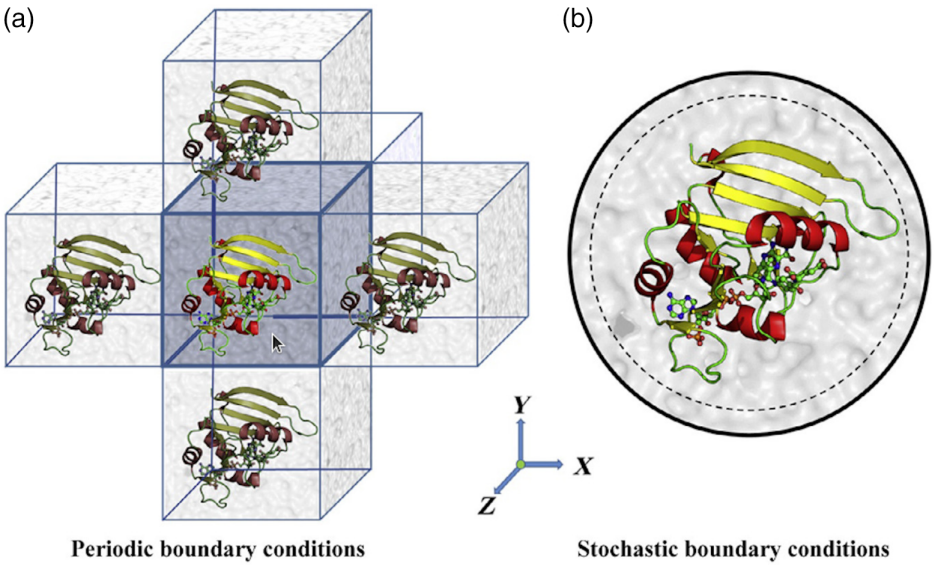


**FIGURE 4** Boundary conditions commonly used in molecular dynamics simulations.[65] (a) Periodic boundary conditions (PBCs). Replicas are made in all directions to mimic an infinite system. (b) Stochastic boundary conditions (SBCs). Atoms are restrained within a certain distance from the boundary (dashed line). Reprinted from Reference 65, with permission from Elsevier

$$G = E_{bond} + E_{el} + E_{vdW} + G_{pol}^{solv} + G_{np}^{solv} - TS. \quad (13)$$

$E_{bond}$, $E_{el}$, and $E_{vdW}$ are the potential energies associated with bonds, electrostatic and van der Waals (vdW) interactions in each molecule. These are available directly from MD. In practice, $E_{bond}$ is not necessary (as long as only a single simulation is used for each variant) as the protein and ligand are not normally covalently bonded. If they are, simple MM/PBSA or MM/GBSA calculations are not adequate, as it is no longer possible to separate the terms in Equation (13).

$G_{pol}^{solv}$ and $G_{np}^{solv}$ together account for the solvation energy. The first term is the (normally much larger, in absolute values) polar contribution to the solvation energy. In MM/PBSA, it is calculated by solving the Poisson–Boltzmann equation (PBE) numerically. In MM/GBSA, it is approximate by using generalized Born (GB) solvation, a computationally cheaper approximation; several GB approximations exist.[69] In both cases, many parameters need to be considered including the approach (MM/PBSA or MM/GBSA), choice of atomic partial charges and radii, the dielectric coefficient of the solutes and how the actual calculation takes place (the numerical approach to solve the PBE or which model to use for MM/GBSA).

$G_{np}^{solv}$ is almost always approximated as

$$G_{np}^{solv} = \gamma A + b, \quad (14)$$

where $\gamma$ is a surface tension term, $A$ is the solvent accessible surface area of the molecule, and $b$ is a correction parameter. The theory on which Equation (14) is based is built upon early developments dealing with the solubility of gases,[70,71] and was later found to be applicable for the solubility of hydrophobic compounds.[72] A calculation of hydration energies for functional groups representing amino acids, after parametrizing the atomic radii and charges, have shown an overall good agreement with the experimental reference ($R = 0.93$).[73] Errors with respect to the experiment were smaller than 0.5 kcal/mol. Such errors might however be too high when a whole protein is considered (as when calculating protein–drug binding energies), since a protein includes many amino acid residues and the errors gather. It might be argued that, since in $\Delta\Delta G_{b(S\to R)}$ calculations only a single side-chain is modified, the error might not be so high. However, as discussed above (Section 3.1.4), a change as small as 1 kcal/mol might lead to drug resistance. Moreover, the errors calculated for the hydration of some organic compounds, namely 2-methylpyrazine, 1,2-ethanediamine, and 1,2-ethanediol were 2.3, 3.4, and 4.3 kcal/mol, respectively,[73] and the approximation of the $G_{np}^{solv}$ for drugs, which are small molecules but often with several organic groups is thus questionable. Thus, if the difference in the $G_{np}^{solv}$ term is large between the resistant and sensitive variants, other methods should be used to calculate $\Delta\Delta G_{b(S\to R)}$.

The last term that needs to be accounted for is the change in conformational entropy upon ligand binding. The formation of a complex often limits the number of microstates that are available for sampling by the drug and by the protein, especially when considering the side chains that interact directly with the drug. Accounting for the conformational entropy in MM/PBSA and MM/GBSA is often performed by extracting a given number of snapshot structures from the simulations and then performing normal mode analysis (NMA) subject to some preparation,[68,69] for example, considering only a subset of protein atoms closer to the drug. Another approach is to run separate simulations for the complex, protein and ligand, from which the internal energy and entropy can be calculated directly (the entropy can be estimated from either NMA or covariance analysis). Performing NMA involves a robust energy minimization (EM), and then obtaining and diagonalizing the Hessian matrix, that is, the matrix of second derivatives of the energy. There are several difficulties with this. First, it is required that the maximum forces are very small after EM, otherwise the modes are not accurate. This is not always easily achievable and might require several cycles of minimization that make the process slow. Second, obtaining and diagonalizing the Hessian matrix is a lengthy process. Hence, the conformational entropic term ($TS$) is often ignored, the justification for this being that it is smaller than other terms. Considering calculations of $\Delta\Delta G_{b(S\to R)}$, it may be argued that the effect of a single (or few) mutations on the entropy of the complex with respect to the protein is indeed negligible. However, this is certainly not true in many cases! We have shown that there might be modifications to the number of distinct configurations that a protein can attain as a function of drug binding for specific mutations[74] (which implies changes in the entropy) and that drug resistance can depend primarily on changes of the conformational entropy (resistance to dasatinib owing to the F317L mutation in Abl1[31]). Thus, ignoring the entropic term when calculating $\Delta\Delta G_{b(S\to R)}$ cannot always be justified. Recognizing the

computational cost involved with calculating the entropic term, a method, termed "interaction entropy" was developed[75] that approximates $TS$ from fluctuations in the protein–ligand interaction energy $\Delta E_{pl}^{int}$:

$$T\Delta S = -kT \ln\left\langle e^{\beta \Delta E_{pl}^{int}} \right\rangle. \tag{15}$$

This approximation does not take into account all contributions to the entropy but it has been shown to improve the correlation to experimental values and reduce the mean absolute error when tested on a set of protein–ligand complexes.[76]

Overall, the accuracy of these methods depends on the protocol and the system,[68] there is no "one size fits all" solution. Relying on default parameters without prior knowledge is thus not advisable, despite the availability of tools that are designed to make MM/(G)PBSA calculations fast to set-up and simple to run. Free energy calculations (FEP and TI) include fewer parameters, and are more robust from a theoretical standpoint—but they are slower.

### 4.1.4 | Linear interaction energy

Linear interaction energy (LIE) is an approach developed to calculate protein–ligand binding affinities from equilibrium MD simulations. Its advantage over alchemical methods is that intermediate (unphysical) states that combine the reactant and product (as in FEP and TI) need not be considered. The underlying idea was to consider the free energy change associated with the transfer of a ligand from a solvent to the protein interior. Considering the Gibbs energy of binding:

$$\Delta G_b = \Delta G_{g\to p}^{(lig)} - \Delta G_{solv}^{(lig)}. \tag{16}$$

$\Delta G_{g\to p}^{(lig)}$ is the Gibbs energy associated with the transfer of the ligand from gas-phase to the protein binding pocket, and $\Delta G_{solv}^{(lig)}$ is the Gibbs energy of solvation for the ligand, that is, the Gibbs energy associated with the transfer of the ligand from gas-phase to the solvent.

Based on observations regarding the solvation free energies of nonpolar compounds (see the discussion regarding MM/PBSA and MM/GBSA, Section 4.1.3, and examination of the size of vdW interactions involving the solutes in MD simulations), a formula was suggested to calculate $\Delta G_b$ from MD simulations[77] as:

$$\Delta G_b = \alpha\left(\left\langle E_{vdW}^{l-p}\right\rangle - \left\langle E_{vdW}^{l-s}\right\rangle\right) + \beta\left(\left\langle E_{el}^{l-p}\right\rangle - \left\langle E_{el}^{l-s}\right\rangle\right) + \gamma. \tag{17}$$

In this equation,$\alpha$, $\beta$, and $\gamma$ are system-dependent parameters. $\left\langle E_{vdW}^{l-\square}\right\rangle$ and $\left\langle E_{el}^{l-\square}\right\rangle$ are the vdW and Coulomb interactions energy between the ligand ($l$) and the protein or solvent ($p$ and $s$, respectively). These are calculated by running two simulations—one for the protein–ligand complex and one for the ligand in water (which should in most cases be run anyhow to ensure that the ligand parameters are correct). From the linear response approximation, $\beta$ should be equal to 0.5,[78] but it has been shown that treating it as an independent parameter may lead to better agreement with experimental measurements.

To calculate $\Delta\Delta G_{b(S\to R)}$, simulations of all complexes need to be run, and in addition a single simulation of the ligand in water. Thereafter, binding energies are calculated from Equation (17), and then relative free energies with Equation (6). Parametrization of the multipliers in Equation (6) based on available data point(s) should be performed and tested to obtain reliable predictions.

LIE has been used extensively with the Q modeling package[64] and SBC (Figure 4b). When using SBC, titrable protein residues that are close to the boundary (and residues outside the solvation sphere, if any) should be neutralized to avoid artifacts due to reduction or lack of the dieletric screening.[79] Likewise, the same boundary conditions should be used for the protein and ligand.

