Review

Dynamics, a Powerful Component of Current and Future in Silico Approaches for Protein Design and Engineering

Bartłomiej Surpeta [1,2] **, Carlos Eduardo Sequeiros-Borja** [1,2] **and Jan Brezovsky** [1,2,*]

[1] Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland; bartlomiej.surpeta@amu.edu.pl (B.S.); carseq@amu.edu.pl (C.E.S.-B.)

[2] International Institute of Molecular and Cell Biology in Warsaw, Ks Trojdena 4, 02-109 Warsaw, Poland

[*] Correspondence: janbre@amu.edu.pl or jbrezovsky@iimcb.gov.pl

check for
updates

Abstract: Computational prediction has become an indispensable aid in the processes of engineering and designing proteins for various biotechnological applications. With the tremendous progress in more powerful computer hardware and more efficient algorithms, some of in silico tools and methods have started to apply the more realistic description of proteins as their conformational ensembles, making protein dynamics an integral part of their prediction workflows. To help protein engineers to harness benefits of considering dynamics in their designs, we surveyed new tools developed for analyses of conformational ensembles in order to select engineering hotspots and design mutations. Next, we discussed the collective evolution towards more flexible protein design methods, including ensemble-based approaches, knowledge-assisted methods, and provable algorithms. Finally, we highlighted apparent challenges that current approaches are facing and provided our perspectives on their further development.

Keywords: protein dynamics; protein engineering; hotspot prediction; mutational analysis; computational design; ligand transport; ensemble-based approach; flexible backbone; de novo design; rational design

1. Introduction

Due to their unique structural and functional properties, proteins constitute an essential element of life as well as various branches of the emerging sustainable economy [1–5]. However, only a few proteins are natively equipped with functional parameters and sufficient stability that are required for their industrial and medical utilization. Hence, protein engineering methods gained popularity as an efficient way to deliver new protein variants with desirable properties for a diverse range of tasks [6,7]. Directed evolution and rational design represent the mainstream approaches introduced in the last decades to deliver enhanced protein variants [8]. In essence, the directed evolution enables the generation of rather extensive mutant libraries by randomly introducing mutations in gene-encoding proteins. Generated variants are then evaluated, focusing on the property of interest [9,10]. The rational design originally incorporated expert knowledge and models of proteins from X-ray crystallography to successfully design a handful of mutations enhancing protein stability, function or solubility [11–13]. With the advent of high-performance computing, rational design processes have progressively relied on computational analyses of these static structures [14–16]. So far, computational protein designs have managed to predict not only smart libraries of improved proteins but also massive modifications of proteins towards novel functions [17–19].

However, proteins are known to be dynamical entities, performing their function as an ensemble of diverse conformations rather than a single static structure. Protein dynamics is a highly complex phenomenon comprising numerous contributions from motions with different mechanisms of action and happening with diverse timescales and amplitudes (Figure 1) that highly depend on the system and the local environment [20,21]. Subangstrom vibrations of covalent bonds represent the fastest of those movements. The exploration of various rotamers of side-chains and fluctuations of the protein backbone involve nontrivial moves that span the space of several angstroms. In protein cores, such moves can require several nanoseconds to execute due to the necessity to synchronize with changes in surrounding residues [22–24]. Many conformational changes involve slower and more prominent coordinated movements of several residues in a sequence that manifests as, for example, gating movement executed by loops surrounding the active sites of many proteins [25]. In ligand binding and unbinding events, especially when the binding site is deeply buried in the protein structure, ligands often have to travel tens of angstroms. Such a transport process requires a series of systematic adjustments of protein side-chains and backbones along the traversed paths that might take up hundreds of milliseconds to occur [26]. Among the slowest principal motions performed by proteins are highly organized collective translocations of whole domains, starting on microsecond timescales and with amplitudes reaching nanometers. Finally, the most extensive conformational change transpires during the protein (un)folding processes, which can take hours and even days, and as such, is out of the scope of this review [22–24].
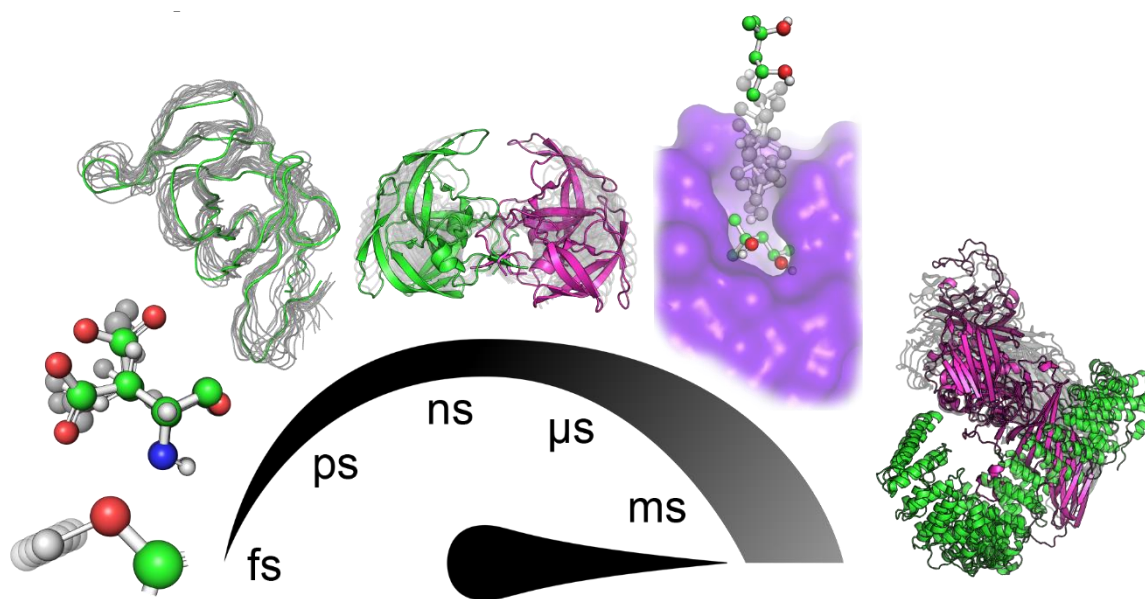

Figure 1. Hierarchy of principal motions in protein dynamics. From left to right: bond vibrations (fs–ps), side-chain rotations (ps–ns), backbone fluctuations (ns), loop motion/gating (ns–ms), ligand binding/unbinding events (>100 ns), and collective domain movement (>μs).

When we consider the reliable treatment of protein dynamics as an essential component of a successful protein design, it is natural to resort to the molecular dynamics (MD) simulation technique as a golden standard to investigate the conformational behavior of a protein. Nowadays, various MD simulation protocols can be utilized to deliver insights into protein dynamics on millisecond timescales with the growing utilization of graphics processing unit (GPU)-enabled parallelism and the development of more efficient software, gradually making such simulations even more affordable [27–31]. Despite all these improvements, MD simulations are not without errors in reproducing a realistic protein ensemble and, hence their experimental confirmation is necessary. Among the major limitations is the accuracy of force fields used to calculate interatomic interactions and the tractable sampling of the ensemble discussed above. The quality of traditionally applied force fields is intrinsically limited

by numerous approximations like the lack of particular interaction types [32], neglect of electronic polarizability [33], and fixed protonation states of titrable residues [34]. At the expense of increased computational demands, some of those limitations can be partially overcome by improving potential models [35], resorting to polarizable force fields [36], and constant pH simulations [37]. Nonetheless, even without these advances, MD simulations relying on the latest force fields have been shown to reach chemical accuracy in their predictions for many different scenarios [38–40].

Regarding their utilization for protein engineering, MD simulations are commonly incorporated into different stages of the design process in order to modulate protein stability, alter interactions of proteins with cognate ligands or perturb dynamics of functional sites [41]. Next, the behavior of protein variants can be closely followed by MD simulations, allowing for the identification, ranking, and selection of promising candidates for experimental validation [42]. In recent years, efforts towards the possibility of also exploiting more distal positions during protein engineering have been gaining momentum [43–47]. By allosteric action, mutations at these positions often affect the preference of proteins to adopt a dominant conformational state, enabling the engineering of proteins with altered selectivity [48,49] or even adopting novel functions [50,51]. As showcased by those mentioned above and other studies [52–57], the crucial role of more comprehensive treatments of protein dynamics for the success of de novo designs, as well as the modification of existing proteins, is well recognized by now.

In this review, we focus on the recent developments in computational methods and tools, which aim to overcome significant challenges brought by integrating protein dynamics into predictions. First, we discuss tools developed for analyzing the fluid nature of interactions in protein ensembles and the elusive transport of ligands in a user-friendly way. In the second part, we critically review the efforts towards the efficient integration of protein flexibility on the backbone level into protein designs and engineering algorithms that are available in established software packages.

2. Tools to Facilitate Analyses of MD Simulation

Accessing information embedded in trajectories produced by MD simulations is a nontrivial task, especially when we focus on phenomena as complex as the networks of interacting residues and their correlated motions or as rare as the events connected with small molecules permeating through protein structures. To alleviate these challenges, we provide an overview of four recently developed tools aiming at understanding and controlling protein allostery and two tools that provide insights into the transport of small molecules (Table 1).

Table 1. Computational tools to extract valuable information for protein engineering from molecular dynamics (MD) simulations.

