

Review

Cite this article: Wang M, Ma A, Wang H and Lou X (2024). Atomic molecular dynamics simulation advances of *de novo*-designed proteins. *Quarterly Reviews of Biophysics*, **57**, e14, 1–9
<https://doi.org/10.1017/S0033583524000131>

Received: 14 January 2024
Revised: 27 June 2024
Accepted: 10 September 2024


Keywords:

Advances; *de novo* design proteins; force field; molecular dynamics; sampling methods; stability; water models

Corresponding authors:

Hongjiang Wang and Xiaotong Lou;
Emails: wujiang_0706@163.com;
louxt@sina.com

Atomic molecular dynamics simulation advances of *de novo*-designed proteins

Moye Wang , Anqi Ma, Hongjiang Wang and Xiaotong Lou

Research Department, PLA Strategic Support Force Medical Center, Beijing, China

Abstract

Proteins are vital biological macromolecules that execute biological functions and form the core of synthetic biological systems. The history of *de novo* protein has evolved from initial successes in subordinate structural design to more intricate protein creation, challenging the complexities of natural proteins. Recent strides in protein design have leveraged computational methods to craft proteins for functions beyond their natural capabilities. Molecular dynamics (MD) simulations have emerged as a crucial tool for comprehending the structural and dynamic properties of *de novo*-designed proteins. In this study, we examined the pivotal role of MD simulations in elucidating the sampling methods, force field, water models, stability, and dynamics of *de novo*-designed proteins, highlighting their potential applications in diverse fields. The synergy between computational modeling and experimental validation continued to play a crucial role in the creation of novel proteins tailored for specific functions and applications.

Table of contents

Introduction	1
MD simulations of <i>de novo</i>-designed proteins	2
Sampling methods	2
Force field	3
Water models	3
MD Simulation to study protein thermal stability	4
MD Simulation to study designed protein dynamics	5
Conclusions and outlook	6

Introduction

Proteins are the main biological macromolecules that carry out biological functions and are also the major components for building synthetic biological systems. The function of proteins depended on their specific three-dimensional spatial structure and specific intermolecular interactions (Anfinsen *et al.*, 1961). The amino acid sequence determined the three-dimensional structure and interactions of the mass, thus determining the function of the protein (Zhang and Wang, 2023). The amino acid sequences of natural proteins have been selected over time by evolution and adapted to the functional needs of the corresponding organism (Marsh and Teichmann, 2015). The function of natural proteins might not be adequate for other required purposes, and even sometimes the available natural proteins cannot be found (Zhu *et al.*, 2021). Therefore, design proteins with the target properties and functions are a significant breakthrough in the history of biology. Protein folding mechanism studies focus on the thermodynamic state and dynamic process from sequence to structure. The aim of *de novo* design of protein is to obtain proteins with a target structure. The process from structure to sequence can be regarded as the reverse process of protein folding.

The concept of *de novo* protein design had a long history. Most of the early protein design successes were based on the design of subordinate secondary structures. In the 1980s, William DeGrado and his colleagues designed a tetrameric structure with a helical sequence connected by three loops. The overall design strategy was relatively simple, and the obtained protein was the $\alpha 4$ protein (Regan and DeGrado, 1988). Following this work, Hetch MH and his team reported their successful design of the four-helix Felix, in which more factors than the $\alpha 4$ protein were considered in the design process, such as hydrophobic interactions inside the helix, helical hydrogen bonding, etc. (Hecht *et al.*, 1990; Patel *et al.*, 2009). These works had great implications for the *de novo* design of protein and accumulated rich experience. Based on these works, scientists then challenged the more difficult *de novo* design proteins. In 1994, Blanco *et al.* reported a successful design of β -hairpin, a motif containing 9 amino acids, which was the first reported successful design of β -hairpin (Blanco *et al.*, 1994). In 1997, Mayo *et al.* reported their first design of a small protein full sequence design 1 (FSD-1), using computer-aided technology

methods (Dahiyat and Mayo, 1997). Another important development was that David Baker and his team designed a $\beta\beta\alpha\beta\beta\beta$ structure, named Top7, whose thermodynamic stability was high. The denaturation temperature reached 69 °C in a guanidine hydrochloride solution, and it still maintained structural stability at 90 °C without denaturant, and its stability was much higher than that of natural proteins of the same scale (Huang *et al.*, 2016).

Significant progresses have been made in *de novo* protein design, and protein scientists were studying the properties of these particular proteins to take advantage of them and provide functions that nature proteins do not have. For example, Correia *et al.* demonstrated that computational protein design could produce small, thermally and conformationally stable protein scaffolds with a neutralization epitope from respiratory syncytial virus, accurately mimicking the viral epitope structure and induce potent neutralizing antibodies (Correia *et al.*, 2014). Marcandalli *et al.* reported that custom-designed vaccines were achieved by designing protein self-assembled nanoparticles as a skeleton to fix and present viral glycoprotein antigen complexes (Marcandalli *et al.*, 2019). Sesterhenn *et al.* established the TopoBuilder system to design proteins that could stabilize complex structural motifs from scratch. Through this system, they obtained proteins that could present three antigens simultaneously (Sesterhenn *et al.*, 2020). Wannier *et al.* obtained red fluorescent proteins that maintained light intensity and were not easy to aggregate through surface mutation design and main-chain dependent side-chain rotamer sampling optimization (Wannier *et al.*, 2015). As we all know, protein functions are closely related to the conformational space and folding process. The commonly used experimental techniques for studying the mechanism of protein folding had restrictions because protein folding typically ranged from microsecond to milliseconds (Dill and MacCallum, 2012). Studying the dynamics of protein folding and conformational exploration on such a fast time scale was extremely challenging, especially for *de novo*-designed proteins whose properties might be entirely different from the natural proteins (Korendovych and DeGrado, 2020). Therefore, molecular modeling and simulation were important complements to experimental characterization for understanding the structural and dynamic properties of *de novo* design proteins at an atomic level (Wang, 2021).

Molecular dynamics (MD) simulation provided the most detailed description of the dynamics and structures of *de novo*-designed proteins (Gonzalez *et al.*, 2022; Leong *et al.*, 2018; Wang *et al.*, 2016). However, big challenges also existed in the MD simulation method and were similar to the experimental biophysical approaches. Due to the special nature of *de novo* design proteins, the folding time scale was longer than that of natural proteins. In this case, the sampling efficiency of classical all-atom simulation methods would be reduced, resulting in difficulties in the study of folding dynamics (Zhu *et al.*, 2013). The stability of designed proteins in the simulations is also an important factor. The choice and efficiency of the sampling method and force field are the prominent elements in the MD simulations. In this review, we will focus on the sampling method and force field, as well as the stability and dynamics of MD simulations of *de novo*-designed proteins.

