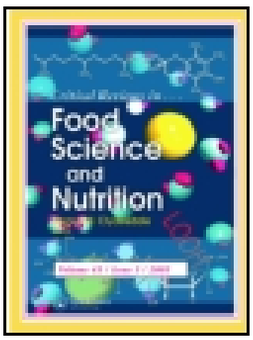

Critical Reviews in Food Science and Nutrition

ISSN: 1040-8398 (Print) 1549-7852 (Online) Journal homepage: https://www.tandfonline.com/loi/bfsn20

# Current methods and applications in computational protein design for food industry

Qiuming Chen, Yaqin Xiao, Wenli Zhang & Wanmeng Mu

**To cite this article:** Qiuming Chen, Yaqin Xiao, Wenli Zhang & Wanmeng Mu (2019): Current methods and applications in computational protein design for food industry, Critical Reviews in Food Science and Nutrition, DOI: 10.1080/10408398.2019.1682513

**To link to this article:** https://doi.org/10.1080/10408398.2019.1682513

Published online: 04 Nov 2019.

CRITICAL REVIEWS IN FOOD SCIENCE AND NUTRITION
https://doi.org/10.1080/10408398.2019.1682513



# Current methods and applications in computational protein design for food industry

Qiuming Chen[a], Yaqin Xiao[a], Wenli Zhang[a], and Wanmeng Mu[a,b]

[a]State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, China; [b]International Joint Laboratory on Food Safety, Jiangnan University, Wuxi, China

**ABSTRACT**

Computational tools for enzyme engineering can readily be used in a broad range of food industrial applications. However, there are too many choices when the enzymologists try to solve their own problems computationally, especially when their studies need to be carried out with a combination of tools. The correct choice of methods requires a broad understanding of the knowledge framework. Therefore, we present a comprehensive overview of the current computational tools and basic principles for enzyme design. The tools can be classified into several groups, including bioinformatics approaches and the calculation methods based on static systems and dynamics systems. In addition, we also provide some successful examples in the food industrial applications to show that the modern tools can dramatically reduce the experimental effort and can help us better understand the catalytic mechanism of food enzymes.

**KEYWORDS**

Computational method; enzyme engineering; rational design; free energy calculation; bioinformatics; food industry

## Introduction

Enzymes are biological molecules that can accelerate the reactions within cells. They are widely used as "greener" catalysts in practical applications than chemical catalysts. Enzymes can decrease the activation barriers of reactions. Some enzymes can be used to transform substrates into products of industrial interest efficiently and selectively. However, because of the delicate feature of protein structures, most natural enzymes cannot survive in the relatively harsh condition for the long-term industrial production. Such a shortcoming of enzymes can be improved by protein engineering. This technology helps to build up enzyme's resistance to extreme environmental conditions. Besides improving the efficiency of the experimental methods, biotechnologists also turn their attention to theoretical guidance for protein engineering.

Computational techniques are powerful tools to unveil the molecular basis of mutations, or to make protein design more rational or both. Plenty of computational methods were developed based on different strategies (Fig. 1). The work to unveil the molecular basis is mainly connected with molecular simulations. With the rapid development of simulation algorithms and the increase of available computational power, molecular simulations, including Monte-Carlo (MC) and molecular dynamics (MD) simulation, are more accessible for enzyme engineers. MC simulation is more suitable for the barrier crossings and is less favorable for the liquid system simulations. Compared to the protein folding and unfolding processes, there are no relatively big energy barriers to cross in most active enzymes. MD simulation becomes method of choice in studies of enzyme-based processes. The aim of identifying mutational hotspots and designing smart libraries can also be achieved by other bioinformatics and molecular modeling methods. In this article, we present recent advances in computational tools for protein engineering with a focus on food industry. Some user-friendly tools along with their applications are discussed to illustrate the advantages provided by computational methods and to enlighten experimentalists in food industry.

The use of hydrolase for debranching, clarification, and improving the solubility in food since ancient times is an age-old process. Nowadays the practical uses of enzymes in food industry have not limited to hydrolase. Oxidoreductase and other types of enzymes, such as isomerase, are also potent tools for food production and processing. Novel enzymes have been explored to synthesize the functional food component beyond primary functions of food. Prebiotics, such as D-Allulose produced by ketose 3-epimerase (Zhang et al. 2016) and lactulose produced by cellobiose 2-epimerase (Chen et al. 2018), show a bright market prospect. The enzyme-applied industry poses a great challenge for protein design to improve cost-effectiveness, that is, to increase the catalytic efficiency and reusability (stability) of the existing enzymes. The strategies of *de novo* protein design for novel enzyme catalysts, which showed the potential in highly-advanced food areas, are not covered in this study.

**CONTACT** Wanmeng Mu wmmu@jiangnan.edu.cn State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu 214122, China.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/bfsn.

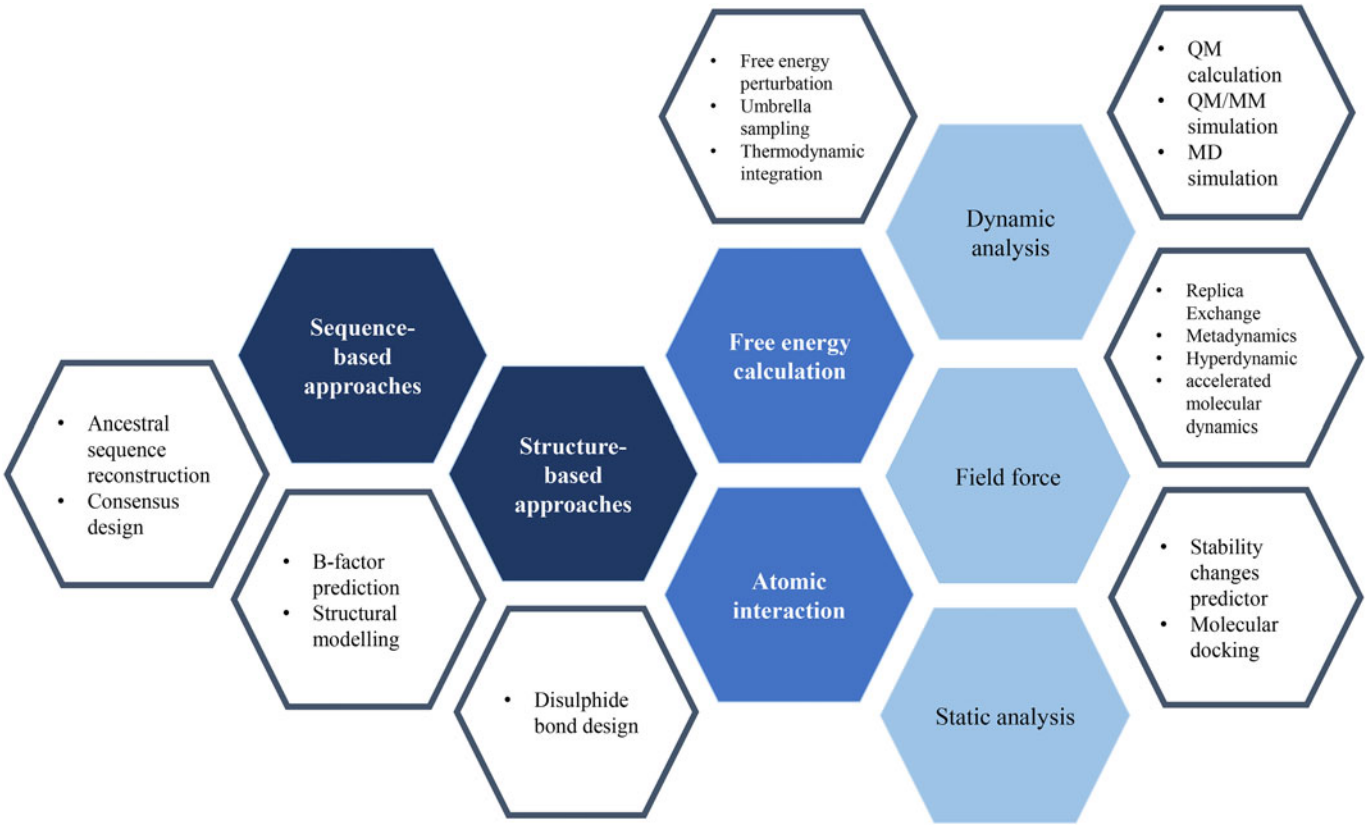


**Figure 1.** Schematic presentation of the architecture of computational tools for enzyme engineering. The methods and strategies in this figure are derived from their adjacent graphics. The darker the graphic is, the more fundamental the strategies in it is.

## Strategies for understanding molecular mechanisms and computational protein design

### *Bioinformatics strategy*

The bioinformatics are knowledge-based approaches that use genomic sequence and structural information to design smart libraries for mutations (Suplatov, Voevodin, and Švedas 2015).