A faster variant of LIE, linear interaction energy-continuum electrostatics (LIECE) that applies robust energy-minimization followed by a calculation of $G_{pol}^{solv}$ by solving the PBE has been developed.[80] The electrostatic contribution is estimated as a sum of the Coulomb energy in vacuo and solvation energy. Being fast, LIECE was used in

computational screening campaign as one of four scoring functions.[81] It can be used for rapid screening of $\Delta\Delta G_{b(S\rightarrow R)}$ values, but the results should thereafter be verified by a more accurate approach.

LIE may be more efficient than MM/GBSA when using a truncated spherical system[82] as implemented in the computer program Q.[64] It has been reported to be less accurate than MM/GBSA for some systems,[83] and conversely for others (in comparison to MM/PBSA).[84] Overall, like MM/(G)PBSA, although it is fairly easy to use out-of-the-box, care should be taken to ensure that the parameters used are representative, and in general TI and FEP are more robust from a theoretical standpoint.

## 4.2 | Methods that involve quantum mechanics

### 4.2.1 | Direct QM calculations using truncated binding site cluster models

QM calculations can capture interactions that are difficult to account for in MM, mainly owing to the inherent limitations of a FF-based treatment (see Reference 85 for an extensive review in the context of protein–ligand interactions). The downside is the much higher computational cost, and (especially for post-Hartree–Fock [HF] methods, such as MP2 and coupled-cluster) the tremendous increase in resources as function of the system size. Whereas density functional theory (DFT) calculations generally scale as $N^3$ where $N$ is the number of basis functions that are used, MP2 scales as $N^5$ and CCSD(T) as $N^7$. Post-HF methods have a great advantage in that the calculations can be systematically improved by using larger basis sets and higher levels of theory, although one should be aware of the limitations of each method; MP3 for example is in general not better than MP2.[86] DFT is often considered as the working horse of computational chemistry. With an appropriate functional, the results can be more accurate than those obtained with methods that are much more demanding. Choosing the right functional however is not an easy task, and often requires trial calculations using cleverly selected small systems (where the DFT results can be compared to those obtained from, for example, CCSD(T) theory[87]).

Another set of methods that can be used for QM calculations are the so-called "semi-empirical methods." In such methods, approximations are used to calculate energies and gradients orders of magnitude faster than ab initio or conventional DFT calculations. The development of semi-empirical methods relies on parametrization based on experimental values and/or other QM calculations (e.g., DFT). With semi-empirical methods, whole proteins and even multiple peptide complexes can be considered.[88] However, at present they cannot be considered accurate enough for calculations of $\Delta\Delta G_{b(S\rightarrow R)}$ if the aim is to study how individual mutations affect binding, even if they can be used for biochemical studies.[89]

When DFT is used, calculations of protein–drug binding energies are often performed with a truncated model that only represents a small part of the system, for example, the drug and few side chains that interact with it directly. This is known as "the cluster approach"[90] or "QM cluster"[85] calculations. QM cluster calculations are used for the following reasons:

1. Computational cost.
2. Accuracy trade-off—while it might be possible to consider a full protein with a bound drug, this would require the use of a smaller basis set and a simpler (local) DFT functional; the associated errors might then be too large.
3. The inherent DFT self-interaction error becomes larger as a function of system size.[91]

Choosing which residues to include in the calculations is important for computational cost and accuracy. Too few, and the model will not capture enough of the interaction; too many, and the computational cost will be prohibitive. Moreover, the larger the model, the more likely it is that the apo (non-drug bound) structure will be different than the drug-bound one, which in turn would result in $\Delta G_b$ values that are too favorable[31] (e.g.,, see Figure 5).

QM calculations always give the internal energy of the studied system as an output, so that, using four calculations the internal binding energy change upon mutation (not Gibbs energy) can be estimated (similar to Equation (6), but with $\Delta E_b$ not $\Delta G_b$). If all the molecules are studied without constraints, it is possible to correct for the zero point energy (ZPE) due to vibrations. However, optimization of a cluster model of the apo-protein without constraints may lead to structures that are not a true representation of the system, because fragments taken from individual side chains may interact in a way that does not reflect the full protein. Hence, constraints on some of the atoms are often used to keep the amino acid residues in place. For example, whereas the drug atoms and protein atoms in the first binding shells are

free to move, constraints are put on the second shell protein atoms keeping them in place. ZPE corrections cannot be used in such case, but might cancel out when binding energy changes upon mutation are of interest.

Another important issue with cluster models, and QM calculations in general in the context of binding energies in solution, is the representation of the solvent. Accounting for the solvent can improve the accuracy, and considering the cost of desolvation of the components formally yields Gibbs energy values. To this end, implicit treatment of the solvent by models such as PCM,[92] COSMO,[93] and SMD[94] is highly useful since it is challenging to model the solvent molecules explicitly.

The choice of the solvent is also of importance. The environment that surrounds the binding site in the protein is not as polar as water, making the case for considering solvents of lower dielectric constant such as tetrahydrofuran ($\epsilon_r = 7.58$)[95] as more representative. On the other hand, the reaction takes place in water, and using the solvated drug in a reference environment of much lower dielectric constant may lead to over-binding (since the reference is too unstable). As a result, considering water as the solvent may be a better choice, especially for polar or charged drugs. The drug is anyhow expected to leave the solvent and reach the protein's interior.

A somewhat different approach[96] is to run the calculation in gas phase and correct for enthalpy, entropy, and solvation:

$$\Delta G_b = \Delta E_b + \Delta H_{corr} - T\Delta S + \Delta\Delta G_{solv}. \tag{18}$$

$\Delta H_{corr}$ is the enthalpy correction from room temperature vibrations (quantum thermochemistry). Similarly, $\Delta S$ is calculated in the gas phase by assuming that each molecule behaves as a pure (ideal) gas. The differences in solvation energies of the complex and components (protein and drug) are calculated separately. Estimating $\Delta G_b$ this way requires full optimization and calculation of the Hessian matrix which is computationally demanding and may lead to errors in the structure (vide supra). An empirical scheme has been suggested[96] that relies on the partition coefficient $P$ of the drug between water and octanol, a property that is used much in medicinal chemistry. In similarity with the LIE approach, the affinity is then calculated from the interaction energy:

$$C = \alpha\Delta E_b + \beta \log P + \gamma. \tag{19}$$

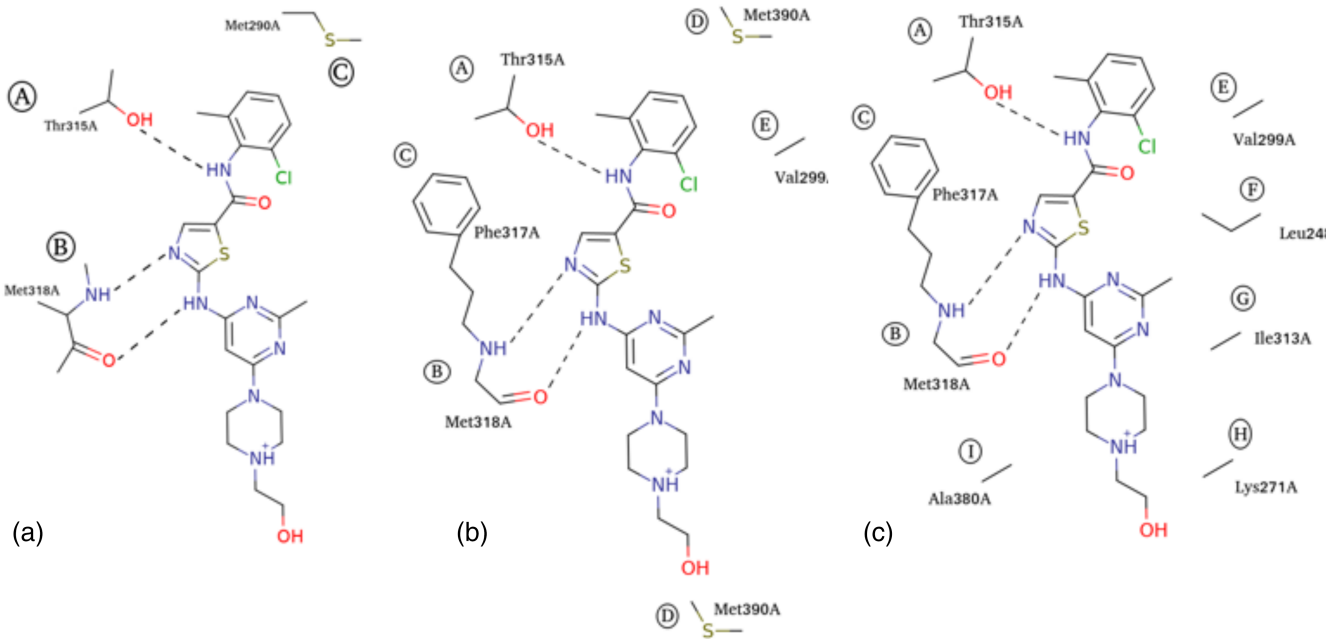


**FIGURE 5** Building a model for QM calculation of protein–drug binding energies. Three models are shown for Abl1 binding to the drug dasatinib (Figure 2), with (a) three (b) five, and (c) nine interacting residues. In this case, the binding energy was within ~1 kcal/mol of the experimental value with the first two models, whereas the third resulted in an over-stabilization of the complex, which may be explained by the apo-structure adopting another (more stable) conformation. Adapted from Reference 31 with permission from The Royal Society of Chemistry

$C$ is the affinity which can be for example $K_d$ whereas $\alpha$, $\beta$, and $\gamma$ are empirical parameters that are specific to the system in question. This approach could in principle be used to calculate $\Delta\Delta G_{b(S\to R)}$ values by considering changes in the affinity upon mutation.

### 4.2.2 | Fragmentation-based methods

As explained above, it is impractical to consider the whole protein as is for QM calculations. In fragmentation-based methods, the whole protein is considered fully quantum mechanically by dividing the system into fragments and summing on the interactions between the fragments in some way. In general, fragment based methods yield $\Delta E_b$ and solvation needs to be considered separately to account for the Gibbs energy. Two methods that have been used to assess protein–ligand interactions are divide and conquer HF (DCHF) and fragment molecular orbital (FMO). Both approaches are started with a division of the protein into smaller fragments. In DCHF,[97] the fragments include a *core* of interest and a *buffer zone* which affects the core due to its proximity to the core, but which should not be included when summing up on the interactions (Figure 6).