MD simulations of *de novo*-designed proteins

Sampling methods

The classical methods of traditional MD simulations have been widely used for sampling conformation methods of peptides and large proteins (Bernardi *et al.*, 2015; Kamberaj, 2018; Lazim *et al.*,

2020; Wang, 2021). To enhance the sampling efficiency and observe the complicated protein activities, one might need enhanced sampling methods rather than the traditional method due to the rugged free energy surface, especially for *de novo*-designed proteins whose protein folding funnels were more challenging than those of natural protein (Dill and Chan, 1997; Leopold *et al.*, 1992). One of the most effective enhanced sampling methods used in the designed proteins was replica exchange molecular dynamics (REMD). REMD was a simulation method that combined the replica exchange method with the MD method. The basic idea was to construct a series of interacting replica systems, one for each temperature, covering a wide range of temperatures from low to high, and then performed a separate MD simulation for each replica at each temperature (Bernardi *et al.*, 2015; Li *et al.*, 2023). The Monte Carlo Metropolis rule controlled the exchange of replicas adjacent in temperature and ensured the detailed balance, and that the structure ensemble at temperature of interested could be generated (Wang, 2021; Zhou, 2007). At adjacent temperatures, every two copies can be exchanged according to the Metropolis standard, and based on this, the REMD method can make the low-temperature conformation space escape the local potential energy lowest point (Chen *et al.*, 2015; Sugita *et al.*, 2019). In practical applications, the minimum temperature of the replica (T_{\min}) was generally set as the temperature of most interest, such as room temperature (300 K), and the maximum temperature (T_{\max}) was the lowest temperature that can help to cross the energy barrier of the system. In order to obtain a better sampling effect, it was necessary to ensure that the total average exchange rate between the replicas was in the range of 0.2–0.5 and ensure that each analog copy can traverse all temperatures (Rosta and Hummer, 2009). The wide application of REMD has also played a significant role in the field of protein folding. Lazim *et al.* explored the mechanism directing the structure variation from $\alpha/4\beta$ -fold protein to 3α -fold protein after mutation by conducting REMD simulation on 42 replicas and provided insights into the folding and unfolding pathway of 3α -fold protein and $\alpha/4\beta$ -fold protein, respectively, paving the way toward the understanding of the ongoing conformational variation (Lazim *et al.*, 2012).

In recent years, there have been significant developments in the optimization of REMD methods, with many related articles being reported. For example, Bernardi *et al.* mentioned mainstream REMD deformation methods such as H-REMD and M-REMD in their 2015 summary of enhanced sampling algorithms (Ostermeier and Zacharias, 2013). Mateusz Marcisz *et al.* also docked 24- and 48-meric glycosaminoglycans using an all-atomic repulsive-scaling Hamiltonian replica exchange molecular dynamics (RS-REMD) method. This advanced methodology was based on replicas with scaled van der Waals radii of interacting molecules and has performed well for proteins complexed with oligomeric GAGs (Marcisz *et al.*, 2021). Danial Pourjafar-Dehkordi studied the flexibility of native and pS111-Rab1b in complex with GTP or GDP using extensive MD simulations and an advanced sampling method called dihedral angle-biasing potential replica exchange molecular dynamics (DIA-REMD), which promoted backbone and side chain dihedral transitions along a series of replica simulations in selected protein segments and through exchanges also improves sampling in an unbiased reference simulation (Pourjafar-Dehkordi and Zacharias, 2021). Thus, advanced REMD methods have made significant contributions to the simulation sampling of protein foldings, especially for proteins with rough potential energy surfaces.

In order to improve the sampling efficiency in protein simulations, there are many other sampling methods in addition to

REMD, such as metadynamics, enhanced sampling method like ossPT-MetaD, population particle MD and other efficient sampling methods.

Metadynamics was a class of methods that improved sampling efficiency by introducing additional biased potential energy or force acting on certain degrees of freedom. These degrees of freedom, often called Collective Variable (CV), were generally functions of points in state space that distinguish between two or more thermodynamic states that needed to be studied. The free energy landscape along the collective variable was a low-dimensional projection of the complete free energy landscape in the direction of the CV, containing all the thermodynamic information we cared about, and the solution of it could be accelerated by the metadynamics (Giberti *et al.*, 2015).

The ossPT-MetaD, as an enhanced sampling method, performed enhanced sampling in both energy and geometric space. By optimizing the sampling efficiency under REMD framework and combining the normal vibration mode as a set variable, the number of replicas could be reduced while maintaining a high exchange rate and simulation convergence. This significantly reduced the degree of freedom of the system and the number of copies required, improved the simulation convergence, and demonstrated high efficiency and rationality in a variety of protein systems (Peng *et al.*, 2021).

The swarm population particle molecular dynamics (SPMD) was an enhanced conformation sampling method for computer simulation of protein folding, particularly suitable for use in combination with replica exchange methods (REM). In the framework of REM, the sampling efficiency was enhanced by introducing the concept of swarm intelligence. In SPMD, individual copies were distinguished not only by temperature but also by particle swarm optimization for efficient search and exchange in the conformation space. In the framework of REM, the sampling efficiency is enhanced by introducing the concept of swarm intelligence. In summary, SPMD was an advanced computational method that significantly improved the efficiency of MD simulations through enhanced sampling techniques, especially in protein folding and conformation space exploration. This method provided a powerful tool for efficient exploration and prediction of protein structure and contributed to a deep understanding of protein dynamic behavior and function (Kamberaj, 2018).

Force field

The molecular mechanics method, also known as molecular force field, is the application of classical physics to explain the microstructure and macroscopic properties of atoms and molecules, and it has become an irreplaceable research tool in scientific exploration. The accuracy of the molecular force field was essential for describing both bonded and nonbond interactions in proteins (Ding *et al.*, 2023; Latour, 2014; Lopes *et al.*, 2015). The energy of the whole molecule was the sum of the bond stretching energy, bending energy, twisting energy, and nonbond interaction (Hwang *et al.*, 2020). The relevant parameters and expressions constituted the force field, that was, the parameters representing the interactions between molecules and atoms (Collier *et al.*, 2020). A complete force field contained a set of functions and related parameters. The parameters in its potential function were derived from experimental results and the methods of quantum mechanics. Different force fields used different parameters and function forms, resulting in different systems and development directions of force fields. For example, the AMBER force field was first proposed by

Kollman *et al.* in 1984 (Weiner *et al.*, 1984), which initially only provided corresponding atomic types for protein and nucleic acid systems. In 1990, the force field parameters suitable for polysaccharide simulation were developed (Homans, 1990), and around 2000 atomic types and parameters suitable for small organic molecules were added (Wang *et al.*, 1999). Amber force field model was a relatively classical force field model in MD. The CHARMM force field was proposed by Karplus group in 1983 and is suitable for the calculation and simulation of various molecular properties, from isolated small molecules to solvated large molecules, but it was not suitable for metal complexes (Brooks *et al.*, 1983). In addition, the CVFF force field proposed by the Dauber-Osguthorpe group in 1988 applied to small molecules and protein systems and could be extended to simulate some inorganic systems, such as silicates, aluminosilicates, and phosphate-aluminum compounds (Brooks *et al.*, 1983), which were mainly used to predict the structure and binding free energy of molecules. The selection of force field should refer to the function form of force field, combine the calculation experience and the characteristics of the system, and select the most suitable force field. The AMBER force field and CHARMM force field are the most used in protein MD, especially in the simulation of *de novo* design proteins.