### *Sequence-based bioinformatics*

The sequence-based approaches, including the ancestral and the consensus methods, can guide protein design without structural information. The first step of the general sequence-based rational engineering is to identify the suitable homologous sequences of the target sequence (Fig. 2). Basic Local Alignment Search Tool (BLAST) algorithm (Altschul et al. 1990) can be used to search the vast amounts of sequence data for sequence similarities. Multiple sequence alignments (MSAs) can subsequentially explore sequence conservation of the homologous sequences (Thompson, Higgins, and Gibson 1994). There are some assumptions about the homologous sequences: (i) amino acid substitution process containing evolutionary history underlies the homologous sequences. Thus, ancestral sequences originating from thermophiles tend to encode more thermostable proteins (Akanuma 2017); (ii) the frequent amino acid at a given position contributes more to the protein stability (Porebski and Buckle 2016); (iii) the high conservative position is more likely to be the hot spots for engineering the catalytic activity or substrate selectivity (Ebert and Pelletier 2017; Jackel et al. 2010; Porebski and Buckle 2016). The ancestral and the consensus methods are developed based on these assumptions.

Ancestral sequence reconstruction is a technique for exploring evolutionary hypotheses. It can also discover the relationship between sequence and molecular phenotype, which can guide protein engineering. The models of amino acid evolution will be built as a starting point. The rationality of the models is the key to success. The phylogenetic hierarchy information from the models can be used to design mutations for improving stability based on assumptions mentioned above. The requirement of significant computational and methodological expertise becomes a barrier to the widespread use of this method among food microbiologists. PhyloBot provides interactive web-based tools to implement ancestral sequence reconstruction. Requiring only a collection of orthologous protein sequences, PhyloBot provides the user with automated ancestral sequence reconstruction analysis and visual tools to analyze results. It also offers mutation suggestions on every phylogenetic branch. PhyloBot greatly simplifies protocol for ancestral reconstruction process and promotes experimental use among food microbiologists (Hanson-Smith and Johnson 2016).

The consensus method is another method to utilize sequence information for protein design. It also needs a set of homologous protein sequences obtained from MSAs as the input. Other than inferring phylogenetic hierarchy, the consensus method uses the aligned sequences to identify

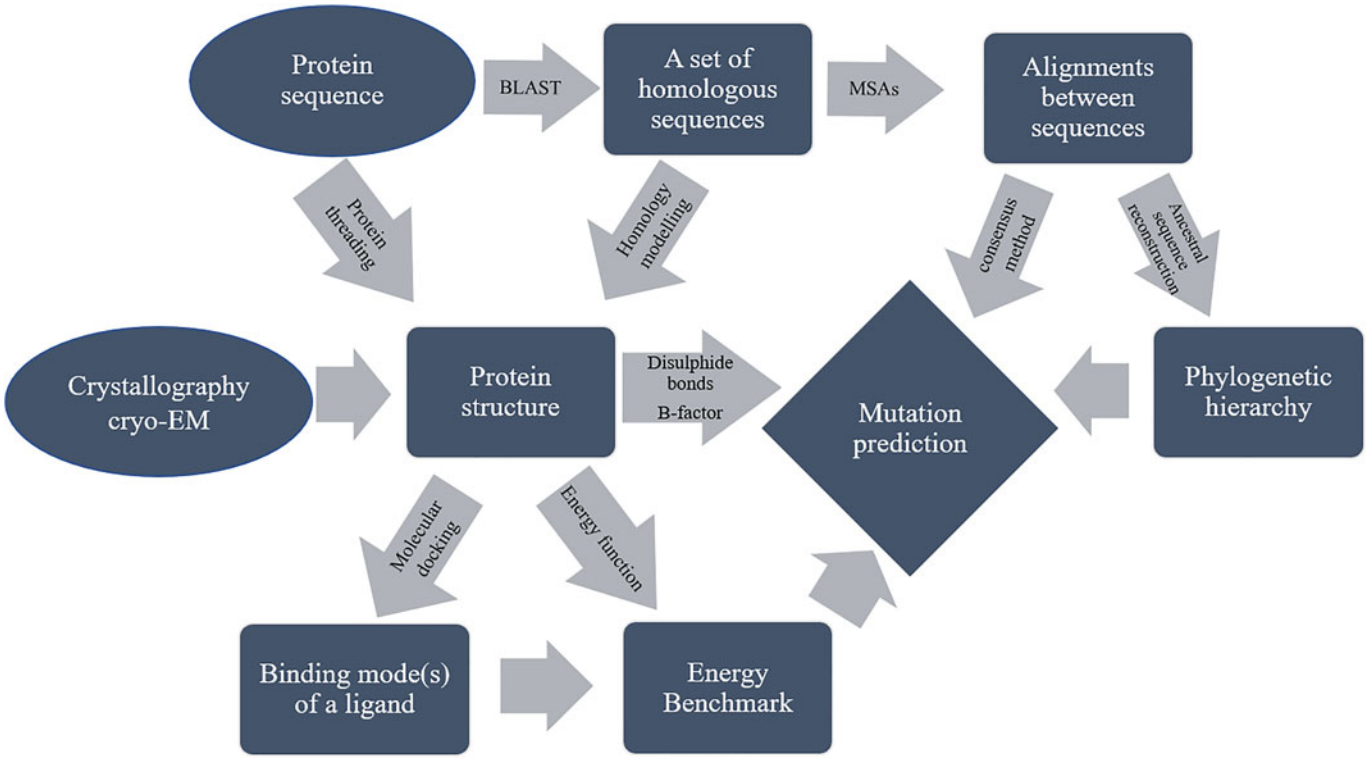


**Figure 2.** The general protocol for computer-aided design of mutations using the information of protein sequence and static structures.

most frequent amino acid at each position in the alignment. These will be the target amino acids for focused mutagenesis. Such computational method can dramatically reduce the library size of mutagenesis (Porebski and Buckle 2016). There is a controversial issue between ancestral sequence reconstruction and consensus method that the inferred ancestor protein may have a bias toward frequent amino acid (Trudeau, Kaltenbach, and Tawfik 2016). And, the consensus method may also be biased by the skew of the MSAs step or the preferences from genome sequencing projects (Porebski and Buckle 2016). Thus, the selection of a proper set of homologous protein sequences is crucial to the prediction accuracy made by consensus method.

Some tools such as INPS (Fariselli et al. 2015) and SIFT (Sim et al. 2012) that can predict the effect of variations on protein function are also developed based on the information of protein sequences. However, these tools are designed to predict the disease-causing mutations, and therefore are out of scope of this article.

When the target enzyme belongs to a well-populated dataset of homologs, it is fast and robust to build smart libraries for engineering enzyme stabilities by using the sequence-based approaches. However, if the current dataset cannot provide unbiased sequential information of the target enzyme, the structure-based strategy should be considered. Compared to sequence-based strategy, the structure-based one gives more insight into the physical basis of mutations and the reaction mechanism.

### *Structure-based bioinformatics*

As the rapid development of structural biology techniques such as X-ray crystallography and cryo-electron microscopy, the availability of structural data increases. However, the number of experimental protein structures is still three orders of magnitude fewer than that of protein sequences (Sumbalova et al. 2018). Therefore, computational structural modeling is an important method to eliminate the information asymmetry between sequential and structural data. The predicted structures can be obtained from sequences by many tools. Based on the results from CAMEO (Haas et al. 2018), the protein structure prediction servers with better performance are listed in Table 1. The listed servers have their own merits and shortages. Researchers should choose the servers according to their individual demands. There is a tendency that machine learning and increasing sequence database are playing important roles in the improvement of these methods (Schaarschmidt et al. 2018). Structure visualizing tools such as Pymol (Schrödinger 2010) and VMD (Humphrey, Dalke, and Schulten 1996) can subsequently assist structure analyses.

