

Computational Approaches to Predict Protein–Protein Interactions in Crowded Cellular Environments

Published as part of *Chemical Reviews* virtual special issue “Molecular Crowding”.

Greta Grassmann, Mattia Miotto, Fausta Desantis, Lorenzo Di Rienzo, Gian Gaetano Tartaglia, Annalisa Pastore, Giancarlo Ruocco, Michele Monti,[○] and Edoardo Milanetti^{*,○}



Cite This: *Chem. Rev.* 2024, 124, 3932–3977



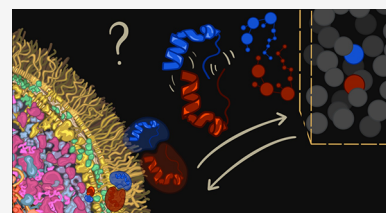
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Investigating protein–protein interactions is crucial for understanding cellular biological processes because proteins often function within molecular complexes rather than in isolation. While experimental and computational methods have provided valuable insights into these interactions, they often overlook a critical factor: the crowded cellular environment. This environment significantly impacts protein behavior, including structural stability, diffusion, and ultimately the nature of binding. In this review, we discuss theoretical and computational approaches that allow the modeling of biological systems to guide and complement experiments and can thus significantly advance the investigation, and possibly the predictions, of protein–protein interactions in the crowded environment of cell cytoplasm. We explore topics such as statistical mechanics for lattice simulations, hydrodynamic interactions, diffusion processes in high-viscosity environments, and several methods based on molecular dynamics simulations. By synergistically leveraging methods from biophysics and computational biology, we review the state of the art of computational methods to study the impact of molecular crowding on protein–protein interactions and discuss its potential revolutionizing effects on the characterization of the human interactome.



CONTENTS

1. Introduction	3933	4.2.3. Looking Closer: All-Atoms Models	3959
2. Biological Background	3934	4.3. Dynamics of Complex formation	3960
2.1. Representative Interactions of Cellular Crowding	3935	4.3.1. Interacting Proteins as Spheres	3960
2.2. Protein Interactions in the Crowded Cellular Environment	3935	4.3.2. Coarse-Grained Models for the Interaction of Proteins	3961
2.3. Effect of Crowding on Protein–Protein Interactions	3936	4.3.3. Atomistic Resolution View	3961
3. Computational Methods to Predict Protein–Protein Interactions in Isolation	3937	4.4. Complex Stability and Dynamics	3961
3.1. Predicting if Two Proteins Interact	3938	5. Conclusions	3963
3.2. Predicting Binding Interfaces and Poses	3940	Author Information	3963
3.3. Predicting the Structural Determinants of Binding	3943	Corresponding Author	3963
3.4. Predicting the Strength of the Interaction	3944	Authors	3963
4. Effect of Crowding across Protein Binding	3947	Author Contributions	3964
4.1. Effect of Crowding on the Single Protein Structure	3952	Notes	3964
4.1.1. At the Beginning of Time There Was a Sphere	3954	Biographies	3964
4.1.2. Increasing the Complexity	3954	Acknowledgments	3965
4.1.3. High Resolution for High Precision	3955	Abbreviations	3965
4.2. Kinetics of the Binding and Reaction Rates	3956	References	3965
4.2.1. Rolling in the Cell	3957		
4.2.2. Kinetics of Mixed Models	3958		

Received: July 31, 2023
Revised: February 20, 2024
Accepted: February 21, 2024
Published: March 27, 2024



1. INTRODUCTION

The cell is a complex world with many layers of biological organization that regulate the interactions and the existence of organelles, proteins, peptides, sugars, DNA, RNA, etc. Its interior is characterized by a high concentration of macromolecules and other cellular components referred to as crowding.^{1,2} Up to ~40% of the cytoplasmic volume is occupied by a concentration of biomolecules between 100 and 450 g/L,^{3–5} and membranes are crowded as well,⁶ having 20–50% of their area occupied by membrane proteins that leave only a few “private” nanometers for individual proteins.⁷ Crowding is ubiquitous in living cells and has important implications for cellular processes, including protein folding and binding, enzyme kinetics, and gene expression, as discussed in more detail in other chapters of this issue.

The cellular environment has long been recognized as crowded, but the interest in this phenomenon has only recently increased, fostered by the development of experimental probes to address crowding directly in the cellular interior,^{8,9} and thanks to the boost of computer power that finally allows molecular simulations of the cytoplasm even at atomic resolution.¹⁰ In the past decade, many reviews on crowding-related computational studies have been proposed, which started from the summary by Zhou et al.¹¹ of all the observations made through simulations between 2004 and 2008. Five years later, Zhou¹² underlined the convergence of experimental and simulation studies, mostly considering artificial crowders. A focus on all-atom simulations of the cytoplasm was proposed by Guin et al.,¹³ while Guigas et al.¹⁴ reviewed simulations of crowded membranes. Shahid et al.¹⁵ summarized the studies on the size-dependent influence of crowders on proteins' properties, whereas a more general overview of the models of crowding agents used in computer modeling investigations of proteins and peptides has been provided by Ostrowska et al.¹⁶

We will focus on protein–protein interactions and provide a critical review of both the works that have investigated, by computational means, the effect of molecular crowding on proteins and of those studies that despite being born outside the crowding field could be employed to strengthen this investigation. This review aims to suggest a stronger connection between the crowding field and the computational methods proposed to study isolated molecules, which in recent years have shown incredible breakthroughs. Given their performance and their rapid advancements (reviewed in Section 3), it is worth asking if these new techniques could foster advancement in the investigation of the crowded cellular environment for which many questions are still to be solved (as discussed in Sections 2 and 4).

While crowding is ubiquitous in all cell components, we opted to focus on the effect of crowding on the interaction between proteins in the cytoplasmic environment. Wider consequences (in terms of molecules distributions in the cell) of the variations of these interactions^{17–19} and the effect of crowding in membranes will not be discussed here. For an in-depth overview of these topics, the reader can look at previous reviews. Heo et al.,²⁰ for example, focused on protein assembly, phase separation, and molecular crowding effects, distinguishing them depending on contact specificity and time-of-life, and highlighted the most recent advances from computer simulation studies. An overview of membrane-associated processes and the computational tools developed to study

them can be found in the reviews by Marrink et al.²¹ and Löwe et al.²²

Understanding and predicting the interaction mechanisms between proteins are of particular importance because over 80% of proteins operate in molecular complexes;²³ the knowledge of protein complexes formation, considered in their crowded environment, can provide crucial information on the physiological and pathological nature of cellular life. These studies can also have important clinical implications, for example, to investigate the mechanisms of action of pharmaceuticals.²⁴

The subject of protein–protein interactions has been widely discussed, and many efforts have been made toward the compilation of this essential part of the so-called interactome, defined as the map of the molecular interactions occurring within the cell. Experimentally, this can be built through a wide set of methodologies, including proteomics techniques when specifically referring to the recovery of the protein–protein interactions network. Mapping the interactome of an organism is fundamental for understanding cellular functions, both under physiological and pathological conditions.²⁵ Often, in the context of proteins, these studies have focused on functional interactions, which are defined as the physical contacts of high specificity established between two or more protein molecules that are related to a known function. However, many interactions in the cell do not correspond to these criteria. It is by now well understood that in the crowded cellular environment, proteins interact with much more than their functionally related proteins (or other macromolecules) in the immediate vicinity.²⁶ For example, nonfunctional proteins can interact after encountering as they diffuse in the cell searching for their specific functional partner/partners. This remark has brought attention to the importance of nonfunctional protein–protein interactions.^{17,27–29} Nonfunctional interactions can lead to the formation of a transient complex between two functionally unrelated proteins but also indirectly influence functional interactions by modifying protein dynamics and association pathways and consequently their folding or binding.

Therefore, taking into account the cellular environment and studying the effects of crowding is particularly crucial for researchers to better understand protein–protein functional interactions and unravel the actual interactome.

The effects of crowding on the behavior of the proteins and other biomolecules immersed in the cellular environment have been under investigation for many years;³⁰ nevertheless, most of the experimental and computational studies concerning protein–protein interactions conducted up to now have considered diluted solutions. The crowding and dynamism characterizing the cell interiors are not easy to replicate both in *in vitro* experiments and *in silico* simulations. The presence of molecules in such high concentration also results in a higher viscosity of the cell interior:³¹ the conditions in which protein interactions are usually measured *in vitro* can differ significantly from those found in the cellular environment where they take place. A consensus on the crowding theory, based on experimental data, is hampered by the intrinsic limitations of studying highly dynamic and interconnected systems in which the single molecule information is lost within the experimentally detected structural ensemble. Even if variations in the structure can be observed over time, the atomic detail is often missing, and its evolution in time—together with the related energy variations—can never be observed. Only the

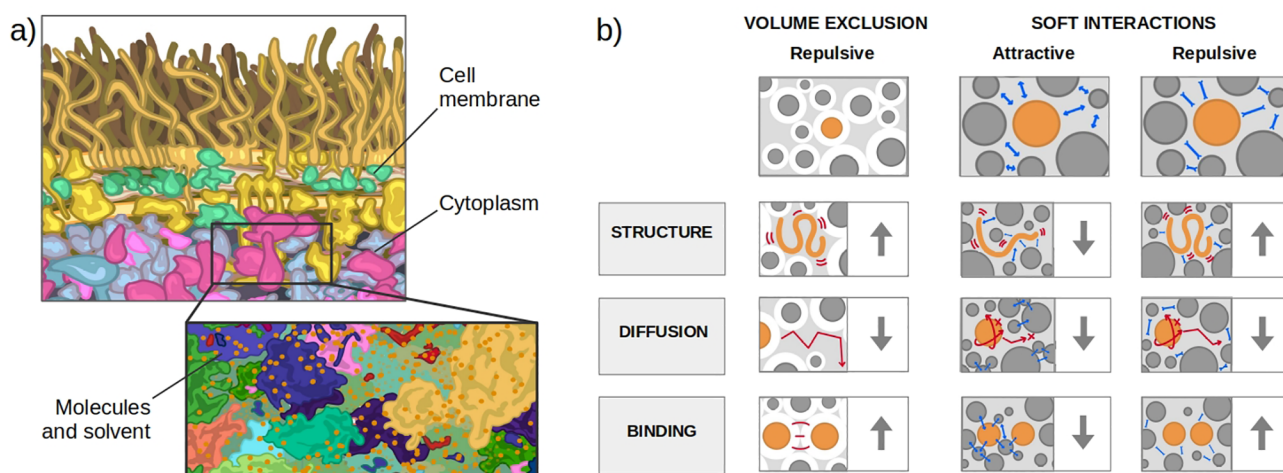


Figure 1. Schematic diagram of crowding inside a cell and of its effect across protein binding. (a) Scientific illustration of the macromolecules inside a cell of *Escherichia coli*, inspired by the figure presented by Goodsell³⁹ with the cytoplasm in blue and purple and the cell membrane in yellow. The magnified portion is a magnification of the cytoplasm constituents. (b) Outline of the effects of crowding on protein folding, diffusion, and binding (from top to bottom) inside the cell. Crowding effects can be divided into volume exclusion and soft interactions. The former has a repulsive nature, which tends to enhance the stability of folding and binding and decrease the translational diffusion, as summarized in the first column. Soft interactions can instead be classified as attractive or repulsive (second and third column, respectively). Attractive soft interactions tend to counteract volume exclusion, which has a destabilizing effect on both folding and binding. Conversely, repulsive soft interactions have a stabilizing effect. Both interactions hinder rotational and translational diffusion.

total variation in interaction energies can be measured, but to understand the general effect of crowding, it is necessary to untangle each component. Moreover, experimental techniques are usually expensive and time-consuming, especially when applied to such complex systems that include thousands of molecules.

In this respect, accurate modeling of protein–crowder interactions and efficient computations can complement experiments by untangling the various effects of crowding. Computational studies—including but not restricted to MD simulations—can, indeed, give direct access to information on single molecules in carefully controlled environments. Low-cost computational methods that screen the interactions both on a large number of crowders and on a high number of conformational states (obtained, for example, from MD simulations) can analyze the interaction of single proteins or complexes with the environment (including the crowders). Even if multimillion atom systems have long since been available, as discussed nearly a decade ago,^{32,33} most molecular dynamics (MD) studies have focused on single molecules/complexes without considering the crowded environment. However, the computational investigation of the crowding effect has mostly relied on simplified models. There are two reasons for this: the computational cost of all-atom simulations of crowded environments and the lack of experimental data to compare. However, the flux of data from high-throughput experiments that can probe the composition of a cell³⁴ and the macromolecule structure and dynamics inside of it^{35,36} has been growing. This increase has encouraged computational studies of realistic models of crowded biological environments that can fill the gaps left by experimental data. The present Review aims to highlight the need for these methods and the importance of bringing together both sides (computational and experimental) of the crowding-focused research world. In particular, it this Review was inspired by the meeting “New Frontiers in Molecular Crowding” organized in July 2022 at the European Synchrotron Radiation Facility (ESRF) of

Grenoble to bring together experimentalists and computational researchers to share and discuss their results.³⁷

After a due introduction to the concept of cellular crowding and protein interactions in Section 2, Section 3 will discuss computational studies that, even if not directly connected to the idea of crowding (in some of them, there is no reference to crowding at all), developed methods that, in our opinion, should be more in contact with this field. Section 4 will then review the computational effort done up to now to characterize the influence of crowding on all the processes leading to the interaction between two proteins; starting from the single monomers that still have to navigate the crowded environment to find each other, we will discuss their conformational variation and diffusion in Sections 4.1 and 4.2, respectively. In Section 4.3, we will focus on the dynamics of complex formation once the proteins are in mutual proximity. We will end with the evaluation of the stability and dynamics of the formed dimer in Section 4.4. Various systems will be presented: crowding acts differently on each of them, for example, because of the presence of a different protein diffusivity. To allow the investigation of these systems characterized by a high number of components, a miscellanea of theoretical models and atomistic simulations have been performed. The obtained computational findings, here reviewed, testify to the importance of considering the crowding effects on structural and conformational dynamics of proteins to better understand the relationship between interactions and biological function.

2. BIOLOGICAL BACKGROUND

The cell is a dynamic and autoregulated system characterized by a network of both inter- and intracellular signaling modulating its main functions. These functions take place in an incredibly complex environment where several nonspecific forces work against or together with specific interactions between the cellular constituents, which include a wide range of different molecules ranging from small ligands to sugars to

large assemblies of RNA and proteins. Among all these biomolecules, proteins are central to most biological functions: almost all biological processes rely on molecular machines formed by proteins that bind to each other. Even if each protein has a molar concentration ranging from nM to μM , under crowded conditions, the distance between neighboring proteins is comparable with the size of the proteins.³⁸ The interactions between these neighboring proteins and how they influence and are influenced by their crowded surroundings will be the focus of this section.

2.1. Representative Interactions of Cellular Crowding

The current picture of the inside of cells is that of a crowded environment, as the one shown in Figure 1a, where biochemical reactions have to take place in a highly concentrated solution with various species of finite sizes, which includes both low molecular weight (molecular crowding) and high molecular weight (macromolecular crowding). Note that the term “crowding agents” is commonly used for molecules similar in size to the considered protein. The influence of water molecules, which are much smaller, is considered as solvation,⁴⁰ while for much bigger molecules (such as macromolecules that can be considered immobile obstacles), the effect is referred to as confinement.⁴¹ This can be understood by picturing the cell as a room full of people: while the motion of a person is influenced by bystanders, smaller entities (like ants or dust) do not change it. The presence of voluminous furniture would delimit the room but not be part of the flow. Historically the effects of crowding agents have been divided into soft chemical/physical interactions and volume exclusion (or hard-core repulsion).^{42–45}

The term “excluded volume” was used by Minton, who coined the expression “macromolecular crowding” in 1981,² to indicate the effect that volume exclusion has on the energetics and transport properties of molecules in a highly volume-occupied medium.⁴⁶ Volume exclusion is a steric effect arising from the impenetrable nature of atoms that can not overlap because of the hard-core or van der Waals repulsion and only involves molecules rearrangement and, thus, affects the entropic component of protein stability. In this setting, the activity of a molecule, for example, a protein in equilibrium between the unfolded state and native folded state, is raised when crowding molecules are added because its excluded volume is inaccessible to the centers of the other crowding molecules.^{46,47} This favors its folded state because it corresponds to a smaller excluded volume. For the same reason, the oligomerization of the protein is promoted, and the association (dissociation) constant for ligand binding is increased (decreased). In general, according to this view, crowding facilitates processes leading to a reduction in excluded volume, like protein folding, oligomerization, complexation, aggregation, and condensation.⁴⁸

To characterize the effects of crowding given by this definition, the past decades have seen an explosion of experiments using large artificial polymers, such as Ficoll, PEG, and dextran to recreate a volume exclusion effect and allow the generalization of effects that may depend on the specific protein crowder. Even if Ficoll and PEG are still often considered as noninteracting with proteins,^{49,50} many studies have argued that Ficoll and PEG show weak chemical interactions with the test protein.^{51–55} In a recent computational study on the behavior of PEG and Ficoll as crowders,

Ostrowska et al.⁵³ investigated how they influence a specific enzyme called NS3/4A. By employing atomistic simulations, they found that while the enzyme’s affinity for its substrate remains similar with or without the crowding agents, the speed at which it performs its catalytic function decreases with PEG and increases with Ficoll. This suggests that these crowding agents may affect the enzyme’s function through specific interactions rather than the more generic volume exclusion effect. The study also showed that both crowding agents made contact with the enzyme and slowed down its movement. Additionally, they found that the crowding agents influenced the structure of a protein component called NS4A, which caused it to adopt helical structures. Overall, PEG had a slightly stronger interaction with NS3/4A, while Ficoll formed more hydrogen bonds with the enzyme. This result was in line with previous experimental measures.⁵⁵

Although for more than 30 years crowding theories have emphasized steric repulsion, in the last decades, the number of studies probing the competition between the excluded volume effects and soft interactions has been continuously rising.^{43,56–62} Soft chemical/physical interactions change the enthalpy of proteins and include water–protein and solute–protein interactions. In this context, some of the key weak interactions are (i) van der Waals interactions, (ii) hydrogen bonds, (iii) ionic interactions, (iv) dipole–dipole interactions, and (v) weak noncovalent hydrophobic interactions. Most soft interactions are attractive and are expected to favor expanded conformations (e.g., the unfolded state of a protein will expose sites for attractive interactions, such as hydrogen bonding and hydrophobic contacts), thereby leading to destabilization.⁴⁵ Thus, they are typically believed to oppose and reverse the hard-core repulsion effect of crowding. However, among soft interactions, there is a strong repulsive one arising from the opposition of charges with the same sign. This repulsion has a stabilizing impact similar to that of the hard-core effect:⁶³ in recent years, it has been experimentally and theoretically shown that chemical interactions can be both stabilizing or destabilizing.^{64–66} The acknowledgment of soft interactions in theories of crowding has brought the conclusion that inert polymers—used in most experiments or simulations aimed at assessing hard-core repulsion effects—are not good mimics of the in-cell environment.^{67,68} To distinguish between these inert polymers and the biomolecules that can, indeed, be found in the cell, we will use the terms “inert crowders” and “crowders,” respectively. To investigate the crowding effects in cell-like conditions, different types of “crowders” with their chemical–physical characteristics have to be considered. The importance of considering more complex systems to obtain a realistic representation of the cellular environment and investigating both the entropic and enthalpic effects that take place *in vivo* has been recently reviewed by Pastore et al.⁶⁹

2.2. Protein Interactions in the Crowded Cellular Environment

Protein–protein interactions are essential for protein function and cellular pathways formation. Since these interactions are also involved in the development of diseases, they are important targets for drug design and the artificial design of protein complexes. Most studies usually distinguish between non functional and functional partners to focus on the latter. “Functional partners” can be defined as proteins that have been evolutionarily selected to have functional interactions with a few partners⁷⁰ and avoid interactions with thousands of other

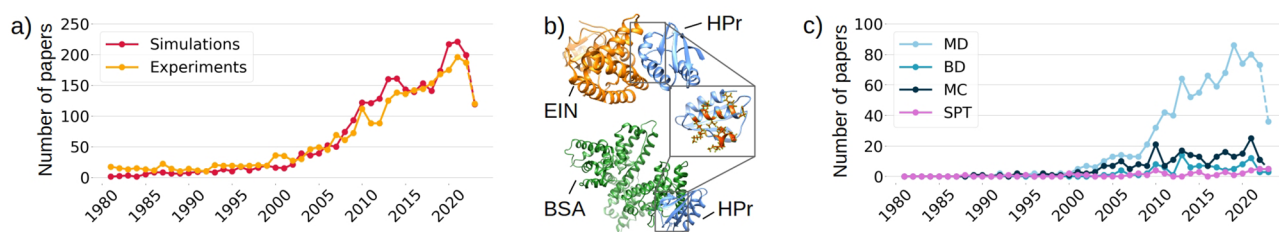


Figure 2. Coexistence and cooperation between experimental and numerical studies. (a) Number of papers whose abstract or title include the terms “crowding” and “simulation” (“experiment”) as a function of time since 1981 (the year in which Minton defined the concept of cellular crowding¹), colored in orange (yellow). The papers were extracted from the Dimensions database.⁹⁰ (b) On top, EIN (orange) and HPr (blue) in complex (PDB ID: 3EZA). In the gray box, the HPr residues interacting with EIN are highlighted in orange. The same residues were experimentally shown by Dong et al.⁷⁴ to interact with BSA crowders. On the bottom is a cartoon representation of the complex formed by HPr (blue) and BSA (green). The structure was obtained by docking the isolated structure of HPr (PDB ID: IPOH) and BSA (PDB ID: 4F5S) with HADDOCK.⁹¹ (c) The same search as in (b) was performed, but in this case, in addition to “crowding,” the words “molecular dynamics” (MD), “Brownian dynamics” (BD), “Monte Carlo” (MC), and “scaled particle theory” (SPT) were looked for. The number of published papers is shown in light blue, blue, dark blue, and violet, respectively.

types of macromolecules in the cell. However, crowding makes “nonfunctional partners” interact despite this evolutionary selection, which leads to a competition between functional and nonfunctional binders.^{26,38,70–72} This competition implies that to predict functional interactions, weak and potentially nonfunctional ones have to be characterized, as well.^{26,73}

In addition to this, in the context of the cellular interior, complexity of the classification, itself, has been challenged.^{42,74–76} For example, sequential metabolic enzymes seem to have evolved to associate weakly,⁷⁷ and similarly, weakly bound functional protein complexes are observed in metabolic, regulatory, and signaling pathways. Sukenik et al.⁷⁸ hypothesized that such complexes create an additional fuzzy network of weak interactions in the cell, which gives the cell the ability to detect, signal, and/or directly initiate regulatory processes quickly and effectively in response to external stresses or internal signals (such as rapid volume variation). According to them, the cell environment is finely tuned to optimize weak interactions networks among its protein machinery. In a case like this, it is difficult to see the involved proteins as functional partners or crowders that are weakly interacting with each other; in a crowded environment, all interactions have to be considered.

The interaction between two molecules is defined as a correlation between their positions, orientations, and motions.⁷⁹ This can happen even when the two partners are distant because of the presence of electrostatic or solvent-mediated interactions. These long-range interactions can be nonspecific or specific.⁷⁹ Nonspecific interactions do not lead to the formation of a specific complex and are, thus, more likely to be independent of the partners’ orientation. Specific interactions are strong and lead the diffusional search of binding partners for one another followed by colocalization and recognition of the compatible binding sites. This determines the formation of a complex with a defined structure and interface. Because of the specific structure that has to be achieved, interactions are expected to depend on the mutual orientations of the partners. In both cases, the importance of a match between different physicochemical properties at a global level and not only between the interacting sites (if any) has been shown: for the partners approach, their total charge, isoelectric point, hydration free energy, and total electrostatic energy have to match.^{80–82} Significant commonalities for the pH of maximal stability⁸³ were found, as well. Simplified model potential of mean forces, including these general features of the

overall intermolecular interaction, have been shown to obtain semiquantitative agreement with experimental measurements of the osmotic second virial coefficient⁸⁴ and of concentration-dependent light scattering and osmotic pressure.⁷⁹

These results underline the importance of taking into account the crowded cellular environment in which the investigated proteins live to understand how their binding can be influenced by these nonspecific interactions.

2.3. Effect of Crowding on Protein–Protein Interactions

Both the interactions among permanent and transient partners are influenced by crowding. This statement would seem pretty obvious, since, for example, a protein has to navigate among many other crowding macromolecules to find its binding partner(s) and fulfill its function. Moreover, crowding influences protein folding and stability, as well, which in turn determines how that protein will interact with the others. Figure 1b schematizes how crowding is today thought to influence protein structure, diffusion, and binding. Nonetheless, the exact effect of crowding on these steps is still being investigated.

Different lines of research have focused, both experimentally and computationally, on determining the predominant effect on protein stability between soft interactions and hard-core repulsion. The latter has a stabilizing effect, by reducing the available volume and enhancing thermal stability through the reduction of water activity, when water molecules are sequestered and immobilized by the macromolecules. Soft interactions can both increase (when repulsive) or decrease (when attractive) stability.^{63,66} Repulsive interactions, such as steric hindrance and electrostatic repulsion from other macromolecules, can stabilize the folded structure of the molecule^{63,66,85} and decrease its susceptibility to denaturation as well. Conversely, weak, nonspecific attractive interactions can decrease the stability.⁶⁶

Several studies have shown that crowding can also alter the dynamics and transport of molecules, thereby leading to changes in binding affinity.^{86–88} When macromolecules are packed tightly in the intracellular space, they can create tortuous pathways that hinder the diffusion of smaller molecules. The decrease in rotational diffusion, in particular, has been linked to the formation of transient protein clusters due to nonspecific contacts generally lasting less than 1 μ s.⁸⁹

In this context, weak attractive interactions—van der Waals interactions—can facilitate the movement of small molecules

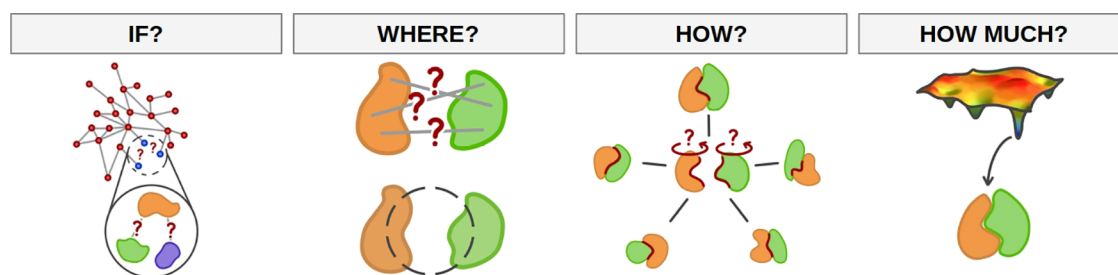


Figure 3. Sketch of the four main questions of protein–protein interaction predictions. Methods can be developed to assess (i) whether and/or (ii) where two proteins functionally interact. Other techniques can assess (iii) the dynamics and outcome of the binding process and (iv) the binding stability of the resulting protein–protein complex. An overview of the methods that have been developed to address these different tasks is provided in Table 1.

by reducing the effective size of the macromolecules and creating temporary voids in the crowded environment.

Other groups have investigated the effect of crowding directly on binding affinity, even in some cases showing that crowding can both increase or decrease it. Dong et al.⁷⁴ showed with nuclear magnetic resonance (NMR) that, in the presence of bovine serum albumin (BSA) crowders, the binding affinity of the EIN–HPr complex is decreased. Interestingly, this was explained by the fact that BSA has a specific interaction with HPr on a binding interface overlapping with that of EIN, which leads to competition. In the same paper, computational techniques were used to rationalize the experimental data and provide quantitative insights into the energetics of protein–crowder interactions. Figure 2a shows a representative reconstruction of the protein–crowder complex, as suggested by Dong et al.

This study shows the importance of considering soft interactions when investigating the effects of crowders. It also attests to the fruitfulness of a collaboration between experimental and numerical observations: Figure 2b illustrates how, starting from when the field was born (1981, thanks to Milton²), *in vitro* and *in silico* studies have grown hand-to-hand. Even if in this review we will focus on MD simulation, it is worth noticing that there are many other computational applications to the study of crowding, as suggested by Figure 2c.

Another experimental study on the effect of crowding on binding affinity was performed by Sudhaharan et al.⁹² They showed that in the cellular environment, the dissociation constant between RhoGTPase CDC42 and three of its effector proteins is decreased approximately by a factor of 2 compared with *in vitro* data. However, the Wohland lab⁹³ found that the dissociation constant of a protein complex was five times larger in the cell compared with *in vitro*. Looking at these results, it would seem safe to assume that the overall effect of crowding varies depending on the specific case.

Chemical interactions have often been found to be more important than hard-core repulsion in physiological conditions,^{43,45,61,62,85,94} but many studies agree that the result of this competition depends on various factors, including temperature,^{95,96} the strength of the affected interaction,⁹⁷ the structure of the affected proteins,⁹⁸ and concentration and type of the crowder.⁹⁹ Many have investigated the difference between molecular and macromolecular crowding in terms of entropy variation, although the differences concerning the enthalpic effect have not been deeply investigated yet. Hard-spheres fluid mixtures theories have predicted that larger molecules are less effective at crowding¹⁰⁰ and that molecules

of medium size (~ 5 Å) have the same effect as much larger ones.¹⁰¹ Both experimental^{102,103} and numerical^{104,105} investigations have confirmed that macromolecular crowding has a weaker entropic stabilizing effect compared with molecular crowding. This inverse dependence between crowders effects and size is to be expected since smaller crowders occupying a given volume will result in a more compact arrangement than when the same volume is filled with bigger molecules. Thus, it will be easier to accommodate a protein in the latter case: since this scenario is more discriminating between the open and closed conformations of a protein, it will produce a stronger effect on the open-to-closed population ratio. Despite this observation and the fact that the cellular cytoplasm is concentrated in smaller molecular weight solutes, up to now the more commonly used polymer crowders have had large molecular weights (e.g., MW > 1000 amu).¹⁰³

3. COMPUTATIONAL METHODS TO PREDICT PROTEIN–PROTEIN INTERACTIONS IN ISOLATION

Interactions between proteins are fundamental to every cellular process from DNA replication to protein degradation and the development of diseases. Therefore, to understand the structure and function of biological pathways and to predict protein function, the compilation of a protein–protein interactions network is fundamental. For what concerns the understanding of single protein–protein associations, many research lines have been working on capturing protein interactions and the resulting complex stability.^{106–108} The opportunities provided by experimental techniques, like NMR, X-ray, and cryo-EM, have brought a big advance in this context. Large-scale networks of some organisms have been compiled thanks to these experimental methods.

Many techniques do not rely on capturing structural information on protein interactions, like yeast two-hybrid (Y2H) screening and affinity purification coupled with mass spectrometry. These methods have been employed, for example, in one of the major objectives in the field of biology, the Human Interactome Project. To obtain a complete reference map of the human protein–protein interactome network, high-quality binary protein–protein interactions have been mapped using a primary Y2H assay followed by orthogonal validation through alternative binary assays.^{109–111} To date, 64006 interactions involving 9094 proteins have been identified.¹¹¹

Despite the significant advances in this project, experimental techniques are typically time-consuming, labor-intensive, and expensive and are not easy to perform on new target organisms or to generalize to different classes of proteins. Moreover,

Table 1. Some of the Available Servers and Software for Studying the Different Aspects of Protein–Protein Interactions in Isolated Conditions^a

if?			where?		
ADVICE ¹¹⁵	2004	coevolution	binding sites and pose prediction		
PIPE2 ¹¹⁶	2008	sequence-based	HDOCK ¹³⁵	2017	homology + FFT docking
LDA-RF ¹¹⁷	2010	sequence-based	ClusPro ¹³⁶	2017	FFT docking
SPRINT ¹¹³	2017	sequence-based	MEGADOCK ¹³⁷	2018	FFT docking
PPI-Detect ¹¹⁸	2019	sequence-based ML	SWISS-MODEL ¹³⁸	2018	homology modeling
DeepFE-PPI ¹¹⁹	2019	sequence-based ML	AlphaFold2 ¹³⁹	2022	sequence-based ML
where?			how?		
binding sites prediction			how?		
BSpred ¹²⁰	2011	sequence/structure-based ML	GROMACS ¹⁴⁰	2005	MD simulations
PAIRpred ¹²¹	2014	sequence/structure-based ML	AMBER ¹⁴¹	2005	MD simulations
Protein Interface Prediction using GCNs ¹²²	2017	structure-based ML	CHARMM ¹⁴²	2009	MD simulations
BIPSPI ¹²³	2019	sequence/structure-based ML	PyEMMA 2 ¹⁴³	2015	MSM construction
DeepPPISP ¹²⁴	2020	sequence-based ML	OpenMM ¹⁴⁴	2017	MD simulations
Zephyros ¹²⁵	2021	geometry-based docking	GENESIS ¹⁴⁵	2017	MD simulations
DELPHI ¹²⁶	2021	sequence-based ML	MSMBuilder ¹⁴⁶	2017	MSM construction
Attention-based-CNNs-for-PPIs-prediction ¹²⁷	2021	sequence-based ML	CANVAS ¹⁴⁷	2023	multiresolution modeling
MaSIF ¹²⁸	2020	structure-based ML	how much?		
PeSTo ¹²⁹	2023	structure-based ML	DFIRE ¹⁴⁸	2005	knowledge-based energy function
binding sites and pose prediction			PRODIGY ¹⁴⁹	2016	empirical function
PatchDock ¹³⁰	2005	geometry-based docking	ATTRACT ¹⁵⁰	2017	MM force field
SymmDock ¹³⁰	2005	geometry-based docking	PIPR ¹⁵¹	2019	sequence-based ML
PIPER ¹³¹	2006	FFT docking	ISLAND ¹⁵²	2020	sequence-based ML
SwarmDock ¹³²	2013	particle swarm optimization	mmCSM-PPI ¹⁵³	2021	structure-based ML
ZDOCK ¹³³	2014	FFT docking	AffPred ¹⁵⁴	2021	knowledge-based energy function
PRISM ¹³⁴	2014	structural similarity	PerSpect-EL ¹⁵⁵	2022	structure-based ML
HADDOCK2 ⁹¹	2016	simulated annealing docking	PPI-Affinity ¹⁵⁶	2022	structure-based ML

^aEach section groups the server/software addressing one of the four main questions of interaction prediction (if, where, how, and how much), as represented in Figure 3. For each server/software, the table reports the name, the year in which it was made available and its core technique. Methods already applied in crowding contests are marked in bold.

many of these methods can not easily identify weak interactions,¹¹² which leaves out many transient interactions: this results in a big disadvantage in the context of crowding, where most interactions are transient.