The overestimation of secondary structures and excessive conformations were the main problems of the early version force field (Best *et al.*, 2008; Piana *et al.*, 2014; Uversky, 2013). Many efforts were made in the development of the traditional force fields to improve protein simulation. Wu *et al.* developed a residue-specific force field (RSFF2) based on conformational free-energy distributions of the 20 amino acid residues from a protein coil library (Zhou *et al.*, 2015). They made modifications to the Amber 99SB force field parameters, and it successfully folded α -helical structures better than RSFF1 and Amber 99SB force field. It also provided folding enthalpies and entropies in reasonably good agreement with available experimental results. We believed that such progress could be very helpful in the simulating designed proteins, particularly those with specific entropies. Shaw *et al.* described a force field that substantially improved on the state-of-the-art accuracy for simulations of disordered proteins without sacrificing accuracy for folded proteins, thereby broadening the range of biological systems amenable to MD simulations (Robustelli *et al.*, 2018). Regarding the CHARMM force field, many computational biologists have made efforts to modify the parameters to perfection. Lin *et al.* optimized the backbone parameters of the CHARMM36 force field, aiming to solve the gas properties of the alanine dipeptide, resulting in the Drude-2019 protein FF (Lin *et al.*, 2020). Their results showed that the updated Drude-2019 protein FF yields smaller overall root-mean-square differences of proteins compared to the additive CHARMM36m and Drude-2013 FFs, as well as similar or improved agreement with experimental NMR properties, allowing for long time scale simulation studies of proteins and more complex biomolecular systems in conjunction with the remainder of the Drude polarizable FF. Apart from force fields mentioned above, examples with improved dihedral angle parameters include Amber ff14SB (Maier *et al.*, 2015), Amber ff99SB-ILDN (Lindorff-Larsen *et al.*, 2010), CHARMM36m (Huang *et al.*, 2017), CHARMM22 (Piana *et al.*, 2011), OPLS3 (Harder *et al.*, 2016), and OPLS-AA/M (Robertson *et al.*, 2015).

Designed proteins had not been selected by evolution, the folding energy barrier was generally lower than that of natural proteins, and the composition of hydrophobic cores might be more complex than that of natural proteins (Cao *et al.*, 2020; Silva *et al.*, 2019). All these features require high accuracy of protein folding

simulation parameters. Therefore, the optimization and development of force field parameters are important aspects of *de novo*-designed proteins.

Water models

In the field of MD, the classification of force fields of water molecules is of unparalleled importance compared to other types of molecules. At the same time, the number and variety of classical force fields for water molecules are unmatched, with more than 100 different types. Similar to force fields for other molecules, different levels of approximation can be used to obtain different types of force fields for water molecules. The classical force fields for water molecules are roughly categorized into the following four types: rigid force fields, flexible force fields, polarizable force fields, and dissociative force fields.

The TIP3P water model has been widely used in MD simulations of proteins, including *de novo* protein design (Park *et al.*, 2022). However, the prevalence of TIP3P may vary depending on the research community and the specific application. In recent years, there has been a trend toward using newer water models, such as TIP4P-2005 and TIP5P, which more accurately reproduce the properties of liquid water. These models may provide better results for specific properties or phenomena that are sensitive to the water model. The TIP4P-2005 and TIP5P water models are more complex than TIP3P, with additional interaction sites that allow for a more accurate description of water's tetrahedral structure and dynamics. These models have been shown to better reproduce the properties of liquid water, such as its density, heat capacity, and diffusion coefficient. Max F Döpke *et al.* demonstrated that the TIP4P-2005 model was among the best rigid water force fields, whereas many of the most successful ion parameters were optimized in combination with SPC-E, TIP3P, or TIP4P-Ew water (Döpke *et al.*, 2020).

The TIP5P water model is a five-site model that was introduced to more accurately capture the tetrahedral structure of liquid water and the properties of water in various phases. One of the most significant features of the TIP5P model is its ability to better reproduce the tetrahedral arrangement of water molecules in the liquid phase. This is achieved by using four charge sites (two hydrogen atoms and two lone pairs) arranged tetrahedrally around the central oxygen atom.

Overall, the choice of water model depends on the specific research question and the desired level of accuracy. While TIP4P-2005 and TIP5P may provide improved accuracy for certain properties or phenomena, they also require more computational resources and may not be necessary for all applications. James Dix *et al.* concluded that, among the MD models tested, the TIP4P-2005 and TIP5P force fields were for now the most reliable when simulating water under confinement (Dix *et al.*, 2018). Researchers should carefully consider the tradeoffs between accuracy and computational efficiency when choosing a water model for protein design and simulation.

MD Simulation to study protein thermal stability

Simulating proteins at different temperatures often results in different average properties and conformational systems. The general approach to studying the protein thermal stability using MD simulations involves performing simulated sampling at temperatures of interest, then comparing the sampled ensemble and analyzing the interactions. This can include calculating structural fluctuations

with root mean square displacement (RMSD), root mean square fluctuation (RMSF), and cyclometric radius (Rg) of atomic positions (Ahamad *et al.*, 2022; Su *et al.*, 2022). Additionally, the solvent accessibility area (SASA) can be used to obtain more information about protein folding and hydrophobicity, while hydrogen bond statistics can be used to reflect local structural changes. Furthermore, free energy calculations can be utilized to determine the most stable conformation in the ensemble. The comprehensive application of these analytical methods is helpful for exploring the possible molecular mechanism of thermal stability.

There have been numerous MD simulation studies analyzing the heat resistance mechanisms of natural proteins in detail. For example, based on the simulation results of firefly luciferase and its mutants, Rahban *et al.* concluded that the increase in thermal stability of the enzyme was due to the increase of salt bridges and hydrogen bonds in the molecule (Rahban *et al.*, 2017). In another study, Kumar *et al.* utilized MD simulation to investigate the molecular basis of the thermal stability of the natural protein SazCA. They calculated hydrogen bond, SASA, Ramchandran plot to trace the folding and unfolding paths of the protein and tried to explain the highest thermal stability at 353 K using free energy calculation (Kumar and Deshpande, 2021). MD simulations can also be used directly to explore potential causes of the macroscopic properties of *de novo*-designed proteins. For example, the design of highly stable proteins often favored the encapsulation of hydrophobic residues (Kuhlman and Baker, 2004). Protein Top7 serves as a classical example (Kuhlman *et al.*, 2003). Bocek *et al.* combined experimental and computational methods to investigate the effects of low pH, high temperature, and the structural stability and dynamics of Top7 and its mutants in order to obtain atomic details of their conformational changes under extreme conditions (Bocek *et al.*, 2009; Soares *et al.*, 2010). The results showed that the mutant of Top7 demonstrated obvious folding under different pH and temperature conditions. Although the helix content was lower at high temperature, the antiparallel β -strand could still fold and form extensive interaction with loop. It was concluded that the molecular basis of tolerance of the Top7 mutant was a very stable hydrophobic β -fold core.