The structure of protein is more conservative than the sequence of it during evolutionary process. It is a more intuitive way to aid protein design with the information from protein structures. A widely used structural information for protein engineering from experimental result is temperature factors, or B-factor. This parameter in Protein Data Bank (PDB) (Berman, 2000) files describes the mobility of the atomic positions. The flexible residues with higher B-factors can be used as candidates for mutation. The information of B-factor can also help the selection of random mutagenesis method (Verma, Schwaneberg, and Roccatano 2012) or the engineering of disulfide bonds (Craig and Dombkowski 2013). The B-factor is obtained from experiments. The fluctuation that B-factor describes is measured under the experimental conditions in crystalline solid state. Therefore, there are several approaches to predict B-factors or flexibility of

**Table 1.** Current computational methods for protein design.

| Strategy and method | General usage | Web server or software tool | Advantage/disadvantage | Exemplar studies for food enzyme |
|---|---|---|---|---|
| *Sequence-based bioinformatics* | | | | |
| BLAST | Sequence searching | National Center for Biotechnology Information (NCBI) and UniProt | Useful for finding homologous proteins; The boundary of identifying the homologs is obscure | – |
| MSAs | Multiple sequence alignment | ClustalW, T-Coffee and MUSCLE | Useful for identifying conserved sequence regions | – |
| Ancestral sequence reconstruction | Ancestral sequence searching | PhyloBot, FastML | Robust and efficient; Require the knowledge about the targeted gene family; Strongly influenced by the evolutionary model | Boucher et al. (2014); Steindel et al. (2016) |
| Consensus method | Conserved residues searching | HotSpot Wizard | Not need structural information; Calculation is time-saving; May be biased by the selection of homologous proteins | Wang et al. (2016) |
| *Structure-based bioinformatics* | | | | |
| B-factor-based method | Predicting flexibility of residues | MAP(2.0)3D, PredyFlexy, PredBF, and HotSpot Wizard | Principle is easy to understand; Less accurate than the B-factor from experimental data | Kim, Quang, and Kim (2010) |
| Disulfide bond design | Introduction of novel disulfide bonds | Disulfide by Design, FRESCO, and SSBOND | Provide considerable stability to proteins; Predicted disulfide bonds may not be as expected | Li et al. (2018); Liu et al. (2014) |
| Structural modeling | Protein structure prediction | HHPred, RaptorX, SWISSMODEL, Rosetta, Robetta, and Sparks-X | Avail of rich abundance of sequence data; Prediction accuracy still needs to be improved; *Ab initio* modeling is time-consuming | Lee et al. (2009); Liu et al. (2014) |
| Structural inspection | Structural visualization | Pymol and VMD | Useful for protein structural visualization | Lee et al. (2009); Park et al. (2017) |
| *Static systems calculations* | | | | |
| Stability changes predictor | Predicting free energy changes upon mutations | FoldX, Rosetta, PoPMuSiC, STRUM, and I-Mutant | High speed of calculations; The predicting power is very limited | Liu et al. (2014); Van Overtveldt et al. (2018); Wijma et al. (2014) |
| Molecular docking | Predicting binding modes of protein and ligand | GOLD, Glide, MOE-Dock, AutoDock Vina, and AutoDock | Fast prediction of the interaction between proteins and ligands. Lack of consideration in the "induced fit" | Pina and Roque (2009); Wang et al. (2014) |
| *Dynamics simulations* | | | | |
| QM calculation and QM/MM | Research on the bond breaking and forming | Gaussian, Orca, and GAMESS | The best predicting power; Capable of energy calculation involving bond breaking and forming; Computationally intensive | Kosugi and Hayashi (2011, 2012); Santiago et al. (2016) |
| MD simulations | Research on the dynamic behaviour of molecules | Amber, CHARMM, Gromacs, NAMD, and CABS-flex | More informative in protein dynamics; Relatively high speed of calculations; Cannot simulate bond breaking and forming | Barbe et al. (2009); Sefidbakht, et al. (2017) |

residues from the sequence (de Brevern et al. 2012; Pan and Shen 2009) or, which can be used to increase protein thermostability (Reetz, Carballeira, and Vogel 2006). The flexibility can also be reflected by the root mean square fluctuations (RMSF) value from MD simulation, which is discussed below.

The introduction of disulfide bonds enhances stability to the folded state of protein. Many successful attempts have utilized engineered disulfide bonds to increase the stability of proteins (Le et al. 2012; Liu et al. 2016; Wijma et al. 2014; Yu et al. 2012). The introduction of a disulfide bond to a protein can increase the unfolding free energy by 2.3–5.2 kcal/mol (Tidor and Karplus 1993). Simply introducing cysteine might not produce an increase in stability. Disulfide by Design 2.0 (Craig and Dombkowski 2013) is a web-based tool to discover possible residue pairs that can be mutated to cysteines. It also provides with visualization tools and takes structural mobility into consideration.

## Calculation methods based on static systems

In a broader sense, this section and the next section "dynamics simulations" also belong to structural bioinformatics. These methods in these sections dig deeper structural information of proteins with the underlying structural properties such as force field. In this article, they were subdivided into two strategies based on dynamic (see the next section) or static system calculation. The calculations based on static system are relatively less computationally expensive and are commonly used for mutant property prediction and molecular docking.

### Predicting free energy changes upon mutations

Food industry requires thermostable enzymes for production because the high reaction temperature can accelerate the

reactions and reduce contamination. The folding (binding) free energy changes caused by the mutations are significant characteristics related to changes of mutant thermostability (ligand affinity). A lot of computational methods have been proposed to predict the changes (ΔΔG) in the folding free energy (ΔG) introduced by mutations (Fig. 3). Each protocol calculates the energy function of ΔG or ΔΔG with different energetic terms, including physical-based potential terms, empirical-based potential items, or a hybrid of them (Table 1). Beyond that, the machine-learning method is a powerful technology to combine the weighted structural, sequence, and dynamical features to predict the effect of point mutations or even multiple-point mutations. However, most of the methods were trained to reproduce experimental results from ProTherm database (Kumar et al. 2006) and the theoretical predictions are usually biased toward the training database (Musil et al. 2019). No one method significantly outperforms the others. Some of the tools obtains a reasonable accuracy at very low computational cost (Khan and Vihinen 2010; Potapov, Cohen, and Schreiber 2009). An attempt of increasing the diversity of conformations features to the Rosetta protocol showed that the ΔΔG prediction was not sensitive to conformational sampling (Kellogg, Leaver-Fay, and Baker 2011). Protocols involving more structural information might be a solution to increasing the accuracy of ΔΔG prediction (Quan, Lv, and Zhang 2016).

### Molecular docking

Molecular docking explores the binding modes of a target protein with some molecules. Many molecular docking methods have been developed over the last two decades, such as GOLD (Jones et al. 1997), Glide (Friesner et al. 2004), MOE-Dock (Corbeil, Williams, and Labute 2012), AutoDock Vina (Trott and Olson 2010), and AutoDock (Morris et al. 2009). All the methods that we introduce belong to flexible docking, in which side chains of the receptor are flexible. This allows the receptor to alter its binding site after the ligand bounds to the protein. AutoDock Vina might not be the best, but it is one of the top-ranking methods (Pagadala, Syed, and Tuszynski 2017). It is user-friendly and extremely faster than AutoDock (Trott and Olson 2010). GOLD showed the highest prediction accuracy among 10 docking tools in a recent comprehensive evaluation (Wang et al. 2016). Absolute energies are not estimated with satisfactory accuracy by the current molecular docking methods. However, these methods can discriminate among different small molecules based on binding affinity. A big limitation of the available docking methods is the inadequate sampling of the receptor. The backbone of the receptor is still rigid in most of the docking protocols. The dynamics simulation methods might be the best solution to this problem.

Although the molecular docking is less accurate and rigorous than MD simulation to study the interaction between receptor and ligand, it can be treated as a simplified form of MD simulation and the computational cost is much less than MD simulation. The molecular docking is mainly used in studies of drug design and enzyme inhibitors. This tool can provide starting structures with different conformations for MD simulation (Zheng et al. 2019). Molecular docking also gives insight into how the substrates bind to the enzymes, which can inspire the rational design of the binding sites.

## Calculation methods based on dynamics simulations

Based on the "induced fit" hypothesis, the enzyme will undergo a conformational change once the ligand is coupled to the active site (Ramakers et al. 2017). Thus, this dynamic process can be better studied by dynamics simulations. Because of the high computational costs, the dynamics simulation-based methods for protein engineering are generally more of methods to rationalize or to understand the roles of target residues in reaction mechanisms than tools to identify mutational hotspots. However, the increase of computational resources and the development of new computational technologies, such as parallel computing and graphics processing units (GPU) acceleration, make long-time simulations more accessible to scientists. Two strategies with different levels of resolution used to describe the atomic interactions are discussed in this section.

### Molecular dynamics simulations

The classical atomic MD simulations employ Newton's laws to evaluate the motions of molecules at the molecular level. The system of interest is described by a force field expressed by predefined functional form instead of solving the Schrödinger equation. The MD simulation generates a time series of conformations of the protein and its environment. Base on the ergodic hypothesis, macroscopic properties of a system can be calculated through enough sampling of phase space of in a fixed duration (Tuckerman and Martyna 2000). Because of the expansion of computational resources and the relatively lower theory level of MD simulation to describe atomic interactions, the timescale of MD simulation can be up to milliseconds, which covers the timescale of many protein events such as ligand recognition, catalysis, even the folding and unfolding of some proteins. The coarse-grained molecular simulation is developed to reduce the degrees of freedom and interaction details. Such strategy can increase the simulated time by orders of magnitude (Jamroz, Kolinski, and Kmiecik 2013; Jamroz, Orozco, et al. 2013; Tozzini 2005). Besides increasing the simulation time, there are some other methods to enhance the conformational state sampling to study biomolecular dynamics. A class of methods is based on replica exchange. A ladder of replicas with different temperatures (Marinari and Parisi 1992) or modified Hamiltonians (Fukunishi, Watanabe, and Takada 2002) is simulated in these methods. Another class of methods is based on a-priori importance sampling, such as umbrella sampling (Torrie and Valleau 1977) and metadynamics (Laio and Parrinello 2002). The third class of methods is achieved by modifying the Hamiltonian or adding a bias potential, including hyperdynamic (Voter

1997) and accelerated molecular dynamics (Hamelberg, Mongan, and McCammon 2004).