In FMO, the total energy of the system is obtained by considering the sum of the energies of monomers, dimers, timers, and so on, up to a given order. Truncating the interactions at dimers yields the following approximation to the system's energy:

$$E^{FMO2} = \sum E_i + \sum_{j>i} \left(E_{ij} - E_i - E_j\right). \tag{20}$$

The system can be further divided into layers that are considered at different levels of theory, for example, by calculating interactions involving the ligand and nearby protein residues at the MP2 level and all other interactions at the HF level.[98]

Different approaches can be used to consider the effect of solvation, since solvating each fragment separately is not an option. For example, empirical approaches such as GBSA or PBSA can be used, sacrificing the theoretical robustness for efficiency and practicality. Another approach, EE-GMFCC-CPCM was developed specifically for solvating proteins in fragmentation calculations[99] and was employed for estimating protein–ligand binding affinities.[100] In EE-GMFCC-CPCM, the fragment-based calculation is run in vacuo first. Thereafter, charges at the

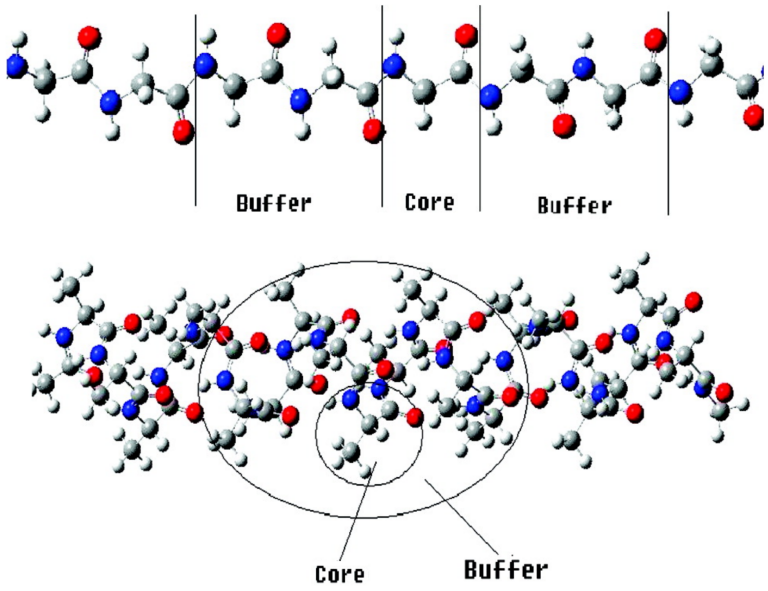


**FIGURE 6** DCHF—an example for a subsetting schemes for the interaction between polyglycine $(Gly)_n$ (upper) and polyalanine in an $\alpha$-helical structure (bottom). Reprinted (adapted) with permission from Reference 97. Copyright (2010) American Chemical Society

surface of the cavity enclosing the solute are calculated in an iterative self-consistent manner, and the energy is corrected accordingly.

The advantage of fragment based division methods is that the energy is calculated with full-QM, but at fraction of the cost of a typical calculation. Moreover, it is possible to consider contributions from individual residues which is useful to explain resistance mutations. However, to use such methods quantitatively it is important to consider not only the QM level of theory but also how the system is divided into fragments and how the effects of solvation are handled.

#### 4.2.3 | QM/MM

Estimation of $\Delta\Delta G_{b(S\rightarrow R)}$ values by MM and QM-based methods has been discussed above. MM-based methods are faster, and in many cases accurate, as modern FFs are well suited for such calculations and sampling can be extensive. QM methods can succeed in modeling interactions that are not captured well by MM, such as when metals in the protein binding site bind directly to the ligand, when other special types of bonds are involved, or when QM effects such as exchange play a significant role in the binding. Moreover, polarization is not accounted by the most commonly used FFs, whereas polarizable FFs have hitherto not been as extensively optimized as non-polarizable once; in QM, contribution from polarization is always included. It might thus be tempting to use a combined QM/MM approach for calculations of $\Delta\Delta G_{b(S\rightarrow R)}$. In principle, methods that are used in MM such as TI and FEP can also be used in QM/MM calculations. However, the bottleneck of the QM/MM calculations is the QM part, and running extensive simulations is hence too demanding. A QM-based LIECE variant (QM-LIECE) has been developed, where the electrostatic contribution to the binding energy, $\Delta E_{el}^{l-p}$, is calculated with a QM approach.[101] To calculate $\Delta E_{el}^{l-p}$ quantum-mechanically, a divide-and-conquer approach was used where the protein was broken down to fragments whose contribution was summed up. It can be envisioned that other empirical approaches (e.g., Equation (19)) might also be used. Another possible approach is to use a much faster but much less accurate QM method, for example, a semi-empirical potential.[102] Alternatively, MM potentials can be fit to reproduce QM/MM interactions, in essence generating an MM-QM/MM correction.[103]

All-in-all, QM/MM has not been used much for calculating $\Delta G_b$ and $\Delta\Delta G_{b(S\rightarrow R)}$ for protein–ligand complexes. On the one hand, free energy methods such as TI and FEP require extensive sampling, where the QM part is limiting. On the other, cluster models such as used in full-QM calculations require the estimation of solvent effects at the same time for the QM and MM parts. QM/MM-P(G)BSA can be performed, but suffers from the same limitations as MM-P(G)BSA (vide supra, Section 4.1.3).

### 4.3 | Machine learning

Machine learning (ML) is getting increasingly popular in chemistry.[104] ML methods are useful when large amount of data is available to train the ML algorithm, and require that the data are formatted in a certain way.[105] In the context of calculating $\Delta\Delta G_{b(S\rightarrow R)}$ values, it should be noted that ML approaches are liable to fail more than physics-based methods when new data are presented,[106] such as a mutation that was not considered before. When studying resistance mutations, the amount of data is more limited and at present it is difficult to see how accurate calculations of $\Delta\Delta G_{b(S\rightarrow R)}$ will be made possible based on ML alone. For calculations of protein–ligand affinities ($K_d$, $\Delta G_b$) from structural features, ML approaches might perform better than other statistics-based scoring functions, but slower, physics based methods (MM/PBSA, MM/GBSA, LIE, FEP, TI, etc.) tend to be more accurate.[107] Yet, even at present, ML can be a useful tool to understand resistance mutations. To gain a realistic model of protein–ligand interactions in solution, MD simulations can be used to gain multiple data points from a single simulation, for example, protein–ligand interactions (number of contacts, different interaction-energies), surface area, orientation, and distances.[108] These can thereafter be used to develop a ML model. On top of improving the performance of ML, extracting features from MD simulations might be useful to shed light on the factors which are affected to a higher degree by resistance mutations, providing a structural and physical understanding of the reasons behind the emergence of resistance.[109] In this sense, ML methods are of great potential since they can highlight the importance of features that are hidden in the data and are important to drug binding and the effect of mutations.

# 5 | LIMITATIONS AND PITFALLS IN COMPUTATIONAL ESTIMATION OF THE AFFINITY CHANGE UPON MUTATION

## 5.1 | Comparison to the experiment (what is measured there, actually?)

Reproducing the results of experimental measurements by computer-aided calculations is often regarded as a sign of success of the computational method. While this is indeed an important prerequisite to show that the modeling is accurate enough, the reconstruction of an experimental measurement is in itself no proof that the computational study is correct in any way. Rather, the prediction should be "right for the right reason," which requires an understanding of the experiment through which a certain measurement is obtained.

Very often, especially when considering a large volume of data points as required for ML and studies involving many targets, there is a tendency to concentrate on the end results without realizing that those are not necessarily measured in the same way. Some factors that can influence the measurements are given here; more details on conformational changes and non-trivial resistance mechanisms are discussed in the following sections (Sections 5.2 and 7.1).

### 5.1.1 | Substituting the dissociation constant

Resistance mutations affect a single protein which, through interactions with other biomolecules, affects cells (be it human or animal cells, or those of pathogens) and eventually the whole organism, leading to a clinical manifestation (the drug does not work as expected any more). Experimental measurements related to proteins ($K_d$, $K_i$) are thus closer in essence to whatever a computational chemist can calculate, but are not always possible to come by, in which case it might be assumed that the observables at the protein and cellular level are related:

$$\frac{K_d^{(R)}}{K_d^{(S)}} \approx \frac{O^{(R)}}{O^{(S)}}, \tag{21}$$

$R$ and $S$ refer to the resistant and sensitive variants, respectively and $O$ is an observable such as $EC_{50}$, $IC_{50}$, or $MIC_{50}$. Here, on top of making sure that the experimental conditions are the same for the measurements if all variants, it is important to consider the reaction. Equation (4), for example, implies that both the concentration of the substrate and $K_M$ are the same, and while the first is likely true for values reported together, the second might differ between mutants.[9] Similarly, when comparing $EC_{50}$ values, it is important to consider that any modification in the mutant protein might also affect these values irrespective of the therapy. Let us consider, for example, a mutation that leads to (1) a lower drug binding affinity and (2) a modified turnover rate for a certain enzyme. In this case, $EC_{50}$ and $K_i$ values reflect both changes, but $K_d$ will only reflect on the binding. On top of this, when considering any type of MIC measurements, it is necessary to verify what are the differences between the microorganisms in question. Bacterial strains may differ in more than one genetic trait, which might also limit the usefulness of Equation (21).

### 5.1.2 | Protein domains

Many proteins include multiple domains, which can be expressed and studied separately. Expression of human proteins in bacterial cells, which is done in order to achieve sufficient quantities of such proteins in reasonable time can be a challenge, and considering only the domain of interest (e.g., the catalytic domain of an enzyme where the drug binds) can lead to better yield (i.e., more protein). Likewise, it may be necessary to consider only some parts of the protein in order to get a crystal structure of a drug-bound protein which is then used as a starting model in calculations. Often times, it is safe to assume that resistance mutations affect the binding through interactions within the same domain of the protein as the one in which they reside. However, this is not always the case. When calculating $\Delta\Delta G_{b(S\rightarrow R)}$ values it is therefore important to consider if the modeled protein includes all the domains that were used in the experiment or not. If a structure of the protein with additional domains is available, it could be interesting to examine if the resistance mutations affect domain-domain interactions within the protein (Figure 7).