In addition to studying hydrophobic interactions, MD simulations have also been used to investigate another class of mechanisms by which *de novo*-designed proteins maintain thermal stability: electrostatic interactions. Missimer *et al.* studied the static mechanism of *de novo*-designed protein (17 peptide) to maintain stability at high temperature through MD simulations (Missimer *et al.*, 2007). The solvent model was used to simulate the protein at three temperatures, and the stability difference between the monomer and the trimer was compared. It was found that the RMSF of the trimer was essentially the same at 278 K and 330 K, while the RMSF of the monomer increased with temperature. It was also observed that the average occupancy of the salt bridge in the trimer helix was higher than that of the monomer, indicating that the salt bridge, as a key electrostatic interaction, significantly contributed to the thermal stability of the trimer. Additionally, the interaction between the trimer and water weakened with increasing temperature, while the monomer showed the opposite trend. This led to the suggestion that desolvation at higher temperatures allowed electrostatic interactions to occur in the form of salt bridges, acting as a major factor in the thermal stability of proteins.

While many of the aforementioned studies aimed to use MD simulations to explore the thermal stability of *de novo*-designed proteins, the majority of research focused on proteins automatically designed by sequences based on physical energy functions. There is

a lack of analysis regarding stability factors for proteins automatically designed based on statistical energy functions. Unlike physical energy models, statistical energy used for sequence design may capture implicit information in the training data and reflect it on the macroscopic properties of the designed protein. By conducting in-depth studies of these designed proteins, it is possible to gain insights into this information, and the new insights obtained can be iteratively incorporated into new design methods to improve and optimize the design process.

Free energy calculations have now become rapid enough to significantly impact protein design. The calculation of free energy represents an important research direction in the field of protein design, particularly for understanding and predicting protein-ligand interactions. In recent years, significant progress has been made in the field of protein design due to the development of computational technology, leading to advancements in free energy calculation. Binding free energy is the most critical physical quantity for describing the process of recognition and assembly. For example, in the process of drug design, the binding free energy of candidate drugs and receptor proteins in the database is often calculated through molecular docking or MD, and potential drugs can be identified based on the results of these calculations.

MD Simulation to study designed protein dynamics

Protein folding describes the process through which one-dimensional structural sequences attain their three-dimensional structure with biological functions, whereas protein design commences from the structure and delves into which sequence can achieve a specific structure. Consequently, research on designing protein folding provides deeper insights into these two aspects and offers more valuable information to the field of protein research.

Membrane proteins play an important role in nature, and the design and simulation of membrane proteins represent essential tools for studying this class of proteins. In recent years, scientists have made significant progress in studying the folding mechanism of newly designed membrane proteins. For instance, in terms of the *de novo* design, DeGrado *et al.* identified a stable packing motif in MD simulations of pentameric α -helical barrel phospholamban and designed a new artificial transmembrane parallel pentamer using the sequence pattern (Mravic *et al.*, 2019). The MD study in this research indicated that van der Waal interaction contributed to the packing of apolar residues between α -helices and demonstrated that the geometry of helices in such packing was strictly determined, showing that MD simulations were useful for evaluating the dynamics and stability of membrane-embedded protein structures (Niitsu and Sugita, 2023). Ulmschneider *et al.* illustrated that membrane binding, insertion, association, and dissociation, all of which were included in the folding dynamics events, can be sampled through MD simulations (Chen *et al.*, 2020). The designed a 14-residue pore-forming peptide with sequence patterning and conducted high-temperature MD simulations for this *de novo*-designed protein, making a comparison of several sequences in the long-time MD simulations. The long-time MD simulations revealed split folding events, aiding scientists in understanding the folding mechanisms of the *de novo*-designed proteins (Chen *et al.*, 2019). The pioneering work of Voth and coworkers involved testing a hypothesis that protons were conducted through dry spots by forming transient water wires, often highly correlated with the presence of the excess proton itself in the water wire. To test this hypothesis, they used MD simulations to design transmembrane

channels with stable water pockets interspersed by apolar segments capable of forming flickering water wires, providing a compelling example of a *de novo* proton channel that allowed exploration of the roles transient water wires through apolar regions play in proton selectivity and conduction using the MD tool (Kratochvil *et al.*, 2023).

In addition to the study of membrane proteins, other *de novo* protein simulation studies have also made significant progress. For example, Schiffer *et al.* studied the dynamics of explicitly solvated designed proteins through all-atom MD simulations to gain insight into the causes leading to the low affinity or instability of most of these designs. Their simulations ranged from 500 to 1000 ns per replicate on 37 designed protein variants encompassing two distinct folds and a range of ligand affinities, resulting in more than 180 μ s of combined sampling. The simulations provided retrospective insights into the properties affecting ligand affinity, proving useful in guiding further steps of design optimization and demonstrating that MD could act as a screening step in protein design, resulting in a more efficient process (Barros *et al.*, 2019).

Additionally, DeGrado, Handel, and colleagues designed a three-helix bundle, α 3D, whose interior sidechains consisted of a diverse set of apolar residues that packed in a geometrically complementary manner (Korendovych and DeGrado, 2020). Due to its relatively simple but cooperatively folded globular structure, α 3D quickly became a widely studied protein for computational and experimental studies of protein folding. Labs that contributed to this work included Shaw, Daggett, and Brooks, and using MD simulations, they assessed the consistency of folding time and protein stability. α 3D exhibited distinct characteristics, and the folding enthalpy obtained by MD simulation was in good agreement with the experimental estimation, highlighting the important role of MD simulation in the study of *de novo* design protein folding (Piana *et al.*, 2014). In a recent paper, Ulas *et al.* used MD and free energy calculations to design low to sub-nM drug-binding proteins with a very high success rate. This contrasts with the very low success rates and weak binders obtained in recent studies in which researchers have relied exclusively on design calculations (Lu *et al.*, 2023).