Hence, the role of fast and inexpensive computational approaches, which can complement experimental methods and have already demonstrated the ability to provide new proteomics-scale protein complex predictions, is crucial.^{112,113}

Protein complex prediction methods can be divided into several classes, including sequence-based, protein–protein interaction network topology-based, function-based, and structure-based. All these methods can answer one or more of the four questions we are interested in when studying protein–protein interactions in isolation, which are schematized in Figure 3: (1) Do the proteins interact? (2) Where are the interacting regions? (3) How will they bind? (4) How strong is the binding? Each of the following sections will summarize the studies centered on one of these investigations. The available servers and software developed by these studies are listed in Table 1, while Figure 4 shows some of the most commonly used techniques for the prediction of protein binding interfaces and poses in isolation.

Although the main interest of this Review is the study of molecular crowding, we will take some time to summarize these methods here because they can improve our comprehension of the mechanisms underlying molecules' interactions and binding. For example, they can overtake the

distinction between functional and nonfunctional interactions, which is particularly ambiguous in the context of crowding. This premise seemed necessary to us because the comprehension of what determines molecular interactions, even when considering the most simple and idealized condition of two isolated partners, is still far from being achieved. Moreover, even if some of these methods have been developed to study proteins in isolation or aqueous solutions, recent advancement in computational modeling have allowed their application to crowded environments, as shown in Table 1. For example, Vakser et al.¹¹⁴ merged MD simulations and Fast Fourier Transform (FFT) docking to perform long dynamics at atomic resolution of systems with many proteins. In general, the current speed of calculations and recently developed techniques have allowed simulations to reach longer time scales and bigger system sizes. The implementation of these methodologies in the context of crowding will be discussed more in-depth in Section 4.

3.1. Predicting if Two Proteins Interact

Many existing techniques for predicting protein–protein interactions and their networks have been developed using only sequence information^{117–119,151,157–167} since they can be applied to proteins for which structural information is unknown or that are intrinsically disordered. Most sequence-based approaches start from the hypothesis that pairs of proteins that are similar to pairs of interacting proteins have a

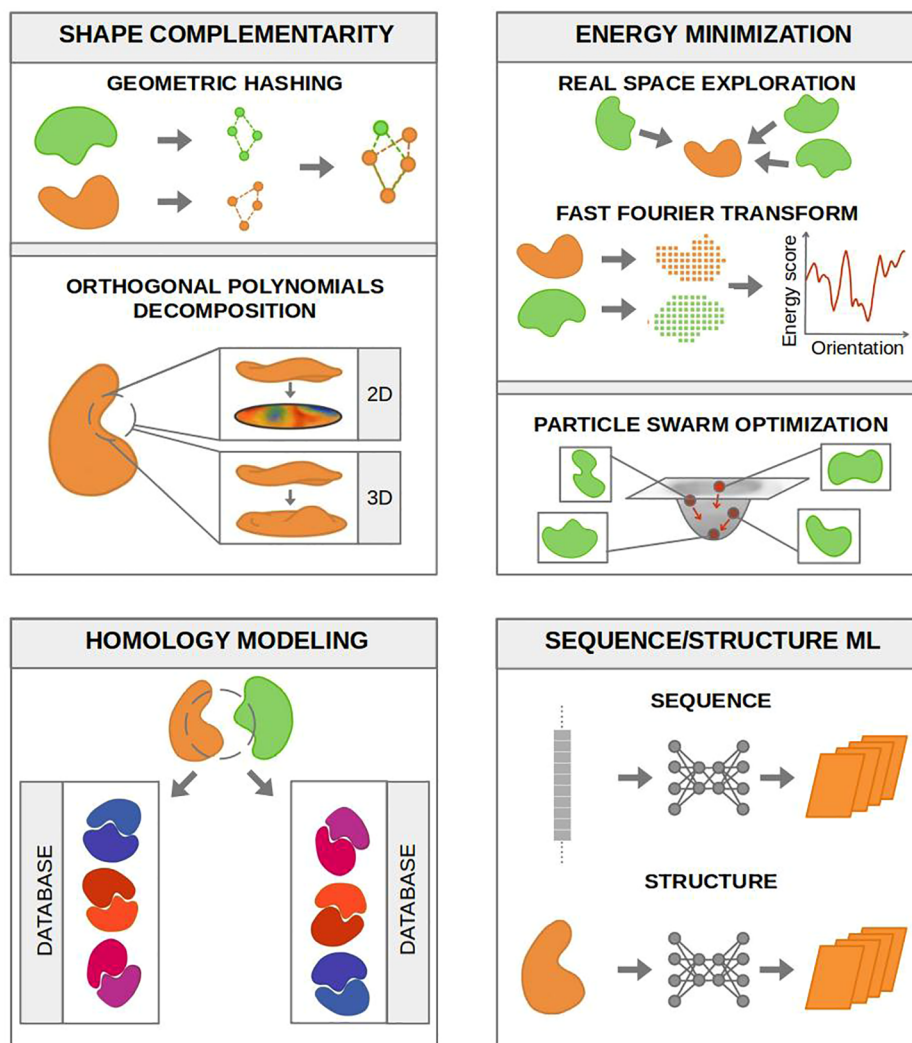


Figure 4. Diagram of the most common techniques used to predict binding interfaces and poses by the methods reported in Table 1. The available servers and codes for binding interfaces and pose prediction are often based on the evaluation of shape complementarity, the minimization of an energy score, sequence and/or structure-based ML, and/or homology modeling. Shape complementarity (top left box) can be searched for with geometric hashing or orthogonal polynomials decomposition. The former defines geometric patches (concave, flat, convex) with discrete points of the protein surface and uses those points to match the partner's patches stored in a hash table. Partners' surfaces can also be compared by expanding the surface patches in terms of orthogonal polynomials (e.g., Zernike polynomials) and computing the distance between the corresponding vectors. Both 2D and 3D expansion have been used. Many methods aim at minimizing an energy expression (including, for example, van der Waals energy, electrostatic interaction energy, and a statistical pairwise potential representing other solvation effects¹³⁶). This minimization can be achieved by testing different orientations and spatial positions of the binding partners and computing the energy term for each step. Such exploration is often performed in real space or through a FFT correlation approach. With this technique, the interaction matrix is approximated by its dominant eigenvectors so that the energy expression is written as the sum of a few correlation functions, and the minimization is solved by repeated FFT calculations. A quick solution to the minimization problem can be achieved with particle swarm optimization (PSO). If homologous structures are available, possible binding interfaces can be obtained by either sequence or structural similarity between the studied proteins surfaces and known binding sites of homologous complexes. Finally, the binding sites scoring problem is often faced with ML methods. ML techniques can be divided into sequence-based, structure-based, and combined. All categories are based on representing the protein features with a vector that is then passed to a network. Many networks and learning algorithms have been used.

higher chance to interact. Different algorithms have been implemented to identify similar regions.

For instance, Shen et al.¹⁶⁸ developed a learning algorithm based on a support vector machine, along with a kernel function and a conjoint triad feature for amino acid description, and trained the model using over 16 000 diverse pairs. In this respect, the effectiveness of current sequence-based methods in constructing interaction networks and understanding disease mechanisms is highly dependent on the availability and reliability of training data (as reviewed by Murakami et al.¹⁶⁹), which means that predictive methods,

especially those based on parameter training (i.e., supervised machine learning methods), need a large amount of data to improve their predictive ability.

The importance of the data set construction has been addressed by Li et al.,¹¹³ who proposed another sequence-based prediction method called Scoring PRotein INTeractions (SPRINT). This algorithm identifies possible binding partners regions that are similar to sites on known complexes with a multiple spaced-seed approach and then eliminates elements that occur too often to be involved in interactions. SPRINT was shown to be several orders of magnitude faster than other

state-of-the-art sequence-based programs, including PIPE2¹¹⁶ and two machine learning (ML)-based methods proposed by Ding et al.¹⁷⁰ and Martin et al.¹⁷¹

Another key step for prediction methods based on ML is the representation of the input data. Romero-Molina et al.¹¹⁸ developed a procedure that transforms pairs of amino acid sequences into a ML-friendly vector whose elements represent numerical descriptors of residues in proteins. This numerical encoding method was then implemented by the same group to develop a support vector machine model, PPI-Detect.¹¹⁸

In the same year, Yao et al.¹¹⁹ proposed a new residue representation method named Res2vec and combined it with deep learning techniques to predict interactions starting only from sequence information (DeepFE-PPI).

Other methods based on sequence information for predicting residues involved in intermolecular binding were developed using coevolution.¹¹⁵ The idea is that many proteins have evolved to form specific molecular complexes, and the specificity of their interactions plays a crucial role in their proper functioning. Therefore, these interactions impose constraints on the protein sequences because the network of inter-residue contacts must be maintained. It is plausible to assume that any sequence change accumulated during the evolution of one interacting protein is counterbalanced by changes in the other protein.

Pazos et al. developed a method for detecting correlated changes in multiple sequence alignments, which is useful to a collection of interacting protein domains.¹⁷² The results revealed that positions exhibiting correlated changes in both interacting molecules tend to be near the protein–protein interfaces. This observation opened up the possibility of applying statistical inference techniques based on the maximum entropy principle^{173–178} to predict pairs of residues that come into contact solely based on their sequence information^{179–181} and, more specifically, in their evolutionary process.

Many methods for interaction prediction focus on protein–protein interaction networks since genome-scale data for different species have been rapidly increasing. By starting from the network of an organism and evaluating the network topology features of proteins, these methods can predict new interactions.¹⁸² In the past, protein–protein interaction networks were thought of as static, defined by interactions constant in any cell location and condition.¹⁸³ The current picture is that interactions are dynamic and that their occurrence depends on a set of conditions, such as spatial or temporal variations.¹⁸⁴ In other words, these networks are now modeled as dynamic systems. Computational approaches can come in aid of experiments in the prediction of the variation and dynamics of these networks in the context of crowding.²⁶ *In silico* methods tend to detect protein complexes by associating them with meaningful clusters in the protein–protein interactions network. They can be classified into three categories:¹⁸⁵ cluster-quality-based,^{186,187} node-affinity-based,^{188,189} and ensemble clustering methods.¹⁹⁰ The first one uses clusters as measuring units, while node-affinity-based methods measure the affinity between nodes inside clusters. Finally, ensemble clustering methods combine the clusters selected by different methods to mine final complexes.

However, all these methods rely on experimentally determined protein complexes data sets, which are afflicted by three main problems: (i) they are incomplete,¹⁹¹ (ii) they contain many false positives caused by experimental con-

ditions,¹⁹² and (iii) they have been measured in conditions that do not replicate the cellular environment.

Other techniques look at the functional similarity between proteins using Gene Ontology (GO) terms.¹⁹³ GO terms give information about a protein's localization within the cells, participation in biological processes, and associated molecular functions.¹⁹³ Interacting proteins belong to the same pathway and, thus, tend to participate in similar processes and/or have similar functions and/or live in similar cellular compartments.^{194–196} Therefore, these methods can classify interactions with a similarity score of GO terms.^{196–198}

3.2. Predicting Binding Interfaces and Poses

Binding takes place on portions of the protein interfaces (or binding sites); for natural dimeric interfaces, they range from a change in solvent-accessible surface area of 850–10 000 Å² (up to 7000 Å² for heterodimers).¹⁹⁹ These protein regions have been vastly characterized, for example, for what concerns their amino acid composition.^{200,201}

Studies of protein–protein interactions can be divided into two categories: (i) those identifying only the interfaces and (ii) those also predicting the complexes 3D structures.

Many methods in the former category only require the sequence of the investigated proteins because structural studies have shown that binding regions are characterized by a combination of geometrical and chemical complementarities^{200,202–204} that can be predicted from the proteins' amino acid sequences.^{205,206} In this context, homology modeling is very effective, as well. An example is BSpred,¹²⁰ which is based on neural networks. The algorithm underwent extensive training using sequence-based features, such as protein sequence profile, secondary structure prediction, and hydrophobicity scales of amino acids. Other recent sequence-based methods for predicting binding sites include simplified long short-term memory (LSTM) network,²⁰⁷ the DeepPPISP method that uses a combination of local and global sequence features,¹²⁴ and the DELPHI method, which combines CNN and recurrent neural network (RNN) in an ensemble structure.¹²⁶

Other protocols take as input the proteins' structures and focus on features that are known to characterize interacting molecular surfaces: the preferentially hydrophobic composition of the binding interfaces or the role of van der Waals interactions.^{107,200,208,209} Milanetti et al.^{210,211} proposed the use of 2D Zernike polynomials to rapidly compare protein molecular iso-electron density surfaces and assess their shape complementarity. The protocol called Zepyros (Zernike Polynomials analysis of pROtein Shapes) has been developed and distributed freely as a web application by Miotto et al.¹²⁵ and applied to characterize different biological systems^{212–214} and to optimize biomolecule interfaces.^{215–218} The work by Grassmann et al.²¹⁹ extended the procedure to account for electrostatic similarity.

Gainza et al.¹²⁸ proposed that proteins involved in similar interactions might possess shared fingerprints, regardless of their evolutionary origins. They introduced MaSIF (molecular surface interaction fingerprinting), a conceptual framework utilizing geometric deep-learning techniques to capture crucial fingerprints relevant to specific biomolecular interactions, including protein pocket–ligand interactions and protein–protein interaction. Two years later, the same framework was used to design novel protein interactions.²²⁰

Several other methods for predicting protein binding sites and residue–residue contacts involve a mixed approach between sequence information and structural information and are based on ML approaches. PAIRpred, a Support Vector Machine approach developed by Asfar Minhas et al.,¹²¹ employs a specific pairwise kernel and a combination of structural and sequence-based features, such as position-specific score matrices (PSSMs), position-specific frequency matrices (PSFMs), and solvent accessibility predictions. The same features used in PAIRpred are used in a graph convolutional neural network (GCNN) proposed by Fout et al.¹²²

Sequence-based and structural features are used also in BIPSPI (xgBoost Interface Prediction of Specific-Partner Interactions), which was developed by Sanchez-Garcia et al.¹²³ More recently, Krapp et al.¹²⁹ introduced PeSTO, which acts directly on protein atoms without the need for parametrization of the system's physics and is provided as a web server.

Techniques based on the structure of the investigated proteins have become powerful tools for predicting protein–protein association because of the success of structure modeling methods; when protein structures are not available from experiments, it is now possible to predict them with a certain degree of reliability starting from the sequence information. The recent advancements in the fields of protein structure and complex predictions have been recently reviewed by Wodak et al.²²¹ The state of the art in structure modeling is assessed by the Critical Assessment of Structure Prediction (CASP).²²² The top-ranked methods in the latest competitions of CASP have reached excellent results. Some software for structure prediction are Rosetta²²³ and D-I-TASSER.²²⁴ Rosetta developers introduced an all-atom force field that focuses on short-range interactions (e.g., van der Waals, hydrogen bonds, and desolvation effects) and discards long-range electrostatics. Predicted structures are refined via a Metropolis Monte Carlo-based refinement protocol. This method reaches a $C\alpha$ root-mean-square deviation (RMSD) = ~ 1.5 Å.²²³ A more recent method is D-I-TASSER, which can predict protein structure and function and was extended from the I-TASSER method developed by the Zhang lab,²²⁵ which integrates threading and deep learning. Its pipeline was ranked in the first place in the CASP15 experiment in all categories of protein structure prediction.²²⁶

Still, the most popular and effective tool for protein structure prediction is AlphaFold, the top-ranked method in the 13th CASP edition in 2018.²²⁷ Two years later, in CASP14, AlphaFold2 was presented²²⁸ and again outclassed the other methods. AlphaFold2 was able to predict the structure of a protein only from its amino acid sequence (we shall remind the reader that the human genome has been mapped since 2001²²⁹) with a median global distance test (GDT) score of 92.4 across all targets.²³⁰ The efficiency of this tool in predicting the coordinates of the heavy atoms of the structure relies on the integration of novel neural network architectures and training procedures based on the evolutionary, physical, and geometric constraints of protein structures. In terms of RMSD₉₅ (i.e., the $C\alpha$ root-mean-square-deviation at 95% residue coverage), it allowed a median backbone prediction accuracy of 0.96 Å against the 2.8 Å achieved by the next best-ranked prediction method. As for all-atom predictions, its accuracy was 1.5 Å RMSD₉₅ compared with the 3.5 Å RMSD₉₅ of the next best-performing approach.²²⁸

Prediction of the interfaces is not always sufficient; to understand the physical mechanisms involving protein complexes, determination of their 3D structures is a critical step. AlphaFold2 has given impressive results even in this context: when applied as a method for complex pose prediction, it exceeded docking approaches for heterodimers prediction,^{139,231} even if the modeling of antibody–antigen complexes was unsuccessful (11% success²³¹). Other limitations were still not resolved: AlphaFold2 can not predict intrinsically disordered proteins or regions.²³² Moreover, it is based on a multiple sequence alignment and, thus, fails in the prediction of novel structures. In October 2021, AlphaFold2 was extended to multiple chains and took the name of AlphaFold-Multimer.²³³ AlphaFold-Multimer was able to predict homodimers even better than heterodimers, despite still failing in the prediction of antigen–antibodies binding. AlphaFold-Multimer can predict heteromeric interfaces with DockQ²³⁴ higher than 0.23 and 0.8 in 70% and 26% of the cases, respectively.²³³ Compared with the performance of AlphaFold, the improvement was +27 and +14 percentage points in each case.²³³ Homomeric interfaces were predicted in 72% of the cases with a high accuracy 36% of the time (+8 and +7 percentage points, respectively).²³³ This year, Liu et al.²³⁵ proposed a further enhancement of AlphaFold2-Multimer based on a novel quaternary structure prediction system, MULTICOM.²³⁶ MULTICOM improves AlphaFold-Multimer predictions by generating diverse multiple sequence alignments and structural templates using both sequence and structure alignments and combining the AlphaFold-Multimer confidence score with the complementary pairwise model similarity score to rank models. It then classifies them through multiple complementary metrics and refines the models via a Foldseek²³⁷ structure-alignment-based refinement.

Many improvements were implemented on AlphaFold-Multimers, which resulted in a better performance compared with existing approaches, but its accuracy in complex prediction was still lower than AlphaFold2's accuracy for tertiary structure prediction.²³⁵

Despite the limitation given by the need to have experimentally resolved homologous structures, methods like AlphaFold that predict protein–protein interactions using sequence homology are among the most promising approaches.²³⁸ Given the importance of the structural characterization of protein complexes, it is not surprising that the number of large complexes deposited each year in the Protein Data Bank (PDB) is growing rapidly.²³⁹ A significant contribution to this trend comes from the advancement of experimental technologies, such as the developments of methods based on electron microscopy (EM), as reviewed by Saibil et al.,²⁴⁰ which are particularly suitable for systems of macromolecular assemblies. Consequently, new computational approaches based on homology information have been developed.^{241–245}

Baspinar et al. proposed PRISM,¹³⁴ which combines structural similarity and accounts for evolutionary conservation in the template interfaces. One of the most used homology-based methods is SWISS-MODEL, which is a pioneered, automated protein homology modeling server that has undergone consistent enhancements over the past 20 years.^{138,246–249} Recently, its modeling capabilities have been expanded to incorporate the modeling of both homo- and heteromeric complexes by utilizing the amino acid sequences of the interacting partners as the initial reference.¹³⁸ Never-

theless, these template-based approaches can offer a satisfactory prediction only when homolog complexes are available.

If the structures of the investigated molecules are available, useful computational tools are docking algorithms, that can take in input two (or more) component proteins and assemble them into putative models of the protein complex. As thoroughly reviewed by Pagadala et al.,²⁵⁰ in the last two decades, more than 60 docking tools have been proposed. In the following, we will provide a quick overview of the most efficient ones. The performances of docking algorithms are usually tested in competitions like CAPRI.²⁵¹

The docking methods participating in CAPRI can be classified into two approaches: the first one uses a matching technique that describes the partner as complementary surfaces,¹³³ and the second one simulates the actual docking process and computes the pairwise interaction energies.^{252–254} The former predicts the static structure of protein complexes, while the latter is based on simulations that also explore the kinetics leading to the binding and will thus be discussed in Section 3.3.

In complementarity-based docking, efficient search strategies are fundamental for global docking software to achieve good performance because of the high cost of the search stage. Many available global docking software rely on FFT correlation search algorithms that enable full systematic searches through translational and rotational degrees of freedom. The FFT approach was originally introduced to evaluate only the shape complementarity²⁵⁵ but has been later expanded to electrostatic interactions (e.g., FTDock²⁰²) or to both electrostatic and solvent terms, for example, in ZDOCK.²⁵⁶ ZDOCK, which also considers atomic contact energy to estimate electrostatic corrections and further improve its results, is one of the best-performing FFT-based docking programs.¹³³ However, simpler scoring functions can reduce the computational costs: MEGADOCK¹³⁷ is a FFT-based rigid-body docking tool similar to ZDOCK²⁵⁶ but ~ 7.5 times faster thanks to the evaluation of only shape complementarity and electrostatic for the scoring.²⁵⁷ Better results have also been obtained with PIPER,¹³¹ the first FFT-based approach to use pairwise structure-based potentials (extracted from structures of protein–protein complexes) called Decoys As the Reference State (DARS).²⁵⁸ Its protocol starts with the generation of a large decoy set of docked conformations that are used as a reference state. Using DARS, the contact frequency between two specific atom types in the native state and in the decoys are compared. By rewarding the occurrence in the interface of the atom pairs that frequently interact in the native complexes, PIPER produced at least 50% more hits in 19 of the 33 tests for enzyme–inhibitor complexes compared with ZDOCK.¹³¹ While still obtaining more hits than ZDOCK in 12 of the 16 test problems, PIPER was less accurate for antigen–antibody.¹³¹ This limitation was addressed six years after its development with the introduction of a new nonsymmetric potential extracted from antibody–protein complexes.²⁵⁹

All the presented docking methods only accept structures as input. Yan et al. proposed HDOCK,¹³⁵ which can accept both sequence and structure inputs. HDOCK is a hybrid docking strategy that combines a FFT-based global docking with template-based modeling. By looking at the available homologous complexes in the protein data bank, it can incorporate binding interface information -if any- into traditional global docking.

Among rigid body docking algorithms, there are also a few global programs that employ alternative search strategies, such as randomized search or local shape matching, across the entire protein structure, like PatchDock¹³⁰ and SymmDock.¹³⁰ The former uses geometric hashing algorithms to perform a global protein–protein docking using local shape descriptors (i.e., surface patches) to predict protein–protein and protein–small molecule complexes. The latter can only predict the structure of a homomultimer with cyclic symmetry given the structure of the monomeric unit.

A common problem of all these docking algorithms is that they can overlook plausible complex structures because of how the poses are ranked: after the search, for a pair of input proteins, the typical docking program generates tens of thousands of different binding poses, including both near-native (i.e., almost correct) and incorrect ones. Good scoring functions are fundamental for identifying the best poses among the proposed set. Even if many types of scoring functions have been developed, they often simplify the interactions on the basis of the shape complementarity and treat solvation effects implicitly.²⁶⁰ Even more advanced programs, like PatchDock, which utilizes geometric hashing for rapid surface patch matching, and ZDOCK, which also employs an advanced pairwise method to efficiently incorporate shape complementarity, do not consider many physical components, including entropy contributions from solvent molecules and proteins. In addition to this, Wass et al.²⁶¹ showed that for interacting proteins even the incorrect poses have relatively favorable scores, although they are lower than the correct ones.

Kozakov et al.¹³⁶ developed the ClusPro docking server, which improves the docking prediction by taking into account entropic effects.²⁶² The protocol starts by implementing PIPER as a rigid body docking step and then performs a RMSD clustering of the 1000 lowest energy structures generated. To choose the most likely near-native clusters, their size is evaluated instead of their energy values: the centers of the largest clusters are selected rather than the lowest energy structures. The selected structures are then refined using energy minimization. A similar procedure is followed by MEGADOCK,¹³⁷ which is based on a FFT-based rigid-docking scheme. MEGADOCK classifies protein pairs as interacting if among the docking-generated poses there are clusters of similar poses with significantly favorable docking scores compared with the others.

All the presented methods allow for only partial protein flexibility. Methods that perform side-chain optimization are able to reach higher-accuracy solutions compared with more rigid methods.²⁶² Desta et al.²⁶² recently tested some of the most used docking servers on the same data set. They showed that ZDOCK and pyDock²⁶³ (another rigid body server similar to ClusPro) obtained the worst results compared with SwarmDock,²⁶⁴ a flexible docking algorithm that was able to produce slightly more models in the top 1 prediction (predicting good models for 23.5% of the 51 targets). This method models flexibility using a linear combination of elastic network normal modes and combines a local search with a variation of the particle swarm optimization (PSO) algorithm.²⁶⁵ In this algorithm, each particle represents a potential solution and moves in a D-dimensional space (including Cartesian space, orientational space as quaternions, and coefficient space for a linear combination of anisotropic elastic network normal modes). During the search, each particle adjusts its position according to its own experience and the

swarm's experience. PSO can simultaneously optimize translational, orientational, and conformational degrees of freedom. It is worth mentioning that SwarmDock was implemented in 2010 in a study on macromolecular crowding⁸⁷ that will be discussed in the next sections. However, the SwarmDock server¹³² has the limitation of handling only files with less than 10 000 atoms.

Another flexible docking protocol, HADDOCK2, was proposed by Van Zundert et al.⁹¹ and has now become one of the most exploited data-driven approaches. Data-driven methods use biochemical and/or biophysical interaction data (e.g., chemical shift perturbation data resulting from NMR titration experiments²⁶⁶) to guide the docking process. In HADDOCK2, experimental data are used to define the simulated annealing in torsion angle space to introduce partial flexibility. Thanks to this flexibility, HADDOCK2 has been shown to yield more accurate models than ClusPro.²⁶²

However, considering only side-chain flexibility has been argued to be insufficient: in the CAPRI experiment, the majority of failures were attributed to the inaccurate prediction of conformational changes in proteins during protein–protein interactions.²⁶⁷

3.3. Predicting the Structural Determinants of Binding

Different theories to describe the kinetics leading to the binding pose have been proposed. The prevailing view has evolved from the early “lock-and-key” hypothesis to the “induced-fit” and “conformational selection” models.²⁶⁸ In the former, partners undergo very small changes of backbone and side-chain atoms upon binding (root-mean-square deviations of 0.6 and 1.7 Å, respectively²⁶⁹). However, it is known that proteins are dynamic: both “induced-fit” and “conformational selection” hypotheses take this aspect into account. According to the former, the interactions between two approaching structures induce conformational changes. Conversely, the latter states that the protein's bound state can be explored even in the absence of the molecular partner.

Potential transition states have been deeply investigated by experimental studies, like double mutant cycles and paramagnetic relaxation enhancement.^{270–272} However, the resulting data are often indirect or limited to specific cases (like metalloproteins or proteins with attached paramagnetic spin labels). Computational methods can instead provide atomic-level observations of the association pathways and explored conformation for a diverse set of protein complexes.

The most employed methods to accomplish this purpose are all-atom (AA) molecular dynamics (MD) simulations, which provide both structural and dynamical insight into protein–protein binding. Some of the most widely used simulation software packages include CHARMM,¹⁴² AMBER,¹⁴¹ GROMOS,²⁷³ and GROMACS.¹⁴⁰

This approach allows us to witness the evolution of the system made of the involved protein partners in terms of association/dissociation events along the simulated trajectory. In this way, it is possible to sample a large number of binding poses while gaining information about intermediate states characterizing the association pathway. Thus, MD simulations are often at the core of simulation-based docking techniques.^{274,275}

In this context, two intermolecular interaction potentials describing Coulombic and van der Waals interactions are present between atoms of the two molecular partners to modulate their interaction. In particular, given two atoms a_i

and a_m holding partial charges q_i and q_m , the Coulombic interaction between them can be computed as

$$E_{lm}^C = \frac{1}{4\pi\epsilon_0} \frac{q_i q_m}{r_{lm}} \quad (1)$$

where r_{lm} is the distance between the two atoms, and ϵ_0 is the vacuum permittivity. van der Waals interactions can instead be calculated by, for example, the Lennard-Jones potential:

$$E_{lm}^{LJ} = \sqrt{\epsilon_l \epsilon_m} \left[\left(\frac{R_{\min}^l + R_{\min}^m}{r_{lm}} \right)^{12} - 2 \left(\frac{R_{\min}^l + R_{\min}^m}{r_{lm}} \right)^6 \right] \quad (2)$$

where ϵ_l and ϵ_m are the depths of the potential wells of a_l and a_m , respectively, and R_{\min}^l and R_{\min}^m are the distances at which the potentials reach their minima.

Commonly available computational power for MD simulation typically allows them to sample only a few binding poses in the short microsecond simulation frame given the long time scale of the interconversion between poses.

Longer time scales (as well as larger systems) can be achieved with coarse-grained (CG) models²⁷⁶ in which several atoms are lumped together in effective interaction sites (beads). Compared with AA representations, CG models have fewer interactions, smoother free energy profiles, and fewer degrees of freedom. Even if CG representations usually need additional terms to preserve the native structure that limits their conformational sampling, alternative structural restraints have been proposed.²⁷⁷

To find a compromise between the benefits and limitations of AA and CG models, many multiple-resolution modeling strategies have been developed. Some of them treat at atomistic resolution the biomolecule's functionally relevant portion,²⁷⁸ while the rest of the system is described with a CG model, which results in a lower computational cost compared with AA model.²⁷⁹ One of the most used force fields is MARTINI.^{280,281} As a downside, these methods often require lengthy reference all-atom simulations and/or the usage of off-shelf coarse-grained force fields to parametrize the CG-modeled part.

The treatment of the interactions between different levels of resolutions has to be determined, as well.¹⁴⁷ To overcome these limitations, Fiorentini et al.¹⁴⁷ proposed a novel multiresolution modeling scheme for proteins, dubbed coarse-grained anisotropic network model for variable resolution simulations (CANVAS). CANVAS, which can be implemented on the most commonly used MD platforms, allows for a user-defined modulation of the resolution level throughout the system structure and a fast parametrization of the potential without reference simulations.

In combination with CG and AA models, structure-based models (SBMs), also called G \bar{o} -type models, can be used.^{282,283} While CG methods model the coarse-grained resolution of the system, SBMs measure the coarse-grained resolution of its force field.

They usually take an experimentally solved native conformation of the protein to determine the native contacts and use them to define the interaction potential for the simulation by introducing bias toward the native state. This reduces the force field complexity without loss of essential information according to energy landscape theory and the principle of minimal frustration,²⁸⁴ which states that a protein energy landscape has a funnel-like shape biased toward the native

state. This ensures the dominance of native over non-native interactions, thereby enabling an efficient folding. The prevalence of native interactions in the folding event has been confirmed by atomistic simulations.²⁸⁵ Thanks to this simplified interaction potential and the consequent reduced computational costs, structure-based models are often used to simulate many cycles of folding and unfolding.

Another popular approach—that can use both atomic and coarse-grained representations—to shed light on molecular processes at time scales that are not accessible with single unbiased MD simulations is the combination of MD simulation with Markov-state modeling (MSM). Popular software packages are pyEMMA¹⁴³ or MSMBuilder.¹⁴⁶

Another approach is speeding up the sampling of rare events by enhancing the conformational sampling performed during the simulation, usually by scaling the temperature or using collective variables (CVs). CVs are identified by dimensionality reduction methods that can determine which properties are of interest for the description of the high-dimensional conformational ensembles.²⁸⁶ In addition to facilitating the interpretation of the huge amount of data produced during a simulation, CVs can direct the conformational sampling to efficiently cover the free energy landscape.²⁸⁷

Some of the algorithms using CVs to enhance the sampling include steered MD,²⁸⁸ umbrella sampling,²⁸⁹ Hamiltonian replica exchange MD (REMD),²⁹⁰ and metadynamics.²⁹¹

Alternatively, the methods using temperature to enhance the sampling include simulated annealing,²⁹² simulated tempering,²⁹³ temperature replica exchange MD (T-REMD),²⁹⁴ and replica exchange with solute tempering (REST²⁹⁵ and REST2²⁹⁶).

In all these cases, a central role is given to the force field, whose quality determines whether the sampled conformations are realistic. Because of this, force fields are constantly being improved.

The most popular atomic-level protein force fields family include AMBER,²⁹⁷ CHARMM,²⁹⁸ OPLS,²⁹⁹ and GROMOS.³⁰⁰ They explicitly model each atom or, in the case of GROMOS, all heavy and nonaliphatic hydrogen atoms, and describe the interactions between bonded and nonbonded atoms with mathematical functions whose parameters are fitted to structural or thermodynamic quantum mechanical or experimental data.

Among the AMBER force fields, ff99SB³⁰¹ has been often recommended for proteins. It has been shown to improve amino-acid-dependent properties and reproduce the differences in amino-acid-specific Ramachandram Map using amino-acid-specific CMAPS.³⁰² Used with the OPC (optimal point charges) water model,³⁰³ it was also found to improve the accuracy of atomistic simulations of disordered proteins.³⁰⁴ However, with time, ff99SB showed weaknesses in some amino acid side chain dihedral parameters;³⁰⁵ in 2019, an updated model with improved backbone profiles for all 20 amino acids, named ff19SB,³⁰⁶ was proposed.

Similar backbone CMAP potentials were used two years prior for CHARMM36m,²⁹⁸ an updated version of CHARMM22,³⁰⁷ which had been for a long time the first choice among CHARMM force fields for proteins simulations. This improvement increased the accuracy of generating ensembles for disordered proteins.

The evaluation of force fields treatment of disordered proteins is commonly done by comparing them^{308,309} because the intrinsic complexity and “roughness” of the disordered

proteins energy landscape can reveal force field deficiencies and strengths. For instance, a comparison between CHARMM36m and ff19SB concerning their ability to describe disordered proteins while retaining proper description of the folded ones has been recently performed by Abriata et al.³¹⁰ Both CHARMM36m and ff19SB were applied with the recommended water model, respectively modified by TIP3P²⁹⁸ and OPC.³⁰⁶ CHARMM36m was found to favor more compact states than ff19SB-OPC, which leads to the overstabilization of aggregates and secondary structures. ff19SB better predicted weak dimerization of a soluble protein while still being able to reproduce the expected aggregation of another system. Conversely, CHARMM36m predicted residue-wise alpha helical propensities better than ff19SB.

In the last decades of protein simulations, other popular force fields have been OPLS-AA³¹¹ and its modification OPLS-AA/L.³¹² In 2015, Robertson et al.³¹³ developed OPLS-AA/M, which proposes a new parametrization of the backbone and side chain dihedral. With a similar strategy, Harder et al.²⁹⁹ presented OPLS3 a year later. The former has an improved ability to simulate disordered proteins, while the latter well simulates protein–ligand binding.

A similar reparameterization against experimental data was proposed for the GROMOS force field.³⁰⁰ Even if it resulted in an improved prediction of secondary structure propensities and structural parameters, investigations on how these new parameters could perform for intrinsically disordered proteins are still ongoing.