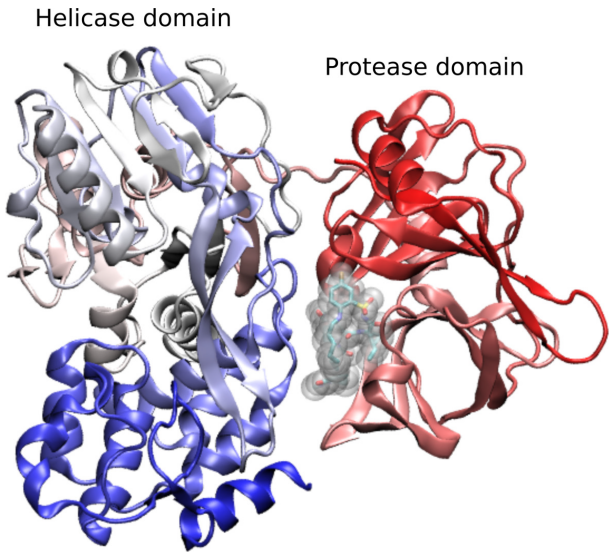


**FIGURE 7** Drugs that bind to multi-domain proteins. The viral NS3/4A protease is a drug target for treatment of hepatitis C. Inhibition is aimed at the protease domain, and many structures with inhibitors include only this domain. However, as shown in the figure, inhibitors (center, stick representation) may also interact with the helicase domain and domain-domain interactions might be important for drug binding. In such cases, it is important to consider both domains when studying resistance mutations

### 5.1.3 | Binding to other proteins

Some proteins adopt a particular conformation upon binding to other proteins, or require other proteins for their activation/inhibition. It can be envisioned that such binding affects the dynamics of the protein in question. Moreover, resistance mutations might interfere with this protein–protein binding. As in the case of a multi-domain protein, care should be taken to understand the nature of interactions between the protein that is the drug target (and where the resistance mutation occurs) and its binding partner(s). On the other hand, it might be possible to gain insights on resistance even when the binding partner is absent, if the ns-scale dynamics of the protein of interest is not perturbed.[109]

### 5.1.4 | Posttranslational modifications

Many proteins undergo posttranslational modifications (PTMs) including phosphorylation, glycosylation, nitrosylation, acetylation, covalent binding to lipid groups (e.g., myristoylation), and more. These change the physico-chemical properties of the protein. Consequently, the protein structure[110] is modified and also its dynamics.[111–115] Consequently, it is important to consider whether the experimental assay, and the relevant protein structure, involve PTMs—these may influence the structure and dynamics of the drug sensitive and drug-resistant variants, not necessarily in the exact same way.

### 5.1.5 | The dynamics of the drug-free protein

Some of the approaches discussed above model the protein–drug interaction without extensive sampling of the free protein whose dynamics and to some extent also local structure might change upon mutation. For example, some mutations modify the size of the drug-binding pocket.[9,116] Such changes are not captured by methods that use only the complex, such as MM/(G)PBSA with a single simulation for each state (Section 4.1.3), LIE (Section 4.1.4), and QM cluster methods (Section 4.2.1). When using alchemical methods (Section 4.1.2) there is a need to have long enough sampling in the simulations to include such changes. The use of unbiased simulations to model the binding (Section 4.1.1) is perhaps the best way to ensure that such effects are captured, but as discussed above it is not practical for real-life situations at present.

## 5.2 | Proteins that adopt multiple states

Many proteins that are drug targets can adopt multiple states that are structurally distinct. Protein kinases for example adopt active and inactive states, and abnormal activation of such proteins leads to disease.[115] Many proteases are normally inactive, as there is no need for them to digest cellular proteins. In such enzymes, cleavage of some part of the protein can trigger activation. Receptors become active upon conformational changes. Overall, the list of proteins that adopt multiple states is long. Regulation of the free energy landscape that governs the transition between the states is often allosteric, that is, it is based on changes that occur at a site that is at a distance from the site of action (the site of action can be a catalytic site, a binding site for an effector protein, or a binding site for a molecule that is transported, to give a few examples).[117] The affinity of a drug to a protein can vary with the state. Importantly, mutations that affect the equilibrium between states (Figure 8) may lead to drug resistance which will often be manifested in experiments

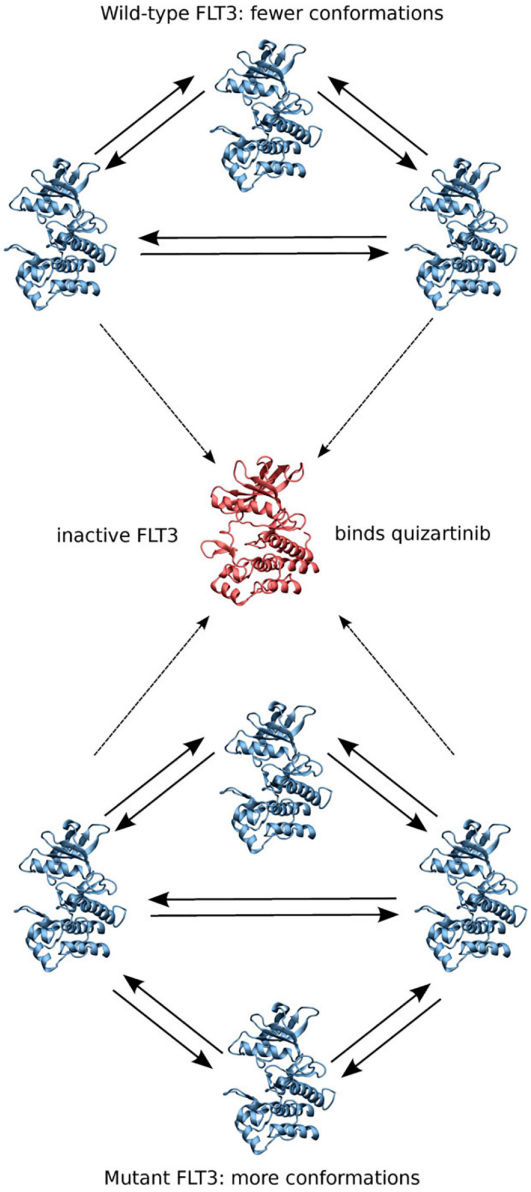


**FIGURE 8** Resistance mutations might affect non-binding conformations. FLT3, an important molecular target in acute myeloid leukemia, binds quizartinib, a highly specific inhibitor, when the protein is in its inactive state. MD simulations however show no difference in protein–drug interactions when considering only the inactive state and comparing the wt protein with two mutants, D835F and Y842H, that lead to resistance. Instead, it was suggested that the mutations affect the active state and make it less probable for the protein to adopt a conformation that is prone to bind quizartinib. Calculations of $\Delta\Delta G_{b(S\rightarrow R)}$ that are based on the inactive state will not correctly account for the full conformational dynamics of the enzyme. Reproduced from Reference 118 with permission from Wiley Periodicals, Inc

(where multiple states are sampled) but not in computational studies that only consider a single state. An example for this is the drug imatinib. Imatinib has the kinase Abl1 as its target and it binds the inactive state. There are several dozen resistance mutations against imatinib. Some of these, such as Abl1/G250E, make the enzyme more efficient thereby leading to resistance. Compound mutations (two mutations that involve the same copy of the gene, e.g., G250E/T315I) are even more potent in increasing the catalytic efficiency.[9] This has an indirect effect on drug resistance that will not be captured by calculations involving the drug in the inactive state (such calculations, if performed correctly, should yield $\Delta\Delta G_{b(S\rightarrow R)}$ values that underestimate the effect of the mutation). See Figure 8 for another example.

### 5.3 | Neutral mutations in the drug target

The use of a drug applies strong evolutionary pressure on entities such as viruses, bacteria, parasites, and tumor cells. All of these typically evolve much faster than normal human cells. Mutations gather all the time, and not all of these lead to resistance. Here, too it is important to consider that not all mutations that appear in the drug target are resistance mutations; some are neutral mutations, whereas another mechanism leads to resistance. An assay that examines the growth of cells or bacteria might not be able to differentiate between these. An accurate computational study can in fact be useful to reveal that the mutation does not have an effect, but we must remember that an error in the order of 1.4 kcal/mol translates to one order of magnitude difference in affinity in room temperature (Table 1), so that accuracy in the calculation is a necessity.

## 6 | BEYOND BINDING AFFINITIES—THE USE OF COMPUTATIONAL MOLECULAR SCIENCE TO EXPLAIN RESISTANCE MUTATIONS

An accurate calculation of binding affinities upon mutation is desired to show that the method that has been used (algorithm, choice of structure, FF or DFT functional, etc.) is adequate, and to enable prediction. However, a molecular modeling study can yield additional information that is of interest but which is not easily (or not at all) accessible to the experiment. Several such studies are summarized below. The aim is to give examples of how the effect of resistance mutations at the molecular level can be studied, rather than to present a complete overview.

### 6.1 | MD simulations explain vandetanib resistance in atomistic detail

Vandetanib is an inhibitor of the RET kinase. Fusion of RET with other genes make it overactive in lung adenocarcinoma, a type of lung cancer. Vandetanib has therefore been studied as a potential therapy for this type of cancer. A RET mutation, S904F, had been observed in a patient who became unresponsive to vandetanib therapy in a clinical trial. Efforts were thereafter been made to estimate whether RET/S904F is a resistance mutation. Once this had been confirmed, it became important to establish how the mutation actually leads to resistance.[119] A kinetic experiment identified that the mutant's catalytic efficiency was twice that of the drug-sensitive variant. The structures of the wild type and mutant enzymes were solved, but did not reveal marked differences (Figure 9a). A decreased flexibility in two key protein regions upon mutation was shown in MD simulations, by analysis of the protein's C$\alpha$ root mean square fluctuations (RMSF), when vandetanib was bound (Figure 9b). This suggests that the affinity of the drug to the mutant protein at its active state is *increased*, which was confirmed by an estimation of the $\Delta\Delta G_{b(S\rightarrow R)}$ value. Further analysis suggested that structural modification is needed to explain the discrepancy between the calculated and observed affinities. The simulations also revealed a modification in a hydrogen-bonding network connecting two of the kinase's regions (the activation loop and the glycine-rich loop, Figure 9c). In summary, the simulations highlighted structural modifications at the protein that could not be observed at the crystal structure and which likely contribute to resistance. This is also a good example of a study that goes beyond a simple reconstruction of the experimental observable: whereas the mutation S904F leads to resistance, this is apparently *not* because it decreases the affinity of the drug directly but due to a structural modification.