In another study, Zhang *et al.* utilized all-atom REMD simulations, through which several comparably stable intermediate states were observed along the folding free-energy landscape of the *de novo* design protein (DS119). MD simulations revealed that when two unfolded DS119 proteins bound together, most of the binding sites of the dimeric aggregates were located in the N-terminal segment, particularly residues 5–10, which were predicted to form β -sheet with its own C-terminal segment. The complex folding behavior of DS119 suggested the significant influence of natural selection on protein-folding kinetics, indicating the need for further improvement in rational protein design (Wang *et al.*, 2016).

From the above examples, it was evident that in the field of *de novo* protein design, MD simulation technology had significantly enhanced people's understanding of the relationship between protein structure and function, offering dynamic information at the atomic level. MD simulation has unveiled the molecular basis of protein function by investigating protein structure and dynamic behavior, capturing the fundamental features of atomic interactions that control movement within proteins to comprehend how proteins move and interact at the atomic level. Furthermore, it has provided dynamic information about conformational transitions, interactions, and control stability and function, which can play a pivotal role in improving the accuracy and efficiency of protein design.

The protein–ligand binding process involved complex structural changes that were challenging to fully capture within the time scales available with MD. To enhance efficiency and reduce the conformational space of the sample, many researchers have utilized geometric constraints, such as funnel-shaped constraints and spherical constraints, to improve the convergence rate of the simulation (Negron *et al.*, 2009). The binding free energy was then accurately calculated by subtracting the influence of constraints through post-processing. Free energy perturbation was a computational method used to estimate the free energy change resulting from small changes in the system (Jespers *et al.*, 2021). The importance sampling method improved computational efficiency by altering the sampling distribution, especially when dealing with rare events. The combination of geometric constraints with free energy perturbation and importance sampling methods can enhance the accuracy of protein–ligand binding free energy calculation. The various methods of free energy computation not only aid in comprehending the fundamental properties of proteins but also guide drug design and protein engineering, providing powerful tools in the field of protein design to understand and manipulate the structure and function of proteins.

Conclusions and outlook

MD simulation is a computational technique used to study the dynamic behavior of proteins at the atomic level, serving as a powerful tool in the simulation field of *de novo*-designed proteins. It has played a crucial role in advancing our understanding of the dynamic behavior, stability, and function of designed proteins, leading to significant advantages in force field modeling and sampling methods.

In the current field of protein molecular simulation, the reliability of simulation results primarily depends on the accuracy of the molecular mechanical force field employed. However, the currently developed force fields are still not without imperfections, and the simulation results of some systems may not align well with experimental data. One method to address this issue is to utilize different force fields or optimize the force field parameters themselves. Additionally, due to limitations in simulation algorithms and computing power, many folding simulation runs are too short to accurately capture the various possible conformational states of artificially *de novo*-designed proteins. This limitation restricts our understanding of the conformational changes of artificially *de novo*-designed proteins at different temperatures. To tackle this problem, it may be valuable to consider using more sophisticated simulation algorithms or more powerful computers.

Numerous research examples have demonstrated that MD simulation can, to a certain extent, replica experimental observations regarding the stability and dynamic properties of *de novo*-designed proteins. It can provide insights into the stability and folding dynamics characteristics of proteins at a microscopic scale, offering a partial reflection on the folding properties of *de novo*-designed proteins. Taking into account the substantial workload and lengthy cycle involved in experimental verification of *de novo*-designed proteins, MD simulations can be conducted on the *de novo*-designed proteins at appropriate temperatures initially. The structural fluctuations can be observed as a preliminary assessment of their structural stability and folding characteristics, allowing for the screening of sequences that may fold into target structures. Subsequently, these sequences can undergo experimental verification.

However, MD simulation methods have limitations when it comes to exploring protein dynamics. One such limitation is

sampling constraints: MD simulations are inherently stochastic, and sampling rare events can be a challenging task. This can result in an incomplete or biased view of the system's dynamics, particularly for processes that occur over long timescales or involve multiple pathways. Another limitation is related to system size: MD simulations are typically restricted to relatively small systems, typically containing a few hundred atoms or less. This can pose a challenge when studying larger proteins or protein complexes, where important dynamic processes may occur on a larger scale. Additionally, computational resource limitations and force-field accuracy present further constraints. However, we believe that with the continuous development of computer software and hardware, along with the widespread adoption of artificial intelligence algorithms, these limitations will gradually diminish in the field of protein simulation. In conclusion, MD simulation has emerged as a powerful tool in the field of *de novo* protein design, playing a crucial role in advancing our understanding of the dynamic behavior, stability, and function of designed proteins. In particular, in terms of structure refinement, MD simulations have been found to be effective in enhancing the structures of *de novo*-designed proteins. They can assist in tasks such as relaxing initial models, resolving steric clashes, and optimizing side-chain conformations, contributing to the generation of more accurate and stable protein structures. MD simulations have provided valuable insights for the design of *de novo* proteins by helping researchers understand the dynamics and stability of protein structures. This technique can be utilized to predict the stability of *de novo* protein designs. By simulating the protein's dynamics over time, researchers can identify regions of the protein that may be unstable or prone to unfolding. Additionally, by simulating the dynamics of protein–protein complexes, researchers can identify key residues involved in the interaction and optimize them for stronger binding. This provides valuable insights into the dynamics and stability of *de novo* proteins and can aid in the design of novel proteins with specific functions and improved stability.

MD simulations provide valuable insights into the stability of designed proteins by monitoring fluctuations in the protein's backbone and side chains. This information is critical for identifying potential weaknesses and improving the design. Continued advancements in enhanced sampling techniques, such as accelerated MD and metadynamics, will enable researchers to explore rare events and long-timescale dynamics more efficiently.

In summary, MD simulation is poised to remain a cornerstone in the field of *de novo* protein design, playing a pivotal role in advancing our ability to create novel proteins with tailored functions and applications in diverse fields, from biotechnology to medicine. As computational methods and experimental techniques continue to evolve, the synergy between simulation and experimentation will undoubtedly lead to groundbreaking discoveries and innovations.

Funding. This research was funded by National Natural Science Foundation of China, grant number 82101922.

Competing interest. The authors declare no conflict of interest.

References

- Ahamad S, Gupta D and Kumar V (2022) Targeting SARS-CoV-2 nucleocapsid oligomerization: Insights from molecular docking and molecular dynamics simulations. *Journal of Biomolecular Structure and Dynamics*, **40**(6), 2430–2443. <http://doi.org/10.1080/07391102.2020.1839563>.