Many software packages have been developed for MD simulation and research. Amber (Salomon-Ferrer, Case, and Walker 2013), Gromacs (Van Der Spoel et al. 2005), NAMD (Phillips et al. 2005), and CHARMM (Brooks et al. 2009) are popular packages. Amber and CHARMM also refer to their own set of molecular mechanical force fields. Most of the force fields including Amber, CHARMM, GROMOS (Schmid et al. 2011), and OPLS-AA (Jorgensen, Maxwell, and Tirado-Rives 1996) can be used in different software programs for MD simulations. The information from MD simulation is mostly stored in trajectory files. Some general but important analyses are root mean square deviation (RMSD), root mean square fluctuations (RMSF), radius of gyration (Rg), and conformational analysis. The chemical and conformational free energy calculations are the most attractive applications using MD simulation to study thermodynamic and kinetic properties of biochemical catalysis. There are several methods to calculate the free energy difference using MD simulations, ranging from high to low prediction accuracy. And the rigorous methods with higher accuracy, such as free energy perturbation (Jorgensen and Thomas 2008), umbrella sampling (Kästner, 2011), and thermodynamic integration (Kästner and Thiel 2005; Mitchell and McCammon 1991), are more time-consuming. Molecular Mechanics/Surface Area methods (Massova and Kollman 2000) are of high speed and reasonable reproductivity.

MD simulation provides atomic details of interactions that contribute to protein stability or protein function, which can guide the rational design of proteins. MD simulation can also give insight into dynamic information unobtainable from static native structures. Moreover, MD simulation can be used to evaluate the protein property or protein designs (Kiss, Pande, and Houk 2013; Xiao et al. 2019). For more details of the MD simulation, we refer the readers to the references (Childers and Daggett 2017; Frenkel and Smit 2001; Likhachev, Balabaev, and Galzitskaya 2016; Zheng et al. 2019). The manuals and tutorials of each MD simulation software package are also helpful for a quick start.

### *Quantum mechanics and hybrid calculations*

Quantum mechanics (QM) calculations including *ab initio* methods and semi-empirical methods can provide useful information about the behavior of molecules. However, the cost of QM-based calculation increases exponentially when the size of the system increases. Based on the current computational resources, it is nearly impossible to simulate the whole bioprocess systems at quantum mechanics level. Thus, only a few active-site residues and the ligand will be considered in QM-based applications (Hotta et al. 2012; Lind and Himo 2013; Siegbahn and Himo 2009).

Quantum mechanics calculations together with MD simulations can be used to study biomolecular systems. Because the classic MD simulation cannot model the process of chemical bond breaking and reactions explicitly, the strategy based on both quantum mechanics and molecular mechanics (QM/MM) is widely used to study enzymatic mechanisms including bond cleavage or bond forming reactions. The active site region will be studied at a high level of theory, and the rest of the system will be calculated using molecular-mechanics (MM) method. The energy or Hamiltonian of MM system is calculated using potential energy functions, namely force field. The molecular mechanics cannot describe electronic changes but is computationally more efficient. Different levels of QM theory can be used to describe the interaction of the key region. And the MM part can be treated as explicit all-atom models or coarse-grained models (see below).

Empirical valence bond (EVB) approach is unique as a QM/MM strategy because EVB theory uses fully classical force fields to describe QM region. In this approach, each valence bond (VB) state represents the bonding patterns of key energy minima along the reaction coordinate. The free energy profile of the studied reaction can be obtained from a linear combination of these different diabatic states. It describes bond forming and breaking for the study of biocatalytic processes.

## Application examples of computational protein design for food enzymes

Enzymes have a wide range of applications in the food industry. There is a huge inclination toward tailoring enzymes to make them apt for applications in food industry. Computational guidance is a potent strategy to rationalize the protein design. This section demonstrates some successful applications of the computational method in food enzymes.

### *Thermostability design of food enzymes*

The stability design of enzymes is more predictable and better studied than the function design. The stability of protein is governed by the overall structure. A single mutation can make a conformational change to the whole enzyme. The calculation of the stability design should take the mutations in entire sequence space into consideration. Therefore, the bioinformatics strategy and the calculation methods based on static system are commonly used in stability design of enzymes because of their smaller amounts of computations (Fig. 2). The basic idea of thermostability design is to increase the energy gap between the natively folded state and lowest-energy misfolded state. Specifically, the main method is to enhance the hydrophobic energy by increasing the numbers of hydrogen bonds, disulfide bonds and the interactions that can stabilize the flexible regions.

α-Amylase (EC 3.2.1.1) is an important industrial enzyme in baking industry. Deng, Yang, Shin, et al. (2014) employed structure-based rational design methods to increase the thermostability of the α-amylase from *Alkalimonas amylolytica*. Seven residues on the enzyme surface were chosen as the targets to introduce arginine residues. Five of seven computational designed single mutants exhibited enhanced

thermostability. Another application of rational design with this enzyme was conducted by the introduction of disulfide bridges in the catalytic domain. The homology modeling of this enzyme was built by the Swiss-Model server. The Disulfide by Design algorithm was then used to predict the possible residues for introducing disulfide bridges. The predicted mutants were selected by structure visualizing tool and showed significantly improved thermostability (Liu et al. 2014). This enzyme has also been engineered by predicting free energy changes upon mutations in another study. PoPMuSiC program was used to predict the stabilizing mutants. The mechanisms responsible for the enhancement of α-amylase was discussed by the analysis of structure models in that study (Deng, Yang, Li, et al. 2014).

Xylanase (EC 3.2.1.8) is of great interest in food industry as it can degrade one of the most abundant bioresources xylan into xylose. The flexible surface residues are possible targets to design thermostable enzymes with no reduction in catalytic activities. FRODA dynamics simulation can also predict flexible region without heavy simulation (Thorpe 2005). Joo et al. (2010) used FRODA dynamics simulation (Thorpe 2005) and RosettaDesign algorithm (Liu and Kuhlman 2006) to engineer the xylanase from *Bacillus circulans*. Three thermostable single mutants were successfully identified by this strategy. The combination of different methods might increase the chance of successful prediction for mutational hotspots. A combined method (Wijma et al. 2014) was employed to engineer xylanase for industrial applications (Bu et al. 2018). Tools including FoldX (Schymkowitz et al. 2005) and Rosetta_ddg (Kellogg, Leaver-Fay, and Baker 2011) were used for energy calculations in that workflow.

Lipases (EC: 3.1.1.3) are ubiquitous enzymes that catalyze the hydrolysis and synthesis of lipids. The enzymes perform essential roles in producing esters from glycerol and fatty acids. The commercial lipases are usually from microbial sources (Aravindan, Anbumathi, and Viruthagiri 2007). Because of the poor thermostability to withstand higher temperatures, the commercial lipases have been widely studied to improve the stability by computational designs. Kim, Quang, and Kim (2010) combined the information of B-factor value from X-ray data and RosettaDesign method to increase in thermal stability of lipases from *Candida antarctica*. The strategy of RosettaDesign is to increase the buried hydrophobic surface area of the target protein. Two of seven mutants showed the improved thermostabilities after experimental validation. Through energy calculation by Rosetta ddg_monomer (Kellogg, Leaver-Fay, and Baker 2011), FoldX (Schymkowitz et al. 2005), and I-Mutant (Capriotti, Fariselli, and Casadio 2005), 60 possible mutation sites on the lipase from *Rhizomucor miehei* were identified by Li et al. (2018). After the exclusion of the residues near the active site, a limited screening library containing 37 mutations were selected to be examined. And 24 out of 37 candidates showed improved thermostabilities. In that study, they also increased the thermostability of lipase by introducing disulfide bonds. It is noticeable that the introduced disulfide bonds also improved the activity of the lipase.

### Function design of food enzymes

The rational design of proteins with desired functional properties is still an elusive goal. The strategy of rational protein design is to identify the optimal residues for mutagenesis. The selection of the residues usually requires a full understanding of the structural information of the target protein, which varies from case to case. A common method for the modification of substrate specificity is to mutate the binding sites of the target protein without changing the active sites. Structural visualization and dynamic simulation are useful for identifying the binding site of proteins.