3.4. Predicting the Strength of the Interaction

Investigation of protein interaction strengths is crucial because they are associated with the protein complex functionality.¹⁶⁶ Their evaluation has many applications, such as predicting and explaining experimental protein–protein dissociation constants and the effect of different mutations on equilibrium constants or ranking the binding poses generated by protein–protein docking algorithms.

The strength related to the binding interaction is commonly quantified by the binding affinity. Usually, the binding affinity is described with the experimental measure of the dissociation constants (K_d)^{314,315} related to the change of Gibbs free energy after binding (ΔG):³¹⁶

$$\Delta G = RT \ln K_d \quad (3)$$

where R is the gas constant ($8.3144 \frac{\text{K}}{\text{J}_{\text{mol}}}$), and T is the absolute temperature.

ΔG can also be expressed as $\Delta G = \Delta H - T\Delta S$, where ΔH and ΔS are the variations in enthalpy and entropy of the system after binding. Enthalpy measures the total energy of the system, given by the sum of the solute and solvent internal energies and the amount of energy required to establish the system's physical dimensions. Its variation upon binding is determined by (i) the formation of noncovalent interactions (van der Waals contacts, hydrogen bonds, ion pairs, and any other polar and apolar interactions) at the interface, (ii) the loss of hydrogen bonds and van der Waals interactions between proteins and solvent, and (iii) the reorganization of the solvent close to the complex. The enthalpic contribution to binding has been well characterized thanks to detailed atomic-resolution structural models, while the factors determining the total binding entropy are still being characterized. Entropy can be interpreted as a measure of the disorder of a system's components (atoms and molecules) or as the number of

configurations that a system can take. Its total variation after binding ΔS can be decomposed in four terms:

$$\Delta S = \Delta S_{\text{conf}} + \Delta S_{\text{sol}} + \Delta S_{r-t} + \Delta S_{\text{other}} \quad (4)$$

where ΔS_{conf} , ΔS_{sol} , and ΔS_{r-t} are the variations in conformational, solvation, and rotational–translational entropy, respectively. ΔS_{other} is related to other processes, such as proton binding or release.³¹⁷ ΔS_{sol} is the most discussed contribution, which is typically framed in terms of hydrophobic effects³¹⁸ and related to the variation of the solvent-accessible surface area (SASA).³¹⁹ Historically, the other components have been neglected because they are not associated with simple experimental measurements. Conformational entropy is particularly challenging because of the potentially high number of local minima, anharmonicity, and high-order coupling among degrees of freedom. However, NMR spectroscopy has emerged as a powerful tool to measure protein internal motions, as reviewed by Wand et al.,³²⁰ and the observation of NMR-derived methyl order parameters has suggested that ΔS_{conf} can be of comparable magnitude to ΔS_{sol} .³²¹ Using a dynamical proxy provided by NMR relaxation methods, it was shown that for a data set of 28 complexes ΔS_{conf} was an essential contributor in ~25% of the cases to the high-affinity binding.³²² However, these experimental techniques do not directly measure correlated dynamics (i.e., intramolecular couplings), protein main chain motions, or angular dynamics, but rather reconstruct them from empirical, linear fits across sets of reference molecules. Even if the existence of a strong linear relationship between ΔS_{conf} and the average change in NMR-observed order parameters has been observed, these indirect observations seem to yield only qualitative estimates of ΔS_{conf} .³²³

This limitation can be solved through computational methods that allow a direct investigation of what the contributions of entropy to protein binding affinity might be.

In silico approaches for binding affinity prediction can be divided into scoring functions and free energy calculation methods. The former are generally faster but often only measure the strength of noncovalent interactions and adopt several simplifications.^{148,149,154,324} The latter need intensive computations but are more precise since they implicitly include both enthalpic (in the form of potential and solvation energy) and entropic components (dynamics and flexibility of the binders and solvent effects).^{325–328}

Scoring functions are often based on empirical, knowledge-based/statistical, or force-field-based energy functions. Empirical methods use regression analysis with known binding affinity data of experimentally determined structures to parametrize different interactions as favorable or unfavorable energy terms. For example, PRODIGY¹⁴⁹ combines the number of contacts at the interface with properties of the noninteracting surface. By doing so, it reaches a Pearson's correlation coefficient of 0.73 between the predicted and measured binding affinity. Empirical methods are fast, but their efficacy depends on the training set. Moreover, they do not consider the intramolecular energy changes occurring after binding, allosteric regulation, and solvent and cofactor effects. Explicit solvent molecules are often excluded, and solvation contributions are estimated using an implicit solvent model. Since they comprise many energy terms, double-counting (overfitting) can be a problem, as well.

Another fast tool that is not hindered as much by the training set quality or overfitting errors are knowledge-based (also called statistical) methods. These techniques assume that

atomic interprotein interactions more frequent than those expected by a random distribution are likely to be energetically favorable and, thus, improve binding affinity.³²⁴ Statistical methods analyze interacting atom pairs from complexes with known structures and convert these data into a pseudopotential (or mean force potential) that describes the preferred geometries of the interacting atoms. For example, the relationship between the protein–protein binding affinity and the buried solvent-accessible (BSA) surface area upon complex formation³²⁹ has been vastly investigated. The BSA, which is defined by subtracting the SASA of the complex from the SASA of the two single proteins considered alone in solution, correlates quite well with the experimentally measured binding affinity when only structures that undergo small changes after binding are considered.^{330,331} Thus, the fraction of accessible surface in the protein interface has been considered by a variety of knowledge-based statistical potentials.^{154,259,332}

Despite the importance of residues belonging to the interface in predicting binding affinity, residues outside of the two binding sites can have a key contribution to binding characterization, as well.^{107,333} Wang et al.¹⁵⁴ recently proposed a knowledge-based potential that considers both the residues at the binding interface and the noninterfacial ones. It assumes that both contributions depend on their local structural environments, such as secondary structural types and SASA; after refinement of the weights for all corresponding components against a large-scale data set, there was a strong improvement (from 0.046 to 0.66) in the correlation between the absolute values of the experimental and calculated binding affinities. Moreover, the method was able to predict the relative changes of binding affinities from mutations.

A correlation coefficient of ~0.73 between experimentally measured and predicted protein–protein (peptide) binding affinities was also obtained by DFIRE, a knowledge-based statistical energy function proposed by Zhang et al.¹⁴⁸

Alternative approaches are force-field-based techniques that consider the energy terms from a molecular mechanics (MM)-type force field description of the binding partners.^{150,334} To reduce the complexity, the classical force field often takes into account only the enthalpic contribution of the complex intermolecular noncovalent interactions, which neglects the variation after the binding in intramolecular interactions, interactions involved in the solvent, and entropic effects. Many terms within the force field are weighted appropriately to provide an optimal correlation with experimental data on a training set of complexes.^{335,336}

Usually, these methods are applied to complexes whose geometry is represented by only one structure. To improve the statistics, free energy calculation-based methods employ MD simulations. By assessing the stability of the system during the observation time, MD simulation-based methods also allow the evaluation of entropic contributions in the calculation of the binding affinity. Even if explicitly considering the effect of interfacial waters has been argued to be fundamental for correctly determining the entropic contribution,³³⁷ a common approach is to use explicit solvent MD simulations first and then replace the surrounding environment with a simpler implicit solvent model for further evaluation. This protocol is often followed by MM Poisson–Boltzmann/surface area (MM-PBSA) or MM/GBSA, which uses the generalized Born method to evaluate the free energy of an ensemble of protein–protein complex conformations taken from simulation trajectories. In MM-PBSA, which was developed by Kollman et

al. in the late 90s,³³⁸ the free energy of a state is estimated from the following sum:³³⁹

$$G = E_{\text{bind}} + E_{\text{el}} + E_{\text{vdW}} + G_{\text{pol}} + G_{\text{np}} - TS \quad (5)$$

where the first three terms are standard energy bonded terms (bond, angle, and dihedral), electrostatic, and van der Waals interactions. G_{pol} and G_{np} are the polar and nonpolar contributions to the solvation free energies, respectively. The last term is obtained by the product between the absolute temperature, T , and the entropy, S , that, when considered, is usually estimated by a normal-mode analysis of the vibrational frequencies³²⁵ that only considers its conformational component, but even this term is often neglected to reduce the computational costs. ΔG_{bind} is estimated from the free energies of the reactants and product, as follows:

$$\Delta G_{\text{bind}} = \langle G_C \rangle + \langle G_{P_1} \rangle + \langle G_{P_2} \rangle \quad (6)$$

where C , P_1 , and P_2 represent the three possible states: (i) the complex given by the structural interaction between the two interacting proteins and (ii) the first and (iii) the second protein partner. Strictly, the averages indicated in eq 6 should be estimated from three separate simulations³⁴⁰ by considering separately both the simulation of the complex and the simulations of the two interacting proteins alone in solution. However, the most common approach is to simulate the complex and create the ensemble average of the free P_1 and P_2 interacting proteins by simply removing the appropriate atoms. To reduce the computational cost, MM-PBSA calculation can overlook entropic contributions. The performance of the MM-PBSA and MM-GBSA methods for binding affinity prediction when not considering entropy were tested by Chen et al.³⁴¹ on a set of 46 protein–protein complexes. To further reduce the computational times, they tested their application on single minimized structures and with implicit water minimization. Using MM-GBSA, in combination with a low interior dielectric constant of 1 and the AMBER ff02 force field, they observed a correlation of -0.647 between the predicted binding affinities and the experimental data. A worse correlation of -0.523 was obtained with MM-PBSA. These results were compared with what could be obtained from a set of MD calculations. The predictions were slightly better than those based on the minimized structures. However, explicit water models did not improve the prediction compared with implicit ones, in contrast with other studies.^{337,342,343}

Others have focused on including the conformational entropy evaluation while preserving reasonable computational costs. The interaction entropy approach³²⁶ tackles this challenge by extracting it directly from MD simulations. It calculates relative entropy changes (for example, following the binding) by looking at the interaction energy fluctuations. Compared with the normal mode analysis, this approach is more efficient and does not use an approximation for the entropy calculation. As suggested by the authors, the calculated interaction entropy can be combined with the solvation free energy obtained through the standard MM-PBSA method to calculate the binding affinities in protein–protein complexes.^{344–346} Since the interaction entropy is mainly determined by the highest spikes of interaction energy, it has been argued that for longer simulations there could be convergence problems³⁴⁷ that make the approach less robust than normal modes or quasiharmonic approximations. Quasiharmonic approximations are one of the most popular

methods for extracting conformational entropy from MD.³⁴⁸ This method diagonalizes the covariance matrix of the atomic coordinates and approximates their distribution as a Gaussian probability distribution of conformations with variances equal to those provided by the MD. However, when Cartesian coordinates are used, the MD frames have to be rotationally and translationally superimposed to obtain a reasonable covariance matrix, and the results depend on how the superposition was done.³⁴⁹ Another drawback is overestimating the entropy for systems with multiple energy minima.³⁴⁹ Good results for molecular systems having multiple energy wells were obtained by Killian et al.³²⁷ with a more direct estimation of probability densities over the conformational space from MD simulations frames. They computed the conformational entropy with a mutual information expansion by taking into account correlated motions among the various internal degrees of freedom (DOF) of a molecule. The mutual information expansion approximation computes the entropies over joint distribution for groups of small numbers of DOF (the so-called marginal entropies). All higher-order terms of couplings are truncated. By assuming that the higher-order terms can be neglected, it avoids the convergence problems that afflict direct estimations over all DOF for increasing system size due to the exponential scaling of sampling requirements. A similar framework was proposed by King et al.³²⁸ that applied the maximum information spanning trees³⁵⁰ to molecular simulations frames. Compared with the mutual information expansion, which includes all couplings of a particular order in the approximation, this approach selects a subset of the same couplings to guarantee a lower bound on the entropic contribution, thereby guaranteeing better results for larger systems and a faster convergence.

All these MD-based methods present some drawbacks related to their computational cost. For instance, kinetic traps may occur, thereby resulting in intermediate non-native bound states featuring lifetimes even longer than the simulation time. The presence of such bottlenecks could strongly affect the sampling of other more kinetically favorable states that may correspond to the more likely structure in physiological conditions.²⁷⁴

Pan et al.²⁷⁴ presented an enhanced sampling method that allowed for the first time the observation of these phenomena during a single simulation trajectory. They proposed a tempered binding simulation in which a simulated Hamiltonian tempering scheme scales at regular time intervals the strength of the interactions between atoms belonging to the same protein and between protein and solvent. This strategy facilitated the observation of complex dissociation. By performing long time scale full-atom simulations of five protein–protein complexes, they found that when protein–protein contacts occur far from the native interface the monomers are more likely to dissociate. Eventually, the complex would associate near to the correct native interface rather than explore all the possible binding regions while maintaining the contact. The most stable complexes resulting from the simulations corresponded to the structures determined crystallographically within atomic resolution, which proved the soundness of the method. Moreover, the fast sampling allowed for the observation of some alternative bound states that could have functional relevance, as well.

Another enhanced sampling technique was implemented by Wang et al.³¹⁵ where a complex is dissociated by performing a standard metadynamics, and the degree of difficulty of pushing

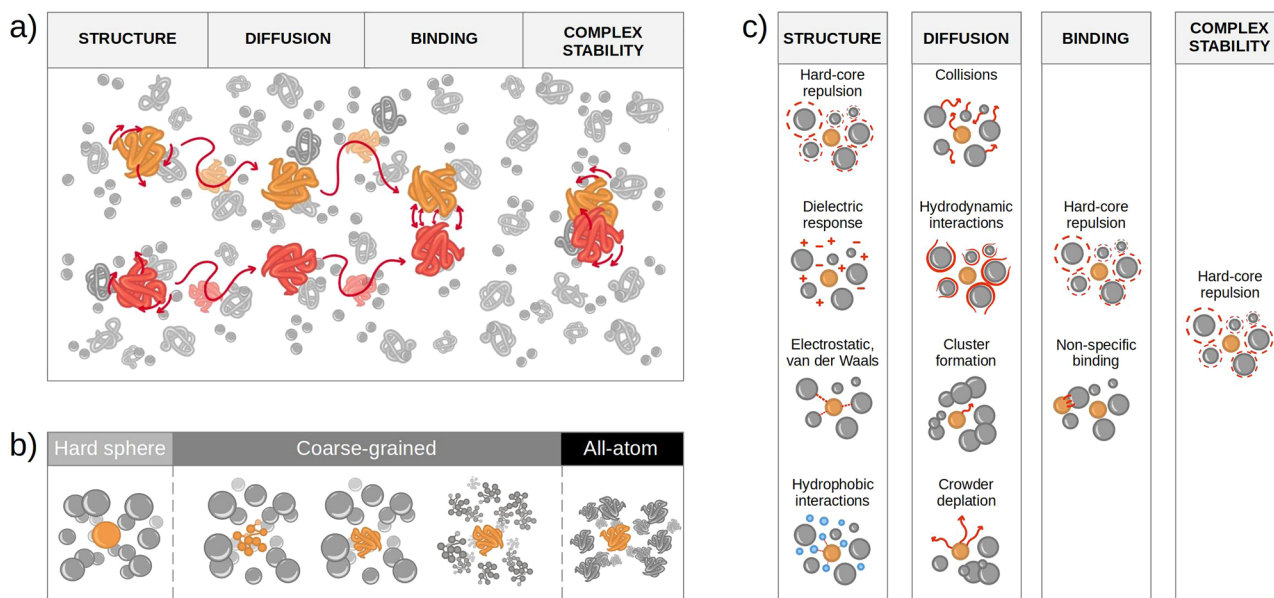


Figure 5. Crowding effects on the different stages of protein binding. (a) Proteins (orange and red) can be found inside the cell as single monomers whose folding and conformational variation are influenced by surrounding crowders (gray). The monomers have to navigate this crowded environment to find each other and bind. The dynamics leading to this binding are affected by the crowders, as well as by the stability and dynamics of the formed dimer. (b) The evaluation of many features of the structure and dynamics of proteins (orange) surrounded by crowders (gray) can be performed with computational methods. The results depend on the level of approximation by which the system has been described. In the most simplified models, both proteins and crowders are described as rigid spheres. More detailed representations (at the cost of higher computational costs) can be obtained with coarse-grained (CG) models. In the lowest-resolution CG systems, crowders are represented as hard spheres and proteins by subsuming multiple atoms into beads or at an atomistic level. More detailed CG studies employ an atomistic description of proteins with crowders as a collection of beads corresponding to multiple atoms. (c) Computer simulations and theoretical calculations have been extensively implemented to study the dynamics of proteins and interacting partners in a crowded environment. The main factors influencing the structure of single proteins characterized by these studies up to now are hard-core repulsion, dielectric response, and electrostatic and hydrophobic interactions. Diffusion, instead, is mainly affected by collisions with the crowders, hydrodynamic interactions, cluster formation, and crowders depletion. In a similar way as for single protein folding, protein binding has been found to depend on hard-core repulsion and nonspecific binding between interacting proteins and crowders. For what concerns the stability of the resulting complex, currently, only the effect of hard-core repulsion has been characterized.

the system out of the bound state region is used to compute the binding affinity.

Even lower computational costs can be achieved with ML methods. These methods can be structure-^{107,153,156,351} or sequence-based.¹⁵²

For example, Vangone and Bonvin proposed a scoring method based on contacts, which incorporated optimal weights for different types of interface contacts (e.g., polar–polar, polar–nonpolar, etc.), and noninteracting surface.¹⁰⁷ The implemented linear regression model demonstrated a correlation with experimental binding affinity data on a benchmark set of 0.73.

A higher correlation (0.75) on a larger data set (SKEMPI³⁵²) was obtained by mmCSM-PPI, proposed by Rodrigues et al.¹⁵³ MmCSM-PPI is a method to predict binding affinity based on graph-based signatures, which describe the distance patterns between atoms on the binding interface.

On the same data set, more recently Wee et al.¹⁵⁵ have tested PerSpect-EL, which obtained a correlation of 0.853.

Another structure-based tool is PPI-Affinity, which relies on a support vector machine (SVM) to predict the protein–protein and protein–peptide binding affinity, as well as to generate and rank mutants of a given structure.¹⁵⁶ Tested again on SKEMPI, it resulted in a correlation of 0.77.

Finally, among the most recent sequence-based methods there are PIPR¹⁵¹ and ISLAND.¹⁵² The former, tested on

SKEMPI, reached a correlation of 0.87, while the latter stopped at a Pearson correlation of 0.44.

4. EFFECT OF CROWDING ACROSS PROTEIN BINDING

Experimental approaches can describe many features of the structure and dynamics of biomolecules, albeit they are not comprehensive enough in terms of space or time resolution. Indeed, often when a good atomistic detail is achieved, a poor time resolution is accomplished and in a setting far from *in vivo* conditions, as in the case of crystallography and cryo-EM. However, the best-performing single-molecule fluorescence methods can track biomolecules' dynamics in conditions close to real-life cell interiors and on a time scale that could reach milliseconds^{353,354} but at only tens of nanometers about spatial resolution.³⁵⁵ Further, NMR spectroscopy can grasp the dynamics at an atomistic level and *in vivo*,³⁵⁶ but the analysis of NMR spectra is time-consuming.³⁵⁷ The development of models that can complement experimental measurements in the difficult task of studying the details of protein behavior has become one of the major challenges that theorists have to face. Many efforts have been devoted to quantitatively predict how crowding influences the main steps determining protein binding, which are summarized in Figure 5a. In this respect, in the past decade, theoretical calculations and molecular simulations have started to consider not only water and ions but also other binding partners, metabolites, or crowders.

Along the line of the historical bipartition of the classification of crowding effects, computational studies about crowding can be divided into two big families: those that consider only volume exclusion and those that add soft interactions. From a methodological point of view, computational approaches can also be classified according to the detail level of the molecular representation or complexity of the interactions they simulate, as shown in Figure 5 panels b and c, respectively.

Theoretical investigations that only consider the excluded volume effect use simplified models that constrain the physical volume available to a molecule by applying a spherical potential or penalizing increased solvent-accessible surface areas. The simplest models describe both proteins and crowders as hard spheres. The typical radius of these crowders is between 10 and 50 Å, with an average of 25 Å, to match the size of a crowding agent of interest, like a folded protein or a crowding inert polymer (e.g., Ficoll). The spherical representation of crowders can be sometimes substituted by other shapes, like dumbbell crowders,⁵⁰ where two spheres are linked by a harmonic bond or spherocylinders. Despite being computationally convenient, these models neglect intramolecular dynamics and conformational sampling.

More complex models simulate a single molecule represented in a CG manner surrounded by spherical crowders. CG models have emerged as powerful tools to investigate protein interactions in crowded environments, thereby offering a simplified representation that captures the essential features of the system while reducing computational complexity.⁸⁶ By incorporating effective interactions between proteins and crowding agents, such as excluded volume effects and steric hindrance, coarse-grained models can mimic the crowded cellular environment and investigate its impact on protein–protein interactions.⁶⁶ Coarse-grained models also facilitate the investigation of the interplay between protein flexibility and macromolecular crowding.⁶⁶ In combination with spherical crowders, coarse-graining of proteins has often been the method of choice when dealing with simulations of crowded environments, as reviewed by Ostrowska et al.¹⁶ Nevertheless, the reparametrization of the interactions is not straightforward, and a loss in the discrimination of similar molecules could occur.³⁵⁸

Recent works have managed to simulate AA models of native proteins despite still using some reduced models of crowders (spherical or CG³⁵⁹). CG and hybrid models can be defined with²⁸¹ or without^{86,359–361} an explicit solvent. The former often involves stochastic dynamics simulations, like Brownian ones,^{362–364} and are used to evaluate the effect of crowding on diffusive properties. The latter are instead employed to study how crowding modulates protein folding, stability, or association equilibria^{96,100,359,360,362} and to estimate binding free energies.⁸⁶ Despite the ability of these models to capture the main features of biomolecular structures and dynamics of cellular environments, they typically fail to capture other aspects, like flexibility.

This is an essential feature to be included in a model aimed at retrieving information about properties such as conformational sampling, stability, and dynamics of biomolecules in the cellular environment. Flexibility should be accounted for through an interaction potential allowing not only to maintain the native states in dilute conditions but also to sample non-native states due to interactions with surrounding cellular components.^{66,365,366} A well-performing model should be able

to provide an accurate physical description of the weak interactions between biomolecules and permit making predictions only relying on the dominating nonspecific interactions. Finally, the model should consider the main features of the solvent together with the effects of reduced dynamics and hydrodynamics interactions that a crowding environment produces, as reviewed by Feig et al.³⁶⁷

In the most recent developments, both proteins and crowders are described at an atom-level resolution. Handling such large numbers of particles is computationally expensive. Nevertheless, enhanced sampling methods allow extensions of several orders of magnitude of the accessible time scales, something that could be achieved also by inferring kinetics from Markov-state models built upon short simulations.¹⁴⁶

In addition to direct simulations where crowders are considered explicitly, postprocessing techniques have also been designed in which the protein (be it in an AA or CG representation) and the crowders are simulated separately. The protein conformations explored in isolation are randomly placed in the snapshots picturing the crowder's trajectory and weighted based on the fraction of successful insertions.^{368,369}

All in all, the physical modeling of cellular environments at an atomic level is becoming more and more affordable thanks to the ever-growing progress of both computational methods and computer hardware devices, which have been recently reviewed by Feig et al.³⁷⁰

One of the major challenges remains the design of an accurate interaction potential that can reproduce the balance between molecular stability, weak interactions, and solvent interactions in this complex scenario. Many investigations have been conducted to check if the existing classical force fields for single-biomolecule simulations are suitable for heterogeneous systems with interactions between different classes of molecules.^{366,371–374}

An often-encountered drawback regards how protein–protein and protein–water interactions are treated. They appear to be imbalanced so that natively folded configurations are favored, which results in poor modelization of IDRs because of a loss of their typical conformational heterogeneity³⁷⁵ and aggregating processes.^{373,376} In this context, GROMOS 54a7³⁷⁷ is one of the preferred force fields since it has a lesser preference for aggregated states.²⁴ Overstabilization of nonspecific sticking between proteins has been discussed, as well: Rickard et al.³⁷² quantified it for three force-fields, CHARMM 36m (C36m),²⁹⁸ CHARMM 22* (C22*),³⁷⁸ and CHARMM 36m with CUFIX (C36mCU),³⁷⁹ by evaluating the nonspecific protein–protein contact size. While experimental studies have indicated an average contact area of ~ 5.7 nm²,³⁸⁰ they found an area of 9.5, 8.1, and 8.5 nm² for C22*, C36m, and C36mCU, respectively. After observing the large contribution of hydrophobic residues to the largest contacts, they suggested that corrections to the hydrophobic interactions, for example, with new water models, may address the issues. Other groups have instead proposed to increase the Lennard-Jones protein–water interactions by about 10% from the existing force fields.^{373,374}

A set of works of a more methodological nature checked the ability of available force fields to reproduce the crowding effect on fundamental types of interactions, such as salt bridges and hydrophobic interactions. Andrews et al.³⁸¹ conducted all-atom explicit solvent MD simulations to investigate the impact of increasing solute concentration on the behavior of zwitterionic amino acids in water. The amino acids studied included

Table 2. Summary of the Most Cited/Recent Simulation-Based Studies Discussed in This Review^a

ref	year	model	software	force field	enhancement	crowder	concentration	protein	solvent	length
373	2017	AA	CHARMM/L GENESIS, NAMD, OpenMM	modified CHARMM36 ($\lambda = 1.1$)	REMD	HP-36	34, 77, 138 g/L	HP-36	TIP3P	20 × 100 ns
374	2019	AA	CHARMM, NAMD, OpenMM	modified CHARMM36 ($\lambda = 1.09$)		HP-36	135 g/L	HP-36	TIP3P	1000 ns
392	2020	AA	MUPHY	OPEP		BSA	200 g/L	SOD1	implicit	1 μ s
96	2021	AA CG	GROMACS MUPHY	AMBER #99SB-ILDN OPEP	REST2	BSA LYZ + BSA	1:1, 2:1 300 g/L	SOD1 C12	TIP3P implicit	24 × 0.5 μ s 1 μ s
365	2010	AA CG (p) S (c)	GROMACS in house	AMBER #99SB-ILDN specific	REST2 REXDMD	LYZ + BSA spheres ($r = 1.5, 2.5, 3.5, 5, 7.5 \text{ \AA}$)	1 BSA + 3 LYZ 5, 13, 26, 39%	C12 Trp-cage, VIL subdo- main, SH3 domain, hemoglobin α -chain	TIP3P no solvent	24 × 1 μ s 10 ⁶ steps
359	2012	AA CG (p) CG (c)	CHARMM	CHARMM22 (CMAP correction)	TREX	protein G	40%	Trp-cage, melittin	implicit (GBMV)	3 × ~20 × 20 ns
393	2015	AA	PROFASI	PROFASI FF	MC REX	BPTI	7%	Trp-cage	implicit	implicit
360	2020	CG (p) S (c)	in house	specific		spheres	up to 40%	polymer chain	implicit	implicit
362	2006	CG (p) S (c)	in house	specific		spheres of spheres	up to 70%	HIV-1 protease	implicit	8 × 20 μ s
100	2019	CG	GROMACS	from eSBMTools		spheres ($r = 0.33, 1.6 \text{ nm}$), polymers	up to 50%	HigA	implicit	250 ns
85	2013	AA	NAMD	CHARMM22 (CMAP correction)		segment B1 of protein G + VIL	10% → 43%	segment B1 of protein G + VIL	TIP3P	300 ns
54	2016	AA	GROMACS	Pam99-SB-ILDN Pam99-SB-ILDN (p), TraPPE-UA (c)		ACTR + NCBD + IRF-3 PEG500	20, 30% PEG500 + water 200 g/L	ACTR+NCBD+IRF-3 ACTR+NCBD+IRF-3	TIP3P TIP3P	0.3 μ s 0.3 μ s
366	2020	AA	NAMD 2.9	CHARMM*, CHARMM36m, CHARMM36m (CUFIX correction)		<i>E. coli</i> cytoplasm: ~16 proteins + ~262 small solutes + ~430 ions + ~50 000 water molecules	300 g/L	GTT WW domain	TIP3P	~200 μ s
394	2016	S	GENESIS	specific		41 types of spheres	302 g/L		HS	10 × ~1 ms
281	2023	S CG CG	ReaDDy GROMACS GROMACS			LYZ LYZ LYZ	up to 35% up to 35% up to 35%	GB1 GB1 GB1	implicit Martini Martini	1.2 ms 10 × 20 μ s
86	2010	CG (p) S (c)	GROMACS in house	Martini Martini	well-tempered metadynamics (PLUMED) PTMetaDWTE REMC	LYZ spheres ($\sim 12 < r < 24 \text{ \AA}$)	up to 35% <30%	GB1 Ubq/UIIM1, Cc/CcP	Martini implicit	4.8 μ s 10 ⁸ MC steps
87	2010	AA	BioSimz	specific		glycolytic pathway molecules	149 g/L	26 complexes	implicit	10 × 200 ns

Table 2. continued

ref	year	model	software	force field	enhancement	crowder	concentration	protein	solvent	length
390	2021	AA	GENESIS	AMBER ff99SB-ILDN		BSA	11, 17, 30%	C-Src kinase, PPI	TIP3P	~ μ s
395	2020	CG (p) S	GROMACS (PLUMED)	SBM (p) repulsive LJ (c)	REMD	spheres	10, 20, 30%	DPO4	no solvent	
63	2005	CG (c) (p) S	AMBER	specific	REM	spheres	<25%	WW domain	implicit	
396	2010	CG (p) S (c)	AMBER	specific	REM	Ficoll70	25, 40%	PGK	implicit	~40 000 conformations
10	2016	AA	GENESIS	CHARMM c36		<i>E. coli</i> cytoplasm: ~10 ³ proteins + ~10 ribosomes + ~10 GroELs + ~10 ² RNAs + ~10 ⁴ metabolites + ~10 ⁵ ions + ~10 ⁷ waters	~300 g/L	PGK, PDHA, NOX, ENO, IF1	SETTLE	~100 ns
389	2019	S AA	GENESIS GROMACS	specific Amber99SB*-ILDN-Q		UBQ, GB3, LYZ, VIL	up to 200 mg/mL	UBQ, GB3, LYZ, VIL	implicit TIP4P-D	10 × 20 μ s ~ μ s
372	2019	AA	in house	CHARMM22*, CHARMM36m, CHARMM36mCU		<i>E. coli</i> cytoplasm	~300 mg/mL	GTT WW domain	TIP3P	20 μ s
38	2010	CG, S	in house	specific		<i>E. coli</i> cytoplasm	250, 300, 350 mg/mL		implicit	30, 50 μ s (no hydrodynamic interactions), 15 μ s (hydrodynamic interactions)
361	2023	CG CG	MUPHY MUPHY	OPEPv7 OPEPv7		BSA BSA,LYZ	~100, 200 g/L 100, 300 g/L	BSA Cl2	implicit implicit	1 μ s 1 μ s
397	2020	CG CG S	MUPHY MUPHY BD-BOX	OPEPv7 OPEPv7 specific		<i>E. coli</i> cytoplasm Ficoll70 + dsDNAs	300 g/L 279 g/L Up to 35%	α -synuclein Ficoll70	implicit implicit implicit	1 μ s 1 μ s 20 μ s

"Most cited papers (number of citations obtained from Google Scholar) published in the last 13 years and focused on molecular dynamics simulations of more or less crowded solutions. For each study, the table reports (from left to right) reference, year of publication, level of protein (p) and crowder (c) representation (spherical, coarse-grained, and full-atom are indicated as S, CG, and AA respectively), implemented software and force field, enhancement method if any was used, type and concentration (measured in g/L or as the percentage of occupied volume) of crowder, investigated protein if a particular one was considered, simulated solvent, and length of the simulations. The acronyms found in the table are reported in the glossary at the end of the manuscript.

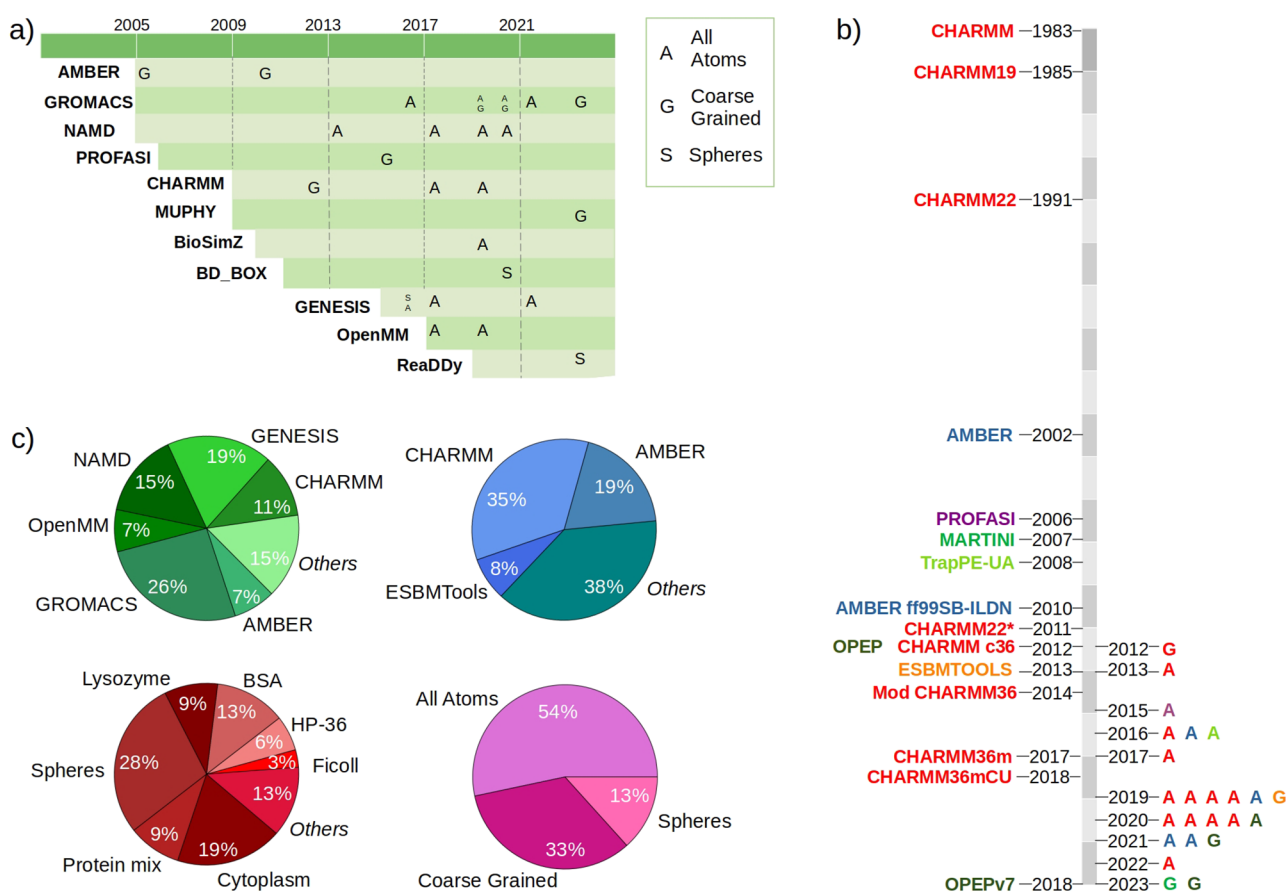


Figure 6. Analysis of techniques used to simulate the crowded environment in the studies reported in Table 2. (a) Each row shows the timeline of one of the most popular software or packages on which crowded systems have been simulated, including AMBER,¹⁴¹ GROMACS,¹⁴⁰ NAMD,³⁹⁸ PROFASI,³⁹⁹ CHARMM,¹⁴² MUPHY,⁴⁰⁰ BioSimZ,⁸⁷ BD-BOX,³⁶⁴ GENESIS,¹⁴⁵ OpenMM,¹⁴⁴ and ReaDDy.⁴⁰¹ Each row starts from the year in which the software or package was made available. The years in which a particularly cited or recent paper used that software are marked with some letters indicating the level of representation chosen for the simulated system, as indicated in the legend. (b) Timeline of the force field implemented in the most cited and recent papers to simulate crowded environments. On the left, each force field is reported in a different color at the year in which it was developed. On the right, the same color is used to highlight the papers implementing that force field. Each paper is indicated with an acronym referring to how the crowders and proteins were represented: all-atom (A) or coarse-grained (G). The year of publication of each paper is reported, as well. (c) Each pie chart refers to the main studies discussed in this review. The top left (right) graph shows the software (force fields) that has been implemented, together with the percentage of papers using each of them. The bottom left (right) plot shows which crowders (representation) have been chosen by what percentage of works.

glycine, valine, phenylalanine, and asparagine with concentrations ranging from 50 to 300 mg/mL. The simulations lasted 1 μ s each to accumulate a total of 128 μ s using eight different force fields and water models (Amber ff99SB-ILDN³⁰² with the TIP3P,³⁸² SPC/E,³⁸³ and TIP4P-Ew³⁸⁴ water models; GROMOS 53A6³⁸⁵ with the SPC³⁸⁶ water model; CHARMM27³⁸⁷ with the TIP3P water model; OPLS-AA/L³¹² with TIP3P, TIP4P, and TIP5P³⁸⁸ water models). They compared simulation-derived density, viscosity, and dielectric properties with experimental data. The results revealed that while all force fields accurately reproduce density changes, disparities are observed in viscosity and dielectric properties, which leads to doubts about the accuracy of the simulation force fields.