Tool	Target Property	Availability		Code	Core Method(s)	Input		Link	Reference
		Web Server	Standalone			Structure	Trajectory		
Residue interaction network in protein molecular dynamics (RIP-MD)	Interaction network	+	+	Python	Residue interaction network	+	+	http://dlab.cl/ripmd/	[58]
Java-based Essential Dynamics (JED)	Essential dynamics	-	+	Java	Principal component analysis (PCA)	-	+	https://github.com/charlesdavid/JED	[59]
DynaComm	Allostery	-	+	Python	Distance and correlation-based graphs, Dijkstra algorithm	+	+	https://silviaosuna.wordpress.com/tools/	[43]
Computation of allosteric mechanism by evaluating residue–residue associations (CAMERRA)	Allostery	-	+	Perl, Python, C	PCA, contact analysis	-	+	shenlab.utk.edu/camerra.html	[60,61]
AQUA-DUCT	Ligand movement	-	+	Python	Geometry analysis	-	+	www.aquaduct.pl	[62,63]
CaverDock	Ligand movement	+	+	Python	Molecular docking	+	+	https://loschmidt.chemi.muni.cz/caverdock/	[64,65]

2.1. Interaction Network and Correlated Motion Analyses

Protein stability and function are dependent on their three-dimensional structures and are frequently conditioned by elaborate networks of noncovalent interactions between numerous residues [66]. Those networks undergo continuous dynamic changes by conformational rearrangement, which can be captured at atomic resolution using MD simulations [67–69] (Figure 2). Due to the inherent complexity in the detection and analysis of those changes, the simultaneous applications of several tools are frequently required. When enumerating a residue interaction network in an ensemble of protein structures from MD simulations, most of the available tools focus on coarse-grained networks consisting of Cα or Cβ atoms only [70,71]. To quantitatively explore the coordinated motions in the network, the use of principal component analysis (PCA)-based methods is considered an efficient strategy [72,73].
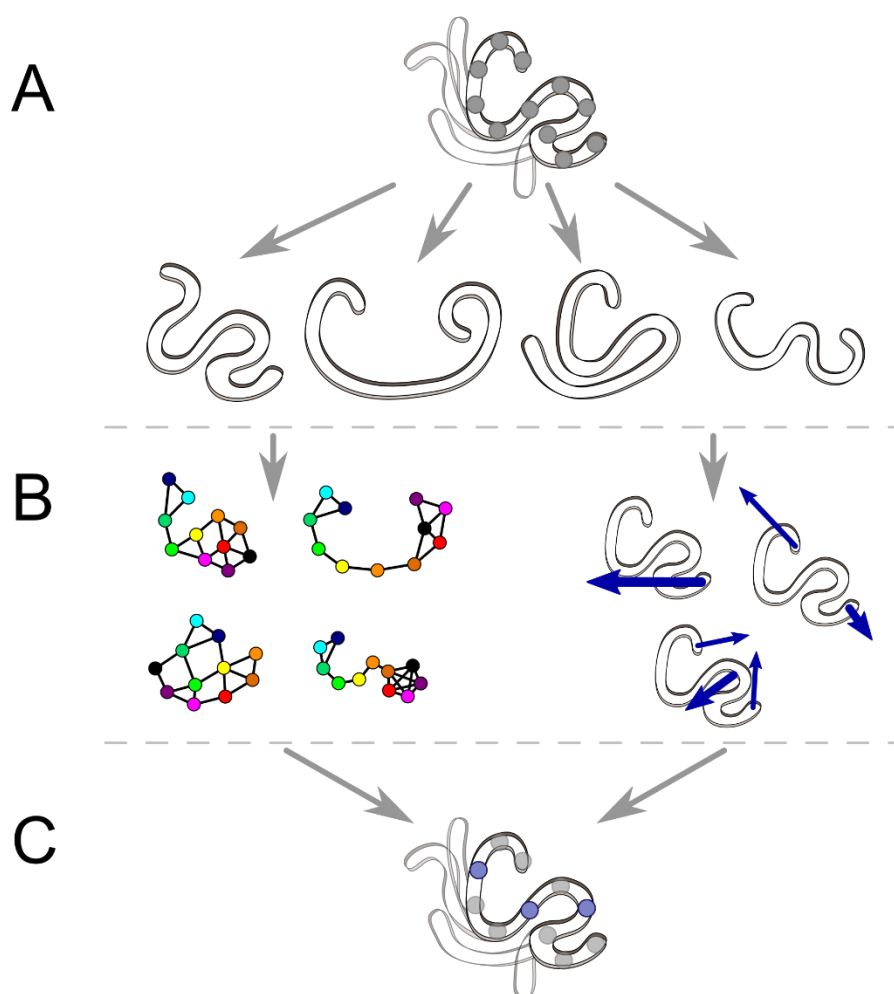

Figure 2. Predicting engineering hotspots for protein dynamics based on analyses of interaction networks and coordinated movements. (**A**) Functional protein dynamics can be represented by a conformational ensemble of a given protein. (**B**) This ensemble can be subjected to contact analysis to identify residue–residue interaction networks (left) or subjected to PCA to reveal coupled movements indicated by blue arrows right). (**C**) Either of these two approaches or their combination and hotspot residues (blue spheres) essential for the dynamics or allosteric communication can be selected for engineering.

To provide a comprehensive view of interactions, the residue interaction network in protein molecular dynamics (RIP-MD) tool was developed [58]. RIP-MD can detect different nonbonded interactions including hydrogen bonds, salt bridges, van der Waals, cation–π, π–π, arginine–arginine,

and Coulomb interactions. As an input, RIP-MD requires a static protein structure in a PDB format (web server) or MD trajectory in a DCD binary format (standalone and VMD plugin). The input is initially processed by removing heteroatoms, adding missing protein atoms and extracting parameters such as partial charges, Lennard-Jones parameters, secondary structure classification, and solvent accessibility. As an output, network files, including residue interaction networks for each interaction type and a combined network, are provided. The network files also store information about the secondary structure and the solvent accessibility. Furthermore, Pearson correlation plots are generated to detect possible behavior relationships between interacting residues. In a case study of the soluble myeloid differentiation-2 protein, RIP-MD was able to detect differences in interactions occurring in different conformational states, suggesting that the closing process increases the number of interactions and reduces the interaction correlations in the closed state. Further work is ongoing to broaden the capabilities of RIP-MD by accounting for interactions with nonprotein species [58]. This addition to the analysis will capture the effect of the environment and interactions with cognate ligands on proteins, which may be beneficial for protein engineering in particular.

A new software package, Java-based Essential Dynamics (JED), was developed to facilitate comparative PCAs of MD simulations of different proteins [59], including their apo- and holoforms, as well as wild-type and mutant variants. In the initial stage, the coarse-grained Cα atoms analysis of an ensemble, provided as PDB files, is performed to generate a pre-PCA output comprising a matrix of atomic coordinates, an overall root-mean-square deviation (RMSD), and an RMSD per residue. Then, the PCA of Cartesian-based coordinates, the PCA of internal distance pairs, or both analyses can be performed, optionally having less relevant modes and outlying PCA variables removed based on user-specified cutoffs. The output consists of files containing displacement vectors, covariance, correlation and partial correlation matrices, eigenvalues, and the most relevant principal components derived from the matrices. The analyses of both covariance and correlation are highly recommended, since they vary in the descriptions of collective motions concerning their amplitudes that are often sensitive to the mutation to a different degree. Finally, essential motions based on the matrices can be visualized, approximating the protein motions at various timescales. To compare dynamics among different proteins or different variants of the same protein, JED can compute cumulative overlaps, root-mean-square inner products, and principle angles. Depending on the degree of the overlaps in these features, the similarity in the protein dynamics can be established. As a case study, the authors analyzed 100 ns long simulations of a single-chain variable-fragment (scFv) antibody and its single-point mutant [59]. The detected disparities in correlation matrices, the PCA results, and the correlated residue pairs indicated that JED is sensitive enough to compare protein design evaluations [59].

Romero-Rivera and coworkers proposed a promising protocol combining information on residues proximity and their correlated movements into the so-called shortest path map (SPM), which can be applied to infer allosteric communication within a protein structure [43]. The first step in generating an SPM is the construction of a graph, in which Cα atoms of residues represent nodes and edges are drawn between pairs of nodes maintaining their distance below 5 Å in the whole MD trajectory. The edge lengths are then assigned based on the correlation coefficient between the connected Cα atoms in an inversed manner, i.e., larger coefficients result in shorter edges and vice versa. Next, the Dijkstra algorithm [74] is used to simplify the graph by identifying the shortest paths throughout the whole protein. Finally, pairs of residues that contribute the most to these paths are located representing central points for the communication. The SPM approach has been implemented in the DynaComm tool, and the development of a web server is ongoing [43]. By combining the SPM approach with PCA, the authors were able to identify the key positions that were previously mutated during the laboratory optimization of a computationally designed retro-aldolase by directed evolution [43]. This indicated rational design guided by SPM and PCA could help to identify distal mutations important for engineering of more efficient proteins akin to those produced by directed evolution experiments.

Similarly, by combining network analyses with PCA, the computation of allosteric mechanism by evaluating residue–residue associations (CAMERRA) tool aims to capture allosteric motions based on the residue–residue contact analysis of protein dynamics [60,61]. The CAMERRA tool is freely available as a set of Perl scripts. The required input for CAMERRA operation is an all-atom ensemble of diverse conformations of the investigated protein supplemented as a set of PDB files. In the beginning, residue–residue and residue–ligand contact matrices describing electrostatic, van der Waals, and hydrogen bond interactions are computed, resulting in contact matrices that are further condensed to form a mean contact matrix. Consequently, the mean contact matrix is exploited to generate a covariance matrix by computing the correlation between a pair of relevant contacts using a four-point correlation. Such an analysis may be able to capture crosstalk between the residues that lead to the formation or disruption of other contacts, therefore providing insight into the mechanisms of an allosteric network. Finally, a PCA is performed on the covariance matrix of the contacts, directly uncovering the displacement modes of the contacts (creations and disruptions), which is advantageous for understanding essential motions of biopolymers. This method was successfully applied to study several novel allosteric mechanisms including a frustrated fit mechanism and negative allostery in a retinoid X receptor complex [75] or the pressure activation of a lipase [76].

2.2. Analyses of Ligand Transport

Detailed tracking and analysis of ligand behavior across MD trajectories of biomolecular systems represent another strategy to enrich the protein design process by highlighting regions crucial for the transport of ligands, i.e., molecular tunnels, channels, and gates [25,77], which determine ligand associations and dissociation mechanisms [78,79]. In such a way, structural hotspot residues can be detected and considered during the protein engineering process to improve protein activity, change selectivity, or stability [80]. Readers interested in current approaches to simulations of ligand transport can refer to the recent expert review by Nunes-Alves and coworkers [81].

AQUA-DUCT [62,63] aims to provide detailed insights into the process of how a given type of molecules, such as water, ions, gasses, or any other kind of ligand, penetrates through the selected region of a protein (Figure 3A). As a minimal input, the user has to provide an MD trajectory and a configuration file describing two important regions for the analysis and defining the traced ligand. The first region is called a scope, which usually covers the whole protein. The second region is called an object and represents a functionally relevant region of interest, for example, the active site of an enzyme. An initial step in the workflow is to detect all traceable residues that reach the object and track their motions within the scope along the trajectory producing the so-called raw paths of ligands. Each path is then analyzed to identify possible repetitive events of a given ligand transiting between the object, scope, and surroundings, thereby dividing the raw paths to separate three types: (i) incoming path, (ii) outgoing path, and (iii) object for paths of ligands residing within the protein. In the following step, separate incoming and outgoing paths are assigned as inlets, i.e., paths connecting the exterior of the scope with the object region in any direction. Finally, the identified inlets are clustered, resulting in the pathways of the protein structure. Additionally, a statistical analysis is performed for all clusters, enumerating the number of the evaluated molecules, paths, inlets, and clusters, and several more specific statistics, including the lengths of the paths or the durations of the transport events. To illustrate the computational demands, the AQUA-DUCT analysis of 100 ns long MD simulations of murine epoxide hydrolase (4992 protein atoms) surrounded by 8488 water molecules requires 8–12 h to execute on a powerful workstation (Intel Core i7 CPU @ 3.50GHz machine, 64 GB RAM) [62]. For visualization purposes, a PyMOL [82] script or session can be generated according to user specifications. The presented method provides an efficient and robust way of detecting the usage of transport pathways in protein structures, including the detailed tracing of a specified ligand type, which is a challenging task, especially when considering thousands of water molecules in a trajectory composed of thousands of snapshots. In a follow-up study, the authors used MD simulations with AQUA-DUCT to examine the internal architecture of epoxide hydrolase from *Solanum tuberosum*, and