Cellobiose 2-epimerase (EC 5.1.3.11) can catalyze the bioconversion of lactose to functional sweeteners and can be applied in the dairy industry (Chen et al. 2018). Cellobiose 2-epimerase from *Caldicellulosiruptor saccharolyticus* has been rationally modified through the redesign of the substrate binding site. The residues involved in interacting with substrates were chosen as target for mutagenesis. The engineered enzyme shifted its product preference toward lactulose (Park et al. 2017).

Hydantoinase can be used for the commercial production of food additives L-amino acids. The amino acids near the pocket of *Bacillus stearothermophilus* hydantoinase was visualized by Pymol, and the size and hydrophobicity of the predicted crucial amino acids were rationally changed. The rationale used in the design was supported by the computational modeling (Lee et al. 2009).

The oxidation of arylamines by laccases is of high industrial interest. QM/MM showed the atomic- and electronic-level detail on substrate binding and electron transfer of a laccase. The rational design protocol, including protein ligand recognition, optimization of donor-acceptor distance, and optimization of SASA, directed the design of mutation, two point mutations were predicted and showed two-fold $k_{cat}$ increase for the oxidation reaction (Santiago et al. 2016).

### Understanding the mechanism of enzyme

The dynamics properties of enzymes are usually studied using MD simulations. The role of a subdomain facing the lid on conformational rearrangements of lipase from *Burkholderia cepacia* was investigated using MD simulations. The lid switches from an open to a closed conformation in MD simulation, which helps to understand the key structural features gearing the closing/opening of the lid (Barbe et al. 2009). The relationship between thermostability and activity is frequently mentioned in thermostability engineering. It is controlled by underlying conformational flexibility. The tradeoffs between stability and function are often observed in the mutagenesis experiments. A lipase mutant with both increased stability and activity was characterized and explained by MD simulation analysis. The MD simulation results demonstrated that catalytically competent active site geometries were more frequently discovered in the mutant conformational space than that of the wild type (Kamal et al. 2012). This investigation shows the potential power of MD simulations in enzyme dynamic motions studies.

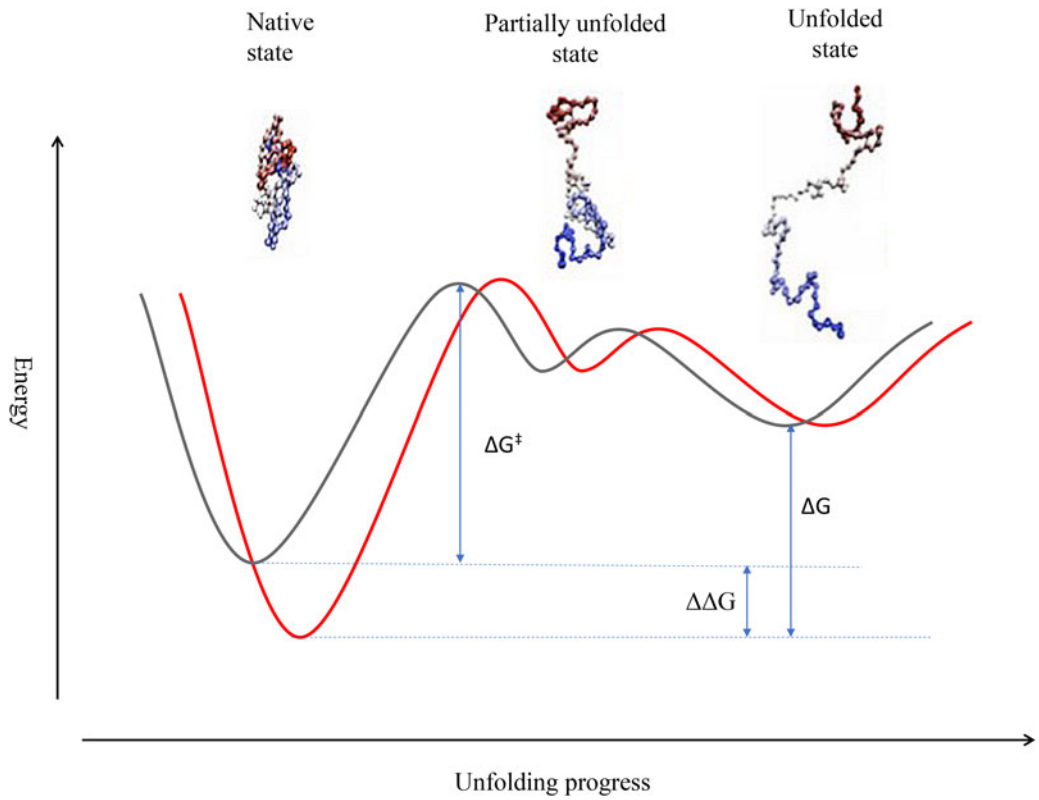


**Figure 3.** Schematic free energy diagram for describing the thermostability change upon a mutation. The free energy is plotted as a function of a generalized unfolding coordinate, which is a one-dimensional abstraction of the unfolding progress of the wild type (*black*) and the mutant (*red*). The ΔΔG was demonstrated by aligning the free energy of the mutant at the unfolding state with that of the wild type. ΔG‡ denotes the free energy barrier of unfolding, which relates to kinetic stability.

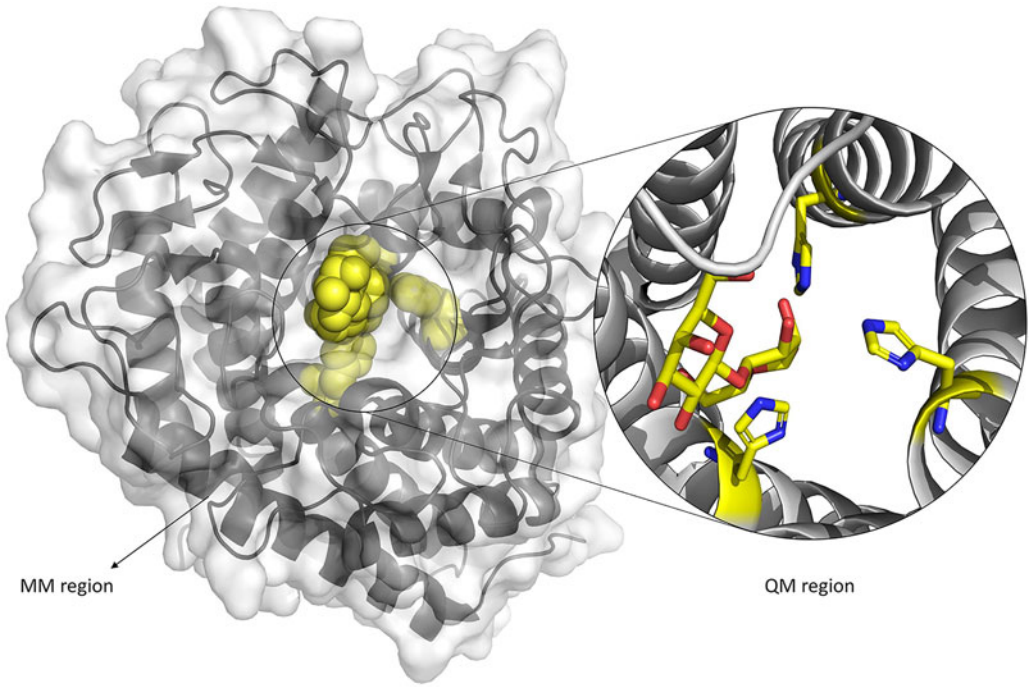


**Figure 4.** Representation of the QM/MM treatment of a cellobiose 2-epimerase with its substrate (epilactose). QM region is shown in yellow and MM region is shown in gray.

MD simulations have been widely used to study the dynamics properties of α-amylases (Sefidbakht, Ranaei Siadat, and Taheri 2017) and their inhibitors (Doruker, Atilgan, and Bahar 2000; Kato-Schwartz et al. 2018; Liao et al. 2018). The formation of the reaction transition state during α-amylase-based catalysis accompanies a movement

of a loop near the catalytic site. The free energy calculations by QM/MM simulation showed that the reorganization of this movement significantly helps to reduce the free energy barrier (Kosugi and Hayashi 2012).

A QM/MM method combined with umbrella sampling was used to gain insights into the epimerization activities of cellobiose 2-epimerase. The QM and MM regions chosen by that study are reproduced and demonstrated in Fig. 4. The free energy calculation help to better understand the detailed epimerization mechanism of cellobiose 2-epimerase (Zhang et al. 2015).