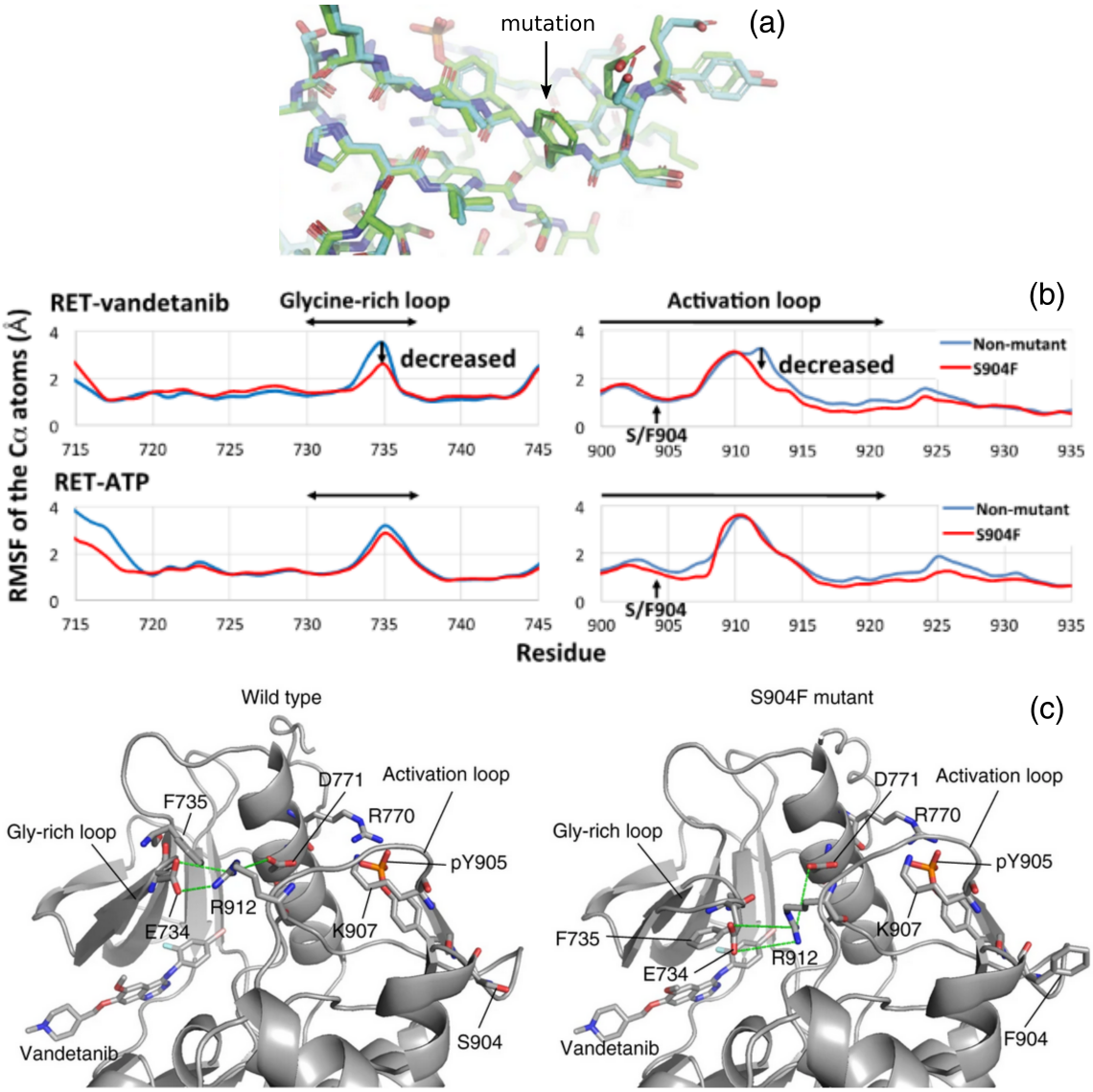


**FIGURE 9** Molecular dynamics simulations explain resistance to vandetanib. (a) Superposition of the structures of wt (cyan) and mutant (green) RET, shown with a close-up into the site of S904F mutation. (b) C$\alpha$ RMSF of two protein regions of the inhibited (top) and ATP-bound protein, calculated from MD simulations. (c) Snapshots from MD simulations showing the hydrogen-bond interactions between $Glu^{734}$ and $Arg^{912}$ in the wt and S904F variants. Reproduced from Reference 119, CC-BY 4.0 license

## 6.2 | Resistance against HCV NS3/4A protease inhibitors

NS3/4A inhibitors are widely used in the treatment of hepatitis C, a viral disease caused by the hepatitis C virus, HCV. Multiple mutations in residue $Asp^{168}$ limit the efficacy of such inhibitors, but also the replication capacity of the virus. Lately, a secondary mutation was observed that increased the replication capacity of some $Asp^{168}$ mutants.[120] The authors then used MD simulations to identify how this secondary mutation, which is not at the substrate binding site, restores the activity of the protease. Using MD simulations, they followed on the hydrogen-bonding network in the substrate binding site and observed a tighter binding between an arginine residue of the protease and a glutamate residue of the substrate upon mutation (Figure 10). This study demonstrate the usefulness of MD simulations in explaining complex phenomena in drug resistance, in this case compensatory effects by secondary mutations, which are common in resistance to antiviral drugs.

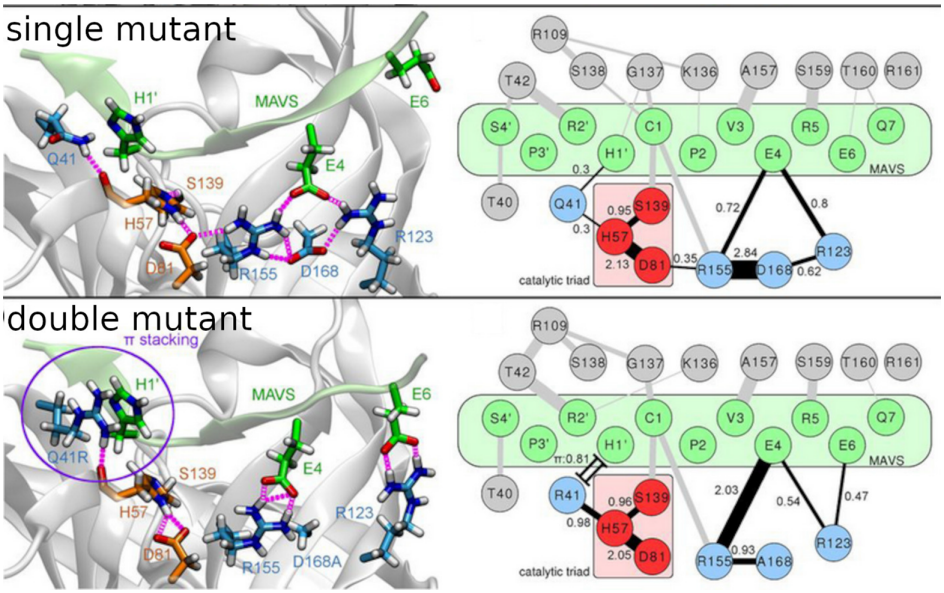


**FIGURE 10** A secondary NS3/4A mutation restores the protein's activity introduced by a resistant mutation through modifying the active site hydrogen-bonding network. The average number of hydrogen bonds between active site residues (including those of the substrate) is shown on the figure for prevalent interactions and demonstrated with lines, whose width is proportional to the average number of h-bonds in the simulations. Substrate residues are shown in green on the right-hand side of the figure, key residues are indicated on the left. The single mutant is shown on the top, the double mutant on the bottom, with a cation-$\pi$ interaction indicated (shown as a double line on the right). Note the strengthening of the charge–charge interaction involving $Arg^{155}$ of the protease and $Glu^{4}$ of the substrate. Adapted from Reference 120, CC-BY 4.0 license

## 6.3 | Compound mutations in Abl1

Abl1 is the first target for specific kinase inhibitors that has become of clinical use. The first marketed Abl1 inhibitor, imatinib, is used for treatment of chronic myeloid leukemia since 2001 and is still a best seller. However, many resistance mutations make imatinib ineffective. Newer drugs are available which are less sensitive to mutations yet the T315I mutation leads to resistance against all but one, ponatinib. No single mutation was identified that drives resistance against ponatinib, but compound mutations, i.e., two mutations in the same copy of the gene, can lead to resistance even against this inhibitor.[25] Almost all of these mutations are a combination of two mutants, where each of these in itself does not lead to ponatinib resistance. For example, ponatinib overcomes the Y253H mutation and the T315I mutation in a clinical setting, but the combination Y253H/T315I leads to resistance, for reasons that were not clear. Kinetic studies provided a hint: many of the double-mutants are more active than a single mutant that in itself does not lead to sufficient resistance. Thus, it is likely that the double mutants shift the conformational space so that a configuration that is prone to bind ponatinib is less often available. MD simulations indeed show changes in the activation loop of the protein. The dynamics of this loop is modified upon mutation, which is likely to affect both catalysis and drug binding,[9] as shown in Figure 11.

## 6.4 | Resistance to an antifungal drug mediated via a distal lysine to glutamate mutation

Antifungal treatment is also subject for resistance mutations as the microscopic pathogenic fungi have ample time to develop resistance to treatment. The oxaborole inhibitor AN2690 is an inhibitor of *Candida albicans* LeuRS, an enzyme that is essential to *C. albicans*' synthesis of Leu-tRNA. AN2690 binds covalently to an AMP cofactor instead of the amino acid. Unfortunately, *C. albicans* can become resistant to treatment with AN2690 through resistance mutations. One of these is K510E. The lysine residue is not located in the vicinity of the bound drug, but about 9 Å away, and it was hence not clear how its mutation leads to drug resistance. Simulations of the wt and mutant variants revealed three configurations, of which one was prone to bind the inhibitor (*I* state) and another one, called *K* state, was the most

stable. Free energy calculations revealed a larger difference in the Gibbs energy between the two states in the mutant protein, suggesting a potential mechanism for drug resistance based on a modification in the conformational energy landscape of the protein,[121] (Figure 12).

## 7 | RESISTANCE MECHANISMS OTHER THAN RESISTANCE MUTATIONS

Although the topic of this article is resistance mutations, a short discussion is included here that covers other resistance mechanisms, for completeness.

### 7.1 | Background on additional resistance mechanisms

Mutations in the drug target are only one of many mechanisms by which resistance to therapy is mediated. Additional resistance mechanisms are briefly summarized in Table 3. Resistance mechanisms are not mutually exclusive, and can occur together with resistance mutations and with other means of drug resistance.

Efflux is mediated by proteins that pump exogenous molecules (including drugs) out of cells, and is an effective mechanism if the molecular drug target is intracellular. Another trivial resistance mechanism is overexpression of the drug target, effectively leading to a need to increase the concentration of the drug in order to bind more of the target molecules. This may lead to a phenomenon known in cancer therapeutics as "tumor drug addiction,"[122] where tumor cells express so much of the drug target that it becomes overactive. Thus, growth of the tumor slows down when the therapy is removed; this is however only a temporary effect as the tumor reacts by growth of cells that do not overexpress the drug target.

A less trivial mechanism is bypass of the drug target through activation of another protein (or pathway) with a biological effect that is similar to that of the inhibited drug target. This may happen via activating mutations, leading to a protein that takes over the role of the drug target, or by another mechanism that leads to overexpression or overactivation of genes that compensate for the drug target's effects.