- Anfinsen CB, Haber E, Sela M and White FH, Jr (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of U S A*, 47(9), 1309–1314. <http://doi.org/10.1073/pnas.47.9.1309>.
- Barros EP, Schiffer JM, Vorobieva A, Dou J, Baker D and Amaro RE (2019) Improving the efficiency of ligand-binding protein design with molecular dynamics simulations. *Journal of Chemical Theory and Computation* 15(10), 5703–5715. <http://doi.org/10.1021/acs.jctc.9b00483>.
- Bernardi RC, Melo MCR and Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta* 1850(5), 872–877. <http://doi.org/10.1016/j.bbagen.2014.10.019>.
- Best RB, Buchete NV and Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophysical Journal* 95(1), L07–09. <http://doi.org/10.1529/biophysj.108.132696>.
- Blanco FJ, Rivas G and Serrano L (1994) A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nature Structural Biology* 1(9), 584–590. <http://doi.org/10.1038/nsb0994-584>.
- Boschek CB, Apiyo DO, Soares TA, Engelmann HE, Pefaur NB, Straatsma TP and Baird CL (2009) Engineering an ultra-stable affinity reagent based on Top7. *Protein Engineering, Design and Selection* 22(5), 325–332. <http://doi.org/10.1093/protein/gzp007>.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S and Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* 4. <https://doi.org/10.1002/jcc.540040211>.
- Cao L, Goreschnik I, Coventry B, Case JB, Miller L, Kozodoy L, Chen RE, Carter L, Walls AC, Park YJ, Strauch EM, Stewart L, Diamond MS, Veesler D and Baker D (2020) De novo design of picomolar SARS-CoV-2 mini-protein inhibitors. *Science* 370(6515), 426–431. <http://doi.org/10.1126/science.abd9909>.
- Chen C, Xiao Y and Huang Y (2015) Improving the replica-exchange molecular-dynamics method for efficient sampling in the temperature space. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics* 91(5), 052708. <http://doi.org/10.1103/PhysRevE.91.052708>.
- Chen CH, Melo MC, Berglund N, Khan A, de la Fuente-Nunez C, Ulmschneider JP and Ulmschneider MB (2020) Understanding and modelling the interactions of peptides with membranes: from partitioning to self-assembly. *Current Opinion in Structural Biology* 61, 160–166. <http://doi.org/10.1016/j.sbi.2019.12.021>.
- Chen CH, Starr CG, Troendle E, Wiedman G, Wimley WC, Ulmschneider JP and Ulmschneider MB (2019) Simulation-guided rational de novo design of a small pore-forming antimicrobial peptide. *Journal of the American Chemical Society* 141(12), 4839–4848. <http://doi.org/10.1021/jacs.8b11939>.
- Collier TA, Piggot TJ and Allison JR (2020) Molecular dynamics simulation of proteins. *Methods in Molecular Biology* 2073, 311–327. http://doi.org/10.1007/978-1-4939-9869-2_17.
- Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, Rupert P, Correnti C, Kalyuzhnyi O, Vittal V, Connell MJ, Stevens E, Schroeter A, Chen M, Macpherson S, Serra AM, Adachi Y, Holmes MA, Li Y, Klevit RE, Graham BS, Wyatt RT, Baker D, Strong RK, Crowe JE, Jr., Johnson PR and Schief WR (2014) Proof of principle for epitope-focused vaccine design. *Nature* 507(7491), 201–206. <http://doi.org/10.1038/nature12966>.
- Dahiyat BI and Mayo SL (1997) De novo protein design: Fully automated sequence selection. *Science* 278(5335), 82–87. <http://doi.org/10.1126/science.278.5335.82>.
- Dill KA and Chan HS (1997) From Levinthal to pathways to funnels. *Nature Structural Biology* 4(1), 10–19. <http://doi.org/10.1038/nsb0197-10>.
- Dill KA and MacCallum JL (2012) The protein-folding problem, 50 years on. *Science* 338(6110), 1042–1046. <http://doi.org/10.1126/science.1219021>.
- Ding Y, Yu K and Huang J (2023) Data science techniques in biomolecular force field development. *Current Opinion in Structural Biology*, 78, 102502. <http://doi.org/10.1016/j.sbi.2022.102502>.
- Dix J, Lue L and Carbone P (2018) Why different water models predict different structures under 2D confinement. *Journal of Computational Chemistry* 39(25), 2051–2059. <http://doi.org/10.1002/jcc.25369>.
- Döpke MF, Moulτος OA and Hartkamp R (2020) On the transferability of ion parameters to the TIP4P/2005 water model using molecular dynamics simulations. *The Journal of Chemical Physics* 152(2), 024501. <http://doi.org/10.1063/1.5124448>.
- Giberti F, Salvalaglio M and Parrinello M (2015) Metadynamics studies of crystal nucleation. *IUCr* 2(Pt 2), 256–266. <http://doi.org/10.1107/s2052252514027626>.
- Gonzalez NA, Li BA and McCully ME (2022) The stability and dynamics of computationally designed proteins. *Protein Engineering, Design and Selection* 35. <http://doi.org/10.1093/protein/gzac001>.
- Harder ED, Damm W, Maple JR, Wu C, Reibou M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, Kaus JW, Cerutti DS, Krilov G, Jorgensen WL, Abel R and Friesner RA (2016) OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *Journal of Chemical Theory and Computation* 12(1), 281–296. <http://doi.org/10.1021/acs.jctc.5b00864>.
- Hecht MH, Richardson JS, Richardson DC and Ogden RC (1990) De novo design, expression, and characterization of Felix: A four-helix bundle protein of native-like sequence. *Science* 249(4971), 884–891. <http://doi.org/10.1126/science.2392678>.
- Homans SW (1990) A molecular mechanical force field for the conformational analysis of oligosaccharides: comparison of theoretical and crystal structures of Man alpha 1-3Man beta 1-4GlcNAc. *Biochemistry* 29(39), 9110–9118. <http://doi.org/10.1021/bi00491a003>.
- Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H and MacKerell AD (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* 14(1), 71–73. <http://doi.org/10.1038/nmeth.4067>.
- Huang PS, Feldmeier K, Parmeggiani F, Velasco DAF, Höcker B and Baker D (2016) De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology* 12(1), 29–34. <http://doi.org/10.1038/nchembio.1966>.
- Hwang SB, Lee CJ, Lee S, Ma S, Kang YM, Cho KH, Kim SY, Kwon OY, Yoon CN, Kang YK, Yoon JH, Nam KY, Kim SG, In Y, Chai HH, Acree WE, Jr., Grant JA, Gibson KD, Jhon MS, Scheraga HA and No KT (2020) PMFF: Development of a physics-based molecular force field for protein simulation and ligand docking. *The Journal of Physical Chemistry B*, 124(6), 974–989. <http://doi.org/10.1021/acs.jpcc.9b10339>.
- Jespers W, Åqvist J and Gutiérrez-de-Terán H (2021) free energy calculations for protein-ligand binding prediction. *Methods in Molecular Biology* 2266, 203–226. http://doi.org/10.1007/978-1-0716-1209-5_12.
- Kamberaj H (2018) Faster protein folding using enhanced conformational sampling of molecular dynamics simulation. *Journal of Molecular Graphics and Modelling* 81, 32–49. <http://doi.org/10.1016/j.jmgm.2018.02.008>.
- Korendovych IV and DeGrado WF (2020) De novo protein design, a retrospective. *Quarterly Reviews of Biophysics* 53, e3. <http://doi.org/10.1017/s0033583519000131>.
- Kratochvil HT, Watkins LC, Mravic M, Thomaston JL, Nicoludis JM, Somberg NH, Liu L, Hong M, Voth GA and DeGrado WF (2023) Transient water wires mediate selective proton transport in designed channel proteins. *Nature Chemistry* 15(7), 1012–1021. <http://doi.org/10.1038/s41557-023-01210-4>.
- Kuhlman B and Baker D (2004) Exploring folding free energy landscapes using computational protein design. *Current Opinion in Structural Biology* 14(1), 89–95. <http://doi.org/10.1016/j.sbi.2004.01.002>.
- Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL and Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302(5649), 1364–1368. <http://doi.org/10.1126/science.1089427>.
- Kumar S and Deshpande PA (2021) Structural and thermodynamic analysis of factors governing the stability and thermal folding/unfolding of SazCA. *PLoS One* 16(4), e0249866. <http://doi.org/10.1371/journal.pone.0249866>.
- Latour RA (2014) Perspectives on the simulation of protein-surface interactions using empirical force field methods. *Colloids and Surfaces B Biointerfaces* 124, 25–37. <http://doi.org/10.1016/j.colsurfb.2014.06.050>.
- Lazim R, Mei Y and Zhang D (2012) Replica exchange molecular dynamics simulation of structure variation from $\alpha/4\beta$ -fold to 3α -fold protein. *Journal of Molecular Modeling* 18(3), 1087–1095. <http://doi.org/10.1007/s00894-011-1147-8>.
- Lazim R, Suh D and Choi S (2020) Advances in molecular dynamics simulations and enhanced sampling methods for the study of protein systems.