based on their experience, they designed a relatively straightforward protocol for the detailed analysis of cavities networks and tunnels capable of pinpointing hotspots for engineering experiments [83]. Such an approach was integrated into the engineering workflows of Subramanian and coworkers on cupin-type phosphoglucose isomerase from *Pyrococcus furiosus* [84] and D-amino acid oxidase (DAAO) [85]. In these studies, the tracking of ligands and water molecules with AQUA-DUCT helped to detected important features related to transport phenomena and to identify remote mutations governing the specificity and activity of these enzymes [84,85].
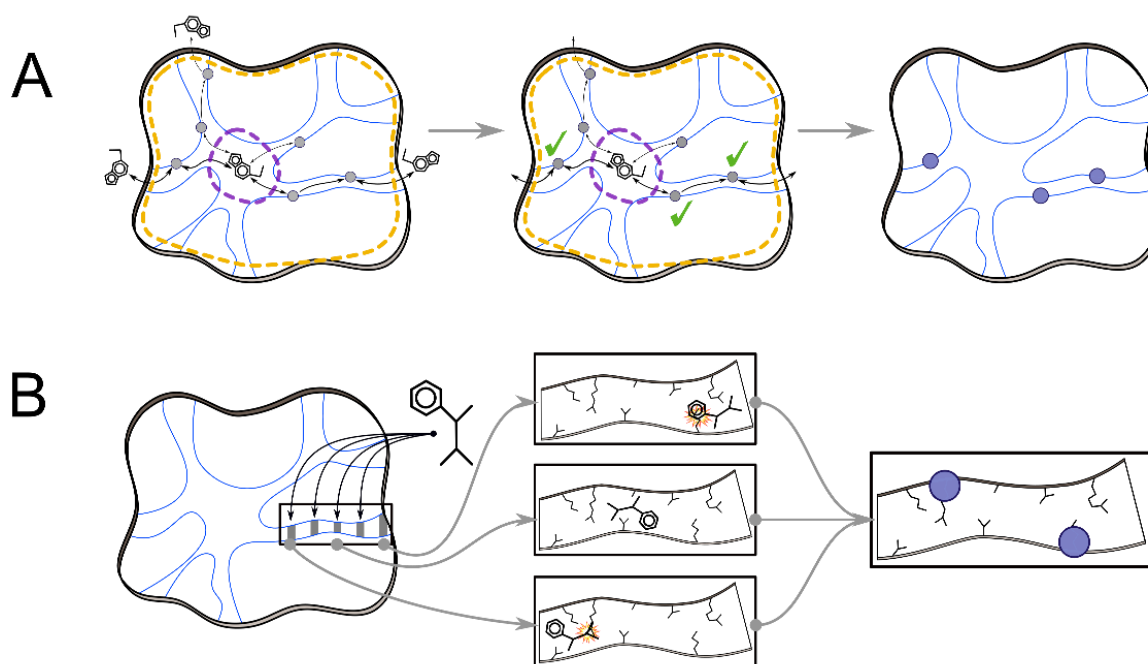

Figure 3. Hotspot detection based on ligand transport analyses. (**A**) AQUA-DUCT tool traces the movement of ligands via void spaces (blue lines) inside the scope region (dotted orange shapes) of the protein moiety throughout an MD trajectory. Only the ligands that reach the functionally important object region (dotted violet ellipses) are considered. The significance of the interactions of transported ligands with residues (grey spheres) along the ligand trajectory (black arrows) can be evaluated to select relevant hotspots (blue spheres) for the modification of the transport kinetics. (**B**) By iteratively docking the ligand along a molecular tunnel, CaverDock estimates the energy profile of a ligand transport, indicating residues that are most likely responsible for energy barriers in the path. These residues represent hotspots (blue spheres) for the design of new protein variants with altered ligand transport.

As an alternative to very costly explicit MD simulations, the passage of ligands through biomolecules can be explored by docking these ligands to an ensemble of precomputed molecular tunnels with CaverDock software [64,65] (Figure 3B). Benefiting from the fast operation of CaverDock calculation, it is possible to run the calculations over such an ensemble for multiple different ligands. For CaverDock operation, tunnels must be represented as sequences of spheres for each given conformation of a macromolecule. Such input data can be easily generated by CAVER 3.0 software [86]. The input spheres of each tunnel are then discretized into a set of discs, which represent planar constrains for the subsequent placement of a ligand with the AutoDock Vina molecular docking tool [87]. Such an approach is, however, inherently noncontinuous, as some bottlenecks can be avoided by the ligand abruptly changing its orientation and/or conformation. A solution to generate a fully continuous trajectory adopted by CaverDock is to restrict conformational changes of the ligand during its transition from one disk to the next. Since the more advanced approach accentuates unrealistically high-energy barriers due to the rigid-protein docking approach, CaverDock can also utilize the flexible docking procedure available in AutoDock Vina. Such flexibility is capable of opening the narrowest sections

of the investigated tunnels connected with the high-energy barriers, enabling the passage of various ligands via tunnels in cytochrome P450 17A1 and leukotriene A4 hydrolase/aminopeptidase [88]. Dealing with flexible residues during docking is more computationally demanding and should be used cautiously, as it can lead to the generation of the unrealistic conformation of flexible residues [65]. Marques et al. benchmarked the capabilities of CaverDock for protein engineering against predictions from sophisticated metadynamics, adaptive sampling, and funnel-metadynamics techniques [89]. In this detailed comparative study, the transport of ligands in two variants of haloalkane dehalogenase was investigated, and based on the analysis of energetic and structural bottlenecks, several residues playing a crucial role in the ligand-transport process were identified, some of them were previously mutated to engineer a very proficient biodegradator of a toxic anthropogenic pollutant 1,2,3-trichloropropane [90,91]. Overall, CaverDock reached good qualitative agreement with the rigorous MD simulations in this model system attesting its applicability for the engineering of ligand transport phenomena [89].

3. Advances in the Integration of Protein Flexibility into Protein Design and Redesign Methods

During the past few years, we have witnessed a surge in the efforts to develop novel design methods capable of robust treatments of protein dynamics (Table 2). These methods can be divided into the following three categories: (i) methods utilizing pregenerated molecular ensembles (Section 3.1; Figure 4A), (ii) knowledge-based approaches to generating more pronounced backbone perturbations effectively (Section 3.2; Figure 4B), and (iii) provable design algorithms with extended backbone flexibility (Section 3.3).
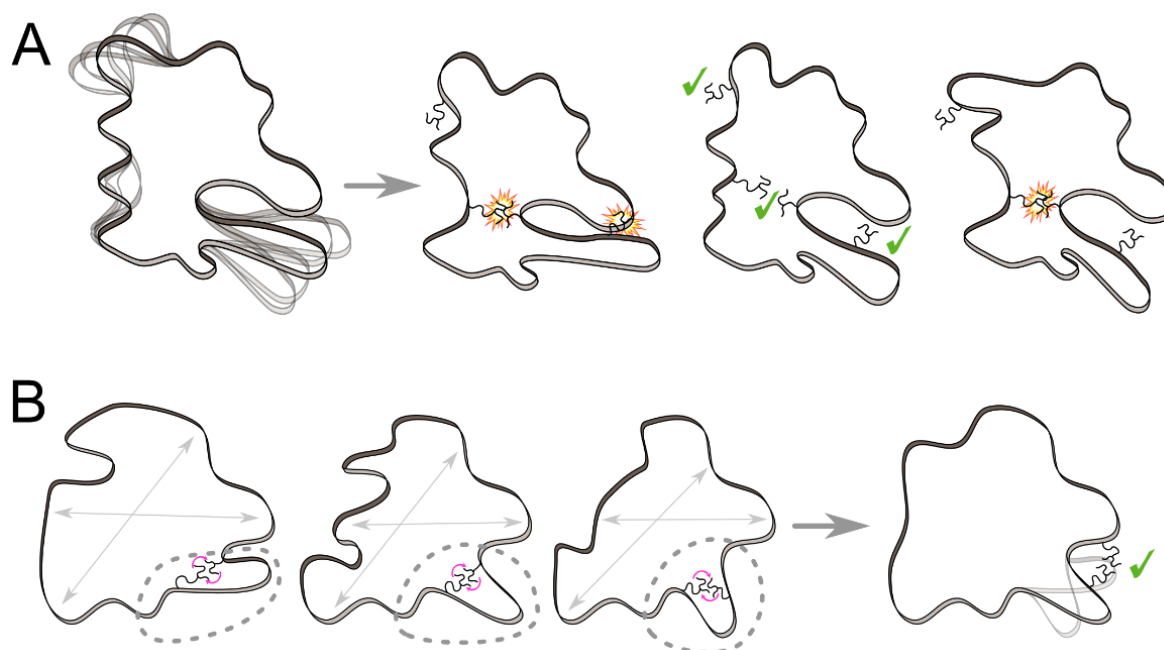

Figure 4. Flexible-backbone approaches facilitating the successful design of more diverse protein variants. (**A**) By employing a structural ensemble of a given protein, a larger variety of residues can be introduced to additional positions (green ticks), including those buried in the protein core, which would otherwise cause steric clashes (orange explosion-like shapes). (**B**) Data on protein dynamics encoded in different experimental structures or predicted ensembles can be extracted in the form of tertiary motifs (grey dotted circle) of interacting residues (pink arrows). Analogously, machine learning methods can learn and generalize the data to inspire novel backbone movements (grey arrows). The derived knowledge then enables the efficient application of more pronounced, yet physically correct, backbone perturbations during the design procedure.

Table 2. Computational protocols implementing protein flexibility for protein design and redesign.