Drug metabolism takes place most often in the liver, and is an important factor in determining the drug's toxicity and efficacy. However, various enzymes can inactivate the drug inside cells, effectively rendering it useless. This is a common means of resistance against antibiotic therapy, but can also occur in cancers.[123]

In the case of an antimicrobial treatment, additional mechanisms may lead to resistance. These include changes in the target organelle (such as ribosome methylation or modification of membrane lipids). In addition, modification to the target organism's cell wall may prevent the entry of the drug in the first place.

### 7.2 | Studying additional resistance mechanisms by methods that stem from computational molecular science

Few of the above-mentioned resistance mechanisms (Table 3) can be studied effectively by molecular simulation methods. Mathematical/evolutionary modeling is often more appropriate to this aim.[124,125] However, even in such models, the use of chemical concepts such as chemical activity[126,127] can be incorporated into the simulations. Moreover, defining the degree of resistance as fold-$IC_{50}$ ($f_{IC_{50}}$) has been used to include measurements of drug-sensitivity directly[128,129]:

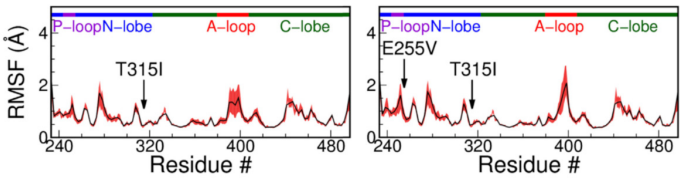


FIGURE 11 The dynamics of T315I and E255V/T315I mutants of Abl1. RMSF, calculated for simulations of Abl1, reveals that there is a decrease in the plasticity of the activation loop (A-loop) residues upon this mutation. In the E255V/T315I double mutant, the flexibility is enhanced (see the highest RMSF for A-loop residues). Since A-loop dynamics is important for activation, this likely stabilizes the active conformation that does not bind ponatinib, which decreases the binding affinity and increases the catalytic efficiency. Reproduced from Reference 9 with permission from Elsevier

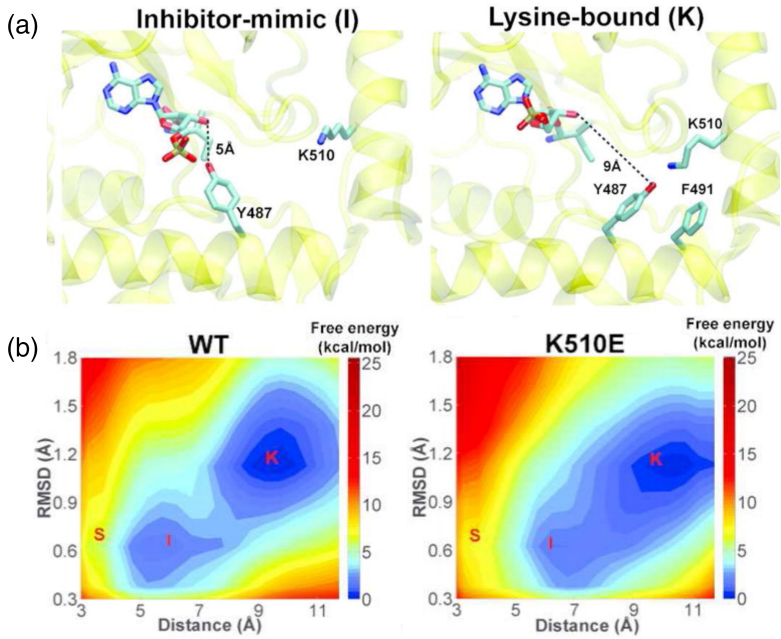


**FIGURE 12** Molecular dynamics simulations and free energy landscape analysis reveals why the K510E mutation in the fungal protein *Candida albicans* LeuRS leads to drug resistance. (a) In the wt protein, the side chain of residue $Tyr^{487}$ can adopt primarily two conformations, one that is similar to the inhibitor-bound conformation, *I*, and another where it is hydrogen bonded to $Lys^{510}$, *K*. A third conformation, *S* is less common in the simulations and is not shown. The simulations were run in the presence of a substrate. (b) Free energy landscapes, extracted from metadynamics simulations of the wt and the mutant show that the *I* conformation is higher in energy with respect to the *K* one, especially in the mutant, suggesting why the inhibitor binding energy is affected despite the large distance between residue 510 and the substrate (or inhibitor). Adapted with permission from Reference 121. Copyright (2015) American Chemical Society

**TABLE 3** Resistance mechanisms other than resistance mutations

| |
|---|
| General resistance mechanisms |
| Efflux of drugs outside of cells (expression of efflux pumps) |
| Overexpression of the drug target |
| Overexpression or activation of genes that compensate for the inactivity of the drug target |
| Mutations in other genes, leading to activation of additional pathways |
| Cellular modification of the drug |
| Resistance mechanisms specific for microorganisms |
| Changes in the target organelle |
| Modification of the cell wall, preventing drug entry |

$$f_{\mathrm{IC}_{50}} = \frac{\mathrm{IC}_{50}^{R}}{\mathrm{IC}_{50}^{S}}. \tag{22}$$

The $f_{IC_{50}}$ values can be used to follow on the emergence of resistance irrespective of the cause of resistance.

When details on the molecular mechanism of resistance are known, MD simulations become a useful tool to understand the causes of drug resistance. Such simulations have been used, for example, to explain how naturally occurring amphipathic amines lead to resistance against peptide antibiotics.[130] Overall, mathematical and evolutionary modeling can be used together with molecular simulations to explain resistance from a clinical and molecular standpoint.

## 8 | CONCLUSIONS

A plethora of tools is available for the estimation of binding affinity changes upon mutations in proteins. Some excel at being fast (ML, MM/GBSA, MM/PBSA, LIE) whereas others are more accurate or specific (FEP, TI, QM-based methods). Given the overall good performance of MM FFs, and the availability of tools to automate running of the simulations, MM-based FEP with BAR or MBAR analysis provides overall good accuracy at affordable speed. QM and QM/MM methods have a potential to succeed where the binding involves difficult cases (change of charge upon mutation, covalent binding to the inhibitor, metals in the binding site). To this aim, however, it is necessary to develop better protocols to consider solvent effects or enable efficient sampling to afford free energy calculations.

It is unlikely that calculations will replace experimental measurements for estimating affinity changes upon mutation any time soon. However, as presented in this article, calculations can be useful to shed light on the molecular mechanisms that underlie resistance due to mutations. MD simulations are the computational method of choice for this, as they provide multiple observables that might be hidden from the experiment. Care should be taken to run such simulations accurately, as too many studies in the literature present analyses that rely on insufficient data (e.g., a single simulation), wrong assumptions about the experimental reference (e.g., missing domains that are important for the interaction), and so on. It is also important to realize that an agreement between the calculated and measured values does not necessarily mean that the calculation is accurate; to know this, we must understand the biological system very well and consider the details of the measurement. Likewise, as highlighted in some of the examples, a disagreement between the experimental value and the calculated one does not imply that the calculations are wrong or inaccurate. Rather, such a disagreement requires an explanation, which might shed light on a molecular mechanism of resistance that could not be foreseen by analyzing the experiment.

In summary, when executed properly and with reference to the experiment, calculations of binding affinity changes upon mutation contribute much to our understanding of drug resistance, which is of utmost importance in contemporary medicine. Simulations help us to know our enemy—explaining resistance mutations is the first step in combating them.

### ACKNOWLEDGMENTS

This work was supported by The Swedish Cancer Society (Cancerfonden, project ID CAN 2018/362). Simulations in the author's group are enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at the PDC and HPC2N centers partially funded by the Swedish Research Council through grant agreement no. 2018-05973. Additional computational resources were obtained from The Center for Scientific and Technical Computing at Lund University (LUNARC).

### CONFLICT OF INTEREST

The author has declared no conflicts of interest for this article.

### AUTHOR CONTRIBUTIONS

**Ran Friedman:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; resources; writing-original draft.

### DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

### ORCID

*Ran Friedman* https://orcid.org/0000-0001-8696-3104

### RELATED WIREs ARTICLES

Protein-ligand interaction databases: advanced tools to mine activity data and interactions on a structural level

Finding the ΔΔG spot: Are predictors of binding affinity changes upon mutations in protein-protein interactions ready for it?

Ligand binding free energy and kinetics calculation in 2020

Quantitative predictions from molecular simulations using explicit or implicit interactions