Another challenge is studying processes with significant kinetic barriers. Biomolecular associations and crowding processes usually require at least time scales on the order of μ s, but until recently, atomistic MD simulations were typically not performed for more than 1–10 μ s.^{372,389,390} These lengths were often achieved thanks to the MD special supercomputer Anton2³⁹¹ and the latest graphics processing unit (GPU)

platforms and were able to provide good statistics on diffusive internal and intermolecular motions. Much longer simulation times have to be achieved to investigate processes that involve folding or condensation.

Enhanced sampling methods, such as REST2 methods,²⁹⁶ that consider specific parts of a system can tackle such issues with affordable computational cost, as recent studies on internal conformations have shown.^{96,392}

The application of these techniques for the description of the crowding effects on protein binding will be the subject of the following subsections. The first three parts will focus on the three steps of the binding process: long-range recognition, approach, and attachment. The section will end with some considerations on the investigation of the stability of the resulting complex. Figure 5c shows for each of these four steps which crowding effects have been investigated through computational means. A list of the most recent/cited studies can be found in Table 2, while Figure 6 shows which software, force field, crowders, and molecular representations have been preferred. Given the complexity and dimension of the studied systems, ~70% of the studies in Table 2 employed techniques

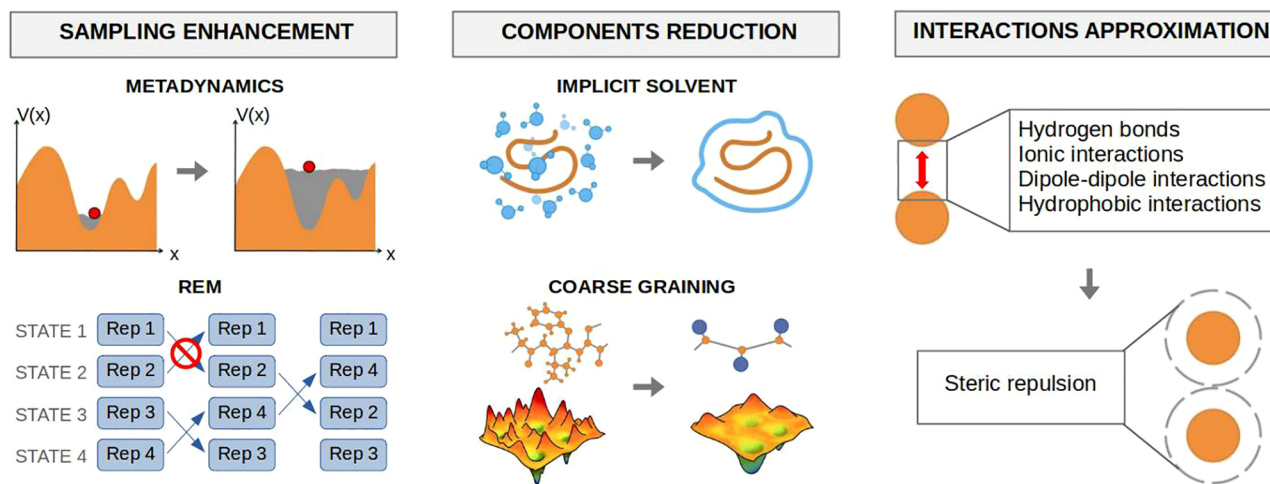


Figure 7. Diagram of the most common techniques used to facilitate the simulation of a crowded environment. On the left are some of the most used sampling enhancement techniques: metadynamics (top) and REM simulations (bottom). In the former, a repulsive bias potential function, shown in gray in the plot, is added to the free energy function $V(x)$, where x is the collective variables (CVs) describing the system (the red dot). This bias discourages the system from revisiting already sampled configurations, which accelerates the exploration of the full energy landscape. In REM simulations, several independent trajectories (called replicas, Rep in the figure) are simultaneously generated. During the simulation, neighboring replicas are exchanged according to specific acceptance criteria. In this way, the trajectory can explore different equilibrium conditions and overcome slow relaxation. Another way to decrease the computational costs is by reducing the number of system components that have to be simulated by either using an implicit solvent or GC model, which are represented in the middle of the figure on top and on bottom, respectively. Implicit solvent models represent the water (blue line) surrounding the protein (orange) as a continuous medium instead of individual molecules. In CG models, molecules are represented by pseudoatoms approximating groups of individual atoms. This results in a smoother energy landscape with fewer local minima, which enables an easier exploration. Finally, faster simulations can be obtained by approximating the interactions: in the context of crowding, many studies chose to only consider steric repulsion to investigate its entropic effect.

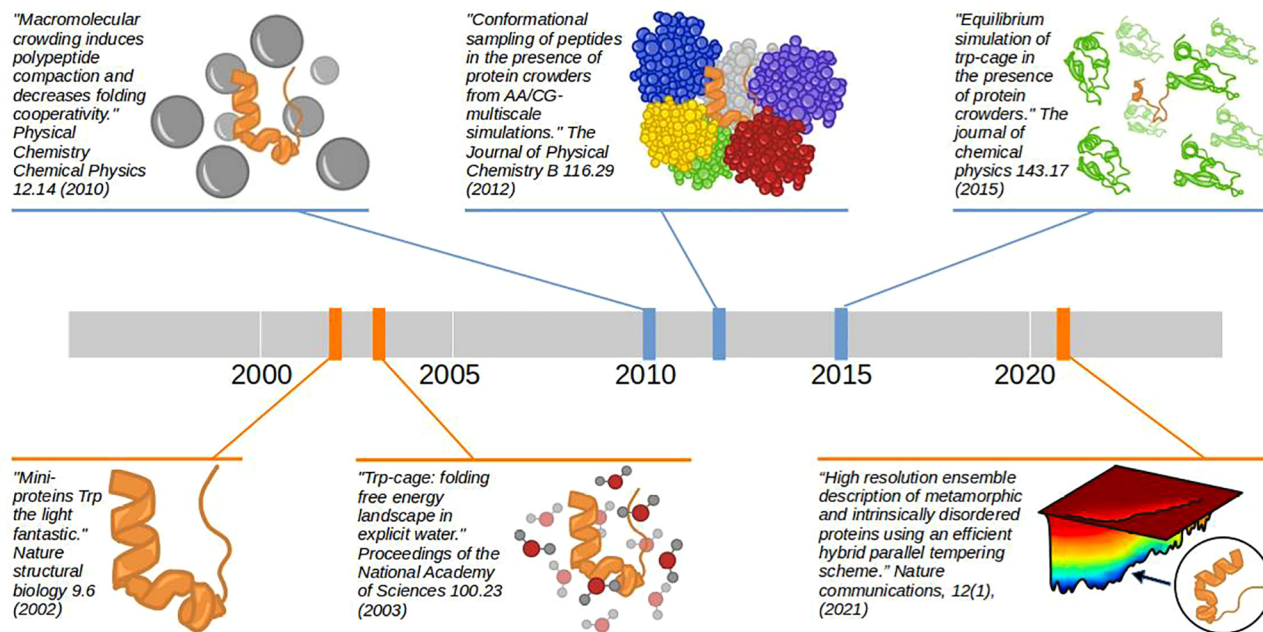


Figure 8. Schematic representation of six different computational studies of the protein Trp-cage structure and dynamics. On top, three articles (from left to right: 359, 365, and 393) that show an example of how the computational investigation of the dynamic in crowded conditions of a case study protein, in this case Trp-cage, has evolved over time. For each paper, the most relevant bibliographic data are reported together with a representative illustration, and the year of publication is marked on the timeline. The same information is shown for the papers on the bottom. The study on the left refers to the insertion of the Trp-cage structure in the PDB databank;⁴⁰² the other two (from left to right: 403 and 404) show the evolution of computational methods that investigated its dynamic without considering crowding.

to either enhance the sampling (metadynamics or REM simulations), reduce the number of components of the system (with CG and/or implicit solvent models), and/or simplify the interactions, as summarized in Figure 7.

4.1. Effect of Crowding on the Single Protein Structure

Protein dynamics link protein structure and function: to meet and interact with other proteins and perform their function, proteins move and explore their conformational space.^{405–407}

Thanks to *in vitro*⁴⁰⁸ and *in silico*⁴⁰⁹ experiments and theoretical models,⁴¹⁰ it has been known for decades that protein folding free energy landscapes are shallow; free energy differences between folded and denatured states and activation free energies are relatively small (0–40 kJ/mol).³⁶⁶ This means that the stability of a protein can be influenced by variations in environmental conditions. Protein folding and stability have been well studied under simplified conditions, mainly through *in vitro* experiments, but it is yet unclear how these protein properties can be modulated by the crowded interior of live cells.⁴¹¹

The effects of crowding on conformational variation have been often described as a shift in equilibrium between native and unfolded states, but the observation of a non-native state that was different from unfolding suggested a more complicated scenario^{66,359,365,366,393,412} in which entropic and enthalpic effects may (co)exist. Harada et al.⁸⁵ hypothesized that, in crowded conditions, entropic constraints should explain the disfavoring of (thermally or chemically) unfolded states, while enthalpic contributions could promote alternative non-native states.^{66,359,365,366,393}

The general take-home message of *in silico* studies that only consider the entropic excluded volume effect is that it enhances stabilization of the native state by increasing the free energy of the unfolded configurations for both single- and multidomain proteins.³⁹⁵ The stabilization of the more compact native states by volume exclusion is consistent with many experiments.^{49,396,413}

Some have measured an enthalpy-driven stabilization, as well; for example, Benton et al.⁴¹⁴ affirmed that different-sized crowders have the same effect on a protein. This stabilization has been presumed to derive from specific interactions with nearby crowder protein surfaces,^{365,366,415} such as edge-to-edge β -sheet interactions,^{366,416} or altered solvent properties.⁴¹⁷ More generically, enthalpic effects have been connected to a destabilization of the folded forms given by protein–protein nonspecific interactions. These interactions counteract the stabilization of the folded forms, which results in a reversed overall effect of the crowding,^{418,419} even if sometimes a stabilizing total effect has been observed both *in vitro*⁴¹⁵ and *in silico*.^{63,365} The importance of soft interactions has been discussed in many *in silico* studies that had to add them to the volume exclusion to match experimental values.^{71,420} These contrasting results could depend on the fact that the crowding effect on protein folding has been shown to depend on a variety of factors, including crowders' occupancy, shape, and size,^{50,98,394} as well as the considered protein.^{98,421}

The inconsistency found in the literature is, of course, also determined by the many different models of crowding that can be chosen. A simple example is given by the evolution of the description of crowding effects on the 20-residue-long Trp-cage protein shown in Figure 8. Trp-cage, whose 3D structure was obtained with NMR in 2002,⁴⁰² is a valuable model system for understanding protein folding dynamics: because of its reduced length and fast folding,⁴⁰³ it has been often tested both inside and outside the crowding field. Figure 8 shows on a timeline some of the studies that have investigated Trp-cage folding in the presence (on top) and in the absence (on bottom) of crowders.^{422–424} One year after the insertion of its 3D structure in the PDB,⁴²⁵ Zhou⁴⁰³ explored the folding free energy landscape of Trp-cage in explicit solvent but without crowders. He observed a two-step folding mechanism with a metastable intermediate state that justifies protein's fast

folding. At 300 K, 70% of the population corresponded to the folded state. One of the first investigations of the effect of crowders on Trp-cage was performed in 2010 by Tsao et al.³⁶⁵ They introduced an implementation of replica-exchange simulations with discrete MD (in which the simulations proceed according to ballistic equations of motion) to study the effect of spherical crowders on an atomistic description of Trp-cage. They only investigated the volume exclusion effect and found that the protein is pushed to adopt a compact state regardless of whether it is the native conformation or not. Two years later, Predeus et al.³⁵⁹ increased the realism of the description by trying to achieve a balance between realistic description and computational efficiency. They introduced a three-component multiscale modeling scheme in which proteins were modeled at an atomistic level, the crowder proteins were modeled at a coarse-grained level (using the PRIMO coarse-grained model⁴²⁶), and the surrounding aqueous solvent was represented as an implicit solvent (using a GB-type implicit solvent model). By combining this model with temperature replica exchange sampling, the simulations enabled the exploration of relatively long time scales that reached into the microsecond range to investigate the impact of protein G crowders on Trp-cage. This approach allowed them to go from simple volume exclusion considerations to including charges and LJ terms. While Tsao et al., by using spherical crowders, indicated a destabilization of extended and unfolded conformations, the interactions between crowders and Trp-cage led to significant populations of partially folded structures that were not observed in corresponding simulations without crowders. This suggests that crowding determines not only a stability shift for Trp-cage but also essential conformational changes. With a dielectric constant of $\epsilon = 80$, Predeus et al. observed two near-native conformations in agreement with what was found by Tsao et al.³⁶⁵ But while Tsao et al. found that the radius of gyration of the second state was larger than that of the native state, for Predeus et al., it was smaller.

A similar system was simulated in 2015 by Bille et al.³⁹³ they chose a system size close to the one studied by Predeus et al. by using the B1 domain of protein G (GB1) and bovine pancreatic trypsin inhibitors (BPTI), which have similar size (but different net charge) compared with protein G, as crowders. Here, crowders were represented in atomistic details, even if with the assumption of a restricted internal dynamic limiting them to the folded state. In this study, they employed Monte Carlo (MC) methods and an AA representation with an implicit solvent force field. As in the previous study, they found that crowders hinder the protein from adopting its global native fold, but this time the interactions between Trp-cage and the crowding agent were found to be specific and involving a select few key residues, therefore denying the generality of the nature of Trp-cage–crowder interactions.

The discrepancy between the findings of Predeus et al. and Bille et al. may, indeed, be partially attributed to the different net charges of BPTI (+6) and GB1 (−4). Since Trp-cage carries a positive net charge (+1), there are more potential attractive electrostatic residue pair interactions between Trp-cage and GB1 than between Trp-cage and BPTI, which likely contributes to the less specific Trp-cage–crowder interactions in the GB1 case. Additionally, while two prolines in BPTI were identified as key residues for interaction with Trp-cage, GB1 lacks any prolines. Therefore, there are at least two reasons to

anticipate differences in the nature of Trp-cage–BPTI and Trp-cage–GB1 interactions.

The comparison between these two crowding studies highlights the importance of selecting appropriate crowders to accurately represent the crowded cellular environment.

Trp-cage has also been used as a model system to test the Replica exchange with hybrid tempering (REHT) method proposed by Appadurai et al.⁴⁰⁴ REHT is a parallel tempering technique to allow efficient and accurate conformational sampling by accelerating water dynamics. Compared with the standard REST2²⁹⁶ method, it guided Trp-cage to the experimentally measured native structure in ~100 ns instead of 300 ns.

Even if in this study crowders were not considered, the improvement of the mapping of the free energy landscapes is an important requirement for simulating crowded conditions. Indeed, when dealing with crowded environments, the exploration of the proteins conformational landscape is slowed by their significant size and by the conformational and dynamical variation of both crowders and proteins. To tackle this challenge, one can either apply this kind of technical improvement to a crowded system or simulate a simplified environment. The latter solution was chosen by the methods presented in the next section.

4.1.1. At the Beginning of Time There Was a Sphere.

The complexity of the detailed protein–crowders interactions is unanimously recognized, but many consider volume exclusion to be the primary effect of crowding. In this context, understanding the excluded volume effects is a necessary step to interpret both experiments and all-atom simulations and to develop more complex models. In general, when considering the stability of single proteins or complexes, one may quantify the effect of crowders by computing the difference in free energy (G) of the process in the presence and absence of the crowded environment, i.e., $\Delta\Delta G = \Delta\mu_F - \Delta\mu_I$, where F and I refer to a generic final and initial state (e.g., complex and monomers, or folded and unfolded protein). We consider a system in which the only varying thermodynamics quantity is the number of molecules so that differences in G are linked to differences in the chemical potential.

To make analytical progress in this framework, scaled particle theory (SPT)⁴²⁷ is often implemented. Worked out by Reiss, Frisch, and Lebowitz in 1959,⁴²⁸ SPT is an equilibrium theory of hard-sphere fluids giving an approximate expression for the equation of state of hard-sphere mixtures and their thermodynamic properties.⁴²⁹ In SPT, the cellular environment is described as a solution filled with crowders and proteins represented as hard spheres. Molecules are considered to be interacting only when touching, and the interacting energy has only a repulsive component that is strongly dependent on distance. The theory provides an estimate of the free energy cost of creating a rigid cavity in the solution on the basis of the ratio (hence the term “scaled”) of the gyration radius of the molecule of interest corresponding to the size of the cavity and the radius of gyration of the crowding agents.⁴³⁰

An approximate analytical solution for $\Delta\mu$ is given by

$$\begin{aligned} \frac{\Delta\mu}{k_B T} = & -\ln(1 - \phi) + \frac{CR_c}{1 - \phi}a \\ & + \left[\frac{CS_c}{1 - \phi} + \frac{(CS_c)^2}{8\pi(1 - \phi)^2} \right] S \\ & + \left[\frac{CS_c}{1 - \phi} + \frac{C^2 R_c S_c}{(1 - \phi)^2} + \frac{(CS_c)^3}{12\pi(1 - \phi)^3} \right] V \end{aligned} \quad (7)$$

where a and V are the radius and volume of the protein, S and S_c are the surface area of the protein and the crowders, and C and ϕ are the number density and volume fraction of the crowders.³⁶⁹ According to this model, macromolecular crowding exerts big effects on the folding free energy, thereby predicting a monotonic increase in stability as the crowder packing increases.

Despite assuming sphericity for both protein and crowders, this theory was often qualitatively consistent with results coming from molecular simulations of more complex systems composed of spherical crowders and CG proteins.^{50,431} Since the conclusions of both SPT theory and simulations of inert spherical crowders were often confirmed by experimental measurements (with inert crowders),⁵⁰ spherical crowders have gained popularity as a swift and straightforward approach and have found widespread application in various types of simulations.^{63,86,100,365,394,395} Because of the limited number of atoms in these systems, the simulations become comparatively rapid, particularly when employing an implicit water model.

However, this approach does not consider chemical interactions, approximates the treatment of solvation, and has an extreme sensitivity to sphere size and density.

4.1.2. Increasing the Complexity. A precise measurement of the entropy reduction is necessary to understand the volume exclusion effect. With this aim, more detailed protein representations have been employed to try to balance computational costs and approximations. Taylor et al.³⁶⁰ directly measured the entropy reduction experienced by a flexible chain molecule in a crowded environment and the resulting impact on the chain-folding transition. To achieve this, they employed a pearl-necklace-type model, which provides a coarse-grained representation of a polymer, with each bead in the chain representing multiple chemical repeat units. This allowed the chain to remain completely flexible at the bead level.^{432,433} Crowders were represented as hard spheres of equal or larger size compared with the bead size of the polymer chain, while the solvent was treated as a continuous background. The investigation utilized the Wang–Landau (WL) simulation method,^{434,435} which grants direct access to the complete thermodynamics of the polymer chain, including conformational entropy. As in previous works,^{436–438} they found an entropic stabilization due to geometric confinement. The study provided validation for the Wang–Landau simulation approach in studying the chain crowder system.

Another study on volume exclusion was proposed by Qin et al.,³⁶⁹ who developed an approach referred to as postprocessing, which paved the way for atomistic modeling of proteins in crowded environments. According to this method, a test protein is simulated at an atomistic level in the absence of crowders. The conformations obtained from this simulation are then used to compute the change in chemical potential when inserted in a solution of spherical crowders. The computed chemical potential variation confirmed the modest

stabilization obtained with experimental measurements using inert polymers as crowders.^{439,440} This stabilizing effect was observed, using the same method as Qin et al., for seven different proteins by Dong et al.⁹⁸ They also found that crowding reduces the open population to several extents and affects the transition rates (as defined by the potentials of mean force along the open–closed reaction coordinate), as well. The magnitude of these effects was shown to be determined by the extent of the conformational changes and the protein size: the influence of crowding becomes stronger as the protein size and the conformational variations increase.

Both studies found that at a given volume fraction, the effects of crowding are larger for smaller crowders than for larger crowders. This result was confirmed by other computational methods^{101,365,441} and experimental measurements.^{103,441} The approach proposed by Qin et al. showed its shortcomings when applied to the HIV-1 protein³⁶⁸ and compared with the direct simulations of the protein in the presence of spherical repulsive crowders performed by Minh et al.³⁶² In the direct simulations, even if close-packed concentrations of repulsive crowding agents were still found to significantly reduce the fraction of open conformations compared with low concentrations, such fraction differed from that found by Qin et al.³⁶⁸ The inconsistency between the two conclusions, that Qin et al. attributed to the simulation box dimension, is presumably coming from the repulsive potential that was added by Minh et al.³⁶² to the volume exclusion effect.

Even if the importance of adding enthalpic effects to volume exclusion to grasp the real effect of crowding has been suggested by this kind of observation made years ago, volume exclusion has remained the focus of many recent investigations. For example, in 2019, Gomez et al.¹⁰⁰ studied the effect of the geometry and size of crowders on protein stability by simulating the protein HigA with a CG structure-based model and modeling crowders as inert spheres or polymers. In particular, they tested spherical crowders of the same size as the $C\alpha$ beads of the considered protein, larger spherical crowders, and polymeric crowders consisting of chains of beads, again, with the same radius as the $C\alpha$ beads. As in the studies by Mittal et al.,⁴³¹ Qin et al.,³⁶⁹ and Minh et al.,³⁶² they confirmed that spherical crowders have a stabilizing effect that decreases with an increase in their size. They also observed that polymeric crowders have a stronger influence than spherical ones; surprisingly, the longer the polymer, the stronger the effect. Even if they employed a structure-based model, whose important disadvantage is not being able to easily capture perturbations of the energy landscape toward unknown conformational variations that may be induced by crowding, their results were consistent with experimental data: the larger crowding effect produced by longer polymers on intrinsically disordered proteins has been shown *in vitro*.⁴⁴² These opposing size effects can be explained by thinking of the interplay between decreasing excluded volume and demixing, whose combination determines the change in the entropy of the crowders during the folding process.¹⁰⁰

To summarize, most studies focused on the volume exclusion effect of crowding have agreed when saying that crowding has a stabilizing effect on protein structure.

However, in the meantime, other studies have addressed the additional interactions between a protein and the surrounding crowders.^{74,359,393}

Enthalpic effects, specifically changes in electrostatic interactions due to crowding, were investigated by the already

mentioned work by Predeus et al.³⁵⁹ that investigated the effect of protein G crowders (GC model) on Trp-cage (AA representation). They compared the multiscale simulations with those of the same peptides in different dielectric environments by varying the dielectric constants from 5 to 80. The results showed that sampling in the presence of crowders resembled sampling with reduced dielectric constants between 10 and 40. On the basis of this work, it is estimated that dielectric constants between 10 and 40 may be good approximations of cellular environments, even if the effective cell dielectric response was probably being overestimated; this work used an implicit solvent, but water, itself, has been hypothesized to have a reduced dielectric response in crowded environments.

A second observation done in the Predeus et al. study was that the presence of explicit protein crowders led to significant populations of partially unfolded structures for Trp-cage. The slightly destabilizing effect of dense cellular environments was consistent with experimental data taken from living cells.^{443,444}

These findings highlight the importance of enthalpic interactions: electrostatic interactions, both in terms of a general reduction in environmental dielectric response and direct interactions with protein crowders, play a crucial role in fully understanding the effects of molecular crowding.

Still, this work had some limitations: the choice to consider only one type of crowder protein raises questions regarding the general validity of the results. Even if the authors justify a generalization based on the fact that they found that the peptide–crowders interactions were not dominated by specific binding interactions, other studies have found that these interactions can be highly specific.^{74,393} Moreover, not all the enthalpic components in the cellular environment are taken into account.

4.1.3. High Resolution for High Precision. Enhancement of the resolution of the simulations allows for a better reconstruction of what happens in the cell: full-atom models enable a more complete study of soft interactions, in addition to volume exclusion effects, and have, thus, led to interesting observation.

Harada et al.⁸⁵ performed a set of 300 ns long simulations using the CHARMM22 all-atom force field of protein mixtures at high total protein concentrations (villin with protein G as crowders) to probe the effect of nonspecific crowder interactions on protein conformational variation. In contrast with what would have been predicted with only an excluded volume effect, they found that native protein stabilities are reduced by crowding. The villin structures become increasingly destabilized upon increasing the concentrations of crowders. Moreover, the authors observed non-native states that do not match the unfolded ensemble, which led to the hypothesis that crowding may significantly alter protein folding landscapes from those observed under dilute conditions. To support the *in silico* results, the authors performed NMR measurements on the same system in both dilute and crowded conditions and showed that the *in vitro* analyses agreed with the simulations.

The total destabilizing effect of crowding was confirmed three years later with much longer atomistic simulations ($\sim 3 \mu\text{s}$) performed by Candotti and Orozco.⁵⁴ They investigated three proteins with different structural levels of disorder surrounded by protein crowders. Prevention of structural collapse was observed for all of them, which was in agreement with the previous study. Moreover, they observed that the impact on nonstructured proteins is more dramatic.

In 2020, Rickard et al.³⁶⁶ achieved more complex and long simulations. They performed an all-atom simulations of a model of the *E. coli* cytoplasm (~16 proteins, ~262 small solutes, ~430 ions, and ~50 000 water molecules) to study the effect of crowding on the folding of the 33-residue GTT WW domain. Their >200 μ s simulations were set in the “protein folding middle ground” between long-time Brownian dynamics (BD) simulations and short-time full-atom simulations of larger cytoplasmic models. Their results confirmed the reduction of folding and diffusion caused by crowding, even if the force fields they employed (C22*, C36m, and C36mCU) are known to exacerbate protein sticking through the hydrophobic interactions at protein–protein interfaces.

Current MD force fields have been often shown to distort protein behavior in biologically relevant crowded environments.^{366,371–374} Petrov et al.³⁷¹ used explicit solvent MD simulations with GROMOS 45a3⁴⁴⁵ and 54a7³⁷⁷ for a total of 6.4 μ s for studying wild-type (folded) and oxidatively damaged (unfolded) forms of villin headpiece at 6 mM and 9.2 mM protein concentration. They demonstrated that the process of aggregation is accompanied by a significant decrease in the overall potential energy, wherein both hydrophobic and polar residues, as well as the protein backbone, made substantial contributions. This phenomenon was directly observed with two other prominent atomistic force fields (AMBER99SB-ILDN³⁰² and CHARMM22-CMAP⁴⁴⁶) and indirectly indicated with two additional force fields (AMBER94⁴⁴⁷ and OPLS-AAL³¹²). This observation suggests a potential issue of current molecular dynamics (MD) force fields wherein there is a tendency to overestimate the potential energy of protein–protein interactions at the expense of water–water and water–protein interactions.

Water behaves differently when considered at the interface with a protein or in the bulk volume.⁴⁴⁸ Therefore, it is reasonable to expect that a crowded environment will cause changes in the dynamical and structural solvation properties of a protein. To capture the features of hydration in a crowded environment, Harada et al.⁴¹⁷ performed MD simulations of protein G and protein G/villin solutions at different concentrations. Both the crowders and the solvent were represented in full-atom detail. The variations in the solvent structure were quantified in terms of radial distribution functions, 3D hydration sites, and retention of tetrahedral coordination. Their study showed that macromolecular crowding poorly affects hydration properties for low levels of crowding (<30% occupied volume), but the structure and dynamics of water change considerably for highly crowded environments (>30% occupied volume). In the latter condition, diffusion rates (including the solvent self-diffusion) and the dielectric constant drop significantly because of the reduced water mobility. The decrease in dielectric constant is a particularly interesting result because it reveals some nontrivial consequences of the presence of crowding agents on the energetics of biomolecules in the cellular environment. On the one hand, an increase in the strength of hydrogen bonds and salt bridges may occur; on the other hand, a reduction of the hydrophobic effect could take place. This would result in competition between secondary structure stabilization and tertiary structure disruption.

To evaluate the effect of crowding on the biophysical properties of a given protein, variation in the thermal stability should also be considered. Such variation can be measured in terms of melting temperature and estimated with computa-

tional approaches starting from the three-dimensional structure.^{204,449,450}

Timr et al.³⁹² used a fully atomistic model in explicit solvent to examine the thermal stability of loop-truncated SOD1 conformations sampled in a crowded environment, which was formed by a solution of bovine serum albumin (BSA). As a first step, GC lattice Boltzmann molecular dynamics (LBMD) simulations were employed to obtain states of local packing (1:1 and 2:1) of BSA around SOD1. Starting from these structures, a set of enhanced sampling simulations based on the REST2 technique^{296,451} were performed (for a total of 24 \times 0.5 μ s), and then these were compared with the dilute solution case. With this approach, which was followed again by Timr et al.,⁹⁶ a slight destabilizing effect of crowding on the SOD1 stability was observed close to ambient temperature. Experimental measures performed by the authors confirmed this result.³⁹² In particular, they observed the propensity of the β 4– β 6 region to early unfolding and its frequent interactions with the crowders. This could be a crucial factor in comprehending the misfolding and aggregation of SOD1^{452,453} in the densely populated cellular environment. The identification of intermediate states during the thermal unfolding, which might be relevant for pathological aggregation, testifies to the most important advantage of computer simulation: the ability to analyze at the atomic level biophysical process.

4.2. Kinetics of the Binding and Reaction Rates

Diffusion and kinetics of diffusion-controlled reactions are fundamental transport mechanisms that play a crucial role in describing a wide range of chemical and biological processes.⁴⁵⁴ In crowded environments, the presence of other molecules and structures can affect transport properties and, consequently, the dynamics of chemical reactions in different ways. Various techniques, such as “single-particle tracking,”⁴⁵⁵ “fluorescence recovery after photobleaching (FRAP),”⁴⁵⁶ and “fluorescence correlation spectroscopy,”⁴⁵⁷ have been used to measure the diffusion constants of macromolecules in the cytoplasm, as well as in membranes. All experiments show that the *in vivo* diffusion of proteins is greatly reduced compared with dilute conditions. For in-depth information about the findings of each of these experimental methods, we refer the reader to other reviews by Prindle et al.,⁴⁵⁸ Cai et al.,⁴⁵⁹ and Yu et al.,⁴⁶⁰ respectively. Many computational studies have been performed to capture the complex interplay between crowding, diffusion, and reaction rates: it is by now well established that crowding can significantly affect the diffusion coefficients of macromolecules,^{1,461,462} the diffusion-controlled reaction rates,⁴⁶³ and cause shifts in chemical equilibria.⁶³ At membrane interfaces, it has been observed with computer simulations⁴⁶⁴ that diffusion is enhanced because of protein depletion. However, diffusion rates provided by simulations under crowded conditions^{10,24,390} have shown that in the cytoplasm diffusion is slowed down by up to a factor of 10. This evidence has also been confirmed by experimental measurements.^{465,466} While the most pronounced effects are observed for macromolecules, crowding impacts the dynamics of small molecules, as well. A widely accepted reason for this is the volume excluded by the crowders that limits free diffusion. However, it has been discussed how the excluded volume effects, alone, could not account for the factor of ~10 reduction.^{467,468}

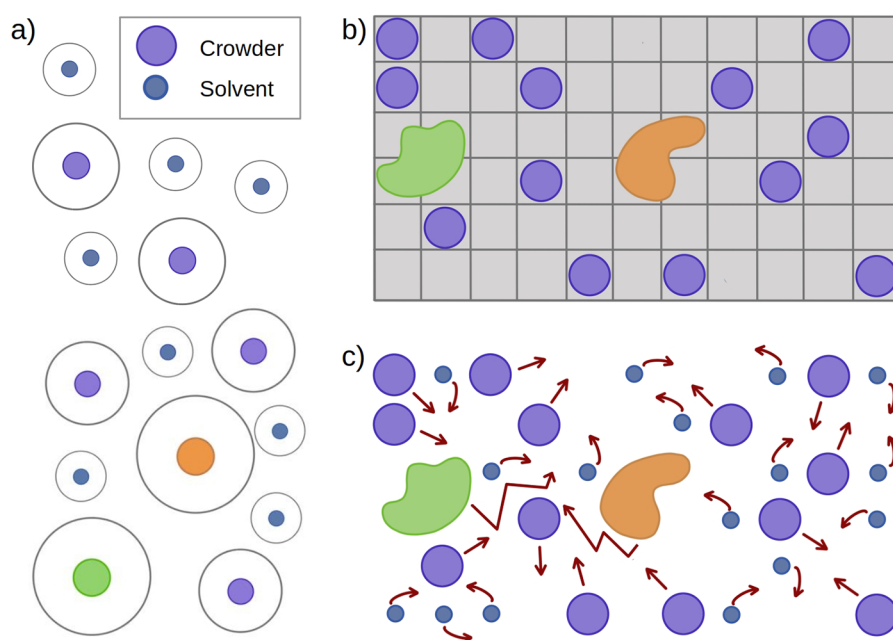


Figure 9. Levels of representation employed to study diffusion and reaction rates in a crowded environment. (a) The GFRD algorithm defines a sphere around proteins (orange and green), crowders (violet), and solvent (blue) and computes for each molecule the probability distribution of the time and the position of leaving this volume. Each molecule is thus associated with a next event type and a next event time, which are put and executed in chronological order. Following execution, the molecule is propagated, and a new sphere with new events is determined. (b) In the simplest lattice model, space is discretized into voxels (gray graph), and proteins (orange and green) and crowders (violet) can only move to a neighboring voxel. The solvent is either modeled implicitly or not considered. (c) A more realistic representation can be obtained when all the system components are left free to move, and the solvent is explicitly modeled.