The sequential and structural information was used to guide the rational design of a GH11 xylanase. The residues with high B-factor predicted by Hotspot Wizard were chosen as mutagenesis targets and then analyzed by MD simulation. The substituted site formed a sandwich structure which might enhance the rigidity of the mutated xylanase (Wang et al. 2017). MD simulations can rationalize the mutation results of a xylanase mutant by analyzing the RMSD differences between mutant and the wild-type xylanase (Joo et al. 2011).

The ancestral sequence reconstruction method can help to understand natural evolution processes, which can provide information for protein design or for seeking high-activity enzymes. Lactate dehydrogenases are involved in food and beverage fermentations. Malate and lactate dehydrogenases are homologous. Using ancestral sequence reconstruction, Steindel et al. (2016) identified the biochemical and evolutionary mechanisms of malate and lactate dehydrogenases. They found the specificity in lactate dehydrogenases involves epistasis and is determined by more than just the charge of the specificity residue.

## Conclusions and outlook

In this review, we provide an introduction of the computational tools for protein design and avoid discussing the developer-specific algorithm in the method. Our desire is to show the nonexperts the power of advanced technologies in this field and to encourage them to apply these methods to food enzyme engineering. The computational approaches to help enzyme design are promising but challenging. The experimental approaches are also important and can confirm or reject computational predictions in turn. The hybrid methods combining two or more computational tools have been proven effective in improving protein design. The machine learning method which can deal with massive data sets will play an increasingly important role in the computational protein design. Although *de novo* design strategy is still far from ideal currently, it is the most exciting field for the future protein design, including food enzyme design.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the National Natural Science Foundation of China (No. 31922073 and 31801583), the Natural Science Foundation of Jiangsu Province (No. BK20180607), and the Postgraduate Research & Practice Innovation Program of Jiangsu Provence (KYCX17_1406).

## References

Akanuma, S. 2017. Characterization of reconstructed ancestral proteins suggests a change in temperature of the ancient biosphere. *Life (Basel)* 7 (3):33. doi: 10.3390/life7030033.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215 (3):403–10. doi: 10.1016/S0022-2836(05)80360-2.

Aravindan, R., P. Anbumathi, and T. Viruthagiri. 2007. Lipase applications in food industry. *Indian Journal of Biotechnology* 6:141–58.

Barbe, S., V. Lafaquiere, D. Guieysse, P. Monsan, M. Remaud-Simeon, and I. Andre. 2009. Insights into lid movements of *Burkholderia cepacia* lipase inferred from molecular dynamics simulations. Proteins: *Structure, Function, and Bioinformatics* 77 (3):509–23. doi: 10.1002/prot.22462.

Berman, H. M. 2000. The protein data bank. *Nucleic Acids Research* 28 (1):235–42. doi: 10.1093/nar/28.1.235.

Boucher, J. I., J. R. Jacobowitz, B. C. Beckett, S. Classen, and D. L. Theobald. 2014. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. *Elife* 3: e02304. doi: 10.7554/eLife.02304.

Brooks, B. R., C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al. 2009. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* 30 (10):1545–614. doi: 10.1002/jcc.21287.

Bu, Y., Y. Cui, Y. Peng, M. Hu, Y. Tian, Y. Tao, and B. Wu. 2018. Engineering improved thermostability of the GH11 xylanase from *Neocallimastix patriciarum* via computational library design. *Applied Microbiology and Biotechnology* 102 (8):3675–85. doi: 10.1007/s00253-018-8872-1.

Capriotti, E., P. Fariselli, and R. Casadio. 2005. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research* 33 (Web Server):W306–W310. doi: 10.1093/nar/gki375.

Chen, Q. M., Y. Q. Xiao, W. L. Zhang, T. Zhang, B. Jiang, T. Stressler, L. Fischer, and W. M. Mu. 2018. Current research on cellobiose 2-epimerase: Enzymatic properties, mechanistic insights, and potential applications in the dairy industry. *Trends in Food Science & Technology* 82:167–76. doi: 10.1016/j.tifs.2018.09.009.

Childers, M. C., and V. Daggett. 2017. Insights from molecular dynamics simulations for computational protein design. Molecular *Systems Design & Engineering* 2 (1):9–33. doi: 10.1039/C6ME00083E.

Corbeil, C. R., C. I. Williams, and P. Labute. 2012. Variability in docking success rates due to dataset preparation. *Journal of Computer-Aided Molecular Design* 26 (6):775–86. doi: 10.1007/s10822-012-9570-1.

Craig, D. B., and A. A. Dombkowski. 2013. Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* 14 (1):346. doi: 10.1186/1471-2105-14-346.

de Brevern, A. G., A. Bornot, P. Craveur, C. Etchebest, and J. C. Gelly. 2012. PredyFlexy: Flexibility and local structure prediction from sequence. *Nucleic Acids Research* 40 (W1):W317–322. (Web Server issue), doi: 10.1093/nar/gks482.

Deng, Z., H. Yang, H. D. Shin, J. Li, and L. Liu. 2014. Structure-based rational design and introduction of arginines on the surface of an alkaline alpha-amylase from Alkalimonas amylolytica for improved thermostability. *Applied Microbiology and Biotechnology* 98 (21): 8937–45. doi: 10.1007/s00253-014-5790-8.

Deng, Z., H. Yang, J. Li, H. D. Shin, G. Du, L. Liu, and J. Chen. 2014a. Structure-based engineering of alkaline alpha-amylase from

alkaliphilic Alkalimonas amylolytica for improved thermostability. *Applied Microbiology and Biotechnology* 98 (9):3997–4007. doi: 10.1007/s00253-013-5375-y.

Doruker, P., A. R. Atilgan, and I. Bahar. 2000. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to alpha-amylase inhibitor. Proteins: *Structure, Function, and Genetics* 40 (3):512–24. doi: 10.1002/1097-0134(20000815)40:3<512::aid-prot180>3.0.co;2-m.

Ebert, M. C., and J. N. Pelletier. 2017. Computational tools for enzyme improvement: Why everyone can – and should – use them. *Current Opinion in Chemical Biology* 37:89–96. doi: 10.1016/j.cbpa.2017.01.021.

Fariselli, P., P. L. Martelli, C. Savojardo, and R. Casadio. 2015. INPS: Predicting the impact of non-synonymous variations on protein stability from sequence. *Bioinformatics* 31 (17):2816–21. doi: 10.1093/bioinformatics/btv291.

Frenkel, D., and B. Smit. 2001. *Understanding molecular simulation: From algorithms to applications[M]*. vol. 1. Netherlands: Elsevier.

Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, et al. 2004. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* 47 (7):1739–49. doi: 10.1021/jm0306430.

Fukunishi, H., O. Watanabe, and S. Takada. 2002. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics* 116 (20):9058–67. doi: 10.1063/1.1472510.

Haas, J., A. Barbato, D. Behringer, G. Studer, S. Roth, M. Bertoni, K. Mostaguir, R. Gumienny, and T. Schwede. 2018. Continuous automated model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12. Proteins: *Structure, Function, and Bioinformatics* 86 (Suppl 1):387–98. doi: 10.1002/prot.25431.

Hamelberg, D., J. Mongan, and J. A. McCammon. 2004. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics* 120 (24):11919–29. doi: 10.1063/1.1755656.

Hanson-Smith, V., and A. Johnson. 2016. PhyloBot: A web portal for automated phylogenetics, ancestral sequence reconstruction, and exploration of mutational trajectories. *PLoS Computational Biology* 12 (7):e1004976. doi: 10.1371/journal.pcbi.1004976.

Hotta, K., X. Chen, R. S. Paton, A. Minami, H. Li, K. Swaminathan, I. I. Mathews, K. Watanabe, H. Oikawa, K. N. Houk, and C.-Y. Kim. 2012. Enzymatic catalysis of anti-Baldwin ring closure in polyether biosynthesis. *Nature* 483 (7389):355–8. doi: 10.1038/nature10865.

Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14 (1):33–8, 27–38. doi: 10.1016/0263-7855(96)00018-5.

Jackel, C., J. D. Bloom, P. Kast, F. H. Arnold, and D. Hilvert. 2010. Consensus protein design without phylogenetic bias. *Journal of Molecular Biology* 399 (4):541–6. doi: 10.1016/j.jmb.2010.04.039.

Jamroz, M., A. Kolinski, and S. Kmiecik. 2013. CABS-flex: Server for fast simulation of protein structure fluctuations. *Nucleic Acids Research* 41 (W1):W427–431. doi: 10.1093/nar/gkt332.

Jamroz, M., M. Orozco, A. Kolinski, and S. Kmiecik. 2013. Consistent view of protein fluctuations from all-atom molecular dynamics and coarse-grained dynamics with knowledge-based force-field. *Journal of Chemical Theory and Computation* 9 (1):119–25. doi: 10.1021/ct300854w.

Jones, G., P. Willett, R. C. Glen, A. R. Leach, and R. Taylor. 1997. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* 267 (3):727–48. doi: 10.1006/jmbi.1996.0897.