Primary Package	Category	Method	Short Description	Input	Sampling of Side-Chain and Backbone Flexibility	Package	Add-Ons	Reference
Rosetta	Ensemble-based	Flex ddG	Estimating interface ΔΔG values upon mutation	Static structure	Backrub, torsion minimization, side-chain repacking	https://www.rosettacommons.org/software/	https://github.com/Kortemme-Lab/flex_ddG_tutorial	[92]
		Rosetta:MSF	Multistate framework using single-state protocols	Ensemble	Genetic algorithm based sequence optimizer and user-defined evaluator from Rosetta protocols	https://www.rosettacommons.org/software/	-	[93]
		Meta-multistate design (meta-MSD)	Engineering protein dynamics by meta-multistate design	Set of ensembles	Fast and accurate side-chain topology and energy refinement algorithm for sequence optimization; backbone-dependent rotamer library optimization for side-chains	https://www.rosettacommons.org/software/	PHOENIX scripts upon request	[94]
	Knowledge-based	Flexible backbone learning by Gaussian processes (FlexiBaL-GP)	Learning global protein backbone movements from multiple structures	Ensemble	Markov Chain Monte Carlo sampling—95% time spent on the side-chain selection and 5% time spent on the generation of the backbone movement	https://www.rosettacommons.org/software/	-	[95]
		Structural homology algorithm for protein design (SHADES)	Protein design guided by local structural environments from known structures	Static structure	Sequence assembly from fragments followed by backbone optimization, side-chains repacking, and structure relaxation	https://www.rosettacommons.org/software/	https://bitbucket.org/satsumaimo/shades/src/master/	[96]
OSPREY 3.0	Provable	Coordinates of atoms by Taylor series (CATS)	Enabling progressive backbone motions during protein design	Static structure	Continuous, strictly localized perturbations of the given segment of the backbone using a new internal coordinate system compatible with dead-end elimination workflows	https://github.com/donaldlab/OSPREY3	-	[97]

3.1. Ensemble-Based Approaches

The generation of molecular ensembles by using MD and Monte Carlo (MC) simulations has become more affordable for a wider group of users, creating a means to face novel protein design challenges. By utilizing conformational ensembles, protein design algorithms can take the dynamic nature of the protein structures into account, providing a biologically sound strategy and frequently improving the performance of the employed methods [98,99].

We start this section by reviewing insights from two studies aiming at benchmarking generic procedures for ensemble generation on the success of protein design or redesign tasks. In the first comparative research by Ludwiczak and colleagues, 10 protocols combining methods from Rosetta software [100] with MD simulations were applied to 12 diverse proteins [54]. For protein redesign, three distinct structural ensembles were obtained using MD simulation, MD simulation followed by the introduction of small backbone perturbations with Rosetta Backrub [101], or Rosetta Backrub alone. Subsequently, the protein sequences were redesigned using either the fixed backbone (FixBB) or design-and-relax (D&R) methods on each ensemble [102,103]. We note here that the employed simulations were run for four ns, although with 50 replicas, representing somewhat limited sampling around the conformational minima even though the target proteins were relatively small (up to 103 residues). The designed sequences were analyzed based on entropy, covariation, profile similarity, and packing quality in the corresponding generated structures. The best performance was observed for the protocol using MD simulation in combination with Rosetta Backrub for the ensemble generation, followed by redesign with the D&R method. This time, analogous protocols were tested for de novo design purposes using only the more efficient D&R method, confirming that the procedure based on the MD simulation coupled with Rosetta Backrub yielded the best results. In the second benchmarking study, Loshbaugh and Kortemme performed a comprehensive evaluation of four different flexible backbone design methods available within the Rosetta software using six datasets [104]. Comparing FastDesign [105,106], Backrub Ensemble Design [107], CoupledMoves with Backrub [52], and CoupledMoves with kinematic closure, the authors concluded that the CoupledMoves method performs better in recapitulating sequences of known proteins compared to the other two alternatives. This finding highlights the importance of incorporating the side-chain and backbone flexibility simultaneously during the design. Interestingly, all methods performed poorly on two deep sequencing datasets, which should be taken with caution when applying Rosetta for such purposes. Overall, both studies emphasize that flexible backbone approaches combined with side-chain flexibility can significantly outperform methods utilizing only a single conformation.

The predictive performance of the Flex ddG method in estimating the change in binding free energy after mutation ($\Delta\Delta G$) at protein–protein interfaces has also been boosted when using a structural ensemble instead of a single static structure [92]. In this method, an ensemble of up to 50 structures is generated by the conformational sampling in the surroundings of mutated sites with the Rosetta Backrub program. Then, the wild-type ensemble is optimized by repacking side-chains and performing energy minimization. To generate a mutant ensemble, the mutation of interest is introduced to each structure before conducting the analogous side-chain repacking and minimization. Finally, both ensembles are scored to calculate the ensemble-averaged $\Delta\Delta G$. The method was validated using the ZEMu dataset of 1240 mutations [108] derived from the SKEMPI database [109]. For this dataset, the Flex ddG method reached a Pearson correlation coefficient (PCC) of 0.63 and an average absolute error of 0.96 Rosetta energy units. The enhanced performance was especially prominent in the case of small-to-large mutations, emphasizing that backbone flexibility constitutes a key factor during the modeling of these mutations. Relevant improvements were also achieved for modeling stabilizing mutations and mutating antibody–antigen interfaces. Interestingly, the enhanced performance over a fixed backbone approach was observed already when averaging over 20–30 conformations, a relatively low number in contrast to by previous ensemble-based methods, for which thousands of structural models were required [110].

Notably, the Flex ddG method was evaluated in three comprehensive benchmarking studies focusing on different engineering scenarios. Aldeghi and coworkers evaluated alchemical free-energy calculations and three Rosetta protocols including Flex ddG in combination with different force fields for the prediction of changes in binding the affinity of ligands upon mutation [111]. In total, 134 mutations were considered for 27 ligands and 17 proteins, showing that Flex ddG can reach quantitative agreement with such experimental data with a root-mean-square error (RMSE) of 1.46 kcal/mol and a PCC of 0.25, which was on par with the best performing alchemical calculations (an RMSE of 1.39 kcal/mol and a PCC of 0.43) [111]. At this point, it is worth comparing the computational resources required for such predictions. The alchemical calculations were reported to take two to five days using 20 CPU threads and one GPU, while Flex ddG computations were usually finished within a day on a single CPU core [111]. The same author collective also evaluated the utilization of these methods for the prediction of 31 drug resistance-conferring mutations for eight tyrosine kinase inhibitors of human kinase ABL [112]. For this dataset, Flex ddG was found to be highly accurate with an RMSE of 0.72 kcal/mol and a PCC of 0.67, even outperforming the much more demanding alchemical calculations [112]. Interestingly, significant improvements in ΔΔG prediction could be reached with a consensus of predictions from Flex ddG and alchemical calculations in both studies [111,112]. Another comparative study investigated the performance of five predictive tools when applied for alanine scanning to identify hotspot residues at protein–protein interfaces [113]. For a dataset of 748 single-point mutations to alanine from the SKEMPI database, Flex ddG ranked the best (PCC of 0.51) from the tools that were not trained using this database [113].

The advantages of incorporating conformational ensembles during design have also been noted during the development of a multistate framework that enables the adoption of reliable methods implemented in the Rosetta package for single-state design (SSD) and also for multistate design (MSD) [93]. Briefly explaining the mode of action, the input for the framework consists of a set of multiple states (structural conformations) and the population of sequences generated by randomly introduced single-point mutations, which are processed and altered by a genetic algorithm. Next, each sequence–state pair is evaluated and scored based on the Rosetta SSD protocol of the user's choice. The score of each sequence are communicated back to a sequence optimizer to perform the next iteration, until the fitness criteria are satisfied, finally giving a population of the optimized sequences. This is opposite to the standard SSD, which uses an MC algorithm and produces only a single sequence. The performance of MSD was evaluated on several design perspectives. Firstly, the performances of MSD and SSD in the task of recapitulating the binding site in the human intestinal fatty acid-binding protein was compared utilizing its ensemble obtained by NMR spectroscopy. Here, the SSD approach was used separately for each conformation, while the MSD was run on the whole ensemble at once. The MSD procedure achieved higher average native sequence recovery (NSR) and native sequence similarity recovery (NSSR) rates. Additionally, de novo ligand-binding design was performed for 16 proteins using SSD and MSD, where conformational ensembles of 20 and 1000 structures were generated by the Rosetta Backrub algorithm and a 10 ns long MD simulation, respectively. In this comparison, the MSD approach primarily produced sequences with higher NSR and NSSR rates and slightly lower energies, proving the advantages of the ensemble utilization. Interestingly, the quality of the designs originating from Rosetta Backrub and MD simulations were comparable, even though the mean Cα RMSDs over the ensembles differed notably, which were 0.17 and 0.62 Å, respectively. Finally, the multistate framework was tested by introducing retro-aldolase activity into protein scaffolds, which revealed nine proteins with experimentally confirmed activities [93].

A similar idea of combining an ensemble-based design and a multistate approach was behind the development of a meta-multistate design procedure (meta-MSD) to rationally design proteins that spontaneously switch between conformational states [94]. In this case, the procedure started with the generation of an ensemble of backbone templates with Rosetta Backrub and PertMin approaches [99,114] to cover the conformational landscape, including all transition states of interest. Next, the whole ensemble was split into microstates that were energy-minimized. Then, these microstates were assigned

to major, transition, and minor states based on their structural features. Finally, the sequences expected to transit between the states were identified based on their relative energies. Based on meta-MSD, several Streptococcal protein G domain β1 variants were engineered to obtain structures that can exchange conformations between two states spontaneously, producing experimentally validated protein exchangers capable of switching between the states on a millisecond timescale [94], thereby highlighting the importance of the accurate modeling of a local energy landscape for designing protein dynamics.

3.2. Knowledge-Based Approaches

Following the expansion of protein structure databases, which contain a considerable amount of data related to structure–dynamics–function relationships in proteins, new methods to assess backbone flexibility have been designed, benefiting from this wealth of knowledge. The methods introduced here are implemented in the Rosetta software and represent an exciting direction for improving protein design processes by more efficiently exploring alternative backbone conformations.

The first among the reviewed data-driven approaches is the flexible backbone learning by Gaussian processes (FlexiBaL-GP) method [95] that uses multiple structures of a given protein to learn the most probable global backbone movements specific for training structures using the Gaussian process latent variable model as a machine learning method. These learned movements are then applied to guide the search for proteins with alternative backbone conformations by Markov Chain Monte Carlo sampling, where at each step 95% of the time is spent on the selection of the optimal side-chain rotamers and 5% of the time is spent on the generation of the protein backbone movements. FlexiBaL-GP can utilize various sources of training data including X-ray structures, NMR models, and MD simulations. When learning from a set of 28 crystal structures of ubiquitin and using two latent variables, the FlexiBal-GP method generated an ensemble of structures for native ubiquitin with an RMSD range of 0.5–0.65 Å from a reference structure. Notably, the ensemble recovered over 40% of the conformational diversity of the ensemble obtained by NMR spectroscopy. Moreover, the method's ability to enrich a library of ubiquitin variants towards those with improved affinity to ubiquitin carboxyl-terminal hydrolase 21 was evaluated. For this task, the FlexiBal-GP method was trained on two wild-type complexes only or combined with either a structure of a tightly binding mutant or MD-based ensembles starting from the two wild-type structures. All three derived models outperformed flexible designs with Rosetta Backrub, as well as designs based on ensembles generated with MD simulations and the constraint-based method, CONCOORD [115].

