Received: 29 February 2024 | Revised: 6 May 2024 | Accepted: 10 June 2024

DOI: 10.1002/pro.5090

TOOLS FOR PROTEIN SCIENCE

THE PROTEIN SOCIETY WILEY

Exploring protein functions from structural flexibility using CABS-flex modeling

Chandran Nithin[1] | **Rocco Peter Fornari**[1] | **Smita P. Pilla**[1] | **Karol Wroblewski**[1] | **Mateusz Zalewski**[1] | **Rafał Madaj**[2] | **Andrzej Kolinski**[1] | **Joanna M. Macnar**[1] | **Sebastian Kmiecik**[1]

[1]Biological and Chemical Research Centre, Faculty of Chemistry, University of Warsaw, Warsaw, Poland

[2]Institute of Evolutionary Biology, Biological and Chemical Research Centre, Faculty of Biology, University of Warsaw, Warsaw, Poland

Correspondence
Sebastian Kmiecik, Biological and Chemical Research Center, Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland.
Email: sekmi@chem.uw.edu.pl

Present address
Joanna M. Macnar, Ryvu Therapeutics, Cracow, Poland.

Funding information
National Science Centre, Poland, Grant/Award Numbers: OPUS 2020/39/B/NZ2/01301, 2021/40/Q/NZ2/00078; PLGrid, Grant/Award Number: PLG/2024/016931

Review Editor: Nir Ben-Tal

Abstract

Understanding protein function often necessitates characterizing the flexibility of protein structures. However, simulating protein flexibility poses significant challenges due to the complex dynamics of protein systems, requiring extensive computational resources and accurate modeling techniques. In response to these challenges, the CABS-flex method has been developed as an efficient modeling tool that combines coarse-grained simulations with all-atom detail. Available both as a web server and a standalone package, CABS-flex is dedicated to a wide range of users. The web server version offers an accessible interface for straightforward tasks, while the standalone command-line program is designed for advanced users, providing additional features, analytical tools, and support for handling large systems. This paper examines the application of CABS-flex across various structure–function studies, facilitating investigations into the interplay among protein structure, dynamics, and function in diverse research fields. We present an overview of the current status of the CABS-flex methodology, highlighting its recent advancements, practical applications, and forthcoming challenges.

KEYWORDS

coarse-grained model, integrative modeling, molecular modeling, Monte Carlo simulation, multiscale simulation, protein aggregation, protein flexibility

1 | INTRODUCTION

The flexibility of a protein's structure can be pivotal to its biological function. Unfortunately, experimental investigation of protein dynamics is often challenging or unfeasible. Consequently, computer simulations play an important role here; however, the computational cost restricts the application of classical simulations like all-atom molecular dynamics (MD) to small systems and short timescales (Kmiecik et al., 2016). In practice, for most biologically relevant proteins, classical simulations of structural flexibility demand significant computational resources, such as high-performance computer clusters.

In the age of AlphaFold, the field of structural biology has witnessed a profound transformation in our ability to predict protein structures with remarkable accuracy

(Varadi et al., 2022). However, the importance of dynamic and flexibility predictions cannot be overstated. Thanks to AlphaFold, a myriad of previously elusive systems (Varadi et al., 2022), including disordered regions (Ruff and Pappu, 2021), are now accessible for comprehensive dynamic simulations. The confidence scores provided by AlphaFold not only inform us about the accuracy of structural predictions but also offer invaluable insights into protein flexibility (Schwarz et al., 2021). Numerous pioneering studies have emerged, aiming to harness AlphaFold's predictive power for flexibility assessments (Guo et al., 2022; Ma et al., 2023). However, these endeavors are not without challenges, such as reconciling the static nature of structural predictions with the inherently dynamic nature of proteins (Saldaño et al., 2022) or better prediction of protein regions that are disordered under ambient conditions (Ma et al., 2023).

A decade ago we proposed CABS-flex, a method for fast simulations of protein flexibility (Jamroz et al., 2013a). This method has demonstrated a speed advantage of three to four orders of magnitude compared to all-atom MD (Jamroz et al., 2013a) and continues to be competitive. The CABS-flex method is built upon the well-established C-alpha, beta, and side chain (CABS) coarse-grained protein model, the applications of which have been comprehensively reviewed (Kmiecik et al., 2016). The dynamics observed in CABS Monte Carlo simulations align well with the dynamics observed in MD simulations of folded globular proteins over nanosecond timescales (Jamroz et al., 2013a, 2013b), fluctuations detected in NMR ensembles (Jamroz et al., 2014), and experimental data on folding and binding dynamics (Ciemny et al., 2016; Kmiecik and Kolinski, 2007, 2008; Kmiecik et al., 2012, 2016; Kurcinski et al., 2014, 2020). CABS-flex was made available as a publicly accessible web server (Jamroz et al., 2013b) and later updated to version 2.0 (Kuriata et al., 2018) available at http://biocomp.chem.uw.edu.pl/CABSflex2. Subsequently, the CABS-flex method was made available as a standalone application (Kurcinski et al., 2019)—a Python package engineered to provide command-line access to CABS computations, granting users full control over the simulation process. CABS-flex also serves as a component responsible for modeling protein flexibility in the Aggrescan3D method (Zambrano et al., 2015) for structure-based predictions of protein aggregation preferences. The CABS-flex methodology was also used in the protocol for flexible docking of protein–peptide complexes, available as CABS-dock (Kurcinski et al., 2020), which enables modeling of large-scale conformational changes during the docking simulation process (Ciemny et al., 2016; Puławski et al., 2023).