A factor whose influence on diffusion has been extensively discussed is hydrodynamic interactions (see, for instance, the review by Zegarra et al.⁴⁶⁹), that is, the interactions between molecules or particles mediated by the fluid in which they are immersed. In crowded environments, the presence of other molecules and structures can significantly modify these interactions. CG simulations and hard-sphere models have often addressed their effect in a regime of short-time diffusion (on nanosecond time scales) where hydrodynamic interactions dominate over negligible protein collisions.⁴⁷⁰ Even if some studies have hypothesized that hydrodynamic interactions play only a minor role^{373,374,389} in short-time diffusion,³⁶¹ other works have argued that they have a central role⁴⁶⁵ and that they even slow down long-time diffusion rates.^{38,470,471} To better investigate their effect, various methods have been developed to incorporate hydrodynamic interactions into MD simulations, for example, Stokesian dynamics. These methods are computationally expensive but highly accurate and are able to model the interactions between solute and solvent molecules. Careful validation of the simulations through comparison with experimental data and/or other numerical methods²⁴ remains important.

The crowding influence of diffusion on a longer time scale has been investigated by recent works based on large-scale atomistic simulations that were able to analyze protein–protein interactions, and it was concluded that the determinant cause for diffusion hindrance is random nonspecific diffusive encounters.^{89,372,389,390} These interactions, called “sticking,” are typically weak and short-lived but have an important effect because of their sheer numbers.⁴⁷² Both in-cell NMR⁴⁷² and computational models³⁷² of the *E. coli* cytoplasm have explored this topic. Sticking can result in the formation of multiprotein complexes through nonspecific contacts lasting less than 1 μ s.

The formation of protein clusters may be the cause for the retarded rotational diffusion compared with translational diffusion,^{373,374} something that could not otherwise be explained using hard-sphere models. The influence of protein sticking grows for longer time transport dynamics;³⁷² by achieving 5 μ s of simulation, Rickard et al.³⁷² measured a smaller diffusion coefficient than other shorter all-atom studies.^{373,389}

4.2.1. Rolling in the Cell. When only considering the volume exclusion effect of crowding, diffusion is expected to be slowed down. Different experiments at different volume occupation fractions φ were found to be described by the phenomenological expression:

$$D(\varphi) = D(\varphi = 0)(1 - \varphi)^\kappa \quad (8)$$

where κ is a free fitting parameter that was found in diffusion experiments in the presence of crowders for the protein carbon monoxide hemoglobin to be $\kappa = 0.36$.⁴⁷³ A similar expression can be applied to the rate k_{diff} at which the reactants get into close contact, which in turn determines the association rate k_{-1} . k_{-1} can, indeed, be expressed as $k_{-1} = k + k_{\text{diff}}$ where k is the intrinsic association rate once the reactants are in contact. The volume exclusion effect of crowding leads to an association rate given by $k_{\text{diff}}(\varphi) = k_{\text{diff}}(\varphi = 0)(1 - \varphi)^\kappa$. $k_{\text{diff}}(\varphi = 0)$ is the classical Smoluchowski equation describing the association rate without crowders, $k_{\text{diff}}(\varphi = 0) = 4\pi Da$, where a is the size of the target molecule.⁴⁷⁴ Thus, as the system’s excluded volume increases, k_{diff} decreases since the diffusion time the reactants need to first encounter becomes large. That said, the overall volume exclusion effect on reaction rates is more complex because once the proteins are in proximity, binding equilibria toward the bound state enhances the overall reaction rate.

One of the primary algorithms to simulate reaction and diffusion in continuous space is the Green's function reaction dynamics (GFRD)^{475,476} represented in Figure 9a. This exact algorithm models the behavior of a diffusive system by cyclically solving the Green's functions that propagate the probability distribution. GFRD breaks down the many-body reaction–diffusion problem into one- and two-body problems, which can be analytically solved using Green's functions.^{475,476} These Green's functions are then utilized to create an event-driven algorithm that allows for large jumps in time and space when particles are far apart. When two particles undergo reaction and diffusion within the same system, higher complexity can be employed to describe their interactions. From simple hard spheres, we can move to two spheres with patches where an analytical solution of the Green's function is still feasible.⁴⁷⁷ Additionally, within the reaction volume where the reactant molecules are present, it becomes possible to locally describe protein binding as a more detailed process. This enables the construction of a multiscale algorithm that computes the spatial dynamics of proteins as hard spheres over large distances while employing more complex methods to describe binding when the proteins share a reaction volume.⁴⁷⁸

Another possible approach is lattice-based methods, represented in Figure 9b, where each site can hold one molecule, and crowders are represented as already occupied lattice points. A lattice model was used by Meinecke⁴⁷⁹ to simulate the diffusion dynamics of spherical particles surrounded by inert and inactive crowders of any size and shape. Meinecke found that shape and size considerably determine how strongly diffusion is hampered: small crowders hinder diffusion more than bigger ones. The effect is also stronger for elongated crowders and larger diffusing molecules.

Despite their higher computational cost—GFRD can, for example, be up to 6 orders of magnitude faster—⁴⁸⁰Brownian dynamics (BD) are another commonly used method to study how a crowded environment affects association rates. As outlined in Długosz et al.,⁴⁸¹ they can directly account for crowded conditions, molecule sizes, and shapes, and compute the rates at which molecules come together. Figure 9c shows a general model in which both proteins and crowders (and solvent molecules, if considered) are free to move.

A stochastic simulation was used by Ridgway et al.⁴⁶⁷ to incorporate the physical dimensions of particles into the diffusion and reaction rates. By combining this model with a proteomic-scale evaluation of protein abundance, researchers could approximate the population and diffusion characteristics of the *E. coli* cytoplasm. This enabled the study of the volumetric impact of macromolecular crowding on biomolecular diffusion and diffusion-limited reactions. However, the authors observed that this effect could not fully account for retarded protein mobility *in vivo*, which suggested that other biophysical factors have to be considered.

To further investigate the relation of diffusion with the solvent, environment, hydrodynamic, and attractive interactions have to be taken into account. Other methods based on BD have been employed with this aim since these kinds of simulations allow the consideration of interactions and other features of the cellular environment, as well.⁴⁸¹

The specific effect of volume exclusion and hydrodynamic interactions were investigated by Grimaldo et al.⁴⁷⁰ by simulating a spherical model of an *Escherichia coli* cytoplasm. The model comprised 15 different macromolecule types at physiological concentrations with each represented by a

different-sized sphere. In this way, they took into account the polydispersity of a cellular environment and overcame standard colloidal modeling. This polydispersity was shown to slow down larger macromolecules more strongly than smaller particles via hydrodynamic interactions already at nanosecond time scales before protein collision. The results were confirmed by experimental measurements in cell lysate. The authors also suggested that the slowdown may also affect molecular mobilities and escape rates of long-time processes.

The same *E. coli* model, studied with BD simulations and the Rotne–Prager–Yamakawa (RPY) tensor, was used by Ando et al.³⁸ to investigate the effect of attractive interactions on diffusion. At first, they only considered a soft repulsive potential and hydrodynamic interactions: without adjustable parameters, the *in vivo* experimental diffusion constant is reproduced. Thus, they hypothesized that excluded volume effects and hydrodynamic interactions are sufficient to explain the large reduction in the diffusion of macromolecules observed *in vivo*. Next, they added nonspecific attractive interactions: reduction in diffusivity becomes very sensitive to the macromolecular radius, with the motion of the largest macromolecules being dramatically slowed down. Moreover, long-lived clusters involving the largest macromolecules form if attractions dominate, whereas hydrodynamic interactions give rise to significant, size-independent intermolecular dynamic correlations.

The importance of considering soft interactions was shown by Blanco et al.,⁴⁸² as well. They developed the chain entanglement softened potential (CESP) to study streptavidin diffusion among dextrans of different sizes and concentrations. The CESP model is based on an interparticle interaction potential that includes both hard-core and soft interactions and takes into account the hydrodynamic interaction effects of protein complexes. The experimental long time diffusion coefficient predicted with this potential was shown to be in better agreement with the experimental measured one than the coefficient predicted when only considering hard-core repulsion.

The fact that the experimentally observed slowdown of diffusion can be achieved only when considering attractive interactions in addition to volume exclusion and repulsion was again confirmed by Skora et al.³⁹⁷ In 2020, they considered a mixture of Ficoll70 and double-stranded DNAs (dsDNAs) to investigate the effect of the crowder's shape/type. Using BD simulations, they observed a stronger slowdown of diffusion by cylindrical double-stranded dsDNAs compared with spherical Ficoll70, which was confirmed by *in vitro* measurements. For a 60% diffusion reduction, a volume fraction of 5% or 35% was needed, respectively, for dsDNA and Ficoll70. The variation in reduction between the two crowders was not only attributed to the larger volume excluded by the former but also to the differences in the attractive interactions.

4.2.2. Kinetics of Mixed Models. The spherical representation of molecules allows the simulation of the collective diffusion in crowded conditions, but to study single proteins, a more detailed representation is needed. With this aim, diffusing proteins can be represented at AA or CG level, and the crowders can be represented as spheres or with CG models.

Wieczorek et al.⁴⁸³ performed BD simulation of two reacting proteins (monoclonal antibody HyHEL-5 and its antigen hen egg lysozyme) at atomic resolution surrounded by spherical crowders and observed an increase in their association rate.

This is in line with the hypothesis that crowding reduces translational diffusion through volume exclusion so that when reactants are in proximity they have more time, compared with dilute solution, to assume a proper mutual orientation. A direct observation of protein mobility, protein–protein contacts, clusters, and their lifetimes in crowded conditions was achieved by Timr et al.³⁶¹ They proposed a mixed particle lattice approach to investigate the effect of solvent-mediated hydrodynamic interactions on protein motions and structures with reduced computational cost and longer time scales ($\sim 1 \mu\text{s}$). With this aim, they coupled the optimized potential for efficient protein structure prediction (OPEP)⁴⁸⁴ to a lattice Boltzmann description of solvent kinetics. OPEP is an implicit solvent protein force field that is applied to combine a detailed description of the protein backbone with a CG description of amino acid side chains and perform rigid body simulations. Since it was able to capture the experimentally observed decrease in protein translational diffusion coefficients only for short trajectories ($< 100 \text{ ns}$) while simulating an excessive slowdown for longer times, the authors proposed a new version, OPEPv7. They investigated the translational diffusion of the folded globular protein chymotrypsin inhibitor 2 (CI2) in crowded solutions of BSA and lysozyme and observed a diffusion slowdown that was similar to the experimental results. Self-clustering was indicated as an important factor. A qualitative agreement with experimental observation was also obtained for an intrinsically disordered protein (α -synuclein) with BSA crowders: both *in vitro* and *in silico*, the protein experiences a less drastic slowdown compared with the folded one, even if *in vitro* the difference is less marked. This suggested that this CG model produces a too-compact structure for disordered proteins; to describe IDPs, further modifications are required, such as the fine-tuning of some hydrophobic interactions and the implementation of intramolecular terms for chain extension. OPEPv7 was also tested against a heterogeneous crowded solution modeling the *E. coli* cytoplasm, which counts 197 proteins. The translational diffusion coefficients calculated for each protein showed a good agreement with the model derived from experimental data. These simulations confirmed that direct protein interactions are an important factor determining the diffusion slowdown in crowded protein solutions. Such interactions were also shown to have an important effect in the short-time regime, which was previously thought to be dominated by hydrodynamic interactions.⁴⁶⁵

4.2.3. Looking Closer: All-Atoms Models. Studying diffusion in a crowded environment simulated at atomistic resolution poses a significant computational challenge. Although the dynamics and underlying mathematics are straightforward, the large number of elements can result in an extremely high computational cost. Moreover, particles can react and bind to membranes or microtubules, thereby changing their effective diffusion constant dynamically over time. Transient interactions between proteins can also result in the formation of clusters, which have been proposed as the dominant crowding effect for reduced rotation diffusion.^{373,374} This hypothesis focuses on direct protein–protein contacts, but whether diffusion could be affected to a similar degree by proteins that are nearby even if not in contact is less clear.

To evaluate how close proteins have to come before their diffusion properties are impacted significantly, Nawrocki et al.³⁷⁴ performed MD simulations for over $1 \mu\text{s}$ for a system of 19 chicken villin headpieces with explicit solvent. They showed

that the diffusion slowdown is primarily correlated with direct protein–protein contacts (on a $1\text{--}100 \text{ ns}$ time scale^{373,374}) because of the formation of clusters rather than indirect interactions via shared hydration layers.³⁷⁴ Consistently, with the Stokes–Einstein relations for translational (D_t) and rotational (D_r) diffusion,

$$D_t = \frac{k_B T}{6\pi\eta r} \quad (9)$$

and

$$D_r = \frac{k_B T}{8\pi\eta r^3} \quad (10)$$

where k_B is the Boltzmann constant, T is the temperature, r is the particle radius, and η is the viscosity of the surrounding medium, they found that rotational diffusion showed a stronger slowdown.

In the same year, von Bülow et al.³⁸⁹ confirmed that nonspecific protein binding (with a complex lifetime of $1\text{--}50 \text{ ns}$) and the formation of clusters account for the high viscosity and slow diffusivity in crowded conditions. They performed 500 ns long atomistic MD simulations of ubiquitin (UBQ), the third IgG-binding domain of protein G (GB3), lysozyme (LYZ), and villin headpiece (VIL) and found that at 200 mg/mL , translational and rotational diffusion decreased by a factor of ~ 4 and ~ 6 , respectively. They computed the effective hydrodynamic radii using the Stokes–Einstein relations and found that it was consistent with the diffusion variation, thereby proving that this relation remains valid in concentrated protein solutions. Finally, they showed that clustering interactions contributed to $\sim 40\%$ and $\sim 50\%$ of the total slowdown in translational and rotational diffusion, respectively.

The dominating effect of cluster formation on diffusion was also found by Bashardanesh et al.,⁸⁹ who simulated systems of a few species of proteins at an increasing concentration (up to 70% of the mass) with AA MD lasting $1 \mu\text{s}$. Even if they found a stronger slowdown than von Bülow et al. (probably due to a higher concentration of protein used in their simulations), they agreed in observing that this effect was particularly noticeable for rotational diffusion. The more noticeable hampering of rotational diffusion was explained by the fact that direct interactions from clustering hinder rotational motion and were related to charge distribution on the surface or weak dispersion interactions. Bashardanesh et al. also noticed that rotational diffusion depends on the protein size, unlike translational diffusion and viscosity: at high concentrations, small proteins were slowed down more than big ones. However, the authors did not propose an explanation.

All the mentioned studies explored time scales of $\sim 0.5\text{--}1 \mu\text{s}$, while a longer simulation of $5 \mu\text{s}$, which reaches a time scale relevant for sampling sticking processes, was performed by Rickard et al.³⁷² Their systems more closely replicated the heterogeneity of the cellular environment, as well. They mimicked a small volume of the *E. coli* cytoplasm with 2×10^5 atoms, which corresponds to proteins and RNA selected among the most abundant cytoplasmic components. This was the first simulation of a bacterial cytoplasm model that reached atomic precision and a μs time scale. The formation of clusters and the consequent diffusion slowdown were observed here as well, even if the observed diffusion coefficients ($\sim 1.4\text{--}2 \text{ nm}^2/\mu\text{s}$) are smaller than those observed in experiment⁴⁸⁵ and in other atomistic simulations.^{373,389} The authors hypothesized

that this could be explained by the much shorter time scales explored by those studies, which highlights the potential importance of protein sticking and cluster formation for long-time transport dynamics. Rickard et al. showed that protein–protein contacts are highly dynamic and can be broken and reformed repeatedly with a time scale of contacts on the order of $\sim 1 \mu\text{s}$ that indeed exceeds the simulation time of previous works.^{89,389,464}

4.3. Dynamics of Complex formation

After navigating the crowded cellular environment, two (or more) functional partners can bind together. It has been suggested that the binding interfaces of proteins have coevolved with the cytoplasm to avoid nonspecific interactions;^{72,372} thus, the influence of crowding would be expected to be negligible at this stage of functional complex formation.

That said, some studies that only considered volume exclusion or repulsive interactions found a modest stabilizing effect.^{86,281,369} In this approximation, crowding also seems to favor specific over nonspecific complexes since the former are more tightly bound and, thus, have smaller excluded volumes.⁸⁶ The fact that crowding increases the specificity of protein binding in a crowded cellular environment means that it can have functional effects apart from the relative stability of bound and unbound complexes.

However, adding attractive interactions has resulted in different observations.²⁸¹ *In vitro* experiments⁴⁸⁶ have observed that protein crowders can destabilize protein–peptide complexes because of weak nonspecific interactions between test proteins and crowders. The sticking determined by transient weak interactions can both hinder, via nonfunctional competitive binding,⁵⁴ or promote complex formation.⁸⁷ These interactions can result in transient “quinary structures.” The concept of quinary structures introduces an additional level of protein organization beyond the primary, secondary, tertiary, and quaternary levels. These structures are characterized by functional interactions that exhibit weaker and more transient properties compared with quaternary protein structures. They involve the formation of transient encounter complexes, which serve as the initial step in the assembly of low-affinity and transient protein complexes. These complexes are involved in various biological processes, such as electron transfer, the formation of enzyme–substrate complexes, and the self-association of proteins with weak affinities. The understanding of quinary structures provides valuable insights into the dynamic and diverse nature of protein interactions and their functional implications.⁴⁸⁷

Crowding also has an indirect influence on the dynamics of complex formation by shifting the population of local conformations,³⁹⁰ as discussed in Section 4.1.

4.3.1. Interacting Proteins as Spheres. The models that only take into account the volume exclusion effect between spheres predict a shift of the binding equilibrium of proteins toward the bound state: once the associated state has formed, more volume is available for the crowders, and the total entropy of the associated system increases, which makes the bound state more stable.^{30,488} To provide a simple way of depicting the effect of volume exclusion on binding equilibria, a reaction of the type $A + B \rightleftharpoons AB$, where A and B are reactants, can be considered. The latter reaction is fully reversible and can be characterized by a dissociation constant $K_{d,0} = \frac{k_{\text{off}}}{k_{\text{on}}}$, where k_{off} is the unbinding rate, k_{on} is the reaction

rate, and the subscript 0 indicates an absence of crowders in the solution. More generally, the reaction dissociation constant can be defined as $K_{d,\varphi}$, where φ is the volume occupation fraction of the crowding agents. Thus, the effect on the dissociation constant is purely thermodynamic and can be characterized by the statistics of the states in which the receptor is either free or occupied. Consider a three-dimensional lattice with Ω lattice sites and a total volume $V = \Omega l^3$, where l is the lattice spacing. Set a static target or receptor R , and with L proteins and C crowders diffusing into the lattice with a given diffusion constant, if one partner finds the other, the complex RL is formed. At equilibrium, the probability for the complex to be formed can be calculated as

$$P_b = \frac{S_b e^{-E_b/k_b T}}{S e^{-E_b/k_b T} + S} \quad (11)$$

where k_b is the Boltzmann constant, T is the temperature, S_b is the number of states in which the complex is formed, S_{ub} is the number of states in which the protein is unbound, and E_b is the binding energy. After obtaining S_b and S_{ub} , P_b can be expressed as

$$P = \frac{1}{1 + K_d(\varphi)rL/\Omega} \quad (12)$$

One can then show that the change in the dissociation constant as a function of φ is given by $K_{d,\varphi} = K_{d,0}(1 - \varphi)^r$, where r is a factor that depends on the crowder size.⁴⁸⁹ Thus, volume exclusion shifts binding equilibria toward the bound state. This shift is expected to result in protein aggregation.⁴⁹⁰

To more deeply investigate the thermodynamic effects of crowding on macromolecular associations, hard-sphere theory can be used in conjunction with Brownian dynamics simulations. This has been done by Ando et al.,³⁹⁴ who were the first to simulate a cell-like inhomogeneous system instead of assuming a uniform crowder size, even if the dependence of crowding effects on the composition of the crowders had already been investigated in the past.⁴⁴¹ They mimicked the physiological concentrations of macromolecules in the cytoplasm of *Mycoplasma genitalium* by using 41 different types of macromolecules represented by spheres with different radii. They found, both with theoretical calculation and simulations, that the heterogeneity of sizes decreases the effect of crowding when compared with the use of uniform crowders with the same mean size at a fixed volume fraction of the system. They also confirmed that smaller crowders show the strongest effects. However, these kinds of simplified studies are limited by the exclusion of shapes, chemical properties, interactions, and flexibility of macromolecules. Ando et al. recognized that the stabilization of the native or complex states of macromolecules predicted by their model is in contrast with what was observed by other experimental and computational studies.^{43,85,414,418}

The fact that considering only volume exclusion results in a wrong prediction of complex stabilization has been recently investigated by Pradhan et al.²⁸¹ They simulated the dimerization of model system GB1 in the presence of lysozyme crowders at two different resolutions by assuming both protein and crowders as spherical beads and then retaining residue-specific structural details. With the former, they found stabilization of the dimers in the presence of the lysozyme crowder, in accordance with the SPT model. The latter residue-specific model showed that soft interactions destabi-

lized GB1 dimerization and were in agreement with the free energy variation obtained from experimental observation. The computed value of free energy variation, despite having the same sign, was much larger than the experimental one. This was hypothesized to be due to the overestimation of the protein–protein interaction strength in the employed Martini 3.0 force field.⁴⁹¹

4.3.2. Coarse-Grained Models for the Interaction of Proteins. CG models can better mimic the crowded cellular environment and investigate its impact on the dynamics of complex formation.⁶⁶ they can provide insights on protein binding affinities, binding pathways, and the formation of specific versus nonspecific complexes.⁸⁶

Kim et al.⁸⁶ presented a detailed examination of a residue-level CG model developed to investigate the influence of macromolecular crowding on protein complex formation. The model is applied to two specific protein complexes, ubiquitin-UIM1 (Ubq/UIM1) and cytochrome *c*-cytochrome *c* peroxidase (*Cc/CcP*), in the presence of repulsive spherical crowders of varying sizes. Their results reveal that repulsive crowding agents modestly stabilize the formation of protein complexes, which leads to a reduction in binding free energy of up to $2 k_B T$ as the crowder packing fraction approaches the physiological range.

An increase in binding stability in the presence of repulsive spherical crowders was observed, as well, by Qin et al.³⁶⁹

In both studies, the observed changes in binding free energy could be accurately predicted using an SPT hard-sphere model. Moreover, Kim et al. observed that crowding not only shifts the overall equilibrium toward bound complexes but also favors specific over nonspecific complexes within the coarse-grained representation of proteins and crowders. This preference arises because of the larger excluded volumes of nonspecific intermediate complexes, which are loosely bound compared with the specific complex. Such functional implications hold significant importance for understanding protein–protein interactions and their specificity in crowded cellular environments. However, the changes in binding free energy observed by Kim et al. are influenced by the coarse-grained and implicit solvent nature of the model and the assumption of rigid-body behavior. In reality, the flexibility of proteins, crowding agents, and explicit solvents may play a crucial role in determining the binding free energy during protein complex formation. This study emphasized the need for further investigations considering protein flexibility and attractive interactions (depending on the hydrophobicity of crowders and protein residue) to model solvent-mediated interactions between crowders and proteins. When combined with the findings related to repulsive crowders, these aspects offer valuable insights into the diverse scenarios arising from crowding effects on protein binding, as discussed in the following section.

4.3.3. Atomistic Resolution View. More detailed representations of proteins and crowders can give better insights into the docking process.

Despite still not fully considering protein flexibility, in 2010, Li et al.⁸⁷ developed a Langevin dynamics-based tool, BioSimz, to inspect protein (modeled as rigid-body) dynamics under crowded conditions. The package allows the exploration of the formation and behavior of the complexes and to quantify the frequency and retention time of the interactions. The identified sites of interest were then refined with the flexible docking algorithm SwarmDock.¹³² They showed that when crowders at

near-physiological concentration are added to the simulations, the time spent by binding partners around the true binding regions—compared with the rest of the surface—is increased. Thus, crowding would seem to steer binding partners toward the correct pose. This could be achieved through multiple mechanisms, but two were found to prevail: (i) the rolling and/or sliding of a molecule on the surface of its binding partner and (ii) the elimination of secondary or weak binding spots. Moreover, specific interaction events were observed to prolong even more than their nonspecific counterparts. Li et al. concluded that the inclusion of crowders generates a favorable environment that facilitates the sampling of diverse conformations by the target proteins. As a result, the traversal of energy barriers becomes more efficient, which effectively smooths the reaction coordinate toward the bound state, thereby enhancing the propensity of proteins to form stable complexes with their respective binding partners.

Later on, protein flexibility was taken into account by Candotti et al.,⁵⁴ who again observed the modulation exerted by crowders on the protein conformational space. They showed that this chaperoning effect is particularly noticeable for intrinsically disordered proteins that are biased toward bioactive conformations by the crowded surroundings. In particular, they observed that the intrinsically disordered activator for thyroid hormone and retinoid receptor (ACTR) is stirred to more compact conformations with well-defined α -helices in the regions interacting with the nuclear coactivator binding domain of cyclic adenosine monophosphate (cAMP) response element binding protein (CREB) (NCBD). However, the proposed mechanism differs from the one proposed by Li et al.: Candotti et al. hypothesized that the viscous environment slows down protein flexibility, which limits the exploration of new intra- and interprotein contacts and, thus, restricts the conformational landscape instead of facilitating its exploration. However, they also observed that crowding hinders the formation of functional contacts because of nonspecific quinary contacts hindering specific partner recognition.

The crowding-induced depletion of binding partners has been observed for a different system (*c*-Src kinase and one of its inhibitors in the presence of bovine serum albumin as crowder) by Kasahara et al.³⁹⁰ By comparing full-atom MD simulations in dilute and crowded solutions, they observed a reduction of the inhibitor population with increasing crowder concentration because target protein–crowder interactions sterically block noncanonical binding sites. Ligands might be trapped on the surfaces of crowder proteins. These observations were validated with experimental measurements.

4.4. Complex Stability and Dynamics

What concerns the impact of molecular crowding on the dynamic structural properties of protein dimers (or protein complexes composed of multiple monomers) has not yet been thoroughly addressed from a computational perspective. One of the few computational studies on this topic has been performed by Dong et al.⁷⁴ They observed that BSA weakly but preferentially interacts with the histidine carrier protein (HPr) on the binding interface for HPr's specific partner protein, EIN, which leads to competition. Thus, crowding would decrease the EIN-HPr binding affinity and accelerate the dissociation of the native complex.

However, early experimental investigations utilizing inert polymers have indicated that hard-core repulsions have limited

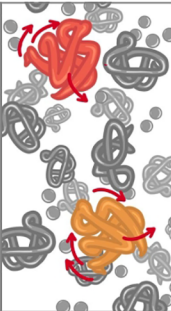
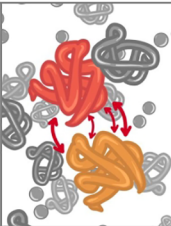

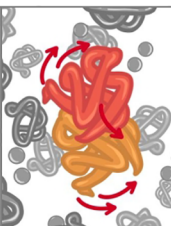
STRUCTURE		VOLUME EXCLUSION	VOLUME EXCLUSION+ REPULSION	VOLUME EXCLUSION+SOFT INTERACTIONS
		<ul style="list-style-type: none"> Open conformations reduction Native state stability Stronger effect for: <ol style="list-style-type: none"> Increasing protein size Bigger conformational variation Smaller spherical crowders Longer crowders 	<ul style="list-style-type: none"> Open conformations reduction Native state stabilization 	<ul style="list-style-type: none"> Native state stability increase/decrease New conformations Dielectric constant reduction Stronger effect for: <ol style="list-style-type: none"> Non-structured proteins
BINDING		VOLUME EXCLUSION	VOLUME EXCLUSION+ REPULSION	VOLUME EXCLUSION+SOFT INTERACTIONS
		<ul style="list-style-type: none"> Complex formation stabilization Specific over non-specific complex formation Stronger effect for: <ol style="list-style-type: none"> Smaller crowder 	<ul style="list-style-type: none"> Complex formation stabilization Specific over non-specific complex formation 	<ul style="list-style-type: none"> Complex formation stabilization/destabilization Stronger effect for: <ol style="list-style-type: none"> Disordered proteins
DIFFUSION		VOLUME EXCLUSION	VOLUME EXCLUSION+ HI	VOLUME EXCLUSION+SOFT INTERACTIONS
		<ul style="list-style-type: none"> Diffusion slow down Stronger effect for: <ol style="list-style-type: none"> Translational diffusion Smaller spherical crowders Elongated crowders Bigger proteins 	<ul style="list-style-type: none"> Long/short time diffusion slow down 	<ul style="list-style-type: none"> Cluster formation slows down long/short time diffusion Stronger effect for: <ol style="list-style-type: none"> Rotational diffusion Bigger proteins
			VOLUME EXCLUSION+HI+SOFT INTERACTIONS	
			<ul style="list-style-type: none"> Stronger effect for bigger proteins 	
COMPLEX STABILITY		VOLUME EXCLUSION	VOLUME EXCLUSION+SOFT INTERACTIONS	
		<ul style="list-style-type: none"> Dimer stabilization Stronger effect for: <ol style="list-style-type: none"> Compact dimer 	<ul style="list-style-type: none"> Dimer destabilization (experimental observation) 	

Figure 10. Summary of the main effects of crowding on protein–protein interactions observed *in silico*. A recap of how crowding influences protein structure, diffusion, complex formation, and stability is reported. In each box, the observations are divided into sub-boxes according to which crowding effects have been considered.

impact:⁴⁹² only a marginal increase in stability was observed due to the slight reduction in volume when two monomers form a dimer.

This is in agreement with SPT theory, which predicts that the influence of hard-core repulsion on a protein dimer depends on its shape; in particular, side-by-side dimers are expected to be only mildly influenced by hard-core repulsion, while more compact dimers should be more stabilized.⁴⁹³ This hypothesis has been confirmed by *in vitro* experiments.^{64,88} For example, Guseman et al.⁶⁴ verified with NMR measurements that among two variants of the B1 domain of protein G, which differ by only three residues but with different shapes, the domain-swapped ellipsoidal dimer was stabilized, whereas the side-by-side dumbbell-shaped dimer showed little to no stabilization.

A pioneering study conducted by Guseman et al.⁸⁸ investigated the effects of small and large cosolutes on the equilibrium stability of the simplest defined protein–protein interactions and specifically focused on the side-by-side homodimer formed by the A34F variant of the 56-residue B1 domain of protein G. The researchers utilized ¹⁹F nuclear magnetic resonance spectroscopy to quantify the influence of urea, trimethylamine oxide, Ficoll, and more physiologically relevant cosolutes on the dissociation constant of the dimer. The findings revealed stabilizing and destabilizing effects from soft interactions.

These experimental results observed crowding effects on complex stability similar to those seen for single protein (i.e., stabilization by volume exclusion and either stabilization or destabilization by soft interactions). However, the study of

protein complexes in crowded conditions through computational methods will allow for a better characterization—in the best case scenario, a prediction—of the crowding effects on the stability of dimers. This could, in turn, have a big impact on the field of protein design.

5. CONCLUSIONS

Both experiments and computational tools have been extensively implemented to study how proteins interact. Most of the computational studies investigating how these interactions are influenced by crowding are centered on simulations by exploiting the improvement of both the computing power and new MD methods of the last decades. Even if the simulation of increasingly complex systems has been made possible, the effects of a variety of crowders and of the many specific and nonspecific interactions that biological molecules have to face when navigating inside the cell have yet to be clarified.

Crowders influence what happens while the partners are approaching in terms of conformational variation and diffusion, but once they are adjacent, proteins bind as if they were isolated. This suggests that even other computational methods applied up to now to protein–protein interaction prediction in diluted or isolated conditions could be applied to the investigation of interactions in crowded environments. Indeed, many of the force fields developed for diluted solutions have been able to reproduce experimental measurements of crowded solutions with simple parameter regulation. It is then worth asking ourselves if the theoretical computational methods developed up to now to describe the physicochemical properties of proteins without considering the crowding effect could be generalized to systems surrounded by crowded environments to better characterize them.

For this reason, in Section 3, we reviewed the most relevant computational methods developed to investigate protein–protein interactions without taking into account the crowding effect. In Section 4, we reported the most significant advances among computational studies in exploring the influence of molecular crowding on the steps leading to protein binding, as well as the contradictions still existing in the field. Both are summarized in Figure 10 starting from the effects on conformational variations: most studies have found novel configurations in crowded environments and a reduction of dielectric constant, but both the increase or decrease of native-state stability have been observed. Stabilizing, as well as destabilizing effects, have been attributed to crowding even when looking at complex formation. More unambiguous observations have been made for what concerns protein diffusion (that is slowed down by crowders) and the formed complex stability (which is reduced by crowders).

The multilevel investigation of computational methods proposed here allows us to conclude that to better understand the crowding effects and the relationship between protein interactions and their biological function it is highly recommended to integrate structural and conformational dynamics with the influence of cellular crowding, considering the numerous interactions occurring at the cellular level under realistic biological conditions. The hope is that with the growth of the knowledge of protein behavior in crowded conditions (through experiments and simulations), the methods developed in the absence of crowding (Section 3) could be extended to this more realistic scenario in two ways: (i) through training on this new data set or (ii) by applying them

to the conformations known to be sampled in crowded conditions. The latter idea has already been partially applied by first performing CG simulations of crowded conditions (that allow for the investigation of protein dynamics and other molecules with a higher level of approximation but extend simulation times considerably) and then studying in more detail with AA simulations the longest-living local configuration of the adjacent crowders binding to the investigated protein.^{96,392} As the hardware capabilities of supercomputers continue to advance, simulations with an atomistic crowded environment representation are progressively becoming a viable option for capturing molecular binding intricacies.^{10,372} Including the effects of crowding on protein binding could in the future play a key role in improving interaction prediction, as well as in achieving reliable prediction of the interactomes of any organism for which the whole genome is known (and consequently, thanks to advances in structural prediction algorithms, the whole proteome is also predictable), including the human interactome.