A different approach to harnessing knowledge from structural databases and to navigating sequence space sampling with a flexible backbone has been explored by the structural homology algorithm for protein design (SHADES) [96]. This approach relies on the libraries of In-contact amino acid residue TErtiary Motifs (ITEMs) derived from curated protein structures, in which local contacts were analyzed for each residue. Analogously, target ITEMs are then identified for each position in the target structure in a position-specific manner and matched to the ITEMs database in order to generate candidate ITEMs libraries. Finally, these libraries are exploited by an iterative population-based optimization method that substitutes all residues in each target ITEM position with all residues from a candidate ITEM. The structure of the altered fragment is then adjusted by optimizing its backbone with the Rosetta Backrub method, repacking the side-chains and minimizing or relaxing the whole structure with or without backbone restraints. Using a dataset of 40 proteins from different families, the SHADES performance in recovering the native sequences of the proteins was evaluated, reaching a 30% average sequence recovery and a 46% sequence similarity between the designed and natural proteins, when candidate ITEMs derived from homologous proteins were excluded. When the homologs were retained in the candidate libraries, the sequence recovery rate increased up to 93%. Notably, rather large conformational diversity was observed for the successfully designed models, in some instances exhibiting more than a 1 Å RMSD from their respective native structures. Overall, these tests indicated

that SHADES could capture sequence–dynamics–structure relationships correctly while spending about 25 times less CPU time than the redesign mode of the Rosetta FastRelax method [116].

3.3. Provable Algorithms

Due to the high complexity of protein design tasks, especially when employing ensemble-based approaches (Section 3.1), the majority of the tools rely on heuristic algorithms as an expedient way to obtain the desired constructs. For more complicated tasks, these approaches are often barred from generating optimal solutions, which in turn can lead to the design of sequences that are not guaranteed to have the lowest energy [117]. In response to those limitations, provable algorithms have been developed, creating a promising alternative for reaching entrenched solutions [117,118]. Here, we briefly outline some of the most compelling developments that led to an advanced description of backbone flexibility. For a more comprehensive overview of provable algorithms and their evolution and application, please see the very insightful reviews published recently [119,120].

The development of provable algorithms started with the adaptation of the dead-end elimination (DEE) method [121] that was later improved by introducing rotamers' minimization before pruning to enable a more continuous description of side-chains, an essential component of several successful designs [118,122]. The initial approach to backbone flexibility was introduced with the dead-end elimination with perturbations (DEEPer) method [123], relying on a predefined set of small local movements extracted from an experimental structure such as Backrub [124] or sheer. However, such motions are mostly restricted to subangstrom dimensions to avoid disruptive changes propagated to a distant region from the segment of the altered backbone. To enable more progressive motions in a predefined contiguous part of the backbone such as the movement of a flexible loop, the coordinates of atoms by Taylor series (CATS) approach was recently introduced [97]. The main idea of the approach lies in the new definition of the backbone internal coordinate system, which enables physically sensible, continuous, and strictly localized perturbations of the given segment of the backbone in a manner that is compatible with the advanced DEE workflows. The CATS method was tested on 28 different proteins with flexible backbone treatment enabled for five to nine-residue long segments. By introducing more pronounced changes in backbone conformations, almost 0.2 Å on average, CATS reached a mean improvement in design energies of 3.5 kcal/mol in comparison to the rigid-backbone approximation. Such an improvement is nearly twice as large as what was observed previously for restricted backbone perturbations introduced by the DEEPer method on the same set.

Owing to persistent optimization efforts [125–128], provable algorithms can nowadays be applied for protein design while simultaneously employing both the continuous flexibility of side-chains and enhanced backbone flexibility efficiently at similar computational costs to more rigid approaches. These methods are available in OSPREY 3.0 [129], in which the analysis speed has been further promoted by the newly supported use of GPUs and multicore CPUs for some of the modeling tasks, which were prohibitively complicated for the previous version of the software. As underlined by several studies featuring various applications of provable algorithms [130–133], these algorithms have matured enough to be of practical utility for protein engineers. This trend will undoubtedly gain further momentum with the recent developments discussed herein, even though their computational demands might still be limiting for some applications.

4. Conclusions, Challenges, and Perspectives

In contrast with proteins evolved through directed evolution, constructs predicted by computational protein engineering methods have so far been focusing mainly on hotspot residues close to functional sites. By considering the proximity of relevant regions, mutations have the highest chance of altering the target function, and at the same time, the number of variants to evaluate is kept tractable. Unfortunately, this restriction often hampers the performance of rationally designed proteins. It is clear that we need more efficient workflows and tools that can pinpoint hotspots at crucial distal sites as well. One class of such hotspots involves residues forming allosteric networks capable of inducing a

shift in populations of protein conformations to support their altered function upon mutation. Here, we would like to highlight the availability of tools for rapid analyses of protein allostery focusing on residue–residue interactions in a single static structure or employing normal mode analysis (NMA) to approximate protein dynamics [134]. However, the performances of these approximate tools are often impeded by two factors: (i) the quality of a single-input structure and the extent, to which this structure represents essential interactions present in the conformational ensemble, and (ii) the limited sensitivity of underlying NMA to mutations that do not produce substantial conformational changes [135]. Those limitations are inherently overcome by ensemble-based approaches, in which network analyses of MD simulations are facilitated by the tools discussed in Section 2.1. The second class of remote hotspots is connected with ligand transport, a phenomenon that is hard to tackle due to its rare nature, which in turn requires extensive sampling. Currently, there are tools suitable for robust analyses of transport events captured by MD simulations and tools capable of the efficient exploration of a precomputed ensemble of transport tunnels in proteins by multiple ligands (Section 2.2). However, there is still a gap to close, before we can rationally design mutations enhancing ligand transport. In particular, effective means to predict how the ligand presence alters the dynamics of transport pathways to factor in ligand-specific effects of mutations [136] still have to be developed together with more efficient methods to sample the passage of ligands through structural ensembles of proteins.

Throughout this review, we have witnessed a consistent success of methods incorporating different degrees of protein dynamics in increasing the accuracy of their predictions owing to the innate ensemble nature of the proteins. These methods frequently require user expertise in complicated computational methods and protocols. Considering that some of fully automated and easy-to-use methods available nowadays originate from very sophisticated and computationally extensive approaches [137–141] and the ongoing rapid development of powerful technologies, in synergy with research on more efficient algorithms, we perceive recent advanced methods and algorithms reviewed here as heralded future automated methods accessible not only to specialists but also to researchers with much broader expertise. As various flexible backbone approaches will, due to their upcoming maturity and indisputable benefits, be gradually joining the mainstream protein design methods, the involvement of dynamics in engineering processes is likely to reveal new challenges to overcome.

First, the successful utilization of molecular ensembles in protein design and redesign is dependent on the quality of input ensembles emphasizing the importance of sufficient and representative sampling. Since this is not a trivial task, but rather an art itself, the ensemble-based approaches reviewed here employ limited sampling. Despite sampling somewhat restricts conformational changes in protein backbones, these approaches achieve substantial advantages over the predictions relying on a single structure. The systematic utilization of a more extensive sampling via much longer, enhanced, or adaptive simulations will be required to thoroughly describe more global conformational transitions [27–31]. Alternatively, with further expansion of the PDB database, the knowledge-based methods similar to those reviewed in Section 3.2 might be trained from data on particular proteins and families, hence providing more global, yet robust, moves compatible with a given fold to be considered during the design. Additionally, there is still largely unexplored potential to derive such system-specific moves from extensive MD simulations that have been shown to recapitulate the conformational behavior of many structured proteins [40,142].

Second, with the increasing amplitude of introduced perturbations, the protein structures will more frequently be drawn from the conformational space further away from the structures produced by protein crystallography. Following the precedent of unsatisfactory performance observed for simulations of intrinsically disordered proteins using standard force fields, which were developed for folded and stable protein structures [143,144], to what degree all energy terms of currently employed scoring functions will be applicable for the ranking of very flexible designs remains to be seen. In parallel, it is evident that the flexible-backbone approaches are more successful in introducing the bulkier and often more hydrophobic residues. This success, however, accentuates a well-known tendency of design methods to improve hydrophobic packing but not polar interaction networks,

since hydrophobic interactions are more straightforward to sample than directional polar ones [145], which regularly results in the problematic solubility of the design proteins. To help to reverse this trend, the utilization of methods for the efficient prediction of hydrogen bond networks, akin to the recently developed MC HBNet protocol [146], would be required, especially when coupled with more continuous descriptions of side-chains to increase the number of accessible solutions.

Author Contributions: Conceptualization, B.S., C.E.S.-B., and J.B.; investigation, B.S., C.E.S.-B., and J.B.; writing of the manuscript, B.S. and J.B.; editing of the manuscript, C.E.S.-B.; visualization, C.E.S.-B.; supervision, J.B.; funding acquisition, J.B. All authors have read and agreed to the published version of the manuscript.

Abbreviations

MD	molecular dynamics
GPU	graphics processing unit
PCA	principal component analysis
RIP-MD	residue interaction network in protein molecular dynamics
VMD	visual molecular dynamics
JED	Java-based Essential Dynamics
scFv	single-chain variable-fragment
PDB	protein data bank
RMSD	root-mean-square deviation
SPM	shortest path map
CAMERRA	computation of allosteric mechanism by evaluating residue–residue associations
DAAO	D-amino acid oxidase
MC	Monte Carlo
FixBB	fixed backbone
D&R	Design-and-relax
PCC	Pearson correlation coefficient
$\Delta\Delta G$	change in binding free energy
RMSE	root-mean-square error
SSD	single-state design
MSD	multistate design
NMR	nuclear magnetic resonance
NSR	native sequence recovery
NSSR	native sequence similarity recovery
FlexiBaL-GP	flexible backbone learning by Gaussian processes
SHADES	structural homology algorithm for protein design
ITEM	In-contact amino acid residue tertiary motif
DEE	dead-end elimination
DEEPer	dead-end elimination with perturbations
CATS	coordinates of atoms by Taylor series
NMA	normal mode analysis

References

1. Kirk, O.; Borchert, T.V.; Fuglsang, C.C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **2002**, *13*, 345–351. [CrossRef]
2. Bodansky, O. Diagnostic applications of enzymes in medicine. General enzymological aspects. *Am. J. Med.* **1959**, *27*, 861–874. [CrossRef]

3. Singh, R.; Kumar, M.; Mittal, A.; Mehta, P.K. Microbial enzymes: Industrial progress in 21st century. *3 Biotech* **2016**, *6*, 174. [CrossRef]

4. Sizer, I.W. Medical Applications of Microbial Enzymes. *Adv. Appl. Microbiol.* **1972**, *15*, 1–11. [CrossRef]

5. Piotrowska-Długosz, A. Significance of Enzymes and Their Application in Agriculture. *Biocatalysis* **2019**, 277–308. [CrossRef]

6. Brannigan, J.A.; Wilkinson, A.J. Protein engineering 20 years on. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 964–970. [CrossRef] [PubMed]

7. Bornscheuer, U.T.; Huisman, G.W.; Kazlauskas, R.J.; Lutz, S.; Moore, J.C.; Robins, K. Engineering the third wave of biocatalysis. *Nature* **2012**, *485*, 185–194. [CrossRef] [PubMed]