- International Journal of Molecular Sciences*, **21**(17).<http://doi.org/10.3390/ijms21176339>.
- Leong SW, Lim TS, Ismail A and Choong YS** (2018) Integration of molecular dynamics simulation and hotspot residues grafting for de novo scFv design against Salmonella Typhi TolC protein. *Journal of Molecular Recognition* **31**(5), e2695.<http://doi.org/10.1002/jmr.2695>.
- Leopold PE, Montal M and Onuchic JN** (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences of the United States of America*, **89**(18), 8721–8725.<http://doi.org/10.1073/pnas.89.18.8721>.
- Li X, Yang Z, Chen Y, Zhang S, Wei G and Zhang L** (2023) Dissecting the molecular mechanisms of the co-aggregation of A β 40 and A β 42 peptides: A REMD simulation study. *The Journal of Physical Chemistry B* **127**(18), 4050–4060.<http://doi.org/10.1021/acs.jpcc.3c01078>.
- Lin FY, Huang J, Pandey P, Rupakheti C, Li J, Roux BT and MacKerell AD, Jr** (2020) Further optimization and validation of the classical drude polarizable protein force field. *Journal of Chemical Theory and Computation* **16**(5), 3221–3239.<http://doi.org/10.1021/acs.jctc.0c00057>.
- Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO and Shaw DE** (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**(8), 1950–1958.<http://doi.org/10.1002/prot.22711>.
- Lopes PE, Guvench O and MacKerell AD, Jr** (2015) Current status of protein force fields for molecular dynamics simulations. *Methods in Molecular Biology* **1215**, 47–71.http://doi.org/10.1007/978-1-4939-1465-4_3.
- Lu L, Gou X, Tan SK, Mann SI, Yang H, Zhong X, Gazgalis D, Valdiviezo J, Jo H, Wu Y, Diolaiti ME, Ashworth A, Polizzi NF and DeGrado WF** (2023) De novo design of drug-binding proteins with predictable binding energy and specificity. *bioRxiv*.<http://doi.org/10.1101/2023.12.23.573178>.
- Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE and Simmerling C** (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation* **11**(8), 3696–3713.<http://doi.org/10.1021/acs.jctc.5b00255>.
- Marcandalli J, Fiala B, Ols S, Perotti M, de van der Schueren W, Snijder J, Hodge E, Benhaim M, Ravichandran R, Carter L, Sheffler W, Brunner L, Lawrenz M, Dubois P, Lanzavecchia A, Sallusto F, Lee KK, Veessler D, Correnti CE, Stewart LJ, Baker D, Loré K, Perez L and King NP** (2019) Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for respiratory syncytial virus. *Cell* **176**(6), 1420–1431.e1417.<http://doi.org/10.1016/j.cell.2019.01.046>.
- Marcisz M, Maszota-Zieleniak M, Huard B and Samsonov SA** (2021) Advanced molecular dynamics approaches to model a tertiary complex APRIL/TACI with Long glycosaminoglycans. *Biomolecules* **11**(9).<http://doi.org/10.3390/biom11091349>.
- Marsh JA and Teichmann SA** (2015) Structure, dynamics, assembly, and evolution of protein complexes. *Annual Review of Biochemistry* **84**, 551–575.<http://doi.org/10.1146/annurev-biochem-060614-034142>.
- Missimer JH, Steinmetz MO, Baron R, Winkler FK, Kammerer RA, Daura X and van Gunsteren WF** (2007) Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Science* **16**(7), 1349–1359.<http://doi.org/10.1110/ps.062542907>.
- Mrvac M, Thomaston JL, Tucker M, Solomon PE, Liu L and DeGrado WF** (2019) Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science* **363**(6434), 1418–1423.<http://doi.org/10.1126/science.aav7541>.
- Negron C, Fufezan C and Koder RL** (2009) Geometric constraints for porphyrin binding in helical protein binding sites. *Proteins* **74**(2), 400–416.<http://doi.org/10.1002/prot.22143>.
- Niitsu A and Sugita Y.** (2023). Towards de novo design of transmembrane α -helical assemblies using structural modelling and molecular dynamics simulation. *Phys Chem Chem Phys*, **25**(5), 3595–3606.<http://doi.org/10.1039/d2cp03972a>.
- Ostermeir K and Zacharias M** (2013) Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochimica et Biophysica Acta*, **1834**(5), 847–853.<http://doi.org/10.1016/j.bbapap.2012.12.016>.
- Park C, Robinson F and Kim D** (2022) On the choice of different water model in molecular dynamics simulations of nanopore transport phenomena. *Membranes (Basel)* **12**(11).<http://doi.org/10.3390/membranes12111109>.
- Patel SC, Bradley LH, Jinadasa SP and Hecht MH** (2009) Cofactor binding and enzymatic activity in an unevolved superfamily of de novo designed 4-helix bundle proteins. *Protein Science* **18**(7), 1388–1400.<http://doi.org/10.1002/pro.147>.
- Peng C, Wang J, Shi Y, Xu Z and Zhu W** (2021) Increasing the sampling efficiency of protein conformational change by combining a modified replica exchange molecular dynamics and normal mode analysis. *Journal of Chemical Theory and Computation*, **17**(1), 13–28.<http://doi.org/10.1021/acs.jctc.0c00592>.
- Piana S, Klepeis JL and Shaw DE.** (2014). Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations. *Current Opinion in Structural Biology* **24**, 98–105.<http://doi.org/10.1016/j.sbi.2013.12.006>.
- Piana S, Lindorff-Larsen K and Shaw DE** (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophysical Journal* **100**(9), L47–L49.<http://doi.org/10.1016/j.bpj.2011.03.051>.
- Pourjafar-Dehkordi D and Zacharias M** (2021) Influence of a Ser111-phosphorylation on Rab1b GTPase conformational dynamics studied by advanced sampling simulations. *Proteins* **89**(10), 1324–1332.<http://doi.org/10.1002/prot.26153>.
- Rahban M, Salehi N, Saboury AA, Hosseinkhani S, Karimi-Jafari MH, Firouzi R, Rezaei-Ghaleh N and Moosavi-Movahedi AA** (2017) Histidine substitution in the most flexible fragments of firefly luciferase modifies its thermal stability. *Archives of Biochemistry and Biophysics* **629**, 8–18.<http://doi.org/10.1016/j.abb.2017.07.003>.
- Regan L and DeGrado WF** (1988) Characterization of a helical protein designed from first principles. *Science* **241**(4868), 976–978.<http://doi.org/10.1126/science.3043666>.
- Robertson MJ, Tirado-Rives J and Jorgensen WL** (2015) Improved peptide and protein torsional energetics with the OPLSAA force field. *Journal of Chemical Theory and Computation* **11**(7), 3499–3509.<http://doi.org/10.1021/acs.jctc.5b00356>.
- Robustelli P, Piana S and Shaw DE** (2018) Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(21), E4758–e4766.<http://doi.org/10.1073/pnas.1800690115>.
- Rosta E and Hummer G** (2009) Error and efficiency of replica exchange molecular dynamics simulations. *The Journal of Chemical Physics* **131**(16), 165102.<http://doi.org/10.1063/1.3249608>.
- Sesterhenn F, Yang C, Bonet J, Cramer JT, Wen X, Wang Y, Chiang CI, Abriata LA, Kucharska I, Castoro G, Vollers SS, Galloux M, Dheilly E, Rosset S, Corthésy P, Georgeon S, Villard M, Richard CA, Descamps D, Delgado T, Oricchio E, Rameix-Welti MA, Más V, Ervin S, Eléouët JF, Riffault S, Bates JT, Julien JP, Li Y, Jardetzky T, Krey T and Correia BE** (2020) De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science* **368**(6492).<http://doi.org/10.1126/science.aay5051>.
- Silva DA, Yu S, Ulge UY, Spangler JB, Jude KM, Labão-Almeida C, Ali LR, Quijano-Rubio A, Ruterbusch M, Leung I, Biary T, Crowley SJ, Marcos E, Walkey CD, Weitzner BD, Pardo-Avila F, Castellanos J, Carter L, Stewart L, Riddell SR, Pepper M, Bernardes GJL, Dougan M, Garcia KC and Baker D** (2019) De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**(7738), 186–191.<http://doi.org/10.1038/s41586-018-0830-7>.
- Soares TA, Boschek CB, Apiyo D, Baird C and Straatsma TP** (2010) Molecular basis of the structural stability of a Top7-based scaffold at extreme pH and temperature conditions. *Journal of Molecular Graphics and Modelling* **28**(8), 755–765.<http://doi.org/10.1016/j.jmgm.2010.01.013>.
- Su J, Sun T, Wang Y and Shen Y** (2022) Conformational dynamics of glucagon-like Peptide-2 with different electric field. *Polymers (Basel)* **14**(13).<http://doi.org/10.3390/polym14132722>.
- Sugita Y, Kamiya M, Oshima H and Re S** (2019) replica-exchange methods for biomolecular simulations. *Methods in Molecular Biology* **2022**, 155–177.http://doi.org/10.1007/978-1-4939-9608-7_7.
- Uversky VN** (2013) Intrinsic disorder-based protein interactions and their modulators. *Current Pharmaceutical Design* **19**(23), 4191–4213.<http://doi.org/10.2174/1381612811319230005>.
- Wang JM, Cieplak P and Kollman P** (1999) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of