AUTHOR INFORMATION

Corresponding Author

Edoardo Milanetti – Department of Physics, Sapienza University, Rome 00185, Italy; Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Rome 00161, Italy; orcid.org/0000-0002-3046-5170; Email: edoardo.milanetti@uniroma1.it

Authors

Greta Grassmann – Department of Biochemical Sciences “Alessandro Rossi Fanelli”, Sapienza University of Rome, Rome 00185, Italy; Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Rome 00161, Italy

Mattia Miotto – Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Rome 00161, Italy; orcid.org/0000-0002-0043-8921

Fausta Desantis – Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Rome 00161, Italy; The Open University Affiliated Research Centre at Istituto Italiano di Tecnologia, Genoa 16163, Italy

Lorenzo Di Rienzo – Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Rome 00161, Italy

Gian Gaetano Tartaglia – Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Rome 00161, Italy; Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genoa 16163, Italy; Center for Human Technologies, Genoa 16152, Italy; orcid.org/0000-0001-7524-6310

Annalisa Pastore – Experiment Division, European Synchrotron Radiation Facility, Grenoble 38043, France; orcid.org/0000-0002-3047-654X

Giancarlo Ruocco – Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Rome 00161, Italy; Department of Physics, Sapienza University, Rome 00185, Italy

Michele Monti – RNA System Biology Lab, Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genoa 16163, Italy

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.chemrev.3c00550>

Author Contributions

○M.Monti and E.M. contributed equally to the present work. CRediT: **Greta Grassmann** conceptualization, investigation, writing-original draft, writing-review & editing; **Mattia Miotto** conceptualization, writing-original draft, writing-review & editing; **Fausta Desantis** writing-original draft, writing-review & editing; **lorenzo di rienzo** conceptualization, writing-review & editing; **Gian Gaetano Tartaglia** conceptualization, writing-review & editing; **Annalisa Pastore** conceptualization, writing-review & editing; **Giancarlo Ruocco** conceptualization, writing-review & editing; **Michele Monti** conceptualization, investigation, supervision, writing-original draft; **Edoardo Milanetti** conceptualization, investigation, project administration, supervision, writing-original draft, writing-review & editing.

Notes

The authors declare no competing financial interest.

Biographies

Greta Grassmann obtained in September 2021 a Master's degree in Applied Physics at the University of Bologna and in the same year started her Ph.D. with the Sapienza University at the Italian Institute of Technology in Rome. For her project, she is focusing on the investigation of protein interactions. She has been contributing to the development of a computational approach on the basis of Zernike polynomials that is able to find complementary regions on molecular surfaces and predict complex formation.

Mattia Miotto obtained a Master's degree in Physics at the Sapienza University of Rome in 2016. In the same year, he started his Ph.D. in Physics with a thesis on biological heterogeneities and noise in living systems addressed from a statistical physics viewpoint. Since 2021, he has been working as Postdoctoral Researcher at the Italian Institute of Technology. His current research delves into the physics of living systems across multiple scales from the molecular level to single cells and populations. In particular, he is interested in the functional role of heterogeneities, the interplay between noise and information, the structure and functions of biomolecules, and the emergence of multicellular and population-level behavior.

Fausta Desantis obtained a Master's degree in Physics from the Sapienza University of Rome in March 2021 and started a Ph.D. at the Italian Institute of Technology in Rome in the same year. Her studies center on the molecular mechanisms and the physicochemical features underlying the stability of protein folding and binding with the aim of understanding the processes driving the pathway leading to protein-protein assembly in aggregative diseases with a particular focus on antibody light chain (AL) amyloidosis.

Lorenzo Di Rienzo obtained a Master's degree in Physics from the Sapienza University of Rome in 2015. Right after, he was enrolled as Ph.D. student in Sapienza University. He is currently a Postdoctoral Researcher at the Italian Institute of Technology. His research focuses on the quantitative characterization of physical mechanisms underlying molecular recognition between biomolecules. This aspect is investigated by analyzing the structural and dynamic features of biomolecules. The findings in this field are meant to allow the development of predictive algorithms regarding elements of relevant biomedical importance, such as protein engineering or biological macromolecules binding sites identification.

Gian Gaetano Tartaglia studied theoretical Physics (Sapienza University of Rome) and Biochemistry (University of Zurich, Switzerland). After postdoctoral studies in Cambridge, United Kingdom (Chemistry Department), he became PI at the Centre for

Genomic Regulation (Barcelona, Spain). In 2013, Tartaglia was awarded an European Research Council grant for studies on the role of coding and noncoding RNAs in regulation of amyloid genes. In 2014, he was tenured in Catalonia as ICREA professor of Life and Medical Sciences and in December 2018 became full professor of Biochemistry in the Department of Biology at the Sapienza University of Rome. In 2019, he started to work at the Italian Institute of Technology where he currently leads the "RNA Systems Biology Lab." In 2020, he was awarded an ERC grant for the study on the composition of phase-separated assemblies. Since September 2020, he has been a MAE (Member of Academia Europaea).

Annalisa Pastore is a member of the Academia Europea (MAE) and a Professor of Chemistry and Molecular Biology at King's College London. She studied chemistry at the University of Naples Federico II to earn a master's degree in 1981 and her Ph.D. in 1987. She worked as an exchange student with Richard R. Ernst at ETH Zurich. In 1988, she was appointed a staff scientist position at European Molecular Biology Laboratory. In 1991, she was made group leader of the Structures Program. In 1997, she joined the National Institute for Medical Research of the Medical Research Council. In 2000, she was made a member of the European Molecular Biology Organization (EMBO). From 2013 until 2018, Pastore worked at King's College in London. Her studies focus on the molecular basis of neurodegeneration and the diseases caused by protein aggregation, as well as pathologies resulting from a malfunctioning of the iron metabolism. In 2018, Pastore was the first woman to be appointed full professor at the Faculty of Science of the Scuola Normale Superiore di Pisa. From 2022–2023, she was the Director of Research for Life Sciences at the European Synchrotron Radiation Facility in Grenoble.

Giancarlo Ruocco is full professor of Structure of Matter at the Physics Department of the Sapienza University of Rome and coproposer and, since 2014, senior scientist, PI, and center coordinator at the Center For Life Nano- & Neuro-science (CLNS) at the Italian Institute of Technology (IIT). In the past, he was a member of the Scientific Council of the ESRF, of the board of directors of ESRF (head of the Italian delegation), and member of the LENS board of directors. He has directed the SOFT-INFM research center at the Sapienza University of Rome. Starting in 2007, he was appointed director of the Physics Department of La Sapienza University for six years. From 2010 to 2014, he was vice-rector of Research Policies at the Sapienza University of Rome. His research activity has been devoted to the study of the dynamics of disordered matter, mostly through experimental and numerical techniques, by developing different spectroscopic instrumentation for Brillouin scattering, Raman scattering, inelastic X-ray scattering, impulse stimulated scattering, photon correlation spectroscopy, and nonlinear optics. In the last period, he has been involved in the development and construction of new microscopy-based setup and the development of methods/systems for the investigation of biomatter. Ruocco has been an invited speaker at about 50 international conferences, coauthor of more than 470 publications in international refereed journals, and a Referee for the major journals in the field. He has been and currently is the Principal Investigator for several International and National projects, among which include a ERC-SyG and a EIC-PFopen. Since 2021, he has been a member of the Ukraine Academy of Science and, since 2022, of the Academia Europaea.

Michele Monti is a Postdoctoral Researcher in the RNA Systems Biology Lab at the Italian Institute of Technology. He studied theoretical physics at the University La Sapienza of Rome. He completed his Ph.D. at AMOLF in The Netherlands where his research centered on the dynamics of information flow within biological systems. Specifically, his focus was on how circadian

bacterial cells interpret temporal information to optimize their growth. Following his doctoral studies, Michele transitioned to the Centre for Genomic Regulation (CRG) in Barcelona to collaborate with Gian Gaetano Tartaglia. His work at CRG delved into molecular biology challenges, such as investigating protein aggregation, phase condensation, riboregulation, and protein binding. In recent years, Michele has been dedicated to his research at the Italian Institute of Technology where he explores mutations influencing protein aggregation and phase condensation behaviors. His primary focus involves employing computational approaches to address diverse challenges within molecular biology.

Edoardo Milanetti obtained his Master's degree in Physics from the Sapienza University of Rome in 2012. During his Ph.D., conducted within the Biocomputing group under the supervision of Prof. Anna Tramontano, he contributed to the development of a novel theoretical computational method for analyzing the hydrophobic properties of small organic compounds. Subsequently, as a first Postdoctoral Researcher, Milanetti focused on developing a new method for predicting protein–protein interactions using computational molecular biophysics approaches. He also developed other computational methods, including a graph-theory-based strategy for estimating protein thermal stability. For several years, he has been dedicated to developing methods for investigating the structural and dynamical properties of biological molecules with a specific emphasis on proteins. In 2020, he was a member of the assessment commission for the “contacts” category in the Critical Assessment of Protein Structure Prediction (CASP14). Currently, Milanetti works as a researcher at the Department of Physics at Sapienza University of Rome.

ACKNOWLEDGMENTS

This research was partially funded by grants from ERC-2019-Synergy Grant (ASTRA, no. 855923), EIC-2022-Pathfinder-Open (ivBM-4PAP, no. 101098989), and Project ‘National Center for Gene Therapy and Drugs based on RNA Technology’ (CN00000041) financed by NextGeneration EU PNRR MUR—M4C2—Action 1.4—Call ‘Potenziamento strutture di ricerca e creazione di campioni nazionali di R&S’ (CUP J33C22001130001).

ABBREVIATIONS

AA	all-atom
ACTR	activator for thyroid hormone and retinoid receptor
BD	Brownian dynamics
BPTI	bovine pancreatic trypsin inhibitor
BSA	buried solvent accessible
BSA	bovine serum albumin
C22*	CHARMM 22*
C36m	CHARMM 36m
C36mCU	CHARMM 36m with CUFIX
Cc	cytochrome c
CcP	cytochrome c peroxidase
CG	coarse-grained
CI2	chymotrypsin inhibitor 2
CNN	convolutional neural network
CV	collective variable
DOF	degrees of freedom
DARS	Decoys As the Reference State
DMD	discrete molecular dynamics
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
EM	electron microscopy

ENO	enolase
FFT	Fast Fourier Transform
GDT	global distance test
GB1	protein G B1 domain
GB3	third IgG-binding domain of protein G
GBMV	generalized Born using molecular volume
GCN	graph convolutional network
GO	Gene Ontology
HI	hydrodynamic interaction
HIV	human immunodeficiency virus
HP-36	chicken villin headpiece subdomain
HS	hard spheres
IF1	translation initiation factor 1
IRF-3	interferon regulatory transcription factor
LJ	Lennard-Jones
LSTM	long short-term memory
LYZ	lysozyme
MaSIF	molecular surface interaction fingerprinting
MC	Monte Carlo
MC REX	Monte Carlo replica exchange
MD	molecular dynamics
ML	machine learning
MM-PBSA	molecular mechanics energies combined with the Poisson–Boltzmann surface area
MM-GBSA	molecular mechanics energies combined with the generalized Born surface area
NCBD	nuclear coactivator-binding domain of CREB
NMR	nuclear magnetic resonance
NOX	NADH oxidase
PDB	Protein Data Bank
PDHA	pyruvate dehydrogenase E1.a
PEG	polyethylene glycol
PGK	phosphoglycerate kinase
PP1	protein phosphatase 1
PPI	protein–protein interaction
PSO	particle swarm optimization
REM	replica exchange method
REMC	replica exchange Monte Carlo
REMD	replica exchange molecular dynamics
REST	replica exchange with solute tempering
REXDMD	replica exchange discrete molecular dynamics
RMSD	root-mean-square deviation
RNA	ribonucleic acid
S	spheres
SASA	solvent-accessible surface area
SOD1	superoxide dismutase
SVM	support vector machine
T-REMD	temperature replica exchange molecular dynamics
TREX	temperature replica exchange
UBQ	ubiquitin
VIL	villin headpiece
Y2H	yeast two hybrid

REFERENCES

- (1) Minton, A. P. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers: Original Research on Biomolecules* **1981**, *20*, 2093–2120.
- (2) Minton, A.; Wilf, J. Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry* **1981**, *20*, 4821–4826.
- (3) Ellis, R. J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* **2001**, *11*, 114–119.

- (4) Ellis, R. J.; Minton, A. P. Join the crowd. *Nature* **2003**, *425*, 27–28.
- (5) Zimmerman, S. B.; Trach, S. O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **1991**, *222*, 599–620.
- (6) Engelman, D. M. others Membranes are more mosaic than fluid. *Nature* **2005**, *438*, 578–580.
- (7) Lindén, M.; Sens, P.; Phillips, R. Entropic tension in crowded membranes. *PLoS Comput. Biol.* **2012**, *8*, e1002431.
- (8) Gnutt, D.; Gao, M.; Brylski, O.; Heyden, M.; Ebbinghaus, S. Excluded-volume effects in living cells. *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 2548–2551.
- (9) Boersma, A. J.; Zuhorn, I. S.; Poolman, B. A sensor for quantification of macromolecular crowding in living cells. *Nat. Methods* **2015**, *12*, 227–229.
- (10) Yu, I.; Mori, T.; Ando, T.; Harada, R.; Jung, J.; Sugita, Y.; Feig, M. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife* **2016**, *5*, e19274.
- (11) Zhou, H.-X.; Rivas, G.; Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **2008**, *37*, 375–397.
- (12) Zhou, H.-X. Influence of crowded cellular environments on protein folding, binding, and oligomerization: biological consequences and potentials of atomistic modeling. *FEBS Lett.* **2013**, *587*, 1053–1061.
- (13) Guin, D.; Gruebele, M. Weak chemical interactions that drive protein evolution: Crowding, sticking, and quinary structure in folding and function. *Chem. Rev.* **2019**, *119*, 10691–10717.
- (14) Guigas, G.; Weiss, M. Effects of protein crowding on membrane systems. *Biochim. Biophys. Acta, Biomembr.* **2016**, *1858*, 2441–2450.
- (15) Shahid, S.; Hassan, M. I.; Islam, A.; Ahmad, F. Size-dependent studies of macromolecular crowding on the thermodynamic stability, structure and functional activity of proteins: in vitro and in silico approaches. *Biochim. Biophys. Acta, Gen. Subj.* **2017**, *1861*, 178–197.
- (16) Ostrowska, N.; Feig, M.; Trylska, J. Modeling crowded environment in molecular simulations. *Front. Mol. Biosci.* **2019**, *6*, 86.
- (17) Bergeron-Sandoval, L.-P.; Safaee, N.; Michnick, S. W. Mechanisms and consequences of macromolecular phase separation. *Cell* **2016**, *165*, 1067–1079.
- (18) Julius, K.; Weine, J.; Gao, M.; Latarius, J.; Elbers, M.; Paulus, M.; Tolan, M.; Winter, R. Impact of macromolecular crowding and compression on protein–protein interactions and liquid–liquid phase separation phenomena. *Macromolecules* **2019**, *52*, 1772–1784.
- (19) Savojardo, C.; Martelli, P. L.; Casadio, R. Protein–protein interaction methods and protein phase separation. *Annu. Rev. Biomed. Data Sci.* **2020**, *3*, 89–112.
- (20) Heo, L.; Sugita, Y.; Feig, M. Protein assembly and crowding simulations. *Curr. Opin. Struct. Biol.* **2022**, *73*, 102340.
- (21) Marrink, S. J.; Corradi, V.; Souza, P. C.; Ingólfsson, H. I.; Tieleman, D. P.; Sansom, M. S. Computational modeling of realistic cell membranes. *Chem. Rev.* **2019**, *119*, 6184–6226.
- (22) Löwe, M.; Kalacheva, M.; Boersma, A. J.; Kedrov, A. The more the merrier: effects of macromolecular crowding on the structure and dynamics of biological membranes. *FEBS J.* **2020**, *287*, 5039–5067.
- (23) Rao, V. S.; Srinivas, K.; Sujini, G.; Kumar, G. Protein–protein interaction detection: methods and analysis. *Int. J. Proteomics* **2014**, *2014*, 147648.
- (24) Pedebos, C.; Smith, I. P. S.; Boags, A.; Khalid, S. The hitchhiker’s guide to the periplasm: Unexpected molecular interactions of polymyxin B1 in *E. coli*. *Structure* **2021**, *29*, 444–456.
- (25) Snider, J.; Kotlyar, M.; Saraon, P.; Yao, Z.; Jurisica, I.; Stagljar, I. Fundamentals of protein interaction network mapping. *Mol. Syst. Biol.* **2015**, *11*, 848.
- (26) Lopes, A.; Sacquin-Mora, S.; Dimitrova, V.; Laine, E.; Ponty, Y.; Carbone, A. Protein–protein interactions in a crowded environment: an analysis via cross-docking simulations and evolutionary information. *PLoS Comput. Biol.* **2013**, *9*, e1003369.
- (27) Krishnan, R.; Ranganathan, S.; Maji, S. K.; Padinhateeri, R. Role of non-specific interactions in the phase-separation and maturation of macromolecules. *PLoS Comput. Biol.* **2022**, *18*, e1010067.
- (28) Tarassov, K.; Messier, V.; Landry, C. R.; Radinovic, S.; Molina, M. M. S.; Shames, I.; Malitskaya, Y.; Vogel, J.; Bussey, H.; Michnick, S. W. An in vivo map of the yeast protein interactome. *Science* **2008**, *320*, 1465–1470.
- (29) Landry, C. R.; Levy, E. D.; Michnick, S. W. Weak functional constraints on phosphoproteomes. *Trends Genet.* **2009**, *25*, 193–197.
- (30) Asakura, S.; Oosawa, F. On interaction between two bodies immersed in a solution of macromolecules. *J. Chem. Phys.* **1954**, *22*, 1255–1256.
- (31) Szymański, J.; Patkowski, A.; Wilk, A.; Garstecki, P.; Holyst, R. Diffusion and viscosity in a crowded environment: from nano-to macroscale. *J. Phys. Chem. B* **2006**, *110*, 25593–25597.
- (32) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64–74.
- (33) Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpi, J. L. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinf. Chem.* **2015**, 37–47.
- (34) Gavin, A.-C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Bastuck, S.; Dümpelfeld, B.; et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **2006**, *440*, 631–636.
- (35) König, I.; Zarrine-Afsar, A.; Aznauryan, M.; Soranno, A.; Wunderlich, B.; Dingfelder, F.; Stüber, J. C.; Plückthun, A.; Nettels, D.; Schuler, B. Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat. Methods* **2015**, *12*, 773–779.
- (36) Monteith, W. B.; Cohen, R. D.; Smith, A. E.; Guzman-Cisneros, E.; Pielak, G. J. Quinary structure modulates protein stability in cells. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 1739–1742.
- (37) Cammarata, M.; Piazza, F.; Rivas, G.; Schiro, G.; Temussi, P. A.; Pastore, A. Revitalizing an important field in biophysics: The new Frontiers of Molecular Crowding. *Front. Mol. Biosci.* **2023**, *10*, 121.
- (38) Ando, T.; Skolnick, J. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 18457–18462.
- (39) Goodsell, D. S. *The Machinery of Life*; Springer, 2009; pp 371–402.
- (40) Lucent, D.; Vishal, V.; Pande, V. S. Protein folding under confinement: a role for solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10430–10434.
- (41) Minton, A. P. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* **2001**, *276*, 10577–10580.
- (42) Qin, S.; Zhou, H.-X. Protein folding, binding, and droplet formation in cell-like conditions. *Curr. Opin. Struct. Biol.* **2017**, *43*, 28–37.
- (43) Wang, Y.; Sarkar, M.; Smith, A. E.; Krois, A. S.; Pielak, G. J. Macromolecular crowding and protein stability. *J. Am. Chem. Soc.* **2012**, *134*, 16614–16618.
- (44) Stadtmiller, S. S.; Pielak, G. J. Protein-complex stability in cells and in vitro under crowded conditions. *Curr. Opin. Struct. Biol.* **2021**, *66*, 183–192.
- (45) Speer, S. L.; Zheng, W.; Jiang, X.; Chu, I.-T.; Guseman, A. J.; Liu, M.; Pielak, G. J.; Li, C. The intracellular environment affects protein–protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, e2019918118.
- (46) Minton, A. P. Macromolecular crowding. *Curr. Biol.* **2006**, *16*, R269–R271.
- (47) Köhn, B.; Schwarz, P.; Wittung-Stafshede, P.; Kovermann, M. Impact of crowded environments on binding between protein and single-stranded DNA. *Sci. Rep.* **2021**, *11*, 17682.

- (48) Richter, K.; Nessler, M.; Lichter, P. Macromolecular crowding and its potential impact on nuclear function. *Biochim. Biophys. Acta, Mol. Cell Res.* **2008**, *1783*, 2100–2107.
- (49) Köhn, B.; Kovermann, M. All atom insights into the impact of crowded environments on protein stability by NMR spectroscopy. *Nat. Commun.* **2020**, *11*, 5760.
- (50) Christiansen, A.; Wang, Q.; Samiotakis, A.; Cheung, M. S.; Wittung-Stafshede, P. Factors defining effects of macromolecular crowding on protein stability: an in vitro/in silico case study using cytochrome c. *Biochemistry* **2010**, *49*, 6519–6530.
- (51) Kulkarni, A. M.; Chatterjee, A. P.; Schweizer, K. S.; Zukoski, C. F. Effects of polyethylene glycol on protein interactions. *J. Chem. Phys.* **2000**, *113*, 9863–9873.
- (52) Crowley, P. B.; Brett, K.; Muldoon, J. NMR Spectroscopy reveals cytochrome c–poly (ethylene glycol) interactions. *Chem-BioChem.* **2008**, *9*, 685–688.
- (53) Ostrowska, N.; Feig, M.; Trylska, J. Varying molecular interactions explain aspects of crowder-dependent enzyme function of a viral protease. *PLoS Comput. Biol.* **2023**, *19*, e1011054.
- (54) Candotti, M.; Orozco, M. The differential response of proteins to macromolecular crowding. *PLoS Comput. Biol.* **2016**, *12*, e1005040.
- (55) Popielec, A.; Ostrowska, N.; Wojciechowska, M.; Feig, M.; Trylska, J. Crowded environment affects the activity and inhibition of the NS3/4A protease. *Biochimie* **2020**, *176*, 169–180.
- (56) Miklos, A. C.; Li, C.; Sharaf, N. G.; Pielak, G. J. Volume exclusion and soft interaction effects on protein stability under crowded conditions. *Biochemistry* **2010**, *49*, 6984–6991.
- (57) Sapir, L.; Harries, D. Is the depletion force entropic? Molecular crowding beyond steric interactions. *Curr. Opin. Colloid Interface Sci.* **2015**, *20*, 3–10.
- (58) Hall, D.; Minton, A. P. Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges. *Biochim. Biophys. Acta, Proteins Proteomics* **2003**, *1649*, 127–139.
- (59) Jiao, M.; Li, H.-T.; Chen, J.; Minton, A. P.; Liang, Y. Attractive protein-polymer interactions markedly alter the effect of macromolecular crowding on protein association equilibria. *Biophys. J.* **2010**, *99*, 914–923.
- (60) Minton, A. P. Quantitative assessment of the relative contributions of steric repulsion and chemical interactions to macromolecular crowding. *Biopolymers* **2013**, *99*, 239–244.
- (61) Kim, Y. C.; Mittal, J. Crowding induced entropy-enthalpy compensation in protein association equilibria. *Phys. Rev. Lett.* **2013**, *110*, 208102.
- (62) Senske, M.; Tork, L.; Born, B.; Havenith, M.; Herrmann, C.; Ebbinghaus, S. Protein stabilization by macromolecular crowding through enthalpy rather than entropy. *J. Am. Chem. Soc.* **2014**, *136*, 9036–9041.
- (63) Cheung, M. S.; Klimov, D.; Thirumalai, D. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4753–4758.
- (64) Guseman, A. J.; Perez Goncalves, G. M.; Speer, S. L.; Young, G. B.; Pielak, G. J. Protein shape modulates crowding effects. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 10965–10970.
- (65) Guseman, A. J.; Pielak, G. J. Protein stability and weak intracellular interactions. *RSC* **2019**, 188–206.
- (66) Kim, Y. C.; Bhattacharya, A.; Mittal, J. Macromolecular crowding effects on coupled folding and binding. *J. Phys. Chem. B* **2014**, *118*, 12621–12629.
- (67) Davis, C. M.; Gruebele, M. Non-steric interactions predict the trend and steric interactions the offset of protein stability in cells. *ChemPhysChem* **2018**, *19*, 2290–2294.
- (68) Gorenssek-Benitez, A. H.; Smith, A. E.; Stadtmiller, S. S.; Perez Goncalves, G. M.; Pielak, G. J. Cosolutes, crowding, and protein folding kinetics. *J. Phys. Chem. B* **2017**, *121*, 6527–6537.
- (69) Pastore, A.; Temussi, P. A. Crowding revisited: Open questions and future perspectives. *Trends Biochem. Sci.* **2022**, *47*, 1048–1058.
- (70) Zhang, J.; Maslov, S.; Shakhnovich, E. I. Constraints imposed by non-functional protein–protein interactions on gene expression and proteome size. *Mol. Syst. Biol.* **2008**, *4*, 210.
- (71) McGuffee, S. R.; Elcock, A. H. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* **2010**, *6*, e1000694.
- (72) Levy, E. D.; De, S.; Teichmann, S. A. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 20461–20466.
- (73) Miotto, M.; Marinari, E.; De Martino, A. Competing endogenous RNA crosstalk at system level. *PLoS Comput. Biol.* **2019**, *15*, e1007474.
- (74) Dong, X.; Qin, L.-Y.; Gong, Z.; Qin, S.; Zhou, H.-X.; Tang, C. Preferential interactions of a crowder protein with the specific binding site of a native protein complex. *J. Phys. Chem. Lett.* **2022**, *13*, 792–800.
- (75) Luh, L. M.; Hansel, R.; Lohr, F.; Kirchner, D. K.; Krauskopf, K.; Pitzius, S.; Schafer, B.; Tufar, P.; Corbeski, I.; Guntert, P.; et al. Molecular crowding drives active Pin1 into nonspecific complexes with endogenous proteins prior to substrate recognition. *J. Am. Chem. Soc.* **2013**, *135*, 13796–13803.
- (76) Ghosh, A.; Smith, P. E.; Qin, S.; Yi, M.; Zhou, H.-X. Both Ligands and macromolecular crowders preferentially bind to closed conformations of maltose binding protein. *Biochemistry* **2019**, *58*, 2208–2217.
- (77) Srivastava, D.; Bernhard, S. A. Metabolite transfer via enzyme-enzyme complexes. *Science* **1986**, *234*, 1081–1086.
- (78) Sukenik, S.; Ren, P.; Gruebele, M. Weak protein–protein interactions in live cells are quantified by cell-volume modulation. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 6776–6781.
- (79) Hoppe, T.; Minton, A. P. Non-specific interactions between macromolecular solutes in concentrated solution: Physico-chemical manifestations and biochemical consequences. *Front. Mol. Biosci.* **2019**, *6*, 10.
- (80) Hlevnjak, M.; Zitkovic, G.; Zagrovc, B. Hydrophilicity Matching—A Potential Prerequisite for the Formation of Protein-Protein Complexes in the Cell. *PLoS One* **2010**, *5*, e11169.
- (81) Bickmore, W. A.; Sutherland, H. G. Addressing protein localization within the nucleus. *EMBO Journal* **2002**, *21*, 1248–1254.
- (82) Sutherland, H. G.; Mumford, G. K.; Newton, K.; Ford, L. V.; Farrall, R.; Dellaire, G.; Cáceres, J. F.; Bickmore, W. A. Large-scale identification of mammalian proteins localized to nuclear sub-compartments. *Hum. Mol. Genet.* **2001**, *10*, 1995–2011.
- (83) Kundrotas, P. J.; Alexov, E. Electrostatic properties of protein-protein complexes. *Biophys. J.* **2006**, *91*, 1724–1736.
- (84) Pusara, S.; Yamin, P.; Wenzel, W.; Krstić, M.; Kozłowska, M. A coarse-grained xDLVO model for colloidal protein–protein interactions. *Phys. Chem. Chem. Phys.* **2021**, *23*, 12780–12794.
- (85) Harada, R.; Tochio, N.; Kigawa, T.; Sugita, Y.; Feig, M. Reduced native state stability in crowded cellular environment due to protein–protein interactions. *J. Am. Chem. Soc.* **2013**, *135*, 3696–3701.
- (86) Kim, Y. C.; Best, R. B.; Mittal, J. Macromolecular crowding effects on protein–protein binding affinity and specificity. *J. Chem. Phys.* **2010**, *133*, 205101.
- (87) Li, X.; Moal, I. H.; Bates, P. A. Detection and refinement of encounter complexes for protein–protein docking: taking account of macromolecular crowding. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 3189–3196.
- (88) Guseman, A. J.; Pielak, G. J. Cosolute and crowding effects on a side-by-side protein dimer. *Biochemistry* **2017**, *56*, 971–976.
- (89) Bashardanesh, Z.; Elf, J.; Zhang, H.; Van der Spoel, D. Rotational and translational diffusion of proteins as a function of concentration. *ACS Omega* **2019**, *4*, 20654–20664.
- (90) Hook, D. W.; Porter, S. J.; Herzog, C. Dimensions: building context for search and evaluation. *Front. Res. Metr. Anal.* **2018**, *3*, 23.
- (91) Van Zundert, G.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastrić, P.; Karaca, E.; Melquiond, A.; van Dijk, M.; De Vries, S.; Bonvin, A. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **2016**, *428*, 720–725.

- (92) Sudhaharan, T.; Liu, P.; Foo, Y. H.; Bu, W.; Lim, K. B.; Wohland, T.; Ahmed, S. Determination of in vivo dissociation constant, KD, of Cdc42-effector complexes in live mammalian cells using single wavelength fluorescence cross-correlation spectroscopy. *J. Biol. Chem.* **2009**, *284*, 13602–13609.
- (93) Shi, X.; Foo, Y. H.; Sudhaharan, T.; Chong, S.-W.; Korzh, V.; Ahmed, S.; Wohland, T. Determination of dissociation constants in living zebrafish embryos with single wavelength fluorescence cross-correlation spectroscopy. *Biophys. J.* **2009**, *97*, 678–686.
- (94) Sahu, S. K.; Biswas, M. Modeling protein association from homogeneous to mixed environments: A reaction-diffusion dynamics approach. *J. Mol. Graphics Model.* **2021**, *107*, 107936.
- (95) Katava, M.; Stirnemann, G.; Pachetti, M.; Capaccioli, S.; Paciaroni, A.; Sterpone, F. Specific interactions and environment flexibility tune protein stability under extreme crowding. *J. Phys. Chem. B* **2021**, *125*, 6103–6111.
- (96) Timr, S.; Sterpone, F. Stabilizing or destabilizing: simulations of chymotrypsin inhibitor 2 under crowding reveal existence of a crossover temperature. *J. Phys. Chem. Lett.* **2021**, *12*, 1741–1746.
- (97) Ghosh, A.; Mazarakos, K.; Zhou, H.-X. Three archetypical classes of macromolecular regulators of protein liquid–liquid phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 19474–19483.
- (98) Dong, H.; Qin, S.; Zhou, H.-X. Effects of macromolecular crowding on protein conformational changes. *PLoS Comput. Biol.* **2010**, *6*, e1000833.
- (99) Stewart, C. J.; Olgenblum, G. I.; Propst, A.; Harries, D.; Pielak, G. J. Resolving the enthalpy of protein stabilization by macromolecular crowding. *Protein Sci.* **2023**, *32*, e4573.
- (100) Gomez, D.; Huber, K.; Klumpp, S. On protein folding in crowded conditions. *J. Phys. Chem. Lett.* **2019**, *10*, 7650–7656.
- (101) Sharp, K. A. Analysis of the size dependence of macromolecular crowding shows that smaller is better. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 7990–7995.
- (102) Verma, P. K.; Kundu, A.; Cho, M. How molecular crowding differs from macromolecular crowding: A femtosecond mid-infrared pump–probe study. *J. Phys. Chem. Lett.* **2018**, *9*, 6584–6592.
- (103) Sung, H.-L.; Sengupta, A.; Nesbitt, D. Smaller molecules crowd better: Crowder size dependence revealed by single-molecule FRET studies and depletion force modeling analysis. *J. Chem. Phys.* **2021**, *154*, 155101.
- (104) Denesyuk, N. A.; Thirumalai, D. Crowding promotes the switch from hairpin to pseudoknot conformation in human telomerase RNA. *J. Am. Chem. Soc.* **2011**, *133*, 11858–11861.
- (105) Kang, H.; Pincus, P. A.; Hyeon, C.; Thirumalai, D. Effects of macromolecular crowding on the collapse of biopolymers. *Phys. Rev. Lett.* **2015**, *114*, No. 068303.
- (106) Erijman, A.; Rosenthal, E.; Shifman, J. M. How structure defines affinity in protein–protein interactions. *PLoS One* **2014**, *9*, e110085.
- (107) Vangone, A.; Bonvin, A. M. Contacts-based prediction of binding affinity in protein–protein complexes. *eLife* **2015**, *4*, No. e07454.
- (108) Réau, M.; Renaud, N.; Xue, L. C.; Bonvin, A. M. DeepRank-GNN: a graph neural network framework to learn patterns in protein–protein interfaces. *Bioinformatics* **2023**, *39*, btac759.
- (109) Rual, J.-F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; et al. Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **2005**, *437*, 1173–1178.
- (110) Rolland, T.; Taşan, M.; Charlotheaux, B.; Pevzner, S. J.; Zhong, Q.; Sahni, N.; Yi, S.; Lemmens, I.; Fontanillo, C.; Mosca, R.; et al. A proteome-scale map of the human interactome network. *Cell* **2014**, *159*, 1212–1226.
- (111) Luck, K.; Kim, D.-K.; Lambourne, L.; Spirohn, K.; Begg, B. E.; Bian, W.; Brignall, R.; Cafarelli, T.; Campos-Laborie, F. J.; Charlotheaux, B.; et al. A reference map of the human binary protein interactome. *Nature* **2020**, *580*, 402–408.
- (112) Ding, Z.; Kihara, D. Computational identification of protein–protein interactions in model plant proteomes. *Sci. Rep.* **2019**, *9*, 8740.
- (113) Li, Y.; Ilie, L. SPRINT: ultrafast protein–protein interaction prediction of the entire human interactome. *BMC Bioinf.* **2017**, *18*, 485.
- (114) Vakser, I. A.; Grudinin, S.; Jenkins, N. W.; Kundrotas, P. J.; Deeds, E. J. Docking-based long timescale simulation of cell-size protein systems at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, e2210249119.
- (115) Tan, S.-H.; Zhang, Z.; Ng, S.-K. ADVICE: automated detection and validation of interaction by co-evolution. *Nucleic Acids Res.* **2004**, *32*, W69–W72.
- (116) Pitre, S.; North, C.; Alamgir, M.; Jessulat, M.; Chan, A.; Luo, X.; Green, J. R.; Dumontier, M.; Dehne, F.; Golshani, A. Global investigation of protein–protein interactions in yeast *Saccharomyces cerevisiae* using re-occurring short polypeptide sequences. *Nucleic Acids Res.* **2008**, *36*, 4286–4294.
- (117) Pan, X.-Y.; Zhang, Y.-N.; Shen, H.-B. Large-Scale prediction of human protein–protein interactions from amino acid sequence based on latent topic features. *J. Proteome Res.* **2010**, *9*, 4992–5001.
- (118) Romero-Molina, S.; Ruiz-Blanco, Y. B.; Harms, M.; Münch, J.; Sanchez-Garcia, E. PPI-detect: A support vector machine model for sequence-based prediction of protein–protein interactions. *J. Comput. Chem.* **2019**, *40*, 1233–1242.
- (119) Yao, Y.; Du, X.; Diao, Y.; Zhu, H. An integration of deep learning with feature embedding for protein–protein interaction prediction. *PeerJ.* **2019**, *7*, e7126.
- (120) Mukherjee, S.; Zhang, Y. Protein–protein complex structure predictions by multimeric threading and template recombination. *Structure* **2011**, *19*, 955–966.
- (121) Afsar Minhas, F. u. A.; Geiss, B. J.; Ben-Hur, A. PAIRpred: partner-specific prediction of interacting residues from sequence and structure. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 1142–1155.
- (122) Fout, A.; Byrd, J.; Shariat, B.; Ben-Hur, A. Protein interface prediction using graph convolutional networks. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, Vol. 30; Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc., 2017.
- (123) Sanchez-Garcia, R.; Sorzano, C. O. S.; Carazo, J. M.; Segura, J. BIPSPi: a method for the prediction of partner-specific protein–protein interfaces. *Bioinformatics* **2019**, *35*, 470–477.
- (124) Zeng, M.; Zhang, F.; Wu, F.-X.; Li, Y.; Wang, J.; Li, M. Protein–protein interaction site prediction through combining local and global features with deep neural networks. *Bioinformatics* **2020**, *36*, 1114–1120.
- (125) Miotto, M.; Di Rienzo, L.; Bo', L.; Ruocco, G.; Milanetti, E. Zepyros: A webserver to evaluate the shape complementarity of protein–protein interfaces. *arXiv*, February 10, 2024, 2402.06960, ver. 1. DOI: 10.48550/arXiv.2402.06960.
- (126) Li, Y.; Golding, G. B.; Ilie, L. DELPHI: accurate deep ensemble model for protein interaction sites prediction. *Bioinformatics* **2021**, *37*, 896–904.
- (127) Lu, S.; Li, Y.; Nan, X.; Zhang, S. Attention-based convolutional neural networks for protein–protein interaction site prediction. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. **2021**, *2021*, 141–144.
- (128) Gainza, P.; Sverrisson, F.; Monti, F.; Rodola, E.; Boscaini, D.; Bronstein, M.; Correia, B. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **2020**, *17*, 184–192.
- (129) Krapp, L. F.; Abriata, L. A.; Cortés Rodríguez, F.; Dal Peraro, M. PeSTo: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. *Nat. Commun.* **2023**, *14*, 2175.
- (130) Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367.