8. Kazlauskas, R.J.; Bornscheuer, U.T. Finding better protein engineering strategies. *Nat. Chem. Biol.* **2009**, *5*, 526–529. [CrossRef]

9. Arnold, F.H. Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew. Chem. Int. Ed.* **2019**, *58*, 14420–14426. [CrossRef] [PubMed]

10. Arnold, F.H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem. Int. Ed.* **2018**, *57*, 4143–4148. [CrossRef] [PubMed]

11. Wilkinson, A.J.; Fersht, A.R.; Blow, D.M.; Carter, P.; Winter, G. A large increase in enzyme-substrate affinity by protein engineering. *Nature* **1984**, *307*, 187–188. [CrossRef] [PubMed]

12. Wells, J.A.; Powers, D.B.; Bott, R.R.; Graycar, T.P.; Estell, D.A. Designing substrate specificity by protein engineering of electrostatic interactions. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 1219–1223. [CrossRef] [PubMed]

13. Thomas, P.G.; Russell, A.J.; Fersht, A.R. Tailoring the pH dependence of enzyme catalysis using protein engineering. *Nature* **1985**, *318*, 375–376. [CrossRef]

14. Barrozo, A.; Borstnar, R.; Marloie, G.; Kamerlin, S.C.L. Computational Protein Engineering: Bridging the Gap between Rational Design and Laboratory Evolution. *Int. J. Mol. Sci.* **2012**, *13*, 12428–12460. [CrossRef] [PubMed]

15. Hellinga, H.W. Computational protein engineering. *Nat. Struct. Biol.* **1998**, *5*, 525–527. [CrossRef]

16. Wijma, H.J.; Janssen, D.B. Computational design gains momentum in enzyme catalysis engineering. *FEBS J.* **2013**, *280*, 2948–2960. [CrossRef]

17. Looger, L.L.; Dwyer, M.A.; Smith, J.J.; Hellinga, H.W. Computational design of receptor and sensor proteins with novel functions. *Nature* **2003**, *423*, 185–190. [CrossRef]

18. Saven, J.G. Computational protein design: Engineering molecular diversity, nonnatural enzymes, nonbiological cofactor complexes, and membrane proteins. *Curr. Opin. Chem. Biol.* **2011**, *15*, 452–457. [CrossRef]

19. Huang, P.-S.; Boyken, S.E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320. [CrossRef]

20. Frauenfelder, H.; Sligar, S.G.; Wolynes, P.G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598–1603. [CrossRef]

21. Agarwal, P.K. Enzymes: An integrated view of structure, dynamics and function. *Microb. Cell Fact.* **2006**, *5*, 2. [CrossRef] [PubMed]

22. Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964–972. [CrossRef] [PubMed]

23. Gáspári, Z.; Perczel, A. Protein Dynamics as Reported by NMR. *Annu. Rep. NMR Spectrosc.* **2010**, *71*, 35–75. [CrossRef]

24. Lewandowski, J.R.; Halse, M.E.; Blackledge, M.; Emsley, L. Direct observation of hierarchical protein dynamics. *Science* **2015**, *348*, 578–581. [CrossRef] [PubMed]

25. Gora, A.; Brezovsky, J.; Damborsky, J. Gates of enzymes. *Chem. Rev.* **2013**, *113*, 5871–5923. [CrossRef] [PubMed]

26. Kokkonen, P.; Sykora, J.; Prokop, Z.; Ghose, A.; Bednar, D.; Amaro, M.; Beerens, K.; Bidmanova, S.; Slanska, M.; Brezovsky, J.; et al. Molecular Gating of an Engineered Enzyme Captured in Real Time. *J. Am. Chem. Soc.* **2018**, *140*, 17999–18008. [CrossRef]

27. Pierce, L.C.T.; Salomon-Ferrer, R.; Augusto, F.; De Oliveira, C.; McCammon, J.A.; Walker, R.C. Routine access to millisecond time scale events with accelerated molecular dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002. [CrossRef]

28. Noé, F. Beating the Millisecond Barrier in Molecular Dynamics Simulations. *Biophys. J.* **2015**, *108*, 228–229. [CrossRef]

29. Sultan, M.M.; Denny, R.A.; Unwalla, R.; Lovering, F.; Pande, V.S. Millisecond dynamics of BTK reveal kinome-wide conformational plasticity within the apo kinase domain. *Sci. Rep.* **2017**, *7*, 15604. [CrossRef]

30. Silva, D.A.; Weiss, D.R.; Avila, F.P.; Da, L.T.; Levitt, M.; Wang, D.; Huang, X. Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7665–7670. [CrossRef]

31. Salomon-Ferrer, R.; Götz, A.W.; Poole, D.; Le Grand, S.; Walker, R.C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888. [CrossRef] [PubMed]

32. Li, P.; Merz, K.M. Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J. Chem. Theory Comput.* **2014**, *10*, 289–297. [CrossRef] [PubMed]

33. Jing, Z.; Liu, C.; Cheng, S.Y.; Qi, R.; Walker, B.D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annu. Rev. Biophys.* **2019**, *48*, 371–394. [CrossRef] [PubMed]

34. Mongan, J.; Case, D.A. Biomolecular simulations at constant pH. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157–163. [CrossRef]

35. Panteva, M.T.; Giambaşu, G.M.; York, D.M. Comparison of structural, thermodynamic, kinetic and mass transport properties of Mg $^{2+}$ ion models commonly used in biomolecular simulations. *J. Comput. Chem.* **2015**, *36*, 970–982. [CrossRef]

36. Wang, A.; Zhang, Z.; Li, G. Higher Accuracy Achieved in the Simulations of Protein Structure Refinement, Protein Folding, and Intrinsically Disordered Proteins Using Polarizable Force Fields. *J. Phys. Chem. Lett.* **2018**, *9*, 7110–7116. [CrossRef]

37. Dobrev, P.; Vemulapalli, S.P.B.; Nath, N.; Griesinger, C.; Grubmüller, H. Probing the accuracy of explicit solvent constant pH molecular dynamics simulations for peptides. *J. Chem. Theory Comput.* **2020**. [CrossRef]

38. Smith, L.G.; Tan, Z.; Spasic, A.; Dutta, D.; Salas-Estrada, L.A.; Grossfield, A.; Mathews, D.H. Chemically Accurate Relative Folding Stability of RNA Hairpins from Molecular Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 6598–6612. [CrossRef]

39. Heo, L.; Feig, M. Experimental accuracy in protein structure refinement via molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 13276–13281. [CrossRef]

40. Tian, C.; Kasavajhala, K.; Belfon, K.A.A.; Raguette, L.; Huang, H.; Migues, A.N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; et al. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528–552. [CrossRef]

41. Childers, M.C.; Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Mol. Syst. Des. Eng.* **2017**, *2*, 9–33. [CrossRef] [PubMed]

42. Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. Computational tools for the evaluation of laboratory-engineered biocatalysts. *Chem. Commun.* **2017**, *53*, 284–297. [CrossRef] [PubMed]

43. Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **2017**, *7*, 8524–8532. [CrossRef]

44. Pabis, A.; Risso, V.A.; Sanchez-Ruiz, J.M.; Kamerlin, S.C. Cooperativity and flexibility in enzyme evolution. *Curr. Opin. Struct. Biol.* **2018**, *48*, 83–92. [CrossRef] [PubMed]

45. Buller, A.R.; Van Roye, P.; Cahn, J.K.B.; Scheele, R.A.; Herger, M.; Arnold, F.H. Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. *J. Am. Chem. Soc.* **2018**, *140*, 7256–7266. [CrossRef] [PubMed]

46. Petrović, D.; Lynn, K.S.C. Molecular modeling of conformational dynamics and its role in enzyme evolution. *Curr. Opin. Struct. Biol.* **2018**, *52*, 50–57. [CrossRef]

47. Maria-Solano, M.A.; Serrano-Hervás, E.; Romero-Rivera, A.; Iglesias-Fernández, J.; Osuna, S. Role of conformational dynamics in the evolution of novel enzyme function. *Chem. Commun.* **2018**, *54*, 6622–6634. [CrossRef]

48. Jiménez-Osés, G.; Osuna, S.; Gao, X.; Sawaya, M.R.; Gilson, L.; Collier, S.J.; Huisman, G.W.; Yeates, T.O.; Tang, Y.; Houk, K.N. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat. Chem. Biol.* **2014**, *10*, 431–436. [CrossRef]

49. Yang, B.; Wang, H.; Song, W.; Chen, X.; Liu, J.; Luo, Q.; Liu, L. Engineering of the Conformational Dynamics of Lipase to Increase Enantioselectivity. *ACS Catal.* **2017**, *7*, 7593–7599. [CrossRef]

50. Hong, N.S.; Petrović, D.; Lee, R.; Gryn'ova, G.; Purg, M.; Saunders, J.; Bauer, P.; Carr, P.D.; Lin, C.Y.; Mabbitt, P.D.; et al. The evolution of multiple active site configurations in a designed enzyme. *Nat. Commun.* **2018**, *9*, 1–10. [CrossRef]

51. Campbell, E.; Kaltenbach, M.; Correy, G.J.; Carr, P.D.; Porebski, B.T.; Livingstone, E.K.; Afriat-Jurnou, L.; Buckle, A.M.; Weik, M.; Hollfelder, F.; et al. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **2016**, *12*, 944–950. [CrossRef] [PubMed]

52. Ollikainen, N.; de Jong, R.M.; Kortemme, T. Coupling Protein Side-Chain and Backbone Flexibility Improves the Re-design of Protein-Ligand Specificity. *PLoS Comput. Biol.* **2015**, *11*, e1004335. [CrossRef] [PubMed]

53. Sevy, A.M.; Jacobs, T.M.; Crowe, J.E.; Meiler, J. Design of Protein Multi-specificity Using an Independent Sequence Search Reduces the Barrier to Low Energy Sequences. *PLoS Comput. Biol.* **2015**, *11*, e1004300. [CrossRef] [PubMed]

54. Ludwiczak, J.; Jarmula, A.; Dunin-Horkawicz, S. Combining Rosetta with molecular dynamics (MD): A benchmark of the MD-based ensemble protein design. *J. Struct. Biol.* **2018**, *203*, 54–61. [CrossRef]

55. Dawson, W.M.; Rhys, G.G.; Woolfson, D.N. Towards functional de novo designed proteins. *Curr. Opin. Chem. Biol.* **2019**, *52*, 102–111. [CrossRef]

56. Marcos, E.; Silva, D.A. Essentials of de novo protein design: Methods and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1374. [CrossRef]

57. Kuhlman, B.; Bradley, P. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 681–697. [CrossRef]

58. Contreras-Riquelme, S.; Garate, J.; Perez-Acle, T.; Martin, A.J.M. RIP-MD: A tool to study residue interaction networks in protein molecular dynamics. *PeerJ* **2018**, *6*, e5998. [CrossRef]