## REFERENCES


1. Gould K. Antibiotics: from prehistory to the present day. J Antimicrob Chemother. 2016;71:572–5.
2. Nicolaou KC, Rigol S. A brief history of antibiotics and select advances in their synthesis. J Antibiot (Tokyo). 2018;71:153–84.
3. Glickman MS, Sawyers CL. Converting cancer therapies into cures: lessons from infectious diseases. Cell. 2012;148:1089–98.
4. Friedman R, Boye K, Flatmark K. Molecular modelling and simulations in cancer research. Biochim Biophys Acta. 2013;1836:1–14.
5. Casás-Selves M, Degregori J. How cancer shapes evolution, and how evolution shapes cancer. Evolution (N Y). 2011;4:624–34.
6. Acar A, Nichol D, Fernandez-Mateos J, Cresswell GD, Barozzi I, Hong SP, et al. Exploiting evolutionary steering to induce collateral drug sensitivity in cancer. Nat Commun. 2020;11:1923.
7. Durão P, Balbontín R, Gordo I. Evolutionary mechanisms shaping the maintenance of antibiotic resistance. Trends Microbiol. 2018;26:677–91.
8. Zhang TH, Dai L, Barton JP, Du Y, Tan Y, Pang W, et al. Predominance of positive epistasis among drug resistance-associated mutations in HIV-1 protease. PLoS Genet. 2020;16:e1009009.
9. Georgoulia PS, Todde G, Bjelic S, Friedman R. The catalytic activity of Abl1 single and compound mutations: implications for the mechanism of drug resistance mutations in chronic myeloid leukaemia. Biochim Biophys Acta Gen Subj. 2019;1863:732–41.
10. Friedman R. Drug resistance missense mutations in cancer are subject to evolutionary constraints. PLoS One. 2013;8:e82059.
11. Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. Evol Appl. 2015;8:273–83.
12. Burlingham BT, Widlanski TS. An intuitive look at the relationship of Ki and $IC_{50}$: a more general use for the Dixon plot. J Chem Educ. 2003;80:214.
13. Neubig RR, Spedding M, Kenakin T, Christopoulos A. International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII. Update on terms and symbols in quantitative pharmacology. Pharmacol Rev. 2003;55:597–606.
14. Yakimchuk K. Protein Receptor-ligand interaction/binding assays. Materials and Methods. 2011;1:199. https://doi.org/10.13070/mm.en.1.199.
15. Kairys V, Baranauskiene L, Kazlauskiene M, Matulis D, Kazlauskas E. Binding affinity in drug design: experimental and computational techniques. Expert Opin Drug Discov. 2019;14:755–68.
16. Rossi AM, Taylor CW. Analysis of protein–ligand interactions by fluorescence polarization. Nat Protoc. 2011;6:365–87.
17. Jelesarov I, Bosshard HR. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. J Mol Recognit. 1999;12:3–18.
18. Gao K, Oerlemans R, Groves MR. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. Biophys Rev. 2020;12:85–104.
19. Copeland RA, Harpel MR, Tummino PJ. Targeting enzyme inhibitors in drug discovery. Expert Opin Ther Targets. 2007;11:967–78.
20. Gross S, Rahal R, Stransky N, Lengauer C, Hoeflich KP. Targeting cancer with kinase inhibitors. J Clin Investig. 2015;125:1780–9.
21. Lou Z, Sun Y, Rao Z. Current progress in antiviral strategies. Trends Pharmacol Sci. 2014;35:86–102.
22. Shlaes DM, Alksne L, Projan SJ. The pharmaceutical industry and inhibitors of bacterial enzymes: implications for drug development. Enzyme-mediated resistance to antibiotics. Hoboken, New Jersey: Wiley Online Library; 2014. p. 215–225. https://onlinelibrary.wiley.com/doi/abs/10.1128/9781555815615.ch13.
23. Jukič M, Gobec S, Sova M. Reaching toward underexplored targets in antibacterial drug design. Drug Dev Res. 2018;80:6–10.
24. Holdgate GA, Meek TD, Grimley RL. Mechanistic enzymology in drug discovery: a fresh perspective. Nat Rev Drug Discov. 2017;17:115–32.
25. Zabriskie MS, Eide CA, Tantravahi SK, Vellore NA, Estrada J, Nicolini FE, et al. BCR-ABL1 compound mutations combining key kinase domain positions confer clinical resistance to ponatinib in Ph chromosome-positive leukemia. Cancer Cell. 2014;26:428–42.
26. Andrews JM. Determination of minimum inhibitory concentrations. J Antimicrob Chemother. 2001;48(Suppl 1):5–16.
27. Schwarz S, Silley P, Simjee S, Woodford N, van Duijkeren E, Johnson AP, et al. Editorial: assessing the antimicrobial susceptibility of bacteria obtained from animals. J Antimicrob Chemother. 2010;65:601–4.
28. Ching C, Zaman MH. Development and selection of low-level multi-drug resistance over an extended range of sub-inhibitory ciprofloxacin concentrations in *Escherichia coli*. Sci Rep. 2020;10:8754.
29. Mobley DL, Gilson MK. Predicting binding free energies: frontiers and benchmarks. Annu Rev Biophys. 2017;46:531–58.
30. Wan S, Bhati AP, Zasada SJ, Coveney PV. Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction. Interface Focus. 2020;10:20200007.
31. Dávila-Rodríguez MJ, Freire TS, Lindahl E, Caracelli I, Zukerman-Schpector J, Friedman R. Is breaking of a hydrogen bond enough to lead to drug resistance? Chem Commun (Camb). 2020;56:6727–30.
32. Malde AK, Le Zuo MB, Stroet M, Poger D, Nair PC, Oostenbrink C, et al. An automated force field topology builder (ATB) and repository: version 1.0. J Chem Theory Comput. 2011;7:4026–37.
33. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem. 2008;29:1859–65.
34. Kim S, Lee J, Jo S, Brooks CL, Lee HS, Im W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. J Comput Chem. 2017;38:1879–86.
35. Suplatov D, Sharapova Y, Švedas V. EasyAmber: a comprehensive toolbox to automate the molecular dynamics simulation of proteins. J Bioinform Comput Biol. 2020;18:2040011.
36. Dodda LS, de Vaca IC, Tirado-Rives J, Jorgensen WL. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. Nucleic Acids Res. 2017;45(W1):W331–6.

37. Zoete V, Cuendet MA, Grosdidier A, Michielin O. SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem. 2011;32:2359–68.
38. Friedman R. Simulations of biomolecules in electrolyte solutions. Adv Theory Simul. 2019;2:1800163.
39. Friedman R. Ions and the protein surface revisited: extensive molecular dynamics simulations and analysis of protein structures in alkali-chloride solutions. J Phys Chem B. 2011;115:9213–23.
40. Friedman R. Specific ion and concentration effects in acetate solutions with $Na^+$, $K^+$ and $Cs^+$. ChemPhysChem. 2019;20:1006–10.
41. Ahlstrand E, Zukerman SJ, Friedman R. Computer simulations of alkali-acetate solutions: accuracy of the forcefields in difference concentrations. J Chem Phys. 2017;147:194102.
42. Ahlstrand E, Spångberg D, Hermansson K, Friedman R. Interaction energies between metal ions ($Zn^{2+}$ and $Cd^{2+}$) and biologically relevant ligands. Int J Quantum Chem. 2013;113:2554–62.
43. Ahlstrand E, Hermansson K, Friedman R. Interaction energies in complexes of Zn and amino acids: a comparison of ab initio and force field based calculations. J Phys Chem A. 2017;121:2643–54.
44. Friedman R, Nachliel E, Gutman M. Molecular dynamics of a protein surface: ion-residues interactions. Biophys J. 2005;89:768–81.
45. Knapp B, Ospina L, Deane CM. Avoiding false positive conclusions in molecular simulation: the importance of replicas. J Chem Theory Comput. 2018;14:6127–38.
46. Friedman R, Pellarin R, Caflisch A. Amyloid aggregation on lipid bilayers and its impact on membrane permeability. J Mol Biol. 2009;387:407–15.
47. Souza PCT, Thallmair S, Conflitti P, Ramírez-Palacios C, Alessandri R, Raniolo S, et al. Protein–ligand binding with the coarse-grained Martini model. Nat Commun. 2020;11:3714.
48. Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H. Calculation of binding free energies. Methods Mol Biol. 2015;1215:173–209.
49. Kirkwood JG. Statistical mechanics of fluid mixtures. J Chem Phys. 1935;3:300–13.
50. Bennett CH. Efficient estimation of free energy differences from Monte Carlo data. J Comput Phys. 1976;22:245–68.
51. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. J Chem Phys. 2008;129:124105.
52. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al., editors. CHARMM: the biomolecular simulation program. J Comput Chem. 2009;30:1545–614.
53. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26:1701–18.
54. Ding X, Vilseck JZ, Brooks CL. Fast solver for large scale multistate Bennett acceptance ratio equations. J Chem Theory Comput. 2019;15:799–802.
55. Shirts MR, Mobley DL. An introduction to best practices in free energy calculations. Methods in molecular biology. Berlin, Germany: Springer; 2012. p. 271–311. https://doi.org/10.1007/978-1-62703-017-5_11.
56. Klimovich PV, Shirts MR, Mobley DL. Guidelines for the analysis of free energy calculations. J Comput Aided Mol Des. 2015;29:397–411.
57. Aldeghi M, Gapsys V, de Groot BL. Accurate estimation of ligand binding affinity changes upon protein mutation. ACS Cent Sci. 2018;4:1708–18.
58. Bhati AP, Wan S, Coveney PV. Ensemble-based replica exchange alchemical free energy methods: the effect of protein mutations on inhibitor binding. J Chem Theory Comput. 2019;15:1265–77.
59. Rocklin GJ, Mobley DL, Dill KA, Hünenberger PH. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. J Chem Phys. 2013;139:184103.
60. Dixit SB, Chipot C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin-streptavidin system revisited. J Phys Chem A. 2001;105:9795–9.
61. Hub JS, de Groot BL, Grubmüller H, Groenhof G. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. J Chem Theory Comput. 2014;10:381–90.
62. Gapsys V, de Groot BL. pmx Webserver: a user friendly Interface for Alchemistry. J Chem Inf Model. 2017;57:109–14.
63. Jespers W, Isaksen GV, Andberg TAH, Vasile S, van Veen A, Åqvist J, et al. QresFEP: an automated protocol for free energy calculations of protein mutations in Q. J Chem Theory Comput. 2019;15:5461–73.
64. Marelius J, Kolmodin K, Feierberg I, Aqvist J. Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. J Mol Graph Model. 1998;16:213–25, 261.
65. Dixit M, Das S, Mhashal AR, Eitan R, Major DT. Practical aspects of multiscale classical and quantum simulations of enzyme reactions. Methods Enzymol. 2016;577:251–86.
66. Ganoth A, Friedman R, Nachliel E, Gutman M. A molecular dynamics study and free energy analysis of complexes between the Mlc1p protein and two IQ motif peptides. Biophys J. 2006;91:2436–50.
67. Reyes CM, Nifosì R, Frankel AD, Kollman PA. Molecular dynamics and binding specificity analysis of the bovine immunodeficiency virus BIV Tat-TAR complex. Biophys J. 2001;80:2833–42.
68. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015;10:449–61.

69. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. Chem Rev. 2019;119:9478–508.
70. Uhlig HH. The solubilities of gases and surface tension. J Phys Chem. 1937;41:1215–26.
71. Eley DD. On the solubility of gases. Part II—a comparison of organic solvents with water. Trans Faraday Soc. 1939;35:1421–32.
72. Hermann RB. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. J Phys Chem. 1972;76:2754–9.
73. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem. 1994; 98:1978–88.
74. Georgoulia PS, Bjelic S, Friedman R. Deciphering the molecular mechanism of FLT3 resistance mutations. FEBS J. 2020;287:3200–20.
75. Duan L, Liu X, Zhang JZH. Interaction entropy: a new paradigm for highly efficient and reliable computation of protein–ligand binding free energy. J Am Chem Soc. 2016;138:5722–8.
76. Huang K, Luo S, Cong Y, Zhong S, Zhang JZH, Duan L. An accurate free energy estimator: based on MM/PBSA combined with interaction entropy for protein–ligand binding affinity. Nanoscale. 2020;12:10737–50.
77. Åqvist J, Medina C, Samuelsson JE. A new method for predicting binding affinity in computer-aided drug design. Protein Eng. 1994;7: 385–91.
78. Åqvist J, Hansson T. On the validity of electrostatic linear response in polar solvents. J Phys Chem. 1996;100:9512–21.
79. de Terán HG, Åqvist J. Linear interaction energy: method and applications in drug design. Methods in molecular biology. New York: Springer; 2011. p. 305–23.
80. Huang D, Caflisch A. Efficient evaluation of binding free energy using continuum electrostatics solvation. J Med Chem. 2004;47: 5791–7.
81. Friedman R, Caflisch A. Discovery of plasmepsin inhibitors by fragment-based docking and consensus scoring. ChemMedChem. 2009; 4:1317–26.
82. Genheden S, Ryde U. Comparison of the efficiency of the LIE and MM/GBSA methods to calculate ligand-binding energies. J Chem Theory Comput. 2011;7:3768–78.
83. Homeyer N, Stoll F, Hillisch A, Gohlke H. Binding free energy calculations for lead optimization: assessment of their accuracy in an industrial drug design context. J Chem Theory Comput. 2014;10:3331–44.
84. Rifai EA, van Dijk M, Vermeulen NPE, Yanuar A, Geerke DP. A comparative linear interaction energy and MM/PBSA study on SIRT1-ligand binding free energy calculation. J Chem Inf Model. 2019;59:4018–33.
85. Ryde U, Söderhjelm P. Ligand-binding affinity estimates supported by quantum-mechanical methods. Chem Rev. 2016;116:5520–66.
86. Grimme S. Improved third-order Møller–Plesset perturbation theory. J Comput Chem. 2003;24:1529–37.
87. Friedman R. Preferential binding of lanthanides to methanol dehydrogenase evaluated with density functional theory. J Phys Chem B. 2021;125:2251–7.
88. Xu Y, Friedman R, Wu W, Su P. Understanding intermolecular interactions of large systems in ground state and excited state by using density functional based tight binding methods. J Chem Phys. 2021;154:194106.
89. Christensen AS, Kubař T, Cui Q, Elstner M. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. Chem Rev. 2016;116:5301–37.
90. Siegbahn PEM, Himo F. The quantum chemical cluster approach for modeling enzyme reactions. WIREs Comput Mol Sci. 2011;1: 323–36.
91. Fouda A, Ryde U. Does the DFT self-interaction error affect energies calculated in proteins with large QM systems? J Chem Theory Comput. 2016;12:5667–79.
92. Miertuš S, Scrocco E, Tomasi J. Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. Chem Phys. 1981;55:117–29.
93. Klamt A, Schüürmann G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans. 1993;2:799–805.
94. Marenich AV, Cramer CJ, Truhlar DG. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phys Chem B. 2009;113:6378–96.
95. Friedman R. Structural and computational insights into the versatility of cadmium binding to proteins. Dalton Trans. 2014;43:2878–87.
96. Lukac I, Abdelhakim H, Ward RA, St-Gallay SA, Madden JC, Leach AG. Predicting protein–ligand binding affinity and correcting crystal structures with quantum mechanical calculations: lactate dehydrogenase A. Chem Sci. 2019;10:2218–27.
97. He X, KM M Jr. Divide-and-conquer Hartree–Fock calculations on proteins. J Chem Theory Comput. 2010;6:405–11.
98. Otsuka T, Okimoto N, Taiji M. Assessment and acceleration of binding energy calculations for protein–ligand complexes by the fragment molecular orbital method. J Comput Chem. 2015;36:2209–18.
99. Jia X, Wang X, Liu J, Zhang JZH, Mei Y, He X. An improved fragment-based quantum mechanical method for calculation of electrostatic solvation energy of proteins. J Chem Phys. 2013;139:214104.
100. Liu J, Wang X, Zhang JZH, He X. Calculation of protein–ligand binding affinities based on a fragment quantum mechanical method. RSC Adv. 2015;5:107020–30.
101. Zhou T, Huang D, Caflisch A. Is quantum mechanics necessary for predicting binding free energy? J Med Chem. 2008;51:4280–8.

102. Nascimento ÉCM, Oliva M, Świderek K, Martins JBL, Andrés J. Binding analysis of some classical acetylcholinesterase inhibitors: insights for a rational design using free energy perturbation method calculations with QM/MM MD simulations. J Chem Inf Model. 2017;57:958–76.
103. Giese TJ, York DM. Development of a robust indirect approach for MM → QM free energy calculations that combines force-matched reference potential and Bennett's acceptance ratio methods. J Chem Theory Comput. 2019;15:5543–62.
104. Tkatchenko A. Machine learning for chemical discovery. Nat Commun. 2020;11:4125.
105. Duran-Frigola M, Fernández-Torras A, Bertoni M, Aloy P. Formatting biological big data for modern machine learning in drug discovery. WIREs Comput Mol Sci. 2018;9:e1408. https://doi.org/10.1002/wcms.1408.
106. Coveney PV, Dougherty ER, Highfield RR. Big data need big theory too. Philos Trans R Soc A Math Phys Eng Sci. 2016;374:20160153.
107. Wang DD, Zhu M, Yan H. Computationally predicting binding affinity in protein–ligand complexes: free energy-based simulations and machine learning-based scoring functions. Brief Bioinform. 2020;22:bbaa107.
108. Wang DD, Ou-Yang L, Xie H, Zhu M, Yan H. Predicting the impacts of mutations on protein–ligand binding affinity based on molecular dynamics simulations and machine learning methods. Comput Struct Biotechnol J. 2020;18:439–54.
109. Hernandez Maganhi S, Jensen P, Caracelli I, Zukerman SJ, Fröhling S, Friedman R. Palbociclib can overcome mutations in cyclin dependent kinase 6 that break hydrogen bonds between the drug and the protein. Protein Sci. 2017;26:870–9.
110. Xin F, Radivojac P. Post-translational modifications induce significant yet not extreme changes to protein structure. Bioinformatics. 2012;28:2905–13.
111. Mittag T, Kay LE, Forman-Kay JD. Protein dynamics and conformational disorder in molecular recognition. J Mol Recognit. 2010;23:105–16.
112. Luckert K, Götschel F, Sorger PK, Hecht A, TO Joos, Pötz O. Snapshots of protein dynamics and post-translational modifications in one experiment—beta-catenin and its functions. Mol Cell Proteomics. 2011;10:M110.007377.
113. Lee HS, Qi Y, Im W. Effects of N-glycosylation on protein conformation and dynamics: protein data bank analysis and molecular dynamics simulation study. Sci Rep. 2015;5:8926.
114. Steinz MM, Persson M, Aresh B, Olsson K, Cheng AJ, Ahlstrand E, et al. Oxidative hotspots on actin promote skeletal muscle weakness in rheumatoid arthritis. JCI Insight. 2019;5:e126347.
115. Friedman R, Bjelic S. Simulations studies of protein kinases that are molecular targets in cancer. Israel J Chem. 2020;60:667–80.
116. Todde G, Friedman R. Pattern and dynamics of FLT3 duplications. J Chem Inf Model. 2020;60:4005–20.
117. Nussinov R, Tsai CJ. Allostery in disease and in drug discovery. Cell. 2013;153:293–305.
118. Friedman R. The molecular mechanism behind resistance of the kinase FLT3 to the inhibitor quizartinib. Proteins. 2017;85:2143–52.
119. Nakaoku T, Kohno T, Araki M, Niho S, Chauhan R, Knowles PP, et al. A secondary RET mutation in the activation loop conferring resistance to vandetanib. Nat Commun. 2018;9:625.
120. Dultz G, Shimakami T, Schneider M, Murai K, Yamane D, Marion A, et al. Extended interaction networks with HCV protease NS3-4A substrates explain the lack of adaptive capability against protease inhibitors. J Biol Chem. 2020;295:13862–74.
121. Zhao H, Palencia A, Seiradake E, Ghaemi Z, Cusack S, Luthey-Schulten Z, et al. Analysis of the resistance mechanism of a benzoxaborole inhibitor reveals insight into the leucyl-tRNA synthetase editing mechanism. ACS Chem Biol. 2015;10:2277–85.
122. Bradley CA. Targeted therapies: understanding tumour drug addiction. Nat Rev Cancer. 2017;17:634–5.
123. Townsend DM, Tew KD. The role of glutathione-S-transferase in anti-cancer drug resistance. Oncogene. 2003;22:7369–75.
124. Sun X, Hu B. Mathematical modeling and computational prediction of cancer drug resistance. Brief Bioinform. 2018;19:1382–99.
125. Knight GM, Davies NG, Colijn C, Coll F, Donker T, Gifford DR, et al. Mathematical modelling for antibiotic resistance control policy: do we know enough? BMC Infect Dis. 2019;19:1011.
126. Buetti-Dinh A, Pivkin IV, Friedman R. S100A4 and its role in metastasis – computational integration of data on biological networks. Mol BioSyst. 2015;11:2238–46.
127. Buetti-Dinh A, Pivkin IV, Friedman R. S100A4 and its role in metastasis – simulations of knockout and amplification of epithelial growth factor receptor and matrix metalloproteinases. Mol BioSyst. 2015;11:2247–54.
128. Lindström HJG, de Wijn AS, Friedman R. Stochastic modelling of tyrosine kinase inhibitor rotation therapy in chronic myeloid leukaemia. BMC Cancer. 2019;19:508.
129. Yang J, Lindström HJG, Friedman R. Combating drug resistance in acute myeloid leukaemia by drug rotations: the effects of quizartinib and pexidartinib. Cancer Cell Int. 2021;21:198.
130. Li J, Beuerman R, Verma CS. Mechanism of polyamine induced colistin resistance through electrostatic networks on bacterial outer membranes. Biochim Biophys Acta Biomembr. 2020;1862:183297.
131. Nič M, Jirát J, Košata B, Jenkins A, McNaught A, editors. IUPAC compendium of chemical terminology. Research Triangle Park, NC: IUPAC; 2009. https://goldbook.iupac.org/.


**How to cite this article:** Friedman R. Computational studies of protein–drug binding affinity changes upon mutations in the drug target. WIREs Comput Mol Sci. 2022;12:e1563. https://doi.org/10.1002/wcms.1563

## APPENDIX A.: REACTION AFFINITY

The formal definition of affinity refers to *reaction affinity*, which is the negative partial derivative of the Gibbs energy with respect to the *extent of reaction*, at constant pressure and temperature[131]:

$$A = -\left(\frac{dG}{d\xi}\right)_{T,p}, \tag{A1}$$

where the change in the extent of reaction is:

$$d\xi = \frac{dn_x}{d\nu_x}. \tag{A2}$$

Here, $n_x$ and $\nu_x$ are the amount and stoichiometric coefficient of $x$, respectively, and $x$ is a reactant or product.