- (131) Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 392–406.
- (132) Torchala, M.; Moal, I. H.; Chaleil, R. A.; Fernandez-Recio, J.; Bates, P. A. SwarmDock: a server for flexible protein–protein docking. *Bioinformatics* **2013**, *29*, 807–809.
- (133) Pierce, B. G.; Wiehe, K.; Hwang, H.; Kim, B.-H.; Vreven, T.; Weng, Z. ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics* **2014**, *30*, 1771–1773.
- (134) Baspinar, A.; Cukuroglu, E.; Nussinov, R.; Keskin, O.; Gursoy, A. PRISM: a web server and repository for prediction of protein–protein interactions and modeling their 3D complexes. *Nucleic Acids Res.* **2014**, *42*, W285–W289.
- (135) Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S.-Y. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **2017**, *45*, W365–W373.
- (136) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro web server for protein–protein docking. *Nat. Protoc.* **2017**, *12*, 255–278.
- (137) Hayashi, T.; Matsuzaki, Y.; Yanagisawa, K.; Ohue, M.; Akiyama, Y. MEGADOCK-Web: an integrated database of high-throughput structure-based protein–protein interaction predictions. *BMC Bioinf.* **2018**, *19*, 61–72.
- (138) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303.
- (139) Bryant, P.; Pozzati, G.; Elofsson, A. Improved prediction of protein–protein interactions using AlphaFold2. *Nat. Commun.* **2022**, *13*, 1265.
- (140) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (141) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (142) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (143) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542.
- (144) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simonnet, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.
- (145) Kobayashi, C.; Jung, J.; Matsunaga, Y.; Mori, T.; Ando, T.; Tamura, K.; Kamiya, M.; Sugita, Y. GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J. Comput. Chem.* **2017**, *38*, 2193–2206.
- (146) Harrigan, M. P.; Sultan, M. M.; Hernández, C. X.; Husic, B. E.; Eastman, P.; Schwantes, C. R.; Beauchamp, K. A.; McGibbon, R. T.; Pande, V. S. MSMBuilder: statistical models for biomolecular dynamics. *Biophys. J.* **2017**, *112*, 10–15.
- (147) Fiorentini, R.; Tarenzi, T.; Potestio, R. Fast, Accurate, and System-Specific Variable-Resolution Modeling of Proteins. *J. Chem. Inf. Model.* **2023**, *63*, 1260–1275.
- (148) Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. A knowledge-based energy function for protein–ligand, protein–protein, and protein–DNA complexes. *J. Med. Chem.* **2005**, *48*, 2325–2335.
- (149) Xue, L. C.; Rodrigues, J. P.; Kastiris, P. L.; Bonvin, A. M.; Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **2016**, *32*, 3676–3678.
- (150) Chéron, J.-B.; Zacharias, M.; Antonczak, S.; Fiorucci, S. Update of the ATTRACT force field for the prediction of protein–protein binding affinity. *J. Comput. Chem.* **2017**, *38*, 1887–1890.
- (151) Chen, M.; Ju, C. J.-T.; Zhou, G.; Chen, X.; Zhang, T.; Chang, K.-W.; Zaniolo, C.; Wang, W. Multifaceted protein–protein interaction prediction based on Siamese residual RCNN. *Bioinformatics* **2019**, *35*, i305–i314.
- (152) Abbasi, W. A.; Yaseen, A.; Hassan, F. U.; Andleeb, S.; Minhas, F. U. A. A. ISLAND: in-silico proteins binding affinity prediction using sequence information. *BioData Min.* **2020**, *13*, 20.
- (153) Rodrigues, C. H.; Pires, D. E.; Ascher, D. B. mmCSM-PPI: predicting the effects of multiple point mutations on protein–protein interactions. *Nucleic Acids Res.* **2021**, *49*, W417–W424.
- (154) Wang, B.; Su, Z.; Wu, Y. Computational assessment of protein–protein binding affinity by reversely engineering the energetics in protein complexes. *Genomics, Proteomics Bioinf.* **2021**, *19*, 1012–1022.
- (155) Wee, J.; Xia, K. Persistent spectral based ensemble learning (PerSpect-EL) for protein–protein binding affinity prediction. *Briefings Bioinf.* **2022**, *23*, bbac024.
- (156) Romero-Molina, S.; Ruiz-Blanco, Y. B.; Mieres-Perez, J.; Harms, M.; Munch, J.; Ehrmann, M.; Sanchez-Garcia, E. PPI-affinity: A web tool for the prediction and optimization of protein–peptide and protein–protein binding affinity. *J. Proteome Res.* **2022**, *21*, 1829–1841.
- (157) Guo, Y.; Yu, L.; Wen, Z.; Li, M. Using support vector machine combined with auto covariance to predict protein–protein interactions from protein sequences. *Nucleic Acids Res.* **2008**, *36*, 3025–3030.
- (158) You, Z.-H.; Lei, Y.-K.; Zhu, L.; Xia, J.; Wang, B. Prediction of protein–protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis. *BMC Bioinf.* **2013**, *14*, 1471–2105.
- (159) Huang, Y.-A.; You, Z.-H.; Gao, X.; Wong, L.; Wang, L. others Using weighted sparse representation model combined with discrete cosine transformation to predict protein–protein interactions from protein sequence. *BioMed. Res. Int.* **2015**, *2015*, 1–10.
- (160) Sun, T.; Zhou, B.; Lai, L.; Pei, J. Sequence-based prediction of protein protein interaction using a deep-learning algorithm. *BMC Bioinf.* **2017**, *18*, 1471–2105.
- (161) Wang, Y.-B.; You, Z.-H.; Li, X.; Jiang, T.-H.; Chen, X.; Zhou, X.; Wang, L. Predicting protein–protein interactions from protein sequences by a stacked sparse autoencoder deep neural network. *Mol. BioSyst.* **2017**, *13*, 1336–1344.
- (162) An, J.-Y.; You, Z.-H.; Zhou, Y.; Wang, D.-F. Sequence-based prediction of protein–protein interactions using gray wolf optimizer–based relevance vector machine. *Evol. Bioinf.* **2019**, *15*, 1–10.
- (163) Chen, K.-H.; Wang, T.-F.; Hu, Y.-J. Protein–protein interaction prediction using a hybrid feature representation and a stacked generalization scheme. *BMC Bioinf.* **2019**, *20*, 308.
- (164) Emamjomeh, A.; Goliaei, B.; Torkamani, A.; Ebrahimpour, R.; Mohammadi, N.; Parsian, A. Protein–protein interaction prediction by combined analysis of genomic and conservation information. *Genes Genet. Syst.* **2014**, *89*, 259–272.
- (165) Hamp, T.; Rost, B. Evolutionary profiles improve protein–protein interaction prediction from sequence. *Bioinformatics* **2015**, *31*, 1945–1950.
- (166) Kamada, M.; Sakuma, Y.; Hayashida, M.; Akutsu, T. Prediction of protein–protein interaction strength using domain features with supervised regression. *Sci. World J.* **2014**, *2014*, 1–7.
- (167) Xu, Y.; Hu, W.; Chang, Z.; DuanMu, H.; Zhang, S.; Li, Z.; Li, Z.; Yu, L.; Li, X. Prediction of human protein–protein interaction by a mixed Bayesian model and its application to exploring underlying cancer-related pathway crosstalk. *J. R. Soc., Interface* **2011**, *8*, 555–567.
- (168) Shen, J.; Zhang, J.; Luo, X.; Zhu, W.; Yu, K.; Chen, K.; Li, Y.; Jiang, H. Predicting protein–protein interactions based only on

- sequences information. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4337–4341.
- (169) Murakami, Y.; Mizuguchi, K. Recent developments of sequence-based prediction of protein–protein interactions. *Biophys. Rev.* **2022**, *14*, 1393.
- (170) Ding, Y.; Tang, J.; Guo, F. Predicting protein–protein interactions via multivariate mutual information of protein sequences. *BMC Bioinf.* **2016**, *17*, 398.
- (171) Martin, S.; Roe, D.; Faulon, J.-L. Predicting protein–protein interactions using signature products. *Bioinformatics* **2005**, *21*, 218–226.
- (172) Pazos, F.; Helmer-Citterich, M.; Ausiello, G.; Valencia, A. Correlated mutations contain information about protein–protein interaction. *J. Mol. Biol.* **1997**, *271*, 511–523.
- (173) Lesne, A. Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics. *Mathematical Structures in Computer Science* **2014**, *24*, e240311.
- (174) Stein, R. R.; Marks, D. S.; Sander, C. Inferring pairwise interactions from biological data using maximum-entropy probability models. *PLoS Comput. Biol.* **2015**, *11*, e1004182.
- (175) De Martino, A.; De Martino, D. An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon* **2018**, *4*, No. e00596.
- (176) Miotto, M.; Monacelli, L. Entropy evaluation sheds light on ecosystem complexity. *Phys. Rev. E* **2018**, *98*, No. 042402.
- (177) Miotto, M.; Monacelli, L. TOLOMEO, a Novel Machine Learning Algorithm to Measure Information and Order in Correlated Networks and Predict Their State. *Entropy* **2021**, *23*, 1138.
- (178) Miotto, M.; Rosito, M.; Paoluzzi, M.; de Turrís, V.; Folli, V.; Leonetti, M.; Ruocco, G.; Rosa, A.; Gosti, G. Collective behavior and self-organization in neural rosette morphogenesis. *Front. Cell Dev. Biol.* **2023**, *11*, 1134091.
- (179) Seno, F.; Trovato, A.; Banavar, J. R.; Maritan, A. Maximum Entropy Approach for Deducing Amino Acid Interactions in Proteins. *Phys. Rev. Lett.* **2008**, *100*, No. 078102.
- (180) Weigt, M.; White, R. A.; Szurmant, H.; Hoch, J. A.; Hwa, T. Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 67–72.
- (181) Morcos, F.; Hwa, T.; Onuchic, J. N.; Weigt, M. Direct coupling analysis for protein contact prediction. *Protein Struct. Predict.* **2014**, *1137*, 55–70.
- (182) Kovács, I. A.; Luck, K.; Spirohn, K.; Wang, Y.; Pollis, C.; Schlabach, S.; Bian, W.; Kim, D.-K.; Kishore, N.; Hao, T.; et al. Network-based prediction of protein interactions. *Nat. Commun.* **2019**, *10*, 2041–1723.
- (183) Przytycka, T. M.; Singh, M.; Slonim, D. K. Toward the dynamic interactome: it's about time. *Briefings Bioinf.* **2010**, *11*, 15–29.
- (184) Chen, B.; Fan, W.; Liu, J.; Wu, F.-X. Identifying protein complexes and functional modules—from static PPI networks to dynamic PPI networks. *Briefings Bioinf.* **2014**, *15*, 177–194.
- (185) Wu, Z.; Liao, Q.; Liu, B. A comprehensive review and evaluation of computational methods for identifying protein complexes from protein–protein interaction networks. *Briefings Bioinf.* **2020**, *21*, 1531–1548.
- (186) Tadaka, S.; Kinoshita, K. NCMine: Core-peripheral based functional module detection using near-clique mining. *Bioinformatics* **2016**, *32*, 3454–3460.
- (187) Pellegrini, M.; Baglioni, M.; Geraci, F. Protein complex prediction for large protein–protein interaction networks with the Core&Peel method. *BMC Bioinf.* **2016**, *17*, 37–58.
- (188) Bader, G. D.; Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf.* **2003**, *4*, 2.
- (189) Adamcsek, B.; Palla, G.; Farkas, I. J.; Derényi, I.; Vicsek, T. CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* **2006**, *22*, 1021–1023.
- (190) Wu, M.; Ou-Yang, L.; Li, X.-L. Protein complex detection via effective integration of base clustering solutions and co-complex affinity scores. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2017**, *14*, 733–739.
- (191) Huttlin, E. L.; Bruckner, R. J.; Paulo, J. A.; Cannon, J. R.; Ting, L.; Baltier, K.; Colby, G.; Gebreab, F.; Gygi, M. P.; Parzen, H.; et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **2017**, *545*, 505–509.
- (192) Yu, H.; Braun, P.; Yildirim, M. A.; Lemmens, I.; Venkatesan, K.; Sahalie, J.; Hirozane-Kishikawa, T.; Gebreab, F.; Li, N.; Simonis, N.; et al. High-quality binary protein interaction map of the yeast interactome network. *Science* **2008**, *322*, 104–110.
- (193) The Gene Ontology Consortium; Aleksander, S. A.; Balhoff, J.; Carbon, S.; Cherry, J. M.; Drabkin, H. J.; Ebert, D.; Feuermann, M.; Gaudet, P.; Harris, N. The Gene Ontology knowledgebase in 2023. *Genetics* **2023**, *224*, iyad031.
- (194) Pellegrini, M.; Marcotte, E. M.; Thompson, M. J.; Eisenberg, D.; Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4285–4288.
- (195) Sharan, R.; Ulitsky, I.; Shamir, R. Network-based prediction of protein function. *Mol. Syst. Biol.* **2007**, *3*, 88.
- (196) Smaili, F. Z.; Gao, X.; Hoehndorf, R. OPA2Vec: combining formal and informal content of biomedical ontologies to improve similarity-based prediction. *Bioinformatics* **2019**, *35*, 2133–2140.
- (197) Wu, X.; Zhu, L.; Guo, J.; Zhang, D.-Y.; Lin, K. Prediction of yeast protein–protein interaction network: insights from the Gene Ontology and annotations. *Nucleic Acids Res.* **2006**, *34*, 2137–2150.
- (198) Ben-Hur, A.; Noble, W. S. Kernel methods for predicting protein–protein interactions. *Bioinformatics* **2005**, *21*, i38–i46.
- (199) Zhanhua, C.; Gan, J. G.-K.; lei, L.; Sakharkar, M. K.; Kanguane, P.; et al. Protein subunit interfaces: heterodimers versus homodimers. *Bioinformatics* **2005**, *1*, 28–39.
- (200) Desantis, F.; Miotto, M.; Di Rienzo, L.; Milanetti, E.; Ruocco, G. Spatial organization of hydrophobic and charged residues affects protein thermal stability and binding affinity. *Sci. Rep.* **2022**, *12*, 12087.
- (201) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot spots—A review of the protein–protein interface determinant amino-acid residues. *Proteins: Struct., Funct., Bioinf.* **2007**, *68*, 803–812.
- (202) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **1997**, *272*, 106–120.
- (203) Skrabanek, L.; Saini, H. K.; Bader, G. D.; Enright, A. J. Computational prediction of protein–protein interactions. *Mol. Biotechnol.* **2008**, *38*, 1–17.
- (204) Miotto, M.; Olimpieri, P. P.; Di Rienzo, L.; Ambrosetti, F.; Corsi, P.; Lepore, R.; Tartaglia, G. G.; Milanetti, E. Insights on protein thermal stability: a graph representation of molecular interactions. *Bioinformatics* **2019**, *35*, 2569–2577.
- (205) Hawkins, T.; Chitale, M.; Luban, S.; Kihara, D. PFP: Automated prediction of gene ontology functional annotations with confidence scores using protein sequence data. *Proteins: Struct., Funct., Bioinf.* **2009**, *74*, 566–582.
- (206) Miskei, M.; Horvath, A.; Vendruscolo, M.; Fuxreiter, M. Sequence-Based Prediction of Fuzzy Protein Interactions. *J. Mol. Biol.* **2020**, *432*, 2289–2303.
- (207) Zhang, B.; Li, J.; Quan, L.; Chen, Y.; Lü, Q. Sequence-based prediction of protein–protein interaction sites by simplified long short-term memory network. *Neurocomputing* **2019**, *357*, 86–100.
- (208) Grassmann, G.; Miotto, M.; Di Rienzo, L.; Gosti, G.; Ruocco, G.; Milanetti, E. A novel computational strategy for defining the minimal protein molecular surface representation. *PLoS One* **2022**, *17*, No. e0266004.
- (209) Di Rienzo, L.; Miotto, M.; Bò, L.; Ruocco, G.; Raimondo, D.; Milanetti, E. Characterizing hydrophobicity of amino acid side chain in a protein environment by investigating the structural changes of water molecules network. *Front. Mol. Biosci.* **2021**, *8*, 626837.

- (210) Milanetti, E.; Miotto, M.; Di Rienzo, L.; Monti, M.; Gosti, G.; Ruocco, G. 2D Zernike polynomial expansion: Finding the protein-protein binding regions. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 29–36.
- (211) Milanetti, E.; Miotto, M.; Di Rienzo, L.; Nagaraj, M.; Monti, M.; Golbek, T. W.; Gosti, G.; Roeters, S. J.; Weidner, T.; Otzen, D. E.; Ruocco, G. In-Silico Evidence for a Two Receptor Based Strategy of SARS-CoV-2. *Frontiers in Molecular Biosciences* **2021**, *8*, 690655.
- (212) Miotto, M.; Di Rienzo, L.; Gosti, G.; Bo', L.; Parisi, G.; Piacentini, R.; Boffi, A.; Ruocco, G.; Milanetti, E. Inferring the stabilization effects of SARS-CoV-2 variants on the binding with ACE2 receptor. *Communications Biology* **2022**, *5*, 20221.
- (213) Piacentini, R.; Centi, L.; Miotto, M.; Milanetti, E.; Di Rienzo, L.; Pitea, M.; Piazza, P.; Ruocco, G.; Boffi, A.; Parisi, G. Lactoferrin Inhibition of the Complex Formation between ACE2 Receptor and SARS CoV-2 Recognition Binding Domain. *International Journal of Molecular Sciences* **2022**, *23*, 5436.
- (214) Monti, M.; et al. Two Receptor Binding Strategy of SARS-CoV-2 Is Mediated by Both the N-Terminal and Receptor-Binding Spike Domain. *J. Phys. Chem. B* **2024**, *128*, 451–464.
- (215) Di Rienzo, L.; Monti, M.; Milanetti, E.; Miotto, M.; Boffi, A.; Tartaglia, G. G.; Ruocco, G. Computational optimization of angiotensin-converting enzyme 2 for SARS-CoV-2 Spike molecular recognition. *Computational and Structural Biotechnology Journal* **2021**, *19*, 3006–3014.
- (216) De Lauro, A.; Di Rienzo, L.; Miotto, M.; Olimpieri, P. P.; Milanetti, E.; Ruocco, G. Shape Complementarity Optimization of Antibody–Antigen Interfaces: The Application to SARS-CoV-2 Spike Protein. *Frontiers in Molecular Biosciences* **2022**, *9*, 874296.
- (217) Di Rienzo, L.; Miotto, M.; Milanetti, E.; Ruocco, G. Computational structural-based GPCR optimization for user-defined ligand: Implications for the development of biosensors. *Computational and Structural Biotechnology Journal* **2023**, *21*, 3002–3009.
- (218) Parisi, G.; Piacentini, R.; Incocciati, A.; Bonamore, A.; Macone, A.; Rupert, J.; Zacco, E.; Miotto, M.; Milanetti, E.; Tartaglia, G. G.; Ruocco, G.; Boffi, A.; Di Rienzo, L. Design of protein-binding peptides with controlled binding affinity: the case of SARS-CoV-2 receptor binding domain and angiotensin-converting enzyme 2 derived peptides. *Frontiers in Molecular Biosciences* **2024**, *10*, 1332359.
- (219) Grassmann, G.; Di Rienzo, L.; Gosti, G.; Leonetti, M.; Ruocco, G.; Miotto, M.; Milanetti, E. Electrostatic complementarity at the interface drives transient protein-protein interactions. *Sci. Rep.* **2023**, *13*, 10207.
- (220) Gainza, P.; Wehrle, S.; Van Hall-Beauvais, A.; Marchand, A.; Scheck, A.; Harteveld, Z.; Buckley, S.; Ni, D.; Tan, S.; Sverrisson, F.; et al. De novo design of protein interactions with learned surface fingerprints. *Nature* **2023**, *617*, 176–184.
- (221) Wodak, S. J.; Vajda, S.; Lensink, M. F.; Kozakov, D.; Bates, P. A. Critical Assessment of Methods for Predicting the 3D Structure of Proteins and Protein Complexes. *Annu. Rev. Biophys.* **2023**, *52*, 183–206.
- (222) Kryshchak, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moutl, J. Critical assessment of methods of protein structure prediction (CASP)—Round XIV. *Proteins: Struct., Funct., Bioinf.* **2021**, *89*, 1607–1617.
- (223) Bradley, P.; Misura, K. M.; Baker, D. Toward high-resolution de novo structure prediction for small proteins. *Science* **2005**, *309*, 1868–1871.
- (224) Zheng, W.; Li, Y.; Zhang, C.; Zhou, X.; Pearce, R.; Bell, E. W.; Huang, X.; Zhang, Y. Protein structure prediction using deep learning distance and hydrogen-bonding restraints in CASP14. *Proteins: Struct., Funct., Bioinf.* **2021**, *89*, 1734–1751.
- (225) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8.
- (226) Elofsson, A. Progress at protein structure prediction, as seen in CASP15. *Curr. Opin. Struct. Biol.* **2023**, *80*, 102594.
- (227) AlQuraishi, M. AlphaFold at CASP13. *Bioinformatics* **2019**, *35*, 4862–4865.
- (228) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (229) International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921.
- (230) Lensink, M. F.; Velankar, S.; Wodak, S. J. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins: Struct., Funct., Bioinf.* **2017**, *85*, 359–377.
- (231) Yin, R.; Feng, B. Y.; Varshney, A.; Pierce, B. G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* **2022**, *31*, e4379.
- (232) Ruff, K. M.; Pappu, R. V. AlphaFold and implications for intrinsically disordered proteins. *J. Mol. Biol.* **2021**, *433*, 167208.
- (233) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Židek, A.; Bates, R.; Blackwell, S.; Yim, J. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, October 4, 2021, ver. 1, 463034.
- (234) Basu, S.; Wallner, B. DockQ: a quality measure for protein-protein docking models. *PLoS One* **2016**, *11*, No. e0161879.
- (235) Liu, J.; Guo, Z.; Wu, T.; Roy, R. S.; Chen, C.; Cheng, J. Improving AlphaFold2-based protein tertiary structure prediction with MULTICOM in CASP15. *Commun. Chem.* **2023**, *6*, 188.
- (236) Wang, Z.; Eickholt, J.; Cheng, J. MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8. *Bioinformatics* **2010**, *26*, 882–888.
- (237) van Kempen, M.; Kim, S. S.; Tumescheit, C.; Mirdita, M.; Lee, J.; Gilchrist, C. L.; Söding, J.; Steinegger, M. Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* **2024**, *42*, 243–246.
- (238) Bhattacharya, S.; Roche, R.; Shuvo, M. H.; Bhattacharya, D. Recent advances in protein homology detection propelled by inter-residue interaction map threading. *Front. Mol. Biosci.* **2021**, *8*, No. 643752.
- (239) RCSB PDB. *PDB Statistics: Overall Growth of Released Structures Per Year*. <https://www.rcsb.org/stats/growth/growth-released-structures> (accessed 2024-02-21).
- (240) Saibil, H. R. Cryo-EM in molecular and cellular biology. *Mol. Cell* **2022**, *82*, 274–284.
- (241) Kundrotas, P. J.; Anishchenko, I.; Dauzhenka, T.; Kotthoff, I.; Mnevets, D.; Copeland, M. M.; Vakser, I. A. Dockground: a comprehensive data resource for modeling of protein complexes. *Protein Sci.* **2018**, *27*, 172–181.
- (242) Mosca, R.; Céol, A.; Aloy, P. Interactome3D: adding structural details to protein networks. *Nat. Methods* **2013**, *10*, 47–53.
- (243) Segura, J.; Sanchez-Garcia, R.; Tabas-Madrid, D.; Cuenca-Alba, J.; Sorzano, C. O. S.; Carazo, J. M. 3DIANA: 3D domain interaction analysis: a toolbox for quaternary structure modeling. *Biophys. J.* **2016**, *110*, 766–775.
- (244) Xue, L. C.; Dobbs, D.; Honavar, V. HomPPI: a class of sequence homology based protein-protein interface prediction methods. *BMC Bioinf.* **2011**, *12*, 244.
- (245) Kundrotas, P. J.; Zhu, Z.; Janin, J.; Vakser, I. A. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 9438–9441.
- (246) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381–3385.
- (247) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22*, 195–201.
- (248) Bordoli, L.; Schwede, T. Automated protein structure modeling with SWISS-MODEL Workspace and the Protein Model Portal. *Methods mol. biol.* **2011**, *857*, 107–136.
- (249) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **2014**, *42*, W252–W258.

- (250) Pagadala, N. S.; Syed, K.; Tuszyński, J. Software for molecular docking: a review. *Biophys. Rev.* **2017**, *9*, 91–102.
- (251) Lensink, M. F.; Brysbaert, G.; Nadzirin, N.; Velankar, S.; Chaleil, R. A.; Gerguri, T.; Bates, P. A.; Laine, E.; Carbone, A.; Grudinin, S.; et al. Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins: Struct., Funct., Bioinf.* **2019**, *87*, 1200–1221.
- (252) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898.
- (253) Bitencourt-Ferreira, G.; de Azevedo, W. F. Docking with SwissDock. *Methods Mol. Biol.* **2019**, *2053*, 189–202.
- (254) Souza, P. C.; Limongelli, V.; Wu, S.; Marrink, S. J.; Monticelli, L. Perspectives on high-throughput ligand/protein docking with Martini MD simulations. *Front. Mol. Biosci.* **2021**, *8*, 199.
- (255) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195–2199.
- (256) Chen, R.; Li, L.; Weng, Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins: Struct., Funct., Bioinf.* **2003**, *52*, 80–87.
- (257) Ohue, M.; Matsuzaki, Y.; Uchikoga, N.; Ishida, T.; Akiyama, Y. MEGADOCK: an all-to-all protein-protein interaction prediction system using tertiary structure data. *Protein Pept. Lett.* **2013**, *21*, 766–778.
- (258) Chuang, G.-Y.; Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. DARS (Decoys As the Reference State) potentials for protein-protein docking. *Biophys. J.* **2008**, *95*, 4217–4227.
- (259) Brenke, R.; Hall, D. R.; Chuang, G.-Y.; Comeau, S. R.; Bohnuud, T.; Beglov, D.; Schueler-Furman, O.; Vajda, S.; Kozakov, D. Application of asymmetric statistical potentials to antibody-protein docking. *Bioinformatics* **2012**, *28*, 2608–2614.
- (260) Lu, J.; Hou, X.; Wang, C.; Zhang, Y. Incorporating explicit water molecules and ligand conformation stability in machine-learning scoring functions. *J. Chem. Inf. Model.* **2019**, *59*, 4540–4549.
- (261) Wass, M. N.; Fuentes, G.; Pons, C.; Pazos, F.; Valencia, A. Towards the prediction of protein interaction partners using physical docking. *Mol. Syst. Biol.* **2011**, *7*, 469.
- (262) Desta, I. T.; Porter, K. A.; Xia, B.; Kozakov, D.; Vajda, S. Performance and its limits in rigid body protein-protein docking. *Structure* **2020**, *28*, 1071–1081.
- (263) Pons, C.; Solernou, A.; Perez-Cano, L.; Grosdidier, S.; Fernandez-Recio, J. Optimization of pyDock for the new CAPRI challenges: Docking of homology-based models, domain-domain assembly and protein-RNA binding. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 3182–3188.
- (264) Moal, I. H.; Bates, P. A. SwarmDock and the use of normal modes in protein-protein docking. *Int. J. Mol. Sci.* **2010**, *11*, 3623–3648.
- (265) Kennedy, J.; Eberhart, R. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks* **1995**, *1995*, 1942–1948.
- (266) van Dijk, A. D.; Kaptein, R.; Boelens, R.; Bonvin, A. M. Combining NMR relaxation with chemical shift perturbation data to drive protein-protein docking. *J. Biomol. NMR* **2006**, *34*, 237–244.
- (267) Kaczor, A. A.; Selent, J.; Poso, A. *Methods Cell Biol.* **2013**, *117*, 91–104.
- (268) Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Nat. Prec.* **2010**, *35*, 539–546.
- (269) Betts, M. J.; Sternberg, M. J. An analysis of conformational changes on protein-protein association: implications for predictive docking. *Protein Eng.* **1999**, *12*, 271–283.
- (270) Frisch, C.; Fersht, A. R.; Schreiber, G. Experimental assignment of the structure of the transition state for the association of barnase and barstar. *J. Mol. Biol.* **2001**, *308*, 69–77.
- (271) Tang, C.; Iwahara, J.; Clore, G. M. Visualization of transient encounter complexes in protein-protein association. *Nature* **2006**, *444*, 383–386.
- (272) Schreiber, G.; Fersht, A. R. Energetics of protein-protein interactions: Analysis of the Barnase-Barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **1995**, *248*, 478–486.
- (273) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; et al. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719–1751.
- (274) Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E. Atomic-level characterization of protein-protein association. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 4244–4249.
- (275) Yu, W.; Jo, S.; Lakkaraju, S. K.; Weber, D. J.; MacKerell, A. D., Jr Exploring protein-protein interactions using the site-identification by ligand competitive saturation methodology. *Proteins: Struct., Funct., Bioinf.* **2019**, *87*, 289–301.
- (276) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822.
- (277) Poma, A. B.; Cieplak, M.; Theodorakis, P. E. Combining the MARTINI and structure-based coarse-grained approaches for the molecular dynamics studies of conformational transitions in proteins. *J. Chem. Theory Comput.* **2017**, *13*, 1366–1374.
- (278) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. Coarse-grained model of proteins incorporating atomistic detail of the active site. *Phys. Rev. Lett.* **2005**, *95*, No. 218102.
- (279) Kar, P.; Feig, M. Hybrid all-atom/coarse-grained simulations of proteins by direct coupling of CHARMM and PRIMO force fields. *J. Chem. Theory Comput.* **2017**, *13*, 5753–5765.
- (280) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (281) Pradhan, S.; Rath, R.; Biswas, M. GB1 Dimerization in Crowders: A Multiple Resolution Approach. *J. Chem. Inf. Model.* **2023**, *63*, 1570–1577.
- (282) Mahmood, M. I.; Poma, A. B.; Okazaki, K.-i. Optimizing Gō-MARTINI coarse-grained model for F-BAR protein on lipid membrane. *Front. Mol. Biosci.* **2021**, *8*, No. 619381.
- (283) Kenzaki, H.; Koga, N.; Hori, N.; Kanada, R.; Li, W.; Okazaki, K.-i.; Yao, X.-Q.; Takada, S. CafeMol: A coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comput.* **2011**, *7*, 1979–1989.
- (284) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct., Funct., Bioinf.* **1995**, *21*, 167–195.
- (285) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 17874–17879.
- (286) Tribello, G. A.; Gasparotto, P. Using dimensionality reduction to analyze protein trajectories. *Front. Mol. Biosci.* **2019**, *6*, 46.
- (287) Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2020**, *2*, 200–212.
- (288) Park, S.; Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **2004**, *120*, 5946–5961.
- (289) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (290) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116*, 9058–9067.
- (291) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (292) Kirkpatrick, S.; Gelatt, C. D., Jr; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.
- (293) Marinari, E.; Parisi, G. Simulated tempering: a new Monte Carlo scheme. *Europhys. Lett.* **1992**, *19*, 451.
- (294) Qi, R.; Wei, G.; Ma, B.; Nussinov, R. Replica exchange molecular dynamics: a practical application protocol with solutions to