Joo, J. C., S. P. Pack, Y. H. Kim, and Y. J. Yoo. 2011. Thermostabilization of Bacillus circulans xylanase: Computational optimization of unstable residues based on thermal fluctuation analysis. *Journal of Biotechnology* 151 (1):56–65. doi: 10.1016/j.jbiotec.2010.10.002.

Joo, J. C., S. Pohkrel, S. P. Pack, and Y. J. Yoo. 2010. Thermostabilization of Bacillus circulans xylanase via computational design of a flexible surface cavity. *Journal of Biotechnology* 146 (1–2):31–9. doi: 10.1016/j.jbiotec.2009.12.021.

Jorgensen, W. L., and L. L. Thomas. 2008. Perspective on free-energy perturbation calculations for chemical equilibria. *Journal of Chemical Theory and Computation* 4 (6):869–76. doi: 10.1021/ct800011m.

Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* 118 (45):11225–36. doi: 10.1021/ja9621760.

Kamal, M. Z., T. A. S. Mohammad, G. Krishnamoorthy, and N. M. Rao. 2012. Role of active site rigidity in activity: MD simulation and fluorescence study on a lipase mutant. *PLoS One* 7 (4):e35188. doi: 10.1371/journal.pone.0035188.

Kästner, J. 2011. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1 (6):932–42. doi: 10.1002/wcms.66.

Kästner, J., and W. Thiel. 2005. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration. *The Journal of Chemical Physics* 123 (14):144104. doi: 10.1063/1.2052648.

Kato-Schwartz, C. G., F. Bracht, G. A. Gonçalves, A. A. Soares, T. F. Vieira, T. Brugnari, A. Bracht, and R. M. Peralta. 2018. Inhibition of α-amylases by pentagalloyl glucose: Kinetics, molecular dynamics and consequences for starch absorption. *Journal of Functional Foods* 44:265–73. doi: 10.1016/j.jff.2018.03.025.

Kellogg, E. H., A. Leaver-Fay, and D. Baker. 2011. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins: Structure, Function, and Bioinformatics* 79 (3):830–8. doi: 10.1002/prot.22921.

Khan, S., and M. Vihinen. 2010. Performance of protein stability predictors. *Human Mutation* 31 (6):675–84. doi: 10.1002/humu.21242.

Kim, H. S., A. T. L. Quang, and Y. H. Kim. 2010. Development of thermostable lipase B from *Candida antarctica* (CalB) through in silico design employing B-factor and RosettaDesign. *Enzyme and Microbial Technology* 47 (1–2):1–5. doi: 10.1016/j.enzmictec.2010.04.003.

Kiss, G., V. S. Pande, and K. N. Houk. 2013. Molecular dynamics simulations for the ranking, evaluation, and refinement of computationally designed proteins. *Methods in Enzymology* 523:145–70. doi: 10.1016/B978-0-12-394292-0.00007-2.

Kosugi, T., and S. Hayashi. 2011. QM/MM reweighting free energy SCF for geometry optimization on extensive free energy surface of enzymatic reaction. *Journal of Chemical Theory and Computation* 8 (1):322–34.

Kosugi, T., and S. Hayashi. 2012. Crucial role of protein flexibility in formation of a stable reaction transition state in an alpha-amylase catalysis. *Journal of the American Chemical Society* 134 (16):7045–55. doi: 10.1021/ja212117m.

Kumar, M. D., K. A. Bava, M. M. Gromiha, P. Prabakaran, K. Kitajima, H. Uedaira, and A. Sarai. 2006. ProTherm and ProNIT: Thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Research* 34 (90001):D204–206. doi: 10.1093/nar/gkj103.

Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* 99 (20):12562–6. doi: 10.1073/pnas.202427399.

Le, Q. A., J. C. Joo, Y. J. Yoo, and Y. H. Kim. 2012. Development of thermostable *Candida antarctica* lipase B through novel in silico design of disulfide bridge. *Biotechnology and Bioengineering* 109 (4):867–76. doi: 10.1002/bit.24371.

Lee, S.-C., Y. J. Chang, D.-M. Shin, J. Han, M.-H. Seo, H. Fazelinia, C. D. Maranas, and H.-S. Kim. 2009. Designing the substrate specificity of d-hydantoinase using a rational approach. *Enzyme and Microbial Technology* 44 (3):170–5. doi: 10.1016/j.enzmictec.2008.10.020.

Li, G., X. Fang, F. Su, Y. Chen, L. Xu, and Y. Yan. 2018. Enhancing the thermostability of *Rhizomucor miehei* lipase with a limited screening library by rational-design point mutations and disulfide bonds. *Applied and Environmental Microbiology* 84 (2):e02129–17. doi: 10.1128/AEM.02129-17.

Liao, S. M., N. K. Shen, G. Liang, B. Lu, Z. L. Lu, L. X. Peng, F. Zhou, L. Q. Du, Y. T. Wei, G. P. Zhou, R. B. Huang. 2018. Inhibition of alpha-amylase activity by Zn2+: Insights from spectroscopy and molecular dynamics simulations. *Medicinal Chemistry* 15 (5):510–20. doi: 10.2174/1573406415666181217114101.

Likhachev, I. V., N. K. Balabaev, and O. V. Galzitskaya. 2016. Available instruments for analyzing molecular dynamics trajectories. The *Open Biochemistry Journal* 10 (1):1–11. doi: 10.2174/1874091X01610010001.

Lind, M. E., and F. Himo. 2013. Quantum chemistry as a tool in asymmetric biocatalysis: Limonene epoxide hydrolase test case. *Angewandte Chemie (International Edition)* 52 (17):4563–7. doi: 10.1002/anie.201300594.

Liu, L., Z. Deng, H. Yang, J. Li, H.-D. Shin, R. R. Chen, G. Du, and J. Chen. 2014. In silico rational design and systems engineering of disulfide bridges in the catalytic domain of an alkaline alpha-amylase from *Alkalimonas amylolytica* to improve thermostability. *Applied and Environmental Microbiology* 80 (3):798–807. doi: 10.1128/AEM.03045-13.

Liu, T., Y. Wang, X. Luo, J. Li, S. A. Reed, H. Xiao, T. S. Young, and P. G. Schultz. 2016. Enhancing protein stability with extended disulfide bonds. *Proceedings of the National Academy of Sciences* 113 (21):5910–5. doi: 10.1073/pnas.1605363113.

Liu, Y., and B. Kuhlman. 2006. RosettaDesign server for protein design. *Nucleic Acids Research* 34 (Web Server):W235–W238. doi: 10.1093/nar/gkl163.

Marinari, E., and G. Parisi. 1992. Simulated tempering: A new Monte Carlo scheme. *Europhysics Letters (EPL)* 19 (6):451–8. doi: 10.1209/0295-5075/19/6/002.

Massova, I., and P. A. Kollman. 2000. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in Drug Discovery and Design* 18 (1):113–35. doi: 10.1023/a:1008763014207.

Mitchell, M. J., and J. A. McCammon. 1991. Free energy difference calculations by thermodynamic integration: Difficulties in obtaining a precise value. *Journal of Computational Chemistry* 12 (2):271–5. doi: 10.1002/jcc.540120218.

Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* 30 (16):2785–91. doi: 10.1002/jcc.21256.

Musil, M., H. Konegger, J. Hon, D. Bednar, and J. Damborsky. 2019. Computational design of stable and soluble biocatalysts. *ACS Catalysis* 9 (2):1033–54. doi: 10.1021/acscatal.8b03613.

Pagadala, N. S., K. Syed, and J. Tuszynski. 2017. Software for molecular docking: A review. Biophysical *Reviews* 9 (2):91–102. doi: 10.1007/s12551-016-0247-1.

Pan, X. Y., and H. B. Shen. 2009. Robust prediction of B-factor profile from sequence using two-stage SVR based on random Forest feature selection. *Protein & Peptide Letters* 16 (12):1447–54. doi: 10.2174/092986609789839250.

Park, A. R., J. S. Kim, S. W. Jang, Y. G. Park, B. S. Koo, and H. C. Lee. 2017. Rational modification of substrate binding site by structure-based engineering of a cellobiose 2-epimerase in *Caldicellulosiruptor saccharolyticus*. *Microbial Cell Factories* 16 (1):224. doi: 10.1186/s12934-017-0841-3.

Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26 (16):1781–802. doi: 10.1002/jcc.20289.

Pina, A., and A. Roque. 2009. Studies on the molecular recognition between bioactive peptides and angiotensin-converting enzyme. *Journal of Molecular Recognition* 22 (2):162–8.

Porebski, B. T., and A. M. Buckle. 2016. Consensus protein design. *Protein Engineering Design and Selection* 29 (7):245–51. doi: 10.1093/protein/gzw015.