59. David, C.C.; Singam, E.R.A.; Jacobs, D.J. JED: A Java Essential Dynamics Program for comparative analysis of protein trajectories. *BMC Bioinform.* **2017**, *18*, 271. [CrossRef]

60. Johnson, Q.R.; Lindsay, R.J.; Shen, T. CAMERRA: An analysis tool for the computation of conformational dynamics by evaluating residue-residue associations. *J. Comput. Chem.* **2018**, *39*, 1568–1578. [CrossRef]

61. Lindsay, R.J.; Siess, J.; Lohry, D.P.; McGee, T.S.; Ritchie, J.S.; Johnson, Q.R.; Shen, T. Characterizing protein conformations by correlation analysis of coarse-grained contact matrices. *J. Chem. Phys.* **2018**, *148*, 025101. [CrossRef] [PubMed]

62. Magdziarz, T.; Mitusińska, K.; Gołdowska, S.; Płuciennik, A.; Stolarczyk, M.; Lugowska, M.; Góra, A. AQUA-DUCT: A ligands tracking tool. *Bioinformatics* **2017**, *33*, 2045–2046. [CrossRef]

63. Magdziarz, T.; Mitusińska, K.; Bzówka, M.; Raczyńska, A.; Stańczak, A.; Banas, M.; Bagrowska, W.; Góra, A. AQUA-DUCT 1.0: Structural and functional analysis of macromolecules from an intramolecular voids perspective. *Bioinformatics* **2019**. [CrossRef]

64. Filipovic, J.; Vavra, O.; Plhak, J.; Bednar, D.; Marques, S.M.; Brezovsky, J.; Matyska, L.; Damborsky, J. CaverDock: A Novel Method for the Fast Analysis of Ligand Transport. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **2019**, 1. [CrossRef]

65. Vavra, O.; Filipovic, J.; Plhak, J.; Bednar, D.; Marques, S.M.; Brezovsky, J.; Stourac, J.; Matyska, L.; Damborsky, J. CaverDock: A molecular docking-based tool to analyse ligand transport through protein tunnels and channels. *Bioinformatics* **2019**, *35*, 4986–4993. [CrossRef]

66. Pace, C.N.; Martin Scholtz, J.; Grimsley, G.R. Forces stabilizing proteins. *FEBS Lett.* **2014**, *588*, 2177–2184. [CrossRef]

67. Feher, V.A.; Durrant, J.D.; Van Wart, A.T.; Amaro, R.E. Computational approaches to mapping allosteric pathways. *Curr. Opin. Struct. Biol.* **2014**, *25*, 98–103. [CrossRef] [PubMed]

68. Dokholyan, N.V. Controlling Allosteric Networks in Proteins. *Chem. Rev.* **2016**, *116*, 6463–6487. [CrossRef] [PubMed]

69. Wodak, S.J.; Paci, E.; Dokholyan, N.V.; Berezovsky, I.N.; Horovitz, A.; Li, J.; Hilser, V.J.; Bahar, I.; Karanicolas, J.; Stock, G.; et al. Allostery in Its Many Disguises: From Theory to Applications. *Structure* **2019**, *27*, 566–578. [CrossRef]

70. Glykos, N.M. Software news and updates carma: A molecular dynamics analysis program. *J. Comput. Chem.* **2006**, *27*, 1765–1768. [CrossRef]

71. Brown, D.K.; Penkler, D.L.; Sheik Amamuddy, O.; Ross, C.; Atilgan, A.R.; Atilgan, C.; Tastan Bishop, Ö. MD-TASK: A software suite for analyzing molecular dynamics trajectories. *Bioinformatics* **2017**, *33*, 2768–2771. [CrossRef] [PubMed]

72. David, C.C.; Jacobs, D.J. Principal component analysis: A method for determining the essential dynamics of proteins. *Protein Dyn. Methods Mol. Biol.* **2014**, *1084*, 193–226. [CrossRef]

73. Peng, J.; Zhang, Z. Simulating large-scale conformational changes of proteins by accelerating collective motions obtained from principal component analysis. *J. Chem. Theory Comput.* **2014**, *10*, 3449–3458. [CrossRef] [PubMed]

74. Dijkstra, E.W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*, 269–271. [CrossRef]

75. Johnson, Q.R.; Lindsay, R.J.; Nellas, R.B.; Fernandez, E.J.; Shen, T. Mapping allostery through computational glycine scanning and correlation analysis of residue-residue contacts. *Biochemistry* **2015**, *54*, 1534–1541. [CrossRef]

76. Johnson, Q.R.; Lindsay, R.J.; Nellas, R.B.; Shen, T. Pressure-induced conformational switch of an interfacial protein. *Proteins Struct. Funct. Bioinform.* **2016**, *84*, 820–827. [CrossRef]

77. Brezovsky, J.; Chovancova, E.; Gora, A.; Pavelka, A.; Biedermannova, L.; Damborsky, J. Software tools for identification, visualization and analysis of protein tunnels and channels. *Biotechnol. Adv.* **2013**, *31*, 38–49. [CrossRef]

78. Marques, S.M.; Daniel, L.; Buryska, T.; Prokop, Z.; Brezovsky, J.; Damborsky, J. Enzyme Tunnels and Gates As Relevant Targets in Drug Design. *Med. Res. Rev.* **2017**, *37*, 1095–1139. [CrossRef]

79. Brezovsky, J.; Babkova, P.; Degtjarik, O.; Fortova, A.; Gora, A.; Iermak, I.; Rezacova, P.; Dvorak, P.; Smatanova, I.K.; Prokop, Z.; et al. Engineering a de Novo Transport Tunnel. *ACS Catal.* **2016**, *6*, 7597–7610. [CrossRef]

80. Kokkonen, P.; Bednar, D.; Pinto, G.; Prokop, Z.; Damborsky, J. Engineering enzyme access tunnels. *Biotechnol. Adv.* **2019**, *37*, 107386. [CrossRef]

81. Nunes-Alves, A.; Kokh, D.B.; Wade, R.C. Recent progress in molecular simulation methods for drug binding kinetics. *arXiv* **2020**, arXiv:2002.08983v2.

82. Schrödinger LLC. *The PyMOL Molecular Graphics System*; Version 2.0; Schrödinger LLC.: New York, NY, USA, 2017.

83. Mitusińska, K.; Magdziarz, T.; Bzówka, M.; Stańczak, A.; Gora, A. Exploring solanum tuberosum epoxide hydrolase internal architecture by water molecules tracking. *Biomolecules* **2018**, *8*, 143. [CrossRef] [PubMed]

84. Subramanian, K.; Mitusińska, K.; Raedts, J.; Almourfi, F.; Joosten, H.J.; Hendriks, S.; Sedelnikova, S.E.; Kengen, S.W.M.; Hagen, W.R.; Góra, A.; et al. Distant non-obvious mutations influence the activity of a hyperthermophilic Pyrococcus furiosus phosphoglucose isomerase. *Biomolecules* **2019**, *9*, 212. [CrossRef] [PubMed]

85. Subramanian, K.; Góra, A.; Spruijt, R.; Mitusińska, K.; Suarez-Diez, M.; Martins dos Santos, V.; Schaap, P.J. Modulating D-amino acid oxidase (DAAO) substrate specificity through facilitated solvent access. *PLoS ONE* **2018**, *13*, e0198990. [CrossRef] [PubMed]

86. Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; et al. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8*, e1002708. [CrossRef] [PubMed]

87. Trott, O.; Olson, A.J. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef]

88. Pinto, G.P.; Vavra, O.; Filipovic, J.; Stourac, J.; Bednar, D.; Damborsky, J. Fast Screening of Inhibitor Binding/Unbinding Using Novel Software Tool CaverDock. *Front. Chem.* **2019**, *7*, 709. [CrossRef]

89. Marques, S.M.; Bednar, D.; Damborsky, J. Computational Study of Protein-Ligand Unbinding for Enzyme Engineering. *Front. Chem.* **2019**, *6*, 650. [CrossRef]

90. Pavlova, M.; Klvana, M.; Prokop, Z.; Chaloupkova, R.; Banas, P.; Otyepka, M.; Wade, R.C.; Tsuda, M.; Nagata, Y.; Damborsky, J. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol.* **2009**, *5*, 727–733. [CrossRef]

91. Marques, S.M.; Dunajova, Z.; Prokop, Z.; Chaloupkova, R.; Brezovsky, J.; Damborsky, J. Catalytic Cycle of Haloalkane Dehalogenases Toward Unnatural Substrates Explored by Computational Modeling. *J. Chem. Inf. Model.* **2017**, *57*, 1970–1989. [CrossRef]

92. Barlow, K.A.; Ó Conchúir, S.; Thompson, S.; Suresh, P.; Lucas, J.E.; Heinonen, M.; Kortemme, T. Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. *J. Phys. Chem. B* **2018**, *122*, 5389–5399. [CrossRef]

93. Löffler, P.; Schmitz, S.; Hupfeld, E.; Sterner, R.; Merkl, R. Rosetta:MSF: A modular framework for multi-state computational protein design. *PLoS Comput. Biol.* **2017**, *13*, e1005600. [CrossRef] [PubMed]

94. Davey, J.A.; Damry, A.M.; Goto, N.K.; Chica, R.A. Rational design of proteins that exchange on functional timescales. *Nat. Chem. Biol.* **2017**, *13*, 1280–1285. [CrossRef] [PubMed]

95. Sun, M.G.F.; Kim, P.M. Data driven flexible backbone protein design. *PLoS Comput. Biol.* **2017**, *13*, e1005722. [CrossRef] [PubMed]

96. Simoncini, D.; Zhang, K.Y.J.; Schiex, T.; Barbe, S. A structural homology approach for computational protein design with flexible backbone. *Bioinformatics* **2018**, *35*, 2418–2426. [CrossRef]

97. Hallen, M.A.; Donald, B.R. CATS (Coordinates of Atoms by Taylor Series): Protein design with backbone flexibility in all locally feasible directions. *Bioinformatics* **2017**, *33*, i5–i12. [CrossRef]

98. Keedy, D.A.; Georgiev, I.; Triplett, E.B.; Donald, B.R.; Richardson, D.C.; Richardson, J.S. The Role of Local Backrub Motions in Evolved and Designed Mutations. *PLoS Comput. Biol.* **2012**, *8*, e1002629. [CrossRef]

99. Davey, J.A.; Chica, R.A. Multistate computational protein design with backbone ensembles. *Comput. Protein Des. Methods Mol. Biol.* **2017**, *1529*, 161–179. [CrossRef]

100. Schreiber, G.; Fleishman, S.J. Computational design of protein–protein interactions. *Curr. Opin. Struct. Biol.* **2013**, *23*, 903–910. [CrossRef]

101. Smith, C.A.; Kortemme, T. Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *J. Mol. Biol.* **2008**, *380*, 742–756. [CrossRef]