- common problems and a peptide aggregation and self-assembly example. *Methods Mol. Biol.* **2018**, *1777*, 101–119.
- (295) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.
- (296) Wang, L.; Friesner, R. A.; Berne, B. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431–9438.
- (297) 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.
- (298) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (299) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
- (300) Diem, M.; Oostenbrink, C. Hamiltonian reweighing To refine protein backbone dihedral angle parameters in the GROMOS force field. *J. Chem. Inf. Model.* **2020**, *60*, 279–288.
- (301) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
- (302) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950–1958.
- (303) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building water models: a different approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.
- (304) Shabane, P. S.; Izadi, S.; Onufriev, A. V. General purpose water model can improve atomistic simulations of intrinsically disordered proteins. *J. Chem. Theory Comput.* **2019**, *15*, 2620–2634.
- (305) Best, R. B.; de Sancho, D.; Mittal, J. Residue-specific α -helix propensities from molecular simulation. *Biophys. J.* **2012**, *102*, 1462–1467.
- (306) Tian, C.; Kasavajhala, K.; Belfon, K. A.; Raguette, L.; Huang, H.; Miguez, 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.
- (307) MacKerell, A. D., Jr; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (308) Miettinen, M. S.; Monticelli, L.; Nedumpully-Govindan, P.; Knecht, V.; Ignatova, Z. Stable polyglutamine dimers can contain β -hairpins with interdigitated side chains—but not α -helices, β -nanotubes, β -pseudohelices, or steric zippers. *Biophys. J.* **2014**, *106*, 1721–1728.
- (309) 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.
- (310) Abriata, L. A.; Dal Peraro, M. Assessment of transferable forcefields for protein simulations attests improved description of disordered states and secondary structure propensities, and hints at multi-protein systems as the next challenge for optimization. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2626–2636.
- (311) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (312) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (313) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved peptide and protein torsional energetics with the OPLS-AA force field. *J. Chem. Theory Comput.* **2015**, *11*, 3499–3509.
- (314) Jost Lopez, A.; Quoika, P. K.; Linke, M.; Hummer, G.; Koofinger, J. Quantifying protein–protein interactions in molecular simulations. *J. Phys. Chem. B* **2020**, *124*, 4673–4685.
- (315) Wang, J.; Ishchenko, A.; Zhang, W.; Razavi, A.; Langley, D. A highly accurate metadynamics-based Dissociation Free Energy method to calculate protein–protein and protein–ligand binding potencies. *Sci. Rep.* **2022**, *12*, 2024.
- (316) Kasttritis, P. L.; Bonvin, A. M. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J. R. Soc., Interface* **2013**, *10*, 20120835.
- (317) Leavitt, S.; Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **2001**, *11*, 560–566.
- (318) Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990**, *29*, 7133–7155.
- (319) Hilser, V. J.; García-Moreno, E. B.; Oas, T. G.; Kapp, G.; Whitten, S. T. A statistical thermodynamic model of the protein ensemble. *Chem. Rev.* **2006**, *106*, 1545–1558.
- (320) Wand, A. J.; Sharp, K. A. Measuring entropy in molecular recognition by proteins. *Annu. Rev. Biophys.* **2018**, *47*, 41–61.
- (321) Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
- (322) Caro, J. A.; Harpole, K. W.; Kasinath, V.; Lim, J.; Granja, J.; Valentine, K. G.; Sharp, K. A.; Wand, A. J. Entropy in molecular recognition by proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 6563–6568.
- (323) Fleck, M.; Polyansky, A. A.; Zagrovic, B. Self-Consistent Framework Connecting Experimental Proxies of Protein Dynamics with Configurational Entropy. *J. Chem. Theory Comput.* **2018**, *14*, 3796–3810.
- (324) Muegge, I. PMF scoring revisited. *J. Med. Chem.* **2006**, *49*, 5895–5902.
- (325) Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf. *J. Comput. Chem.* **2004**, *25*, 238–250.
- (326) Duan, L.; Liu, X.; Zhang, J. Z. Interaction entropy: A new paradigm for highly efficient and reliable computation of protein–ligand binding free energy. *J. Am. Chem. Soc.* **2016**, *138*, 5722–5728.
- (327) Killian, B. J.; Yundenfreund Kravitz, J.; Gilson, M. K. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Phys.* **2007**, *127*, No. 024107.
- (328) King, B. M.; Silver, N. W.; Tidor, B. Efficient calculation of molecular configurational entropies using an information theoretic approximation. *J. Phys. Chem. B* **2012**, *116*, 2891–2904.
- (329) Chen, J.; Sawyer, N.; Regan, L. Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci.* **2013**, *22*, 510–515.
- (330) Kasttritis, P. L.; Moal, I. H.; Hwang, H.; Weng, Z.; Bates, P. A.; Bonvin, A. M.; Janin, J. A structure-based benchmark for protein–protein binding affinity. *Protein Sci.* **2011**, *20*, 482–491.
- (331) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. A dissection of specific and non-specific protein–protein interfaces. *J. Mol. Biol.* **2004**, *336*, 943–955.
- (332) Mintseris, J.; Pierce, B.; Wiehe, K.; Anderson, R.; Chen, R.; Weng, Z. Integrating statistical pair potentials into protein complex prediction. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 511–520.
- (333) Kasttritis, P. L.; Rodrigues, J. P.; Folkers, G. E.; Boelens, R.; Bonvin, A. M. Proteins feel more than they see: fine-tuning of binding affinity by properties of the non-interacting surface. *J. Mol. Biol.* **2014**, *426*, 2632–2652.
- (334) Schindler, C. E.; de Vries, S. J.; Zacharias, M. iATTRACT: Simultaneous global and local interface optimization for protein–

- protein docking refinement. *Proteins: Struct., Funct., Bioinf.* **2015**, *83*, 248–258.
- (335) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048.
- (336) De Vries, S. J.; Van Dijk, A. D.; Krzeminski, M.; van Dijk, M.; Thureau, A.; Hsu, V.; Wassenaar, T.; Bonvin, A. M. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 726–733.
- (337) Chong, S.-H.; Ham, S. Dynamics of Hydration Water Plays a Key Role in Determining the Binding Thermodynamics of Protein Complexes. *Sci. Rep.* **2017**, *7*, 8744.
- (338) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
- (339) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897.
- (340) Swanson, J. M.; Henchman, R. H.; McCammon, J. A. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.* **2004**, *86*, 67–74.
- (341) Chen, F.; Liu, H.; Sun, H.; Pan, P.; Li, Y.; Li, D.; Hou, T. Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking. *Phys. Chem. Chem. Phys.* **2016**, *18*, 22129–22139.
- (342) Ulucan, O.; Jaitly, T.; Helms, V. Energetics of hydrophilic protein–protein association and the role of water. *J. Chem. Theory Comput.* **2014**, *10*, 3512–3524.
- (343) Maffucci, I.; Contini, A. Improved computation of protein–protein relative binding energies with the Nwat-MMGBSA method. *J. Chem. Inf. Model.* **2016**, *56*, 1692–1704.
- (344) Sun, H.; Duan, L.; Chen, F.; Liu, H.; Wang, Z.; Pan, P.; Zhu, F.; Zhang, J. Z.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* **2018**, *20*, 14450–14460.
- (345) Sun, Z.; Yan, Y. N.; Yang, M.; Zhang, J. Z. Interaction entropy for protein–protein binding. *J. Chem. Phys.* **2017**, *146*, 124124.
- (346) Huang, K.; Luo, S.; Cong, Y.; Zhong, S.; Zhang, J. Z.; Duan, L. An accurate free energy estimator: based on MM/PBSA combined with interaction entropy for protein–ligand binding affinity. *Nanoscale* **2020**, *12*, 10737–10750.
- (347) Kohut, G.; Liwo, A.; Bosze, S.; Beke-Somfai, T.; Samsonov, S. A. Protein–ligand interaction energy-based entropy calculations: Fundamental challenges for flexible systems. *J. Phys. Chem. B* **2018**, *122*, 7821–7827.
- (348) Prompers, J. J.; Brüschweiler, R. Reorientational eigenmode dynamics: a combined MD/NMR relaxation analysis method for flexible parts in globular proteins. *J. Am. Chem. Soc.* **2001**, *123*, 7305–7313.
- (349) Chang, C.-E.; Chen, W.; Gilson, M. K. Evaluating the accuracy of the quasiharmonic approximation. *J. Chem. Theory Comput.* **2005**, *1*, 1017–1028.
- (350) King, B. M.; Tidor, B. MIST: Maximum Information Spanning Trees for dimension reduction of biological data sets. *Bioinformatics* **2009**, *25*, 1165–1172.
- (351) Moal, I. H.; Agius, R.; Bates, P. A. Protein–protein binding affinity prediction on a diverse set of structures. *Bioinformatics* **2011**, *27*, 3002–3009.
- (352) Jankauskaitė, J.; Jiménez-García, B.; Dapkūnas, J.; Fernández-Rrecio, J.; Moal, I. H. SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics* **2019**, *35*, 462–469.
- (353) Tinnefeld, P.; Sauer, M. Branching out of single-molecule fluorescence spectroscopy: Challenges for chemistry and influence on biology. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 2642–2671.
- (354) Weiss, S. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.* **2000**, *7*, 724–729.
- (355) Nguyen, D.; Yan, G.; Chen, T.-Y.; Do, L. H. Variations in Intracellular Organometallic Reaction Frequency Captured by Single-Molecule Fluorescence Microscopy. *Angew. Chem., Int. Ed. Engl.* **2023**, *62*, e202300467.
- (356) Milles, S.; Salvi, N.; Blackledge, M.; Jensen, M. R. Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments. *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *109*, 79–100.
- (357) Klukowski, P.; Riek, R.; Güntert, P. Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA. *Nat. Commun.* **2022**, *13*, 6151.
- (358) Souza, P. C.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. Protein–ligand binding with the coarse-grained Martini model. *Nat. Commun.* **2020**, *11*, 3714.
- (359) Predeus, A. V.; Gul, S.; Gopal, S. M.; Feig, M. Conformational sampling of peptides in the presence of protein crowders from AA/CG-multiscale simulations. *J. Phys. Chem. B* **2012**, *116*, 8610–8620.
- (360) Taylor, M. P.; Vinci, C.; Suzuki, R. Effects of macromolecular crowding on the folding of a polymer chain: A Wang–Landau simulation study. *J. Chem. Phys.* **2020**, *153*, 174901.
- (361) Timr, S.; Melchionna, S.; Derreumaux, P.; Sterpone, F. Optimized OPEP Force Field for Simulation of Crowded Protein Solutions. *J. Phys. Chem. B* **2023**, *127*, 3616–3623.
- (362) Minh, D. D.; Chang, C.-e.; Trylska, J.; Tozzini, V.; McCammon, J. A. The influence of macromolecular crowding on HIV-1 protease internal dynamics. *J. Am. Chem. Soc.* **2006**, *128*, 6006–6007.
- (363) Stagg, L.; Zhang, S.-Q.; Cheung, M. S.; Wittung-Stafshede, P. Molecular crowding enhances native structure and stability of α/β protein flavodoxin. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18976–18981.
- (364) Długosz, M.; Zieliński, P.; Trylska, J. Brownian dynamics simulations on CPU and GPU with BD_BOX. *J. Comput. Chem.* **2011**, *32*, 2734–2744.
- (365) Tsao, D.; Dokholyan, N. V. Macromolecular crowding induces polypeptide compaction and decreases folding cooperativity. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3491–3500.
- (366) Rickard, M.; Zhang, Y.; Pogorelov, T.; Gruebele, M. Crowding, sticking, and partial folding of GTT WW domain in a small cytoplasm model. *J. Phys. Chem. B* **2020**, *124*, 4732–4740.
- (367) Feig, M.; Yu, I.; Wang, P.-h.; Nawrocki, G.; Sugita, Y. Crowding in cellular environments at an atomistic level from computer simulations. *J. Phys. Chem. B* **2017**, *121*, 8009–8025.
- (368) Qin, S.; Minh, D. D.; McCammon, J. A.; Zhou, H.-X. Method to predict crowding effects by postprocessing molecular dynamics trajectories: application to the flap dynamics of HIV-1 protease. *J. Phys. Chem. Lett.* **2010**, *1*, 107–110.
- (369) Qin, S.; Zhou, H.-X. Atomistic modeling of macromolecular crowding predicts modest increases in protein folding and binding stability. *Biophys. J.* **2009**, *97*, 12–19.
- (370) Feig, M.; Sugita, Y. Whole-cell models and simulations in molecular detail. *Annu. Rev. Cell Dev. Biol.* **2019**, *35*, 191–211.
- (371) Petrov, D.; Zagrovic, B. Are current atomistic force fields accurate enough to study proteins in crowded environments? *PLoS Comput. Biol.* **2014**, *10*, e1003638.
- (372) Rickard, M. M.; Zhang, Y.; Gruebele, M.; Pogorelov, T. V. In-cell protein–protein contacts: Transient interactions in the crowd. *J. Phys. Chem. Lett.* **2019**, *10*, 5667–5673.

- (373) Nawrocki, G.; Wang, P.-h.; Yu, I.; Sugita, Y.; Feig, M. Slow-down in diffusion in crowded protein solutions correlates with transient cluster formation. *J. Phys. Chem. B* **2017**, *121*, 11072–11084.
- (374) Nawrocki, G.; Karaboga, A.; Sugita, Y.; Feig, M. Effect of protein–protein interactions and solvent viscosity on the rotational diffusion of proteins in crowded environments. *Phys. Chem. Chem. Phys.* **2019**, *21*, 876–883.
- (375) Best, R. B.; Zheng, W.; Mittal, J. Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.
- (376) Abriata, L. A.; Dal Peraro, M. Assessing the potential of atomistic molecular dynamics simulations to probe reversible protein–protein recognition and binding. *Sci. Rep.* **2015**, *5*, 10549.
- (377) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; Van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.
- (378) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **2011**, *100*, L47–L49.
- (379) Yoo, J.; Aksimentiev, A. New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 8432–8449.
- (380) Janin, J.; Rodier, F. Protein–protein interaction at crystal contacts. *Proteins: Struct., Funct., Bioinf.* **1995**, *23*, 580–587.
- (381) Andrews, C. T.; Elcock, A. H. Molecular dynamics simulations of highly crowded amino acid solutions: comparisons of eight different force field combinations with experiment and with each other. *J. Chem. Theory Comput.* **2013**, *9*, 4585–4602.
- (382) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (383) Berendsen, H.; Grigera, J.; Straatsma, T. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (384) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
- (385) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- (386) Berendsen, H.; Postma, J.; Van Gunsteren, W.; Hermans, a. J. Intermolecular forces. *Reidel Dordrecht* **1981**, *14*, 331–342.
- (387) Mackerell, A. D., Jr; Feig, M.; Brooks, C. L., III Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (388) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000**, *112*, 8910–8922.
- (389) von Bülow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 9843–9852.
- (390) Kasahara, K.; Re, S.; Nawrocki, G.; Oshima, H.; Mishima-Tsumagari, C.; Miyata-Yabuki, Y.; Kukimoto-Niino, M.; Yu, I.; Shirouzu, M.; Feig, M.; et al. Reduced efficacy of a Src kinase inhibitor in crowded protein solution. *Nat. Commun.* **2021**, *12*, 4099.
- (391) Shaw, D. E.; Grossman, J.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; et al. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* **2014**, 41–53.
- (392) Timr, S.; Gnutt, D.; Ebbinghaus, S.; Sterpone, F. The unfolding journey of superoxide dismutase 1 barrels under crowding: atomistic simulations shed light on intermediate states and their interactions with Crowders. *J. Phys. Chem. Lett.* **2020**, *11*, 4206–4212.
- (393) Bille, A.; Linse, B.; Mohanty, S.; Irbäck, A. Equilibrium simulation of trp-cage in the presence of protein crowders. *J. Chem. Phys.* **2015**, *143*, 175102.
- (394) Ando, T.; Yu, I.; Feig, M.; Sugita, Y. Thermodynamics of macromolecular association in heterogeneous crowding environments: theoretical and simulation studies with a simplified model. *J. Phys. Chem. B* **2016**, *120*, 11856–11865.
- (395) Chu, X.; Suo, Z.; Wang, J. Confinement and crowding effects on folding of a multidomain Y-family DNA polymerase. *J. Chem. Theory Comput.* **2020**, *16*, 1319–1332.
- (396) Dhar, A.; Samiotakis, A.; Ebbinghaus, S.; Nienhaus, L.; Homouz, D.; Gruebele, M.; Cheung, M. S. Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 17586–17591.
- (397) Skora, T.; Vaghefikia, F.; Fitter, J.; Kondrat, S. Macromolecular crowding: how shape and interactions affect diffusion. *J. Phys. Chem. B* **2020**, *124*, 7537–7543.
- (398) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (399) Irbäck, A.; Mohanty, S. PROFASI: a Monte Carlo simulation package for protein folding and aggregation. *J. Comput. Chem.* **2006**, *27*, 1548–1555.
- (400) Bernaschi, M.; Melchionna, S.; Succi, S.; Fyta, M.; Kaxiras, E.; Sircar, J. K. MUPHY: A parallel Multi PHYsics/scale code for high performance bio-fluidic simulations. *Comput. Phys. Commun.* **2009**, *180*, 1495–1502.
- (401) Hoffmann, M.; Fröhner, C.; Noé, F. ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics. *PLoS Comput. Biol.* **2019**, *15*, e1006830.
- (402) Gellman, S. H.; Woolfson, D. N. Mini-proteins Trp the light fantastic. *Nat. Struct. Biol.* **2002**, *9*, 408–410.
- (403) Zhou, R. Trp-cage: folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.
- (404) Appadurai, R.; Nagesh, J.; Srivastava, A. High resolution ensemble description of metamorphic and intrinsically disordered proteins using an efficient hybrid parallel tempering scheme. *Nat. Commun.* **2021**, *12*, 958.
- (405) Berendsen, H. J.; Hayward, S. Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165–169.
- (406) Hensen, U.; Meyer, T.; Haas, J.; Rex, R.; Vriend, G.; Grubmüller, H. Exploring protein dynamics space: the dynasome as the missing link between protein structure and function. *PLoS One* **2012**, *7*, e33931.
- (407) Schultze, S.; Grubmüller, H. Time-lagged independent component analysis of random walks and protein dynamics. *J. Chem. Theory Comput.* **2021**, *17*, 5766–5776.
- (408) Yang, W. Y.; Gruebele, M. Folding at the speed limit. *Nature* **2003**, *423*, 193–197.
- (409) Freddolino, P. L.; Park, S.; Roux, B.; Schulten, K. Force field bias in protein folding simulations. *Biophys. J.* **2009**, *96*, 3772–3780.
- (410) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3626–3630.
- (411) Danielsson, J.; Mu, X.; Lang, L.; Wang, H.; Binolfi, A.; Theillet, F.-X.; Bekei, B.; Logan, D. T.; Selenko, P.; Wennerström, H.; et al. Thermodynamics of protein destabilization in live cells. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 12402–12407.
- (412) Hong, J.; Gierasch, L. M. Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state. *J. Am. Chem. Soc.* **2010**, *132*, 10445–10452.
- (413) van den Berg, B.; Wain, R.; Dobson, C. M.; Ellis, R. J. Macromolecular crowding perturbs protein refolding kinetics:

implications for folding inside the cell. *EMBO journal* **2000**, *19*, 3870–3875.

(414) Benton, L. A.; Smith, A. E.; Young, G. B.; Pielak, G. J. Unexpected effects of macromolecular crowding on protein stability. *Biochemistry* **2012**, *51*, 9773–9775.

(415) Bille, A.; Jensen, K. S.; Mohanty, S.; Akke, M.; Irbäck, A. Stability and local unfolding of SOD1 in the presence of protein crowders. *J. Phys. Chem. B* **2019**, *123*, 1920–1930.

(416) Bille, A.; Mohanty, S.; Irbäck, A. Peptide folding in the presence of interacting protein crowders. *J. Chem. Phys.* **2016**, *144*, 175105.

(417) Harada, R.; Sugita, Y.; Feig, M. Protein crowding affects hydration structure and dynamics. *J. Am. Chem. Soc.* **2012**, *134*, 4842–4849.

(418) Miklos, A. C.; Sarkar, M.; Wang, Y.; Pielak, G. J. Protein crowding tunes protein stability. *J. Am. Chem. Soc.* **2011**, *133*, 7116–7120.

(419) Schlesinger, A. P.; Wang, Y.; Tadeo, X.; Millet, O.; Pielak, G. J. Macromolecular crowding fails to fold a globular protein in cells. *J. Am. Chem. Soc.* **2011**, *133*, 8082–8085.

(420) Zegarra, F. C.; Homouz, D.; Gasic, A. G.; Babel, L.; Kovermann, M.; Wittung-Stafshede, P.; Cheung, M. S. Crowding-induced elongated conformation of urea-unfolded apoazurin: investigating the role of crowder shape in silico. *J. Phys. Chem. B* **2019**, *123*, 3607–3617.

(421) Bonucci, A.; Palomino-Schätzlein, M.; Malo de Molina, P.; Arbe, A.; Pierattelli, R.; Rizzuti, B.; Iovanna, J. L.; Neira, J. L. Crowding effects on the structure and dynamics of the intrinsically disordered nuclear chromatin protein NUPR1. *Front. Mol. Biosci.* **2021**, *8*, 684622.

(422) Bo, L.; Milanetti, E.; Chen, C. G.; Ruocco, G.; Amadei, A.; D'Abramo, M. Computational Modeling of the Thermodynamics of the Mesophilic and Thermophilic Mutants of Trp-Cage Mini-protein. *ACS Omega* **2022**, *7*, 13448–13454.

(423) Byrne, A.; Williams, D. V.; Barua, B.; Hagen, S. J.; Kier, B. L.; Andersen, N. H. Folding dynamics and pathways of the trp-cage mini-proteins. *Biochemistry* **2014**, *53*, 6011–6021.

(424) D'Abramo, M.; Del Galdo, S.; Amadei, A. Theoretical-computational modelling of the temperature dependence of the folding-unfolding thermodynamics and kinetics: The case of a Trp-cage. *Phys. Chem. Chem. Phys.* **2019**, *21*, 23162–23168.

(425) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(426) Gopal, S. M.; Mukherjee, S.; Cheng, Y.-M.; Feig, M. PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1266–1281.

(427) Lebowitz, J. L.; Helfand, E.; Praestgaard, E. Scaled particle theory of fluid mixtures. *J. Chem. Phys.* **1965**, *43*, 774–779.

(428) Reiss, H.; Frisch, H. L.; Lebowitz, J. L. Statistical Mechanics of Rigid Spheres. *J. Chem. Phys.* **1959**, *31*, 369–380.

(429) O'Keefe, S. P.; Sanchez, I. C. Scaled Particle Theory of solutions: Comparison with Lattice Fluid model. *Fluid Phase Equilib.* **2017**, *433*, 67–78.

(430) Minton, A. P. Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: macromolecular crowding and protein stability revisited. *Biophys. J.* **2005**, *88*, 971–985.

(431) Mittal, J.; Best, R. B. Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys. J.* **2010**, *98*, 315–320.

(432) Binder, K. *Monte Carlo and molecular dynamics simulations in polymer science*; Oxford University Press, 1995.

(433) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.

(434) Wang, F.; Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **2001**, *86*, 2050.

(435) Wang, F.; Landau, D. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E* **2001**, *64*, No. 056101.

(436) Taylor, M. P. Polymer folding in slitlike nanoconfinement. *Macromolecules* **2017**, *50*, 6967–6976.

(437) Taylor, M. P.; Prunty, T. M.; O'Neil, C. M. All-or-none folding of a flexible polymer chain in cylindrical nanoconfinement. *J. Chem. Phys.* **2020**, *152*, 094901.

(438) Taylor, M. P. Note: Rigorous results for the partition function of a square-well chain in hard-sphere solvent. *J. Chem. Phys.* **2017**, *147*, 166101.

(439) Qu, Y.; Bolen, D. Efficacy of macromolecular crowding in forcing proteins to fold. *Biophys. Chem.* **2002**, *101*, 155–165.

(440) Ai, X.; Zhou, Z.; Bai, Y.; Choy, W.-Y. 15N NMR spin relaxation dispersion study of the molecular crowding effects on protein folding under native conditions. *J. Am. Chem. Soc.* **2006**, *128*, 3916–3917.

(441) Batra, J.; Xu, K.; Qin, S.; Zhou, H.-X. Effect of macromolecular crowding on protein binding stability: modest stabilization and significant biological consequences. *Biophys. J.* **2009**, *97*, 906–911.

(442) Soranno, A.; Koenig, I.; Borgia, M. B.; Hofmann, H.; Zosel, F.; Nettels, D.; Schuler, B. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 4874–4879.

(443) Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.; Futaki, S.; Ito, Y.; Hiroaki, H.; et al. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **2009**, *458*, 106–109.

(444) Ignatova, Z.; Krishnan, B.; Bombardier, J. P.; Marcelino, A. M. C.; Hong, J.; Gierasch, L. M. From the test tube to the cell: Exploring the folding and aggregation of a β -clam protein. *Biopolymers* **2007**, *88*, 157–163.

(445) Schuler, L. D.; Daura, X.; Van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **2001**, *22*, 1205–1218.

(446) Chen, J.; Im, W.; Brooks, C. L. Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.

(447) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(448) Martini, S.; Bonechi, C.; Foletti, A.; Rossi, C. Water-protein interactions: The secret of protein dynamics. *Sci. World J.* **2013**, *2013*, 138916.

(449) Miotto, M.; Di Rienzo, L.; Corsi, P.; Ruocco, G.; Raimondo, D.; Milanetti, E. Simulated Epidemics in 3D Protein Structures to Detect Functional Properties. *J. Chem. Inf. Model.* **2020**, *60*, 1884–1891.

(450) Miotto, M.; Armaos, A.; Di Rienzo, L.; Ruocco, G.; Milanetti, E.; Tartaglia, G. G. Thermometer: a webserver to predict protein thermal stability. *Bioinformatics* **2022**, *38*, 2060.

(451) Stirnemann, G.; Sterpone, F. Recovering protein thermal stability using all-atom Hamiltonian replica-exchange simulations in explicit solvent. *J. Chem. Theory Comput.* **2015**, *11*, 5573–5577.

(452) Sheng, Y.; Chattopadhyay, M.; Whitelegge, J.; Valentine, J. S. SOD1 Aggregation and ALS: Role of Metalloion States and Disulfide Status. *Curr. Trends Med. Chem.* **2013**, *12*, 2560–2572.

(453) Kalia, M.; et al. Molecular dynamics analysis of Superoxide Dismutase 1 mutations suggests decoupling between mechanisms underlying ALS onset and progression. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 5296–5308.

(454) Mereghetti, P.; Wade, R. C. Brownian dynamics simulation of protein diffusion in crowded environments. *AIP Conf. Proc.* **2013**, *1518*, 511–516.

- (455) Jin, S.; Verkman, A. Single particle tracking of complex diffusion in membranes: simulation and detection of barrier, raft, and interaction phenomena. *J. Phys. Chem. B* **2007**, *111*, 3625–3632.
- (456) Verkman, A. S. Diffusion in cells measured by fluorescence recovery after photobleaching. In *Methods in Enzymology*, Vol. 360; Elsevier, 2003; pp 635–648.
- (457) Bacia, K.; Kim, S. A.; Schwille, P. Fluorescence cross-correlation spectroscopy in living cells. *Nat. Methods* **2006**, *3*, 83–89.
- (458) Prindle, J. R.; de Cuba, O. I. C.; Gahlmann, A. Single-molecule tracking to determine the abundances and stoichiometries of freely-diffusing protein complexes in living cells: Past applications and future prospects. *J. Chem. Phys.* **2023**, *159*, No. 071002.
- (459) Cai, N.; Lai, A. C.-K.; Liao, K.; Corridon, P. R.; Graves, D. J.; Chan, V. Recent advances in fluorescence recovery after photobleaching for decoupling transport and kinetics of biomacromolecules in cellular physiology. *Polymers* **2022**, *14*, 1913.
- (460) Yu, L.; Lei, Y.; Ma, Y.; Liu, M.; Zheng, J.; Dan, D.; Gao, P. A comprehensive review of fluorescence correlation spectroscopy. *Front. Phys.* **2021**, *9*, 644450.
- (461) Han, J.; Herzfeld, J. Macromolecular diffusion in crowded solutions. *Biophys. J.* **1993**, *65*, 1155–1161.
- (462) Kim, J. S.; Yethiraj, A. Effects of macromolecular crowding on reaction rates: a computational and theoretical study. *Biophys. J.* **2009**, *96*, 1333–1340.
- (463) Kim, Y. C.; Best, R. B.; Mittal, J. Molecular crowding effects on protein-protein binding affinity and specificity. *J. Chem. Phys.* **2010**, *133*, 205101.
- (464) Nawrocki, G.; Im, W.; Sugita, Y.; Feig, M. Clustering and dynamics of crowded proteins near membranes and their influence on membrane bending. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 24562–24567.
- (465) Roosen-Runge, F.; Hennig, M.; Zhang, F.; Jacobs, R. M.; Sztucki, M.; Schober, H.; Seydel, T.; Schreiber, F. Protein self-diffusion in crowded solutions. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11815–11820.
- (466) Li, C.; Wang, Y.; Pielak, G. J. Translational and rotational diffusion of a small globular protein under crowded conditions. *J. Phys. Chem. B* **2009**, *113*, 13390–13392.
- (467) Ridgway, D.; Broderick, G.; Lopez-Campistrous, A.; Ru'aini, M.; Winter, P.; Hamilton, M.; Boulanger, P.; Kovalenko, A.; Ellison, M. J. Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. *Biophys. J.* **2008**, *94*, 3748–3759.
- (468) Roberts, E.; Stone, J. E.; Sepúlveda, L.; Hwu, W.-M. W.; Luthy-Schulten, Z. Long time-scale simulations of in vivo diffusion using GPU hardware. *IEEE International Symposium on Parallel & Distributed Processing* **2009**, 2009, 1–8.
- (469) Zegarra, F. C.; Homouz, D.; Eliaz, Y.; Gasic, A. G.; Cheung, M. S. Impact of hydrodynamic interactions on protein folding rates depends on temperature. *Phys. Rev. E* **2018**, *97*, No. 032402.
- (470) Grimaldo, M.; Lopez, H.; Beck, C.; Roosen-Runge, F.; Moulin, M.; Devos, J. M.; Laux, V.; Hartlein, M.; Da Vela, S.; Schweins, R.; et al. Protein short-time diffusion in a naturally crowded environment. *J. Phys. Chem. Lett.* **2019**, *10*, 1709–1715.
- (471) Skolnick, J. Perspective: On the importance of hydrodynamic interactions in the subcellular dynamics of macromolecules. *J. Chem. Phys.* **2016**, *145*, 100901.
- (472) Mu, X.; Choi, S.; Lang, L.; Mowray, D.; Dokholyan, N. V.; Danielsson, J.; Oliveberg, M. Physicochemical code for quinary protein interactions in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E4556–E4563.
- (473) Muramatsu, N.; Minton, A. P. Tracer diffusion of globular proteins in concentrated protein solutions. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2984–2988.
- (474) Smoluchowski, M. V. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Z. Phys. Chemie* **1918**, *92*, 129–168.
- (475) Van Zon, J. S.; Ten Wolde, P. R. Simulating biochemical networks at the particle level in time and space: Green's Function Reaction Dynamics. *Phys. Rev. Lett.* **2005**, *94*, 128103.
- (476) Sokolowski, T. R.; Pajmians, J.; Bossen, L.; Miedema, T.; Wehrens, M.; Becker, N. B.; Kaizu, K.; Takahashi, K.; Dogterom, M.; Ten Wolde, P. R. eGFRD in all dimensions. *J. Chem. Phys.* **2019**, *150*, No. 054108.
- (477) Vijaykumar, A.; ten Wolde, P. R.; Bolhuis, P. G. Generalised expressions for the association and dissociation rate constants of molecules with multiple binding sites. *Mol. Phys.* **2018**, *116*, 3042–3054.
- (478) Kim, J. S.; Yethiraj, A. Crowding Effects on Protein Association: Effect of Interactions between Crowding Agents. *J. Phys. Chem. B* **2011**, *115*, 347–353.
- (479) Meinecke, L. Multiscale Modeling of Diffusion in a Crowded Environment. *Bull. Math. Biol.* **2017**, *79*, 2672–2695.
- (480) Van Zon, J. S.; Ten Wolde, P. R. Green's Function Reaction Dynamics: A particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys.* **2005**, *123*, 234910.
- (481) Dlugosz, M.; Trylska, J. Diffusion in crowded biological environments: applications of Brownian dynamics. *BMC biophysics* **2011**, *4*, 3.
- (482) Blanco, P. M.; Garcés, J. L.; Madurga, S.; Mas, F. Macromolecular diffusion in crowded media beyond the hard-sphere model. *Soft Matter* **2018**, *14*, 3105–3114.
- (483) Wiczorek, G.; Zielenkiewicz, P. Influence of macromolecular crowding on protein-protein association rates—a Brownian dynamics study. *Biophys. J.* **2008**, *95*, 5030–5036.
- (484) Sterpone, F.; Derreumaux, P.; Melchionna, S. Protein simulations in fluids: Coupling the OPEP coarse-grained force field with hydrodynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1843–1853.
- (485) Nenninger, A.; Mastroianni, G.; Mullineaux, C. W. Size dependence of protein diffusion in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* **2010**, *192*, 4535–4540.
- (486) Stadmler, S. S.; Aguilar, J. S.; Parnham, S.; Pielak, G. J. Protein-peptide binding energetics under crowded conditions. *J. Phys. Chem. B* **2020**, *124*, 9297–9309.
- (487) Wirth, A. J.; Gruebele, M. Quinary protein structure and the consequences of crowding in living cells: Leaving the test-tube behind. *BioEssays* **2013**, *35*, 984–993.
- (488) Morelli, M. J.; Allen, R. J.; ten Wolde, P. R. Effects of macromolecular crowding on genetic networks. *Biophys. J.* **2011**, *101*, 2882–2891.
- (489) Gomez, D.; Klumpp, S. Biochemical reactions in crowded environments: Revisiting the effects of volume exclusion with simulations. *Front. Phys.* **2015**, *3*, 45.
- (490) Zhou, H.-X. Protein folding and binding in confined spaces and in crowded solutions. *J. Mol. Recognit.* **2004**, *17*, 368–375.
- (491) Souza, P. C.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M.; Wassenaar, T. A.; et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382–388.
- (492) Phillip, Y.; Sherman, E.; Haran, G.; Schreiber, G. Common crowding agents have only a small effect on protein-protein interactions. *Biophys. J.* **2009**, *97*, 875–885.
- (493) Berg, O. G. The influence of macromolecular crowding on thermodynamic activity: Solubility and dimerization constants for spherical and dumbbell-shaped molecules in a hard-sphere mixture. *Biomolecules: Original Research on Biomolecules* **1990**, *30*, 1027–1037.