Potapov, V., M. Cohen, and G. Schreiber. 2009. Assessing computational methods for predicting protein stability upon mutation: Good on average but not in the details. *Protein Engineering Design and Selection* 22 (9):553–60. doi: 10.1093/protein/gzp030.

Quan, L., Q. Lv, and Y. Zhang. 2016. STRUM: Structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics* 32 (19):2936–46. doi: 10.1093/bioinformatics/btw361.

Ramakers, L. A. I., G. Hithell, J. J. May, G. M. Greetham, P. M. Donaldson, M. Towrie, A. W. Parker, G. A. Burley, and N. T. Hunt. 2017. 2D-IR spectroscopy shows that optimized DNA minor groove binding of Hoechst33258 follows an induced fit model. The *Journal of Physical Chemistry B* 121 (6):1295–303. doi: 10.1021/acs.jpcb.7b00345.

Reetz, M. T., J. D. Carballeira, and A. Vogel. 2006. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angewandte Chemie International Edition* 45 (46):7745–51. doi: 10.1002/anie.200602795.

Salomon-Ferrer, R., D. A. Case, and R. C. Walker. 2013. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3 (2):198–210. doi: 10.1002/wcms.1121.

Santiago, G., F. de Salas, M. F. Lucas, E. Monza, S. Acebes, Á. T. Martinez, S. Camarero, and V. Guallar. 2016. Computer-aided laccase engineering: Toward biological oxidation of arylamines. *ACS Catalysis* 6 (8):5415–23. doi: 10.1021/acscatal.6b01460.

Schaarschmidt, J., B. Monastyrskyy, A. Kryshtafovych, and A. Bonvin. 2018. Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age. Proteins: *Structure*, *Function*, *and Bioinformatics* 86 (Suppl 1):51–66. doi: 10.1002/prot.25407.

Schmid, N., A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren. 2011. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal* 40 (7):843–56. doi: 10.1007/s00249-011-0700-9.

Schrödinger, L. L. C. 2010. The PyMOL molecular graphics system. Retrieved from https://pymol.org/2/

Schymkowitz, J., J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano. 2005. The FoldX web server: An online force field. *Nucleic Acids Research* 33 (Web Server issue):W382–W388. doi: 10.1093/nar/gki387.

Sefidbakht, Y., O. Ranaei Siadat, and F. Taheri. 2017. Homology modeling and molecular dynamics study on *Schwanniomyces occidentalis* alpha-amylase. *Journal of Biomolecular Structure and Dynamics* 35 (3):574–84. doi: 10.1080/07391102.2016.1154892.

Siegbahn, P. E., and F. Himo. 2009. Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *Journal of Biological Inorganic Chemistry* 14 (5):643–51. doi: 10.1007/s00775-009-0511-y.

Sim, N. L., P. Kumar, J. Hu, S. Henikoff, G. Schneider, and P. C. Ng. 2012. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research* 40 (W1):W452–457. (Web Server issue), doi: 10.1093/nar/gks539.

Steindel, P. A., E. H. Chen, J. D. Wirth, and D. L. Theobald. 2016. Gradual neofunctionalization in the convergent evolution of trichomonad lactate and malate dehydrogenases. *Protein Science* 25 (7):1319–31. doi: 10.1002/pro.2904.

Sumbalova, L., J. Stourac, T. Martinek, D. Bednar, and J. Damborsky. 2018. HotSpot Wizard 3.0: Web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Research* 46 (W1):W356–W362. doi: 10.1093/nar/gky417.

Suplatov, D., V. Voevodin, and V. Švedas. 2015. Robust enzyme design: Bioinformatic tools for improved protein stability. *Biotechnology Journal* 10 (3):344–55. doi: 10.1002/biot.201400150.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and

weight matrix choice. *Nucleic Acids Research* 22 (22):4673–80. doi: 10.1093/nar/22.22.4673.

Thorpe, M. 2005. FRODA: Framework rigidity optimized dynamics algorithm.

Tidor, B., and M. Karplus. 1993. The contribution of cross-links to protein stability: A normal mode analysis of the configurational entropy of the native state. Proteins: *Structure*, *Function*, *and Genetics* 15 (1):71–9. doi: 10.1002/prot.340150109.

Torrie, G. M., and J. P. Valleau. 1977. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* 23 (2):187–99. doi: 10.1016/0021-9991(77)90121-8.

Tozzini, V. 2005. Coarse-grained models for proteins. *Current Opinion in Structural Biology* 15 (2):144–50. doi: 10.1016/j.sbi.2005.02.005.

Trott, O., and A. J. Olson. 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31 (2):455–61. doi: 10.1002/jcc.21334.

Trudeau, D. L., M. Kaltenbach, and D. S. Tawfik. 2016. On the potential origins of the high stability of reconstructed ancestral proteins. *Molecular Biology and Evolution* 33 (10):2633–41. doi: 10.1093/molbev/msw138.

Tuckerman, M. E., and G. J. Martyna. 2000. Understanding modern molecular dynamics: Techniques and applications. *The Journal of Physical Chemistry B* 104 (2):159–78.

Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen. 2005. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* 26 (16):1701–18. doi: 10.1002/jcc.20291.

Van Overtveldt, S., O. Gevaert, M. Cherlet, K. Beerens, and T. Desmet. 2018. Converting galactose into the rare sugar talose with cellobiose 2-epimerase as biocatalyst. *Molecules* 23 (10):2519. doi: 10.3390/molecules23102519.

Verma, R., U. Schwaneberg, and D. Roccatano. 2012. MAP (2.0) 3D: A sequence/structure based server for protein engineering. *ACS Synthetic Biology* 1 (4):139–50. doi: 10.1021/sb200019x.

Voter, A. F. 1997. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters* 78 (20):3908–11. doi: 10.1103/PhysRevLett.78.3908.

Wang, X., H. Huang, X. Xie, R. Ma, Y. Bai, F. Zheng, S. You, B. Zhang, H. Xie, B. Yao, and H. Luo. 2016. Improvement of the catalytic performance of a hyperthermostable GH10 xylanase from *Talaromyces leycettanus* JCM12802. *Bioresource Technology* 222: 277–84. doi: 10.1016/j.biortech.2016.10.003.

Wang, X., R. Ma, X. Xie, W. Liu, T. Tu, F. Zheng, S. You, J. Ge, H. Xie, B. Yao, and H. Luo. 2017. Thermostability improvement of a *Talaromyces leycettanus* xylanase by rational protein engineering. *Scientific Reports* 7 (1):15287. doi: 10.1038/s41598-017-12659-y.

Wang, Y., G. Zhang, J. Yan, and D. Gong. 2014. Inhibitory effect of morin on tyrosinase: Insights from spectroscopic and molecular docking studies. *Food Chemistry* 163:226–33. doi: 10.1016/j.foodchem.2014.04.106.

Wang, Z., H. Sun, X. Yao, D. Li, L. Xu, Y. Li, S. Tian, and T. Hou. 2016. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power. *Physical Chemistry Chemical Physics* 18 (18):12964–75. doi: 10.1039/C6CP01555G.

Wijma, H. J., R. J. Floor, P. A. Jekel, D. Baker, S. J. Marrink, and D. B. Janssen. 2014. Computationally designed libraries for rapid enzyme stabilization. *Protein Engineering Design and Selection* 27 (2):49–58. doi: 10.1093/protein/gzt061.

Xiao, Y., Q. Chen, E. I. Shakhnovich, W. Zhang, and W. Mu. 2019. Simulation-guided enzyme discovery: A new microbial source of cellobiose 2-epimerase. *International Journal of Biological Macromolecules* 139:1002. doi: 10.1016/j.ijbiomac.2019.08.075.

Yu, X. W., N. J. Tan, R. Xiao, and Y. Xu. 2012. Engineering a disulfide bond in the lid hinge region of *Rhizopus chinensis* lipase: Increased thermostability and altered acyl chain length specificity. *PLoS One* 7 (10):e46388. doi: 10.1371/journal.pone.0046388.

Zhang, W. L., S. H. Yu, T. Zhang, B. Jiang, and W. M. Mu. 2016. Recent advances in D-allulose: Physiological functionalities, applications, and biological production. *Trends in Food Science & Technology* 54:127–37. doi: 10.1016/j.tifs.2016.06.004.

Zhang, Y. L., Q. C. Zheng, J. L. Zhang, and H. X. Zhang. 2015. Insights into the epimerization activities of RaCE and pAGE: The quantum mechanics/molecular mechanics simulations. *RSC Advances* 5 (124):102284–93. doi: 10.1039/C5RA14091A.

Zheng, L., A. A. Alhossary, C.-K. Kwoh, and Y. Mu. 2019. Molecular dynamics and simulation. In *Encyclopedia of bioinformatics and computational biology*, ed. S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, 550–566. Oxford: Academic Press.