- organic and biological molecules? *Journal of Computational Chemistry*, **21**, 1049. [http://doi.org/10.1002/1096-987X\(200009\)21:12<1049::AID-JCC3>3.0.CO;2-F](http://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F).
- Wang M, Hu J and Zhang Z** (2016) The folding of de novo designed protein DS119 via molecular dynamics simulations. *International Journal of Molecular Sciences* **17**(5).<http://doi.org/10.3390/ijms17050612>.
- Wang W** (2021) Recent advances in atomic molecular dynamics simulation of intrinsically disordered proteins. *Physical Chemistry Chemical Physics* **23**(2), 777–784.<http://doi.org/10.1039/d0cp05818a>.
- Wannier TM, Moore MM, Mou Y and Mayo SL** (2015) Computational design of the β -Sheet surface of a red fluorescent protein allows control of protein oligomerization. *PLoS One*, **10**(6), e0130582.<http://doi.org/10.1371/journal.pone.0130582>.
- Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S and Weiner P** (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society* **106** (3), 765–784.<http://doi.org/10.1021/ja00315a051>.
- Zhang W and Wang T** (2023) Understanding protein functions in the biological context. *Protein & Peptide Letters*, **30**(6), 449–458.<http://doi.org/10.2174/0929866530666230507212638>.
- Zhou CY, Jiang F and Wu YD** (2015) Residue-specific force field based on protein coil library. RSFF2: modification of AMBER ff99SB. *The Journal of Physical Chemistry B* **119**(3), 1035–1047.<http://doi.org/10.1021/jp5064676>.
- Zhou R** (2007) Replica exchange molecular dynamics method for protein folding simulation. *Methods in Molecular Biology* **350**, 205–223.<http://doi.org/10.1385/1-59745-189-4:205>.
- Zhu C, Dai Z, Liang H, Zhang T, Gai F and Lai L** (2013) Slow and bimolecular folding of a de novo designed monomeric protein DS119. *Biophysical Journal* **105**(9), 2141–2148.<http://doi.org/10.1016/j.bpj.2013.09.014>.
- Zhu J, Avakyan N, Kakkis A, Hoffnagle AM, Han K, Li Y, Zhang Z, Choi TS, Na Y, Yu CJ and Tezcan FA** (2021) Protein assembly by design. *Chemical Reviews* **121**(22), 13701–13796.<http://doi.org/10.1021/acs.chemrev.1c00308>.