102. Kuhlman, B.; Dantas, G.; Ireton, G.C.; Varani, G.; Stoddard, B.L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003**, *302*, 1364–1368. [CrossRef] [PubMed]

103. Murphy, G.S.; Mills, J.L.; Miley, M.J.; Machius, M.; Szyperski, T.; Kuhlman, B. Increasing sequence diversity with flexible backbone protein design: The complete redesign of a protein hydrophobic core. *Structure* **2012**, *20*, 1086–1096. [CrossRef] [PubMed]

104. Loshbaugh, A.L.; Kortemme, T. Comparison of Rosetta flexible-backbone computational protein design methods on binding interactions. *Proteins Struct. Funct. Bioinform.* **2020**, *88*, 206–226. [CrossRef]

105. Khatib, F.; Cooper, S.; Tyka, M.D.; Xu, K.; Makedon, I.; Popović, Z.; Baker, D.; Players, F. Algorithm discovery by protein folding game players. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18949–18953. [CrossRef] [PubMed]

106. Tyka, M.D.; Keedy, D.A.; André, I.; Dimaio, F.; Song, Y.; Richardson, D.C.; Richardson, J.S.; Baker, D. Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **2011**, *405*, 607–618. [CrossRef] [PubMed]

107. Smith, C.A.; Kortemme, T. Predicting the Tolerated Sequences for Proteins and Protein Interfaces Using RosettaBackrub Flexible Backbone Design. *PLoS ONE* **2011**, *6*, e20451. [CrossRef]

108. Dourado, D.F.A.R.; Flores, S.C. A multiscale approach to predicting affinity changes in protein-protein interfaces. *Proteins Struct. Funct. Bioinform.* **2014**, *82*, 2681–2690. [CrossRef]

109. Moal, I.H.; Fernández-Recio, J. SKEMPI: A Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models. *Bioinformatics* **2012**, *28*, 2600–2607. [CrossRef]

110. Benedix, A.; Becker, C.M.; de Groot, B.L.; Caflisch, A.; Böckmann, R.A. Predicting free energy changes using structural ensembles. *Nat. Methods* **2009**, *6*, 3–4. [CrossRef]

111. Aldeghi, M.; Gapsys, V.; De Groot, B.L. Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. *ACS Cent. Sci.* **2018**, *4*, 1708–1718. [CrossRef]

112. Aldeghi, M.; Gapsys, V.; De Groot, B.L. Predicting Kinase Inhibitor Resistance: Physics-Based and Data-Driven Approaches. *ACS Cent. Sci.* **2019**, *5*, 1468–1474. [CrossRef] [PubMed]

113. Ibarra, A.A.; Bartlett, G.J.; Hegedüs, Z.; Dutt, S.; Hobor, F.; Horner, K.A.; Hetherington, K.; Spence, K.; Nelson, A.; Edwards, T.A.; et al. Predicting and Experimentally Validating Hot-Spot Residues at Protein-Protein Interfaces. *ACS Chem. Biol.* **2019**, *14*, 2252–2263. [CrossRef] [PubMed]

114. Davey, J.A.; Chica, R.A. Improving the accuracy of protein stability predictions with multistate design using a variety of backbone ensembles. *Proteins Struct. Funct. Bioinform.* **2014**, *82*, 771–784. [CrossRef]

115. de Groot, B.L.; van Aalten, D.M.F.; Scheek, R.M.; Amadei, A.; Vriend, G.; Berendsen, H.J.C. Prediction of protein conformational freedom from distance constraints. *Proteins Struct. Funct. Genet.* **1997**, *29*, 240–251. [CrossRef]

116. Nivón, L.G.; Moretti, R.; Baker, D. A Pareto-Optimal Refinement Method for Protein Design Scaffolds. *PLoS ONE* **2013**, *8*, e59004. [CrossRef]

117. Simoncini, D.; Allouche, D.; De Givry, S.; Delmas, C.; Barbe, S.; Schiex, T. Guaranteed Discrete Energy Optimization on Large Protein Design Problems. *J. Chem. Theory Comput.* **2015**, *11*, 5980–5989. [CrossRef]

118. Gainza, P.; Roberts, K.E.; Donald, B.R. Protein Design Using Continuous Rotamers. *PLoS Comput. Biol.* **2012**, *8*, e1002335. [CrossRef]

119. Gainza, P.; Nisonoff, H.M.; Donald, B.R. Algorithms for protein design. *Curr. Opin. Struct. Biol.* **2016**, *39*, 16–26. [CrossRef]

120. Hallen, M.A.; Donald, B.R. Protein design by provable algorithms. *Commun. ACM* **2019**, *62*, 76–84. [CrossRef]

121. Desmet, J.; De Maeyer, M.; Hazes, B.; Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **1992**, *356*, 539–542. [CrossRef]

122. Georgiev, I.; Lilien, R.H.; Donald, B.R. The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. *J. Comput. Chem.* **2008**, *29*, 1527–1542. [CrossRef] [PubMed]

123. Hallen, M.A.; Keedy, D.A.; Donald, B.R. Dead-end elimination with perturbations (DEEPer): A provable protein design algorithm with continuous sidechain and backbone flexibility. *Proteins Struct. Funct. Bioinform.* **2013**, *81*, 18–39. [CrossRef] [PubMed]

124. Davis, I.W.; Arendall, W.B.; Richardson, D.C.; Richardson, J.S. The Backrub Motion: How Protein Backbone Shrugs When a Sidechain Dances. *Structure* **2006**, *14*, 265–274. [CrossRef]

125. Hallen, M.A.; Gainza, P.; Donald, B.R. Compact representation of continuous energy surfaces for more efficient protein design. *J. Chem. Theory Comput.* **2015**, *11*, 2292–2306. [CrossRef]

126. Hallen, M.A.; Jou, J.D.; Donald, B.R. LUTE (Local Unpruned Tuple Expansion): Accurate Continuously Flexible Protein Design with General Energy Functions and Rigid Rotamer-Like Efficiency. *J. Comput. Biol.* **2017**, *24*, 536–546. [CrossRef]

127. Hallen, M.A. PLUG (Pruning of Local Unrealistic Geometries) removes restrictions on biophysical modeling for protein design. *Proteins Struct. Funct. Bioinform.* **2019**, *87*, 62–73. [CrossRef]

128. Ojewole, A.A.; Jou, J.D.; Fowler, V.G.; Donald, B.R. BBK* (Branch and Bound Over K*): A Provable and Efficient Ensemble-Based Protein Design Algorithm to Optimize Stability and Binding Affinity Over Large Sequence Spaces. *J. Comput. Biol.* **2018**, *25*, 726–739. [CrossRef]

129. Hallen, M.A.; Martin, J.W.; Ojewole, A.; Jou, J.D.; Lowegard, A.U.; Frenkel, M.S.; Gainza, P.; Nisonoff, H.M.; Mukund, A.; Wang, S.; et al. OSPREY 3.0: Open-source protein redesign for you, with powerful new features. *J. Comput. Chem.* **2018**, *39*, 2494–2507. [CrossRef]

130. Frey, K.M.; Georgiev, I.; Donald, B.R.; Anderson, A.C. Predicting resistance mutations using protein design algorithms. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13707–13712. [CrossRef]

131. Roberts, K.E.; Cushing, P.R.; Boisguerin, P.; Madden, D.R.; Donald, B.R. Computational design of a PDZ domain peptide inhibitor that rescues CFTR activity. *PLoS Comput. Biol.* **2012**, *8*, e1002477. [CrossRef]

132. Reevea, S.M.; Gainzab, P.; Freya, K.M.; Georgievb, I.; Donaldb, B.R.; Andersona, A.C. Protein design algorithms predict viable resistance to an experimental antifolate. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 749–754. [CrossRef]

133. Rudicell, R.S.; Kwon, Y.D.; Ko, S.-Y.; Pegu, A.; Louder, M.K.; Georgiev, I.S.; Wu, X.; Zhu, J.; Boyington, J.C.; Chen, X.; et al. Enhanced Potency of a Broadly Neutralizing HIV-1 Antibody In Vitro Improves Protection against Lentiviral Infection In Vivo. *J. Virol.* **2014**, *88*, 12669–12682. [CrossRef]

134. Sheik Amamuddy, O.; Veldman, W.; Manyumwa, C.; Khairallah, A.; Agajanian, S.; Oluyemi, O.; Verkhivker, G.M.; Tastan Bishop, Ö. Integrated Computational Approaches and Tools for Allosteric Drug Discovery. *Int. J. Mol. Sci.* **2020**, *21*, 847. [CrossRef]

135. Bauer, J.A.; Pavlović, J.; Bauerová-Hlinková, V. Normal Mode Analysis as a Routine Part of a Structural Investigation. *Molecules* **2019**, *24*, 3293. [CrossRef]

136. Kaushik, S.; Marques, S.M.; Khirsariya, P.; Paruch, K.; Libichova, L.; Brezovsky, J.; Prokop, Z.; Chaloupkova, R.; Damborsky, J. Impact of the access tunnel engineering on catalysis is strictly ligand-specific. *FEBS J.* **2018**, *285*, 1456–1476. [CrossRef]

137. Musil, M.; Stourac, J.; Bendl, J.; Brezovsky, J.; Prokop, Z.; Zendulka, J.; Martinek, T.; Bednar, D.; Damborsky, J. FireProt: Web server for automated design of thermostable proteins. *Nucleic Acids Res.* **2017**, *45*, W393–W399. [CrossRef]

138. Kim, D.E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004**, *32*, W526–W531. [CrossRef]

139. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J.C.; Cieplak, P.; Dupradeau, F.-Y. RED Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011**, *39*, W511–W517. [CrossRef]

140. Yang, J.; Zhang, Y. Protein Structure and Function Prediction Using I-TASSER. *Curr. Protoc. Bioinform.* **2015**, *52*, 5.8.1–5.8.15. [CrossRef]

141. Zimmermann, L.; Stephens, A.; Nam, S.Z.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A.N.; Alva, V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **2018**, *430*, 2237–2243. [CrossRef]

142. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef]

143. Rauscher, S.; Gapsys, V.; Gajda, M.J.; Zweckstetter, M.; De Groot, B.L.; Grubmüller, H. Structural ensembles of intrinsically disordered proteins depend strongly on force field: A comparison to experiment. *J. Chem. Theory Comput.* **2015**, *11*, 5513–5524. [CrossRef]

144. Piana, S.; Donchev, A.G.; Robustelli, P.; Shaw, D.E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113–5123. [CrossRef]

145. Stranges, P.B.; Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **2013**, *22*, 74–82. [CrossRef]

146. Maguire, J.B.; Boyken, S.E.; Baker, D.; Kuhlman, B. Rapid Sampling of Hydrogen Bond Networks for Computational Protein Design. *J. Chem. Theory Comput.* **2018**, *14*, 2751–2760. [CrossRef]