In this paper, we provide a concise overview of the CABS-flex method (Section 2) and showcase its practical applications, supported by recent case studies (Section 3). Our demonstrations confirm the efficacy of the CABS-flex method in real-world applications, including analyzing the structural dynamics of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike proteins, modeling protein flexibility, exploring allosteric sites, and predicting regions prone to aggregation or nitrosylation. We also illustrate how CABS-flex can be used to combine flexibility analysis with predicted local distance difference test (pLDDT) predictive confidence scores. Finally, we discuss the important aspect of all-atom reconstruction from coarse-grained models and conclude with remarks and challenges for future research.

2 | CABS-FLEX METHOD

The CABS-flex methodology relies on a high-resolution coarse-grained model to represent protein chains, where each protein residue is characterized by up to four atoms. This representation includes the Cα and Cβ atoms, as well as two virtual pseudo-atoms: one representing the center of mass of the side chain, and the other positioned at the center of the Cα−Cα virtual bond (Kmiecik et al., 2016; Kolinski, 2004). Notably, CABS-flex leverages knowledge-based statistical potentials, encompassing sequence-dependent short-range conformational preferences, context-dependent pairwise interactions of side chains, and a comprehensive model for hydrogen bonds within the protein's main chain. Importantly, our tool incorporates solvent effects implicitly into its calculations, enhancing the accuracy of its simulations. Through the implementation of a Monte Carlo-based scheme that controls a random series of small local movements, CABS-flex captures the intricate dynamics of proteins in long-term evolution. CABS-flex simulations demonstrate computational efficiency, with simulation times dependent on the size of the system. For example, 20 amino acid peptides require just over a minute, 500 amino acid proteins take less than an hour, and 1000 amino acid proteins about 2 h on an Intel Xeon 2.50 GHz processor. Coarse-grained structures produced during CABS-flex simulations can be rapidly translated onto atomistic models of reasonable fidelity (see discussion in Section 3.7). This unique combination of techniques makes CABS-flex an invaluable tool for protein research.

2.1 | CABS-flex web server

The CABS-flex web server prioritizes accessibility and user-friendliness for nonspecialists in programming (Jamroz et al., 2013b; Kuriata et al., 2018). Simply providing a file containing a full-atom model of a protein in PDB format is

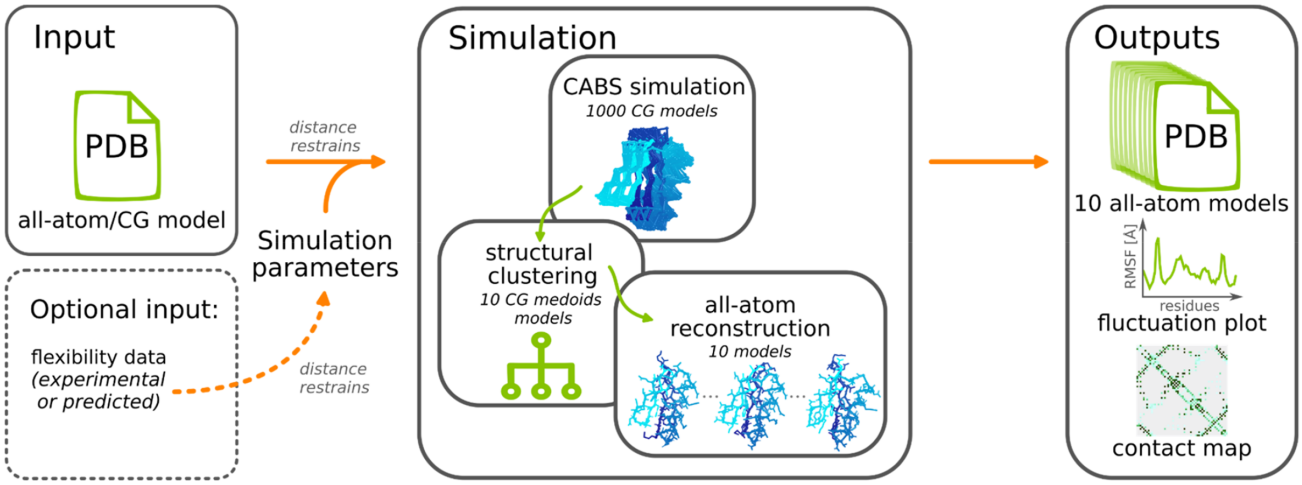

FIGURE 1 CABS-flex pipeline in default mode. The input protein structure serves as the starting point for the simulation and defines distance restraints based on either a default or a selected scheme. Optional inputs may include user-defined distance restraints. The simulation employs a coarse-grained CABS model. The protein dynamics trajectory is analyzed and clustered into 10 representative models, each capturing the essential variability of its cluster. These representative models are then reconstructed into all-atom models. The outputs of this process include PDB structures of the representative models, fluctuation profiles (illustrating RMSF vs. amino acid residue number), and contact maps derived from the trajectory. RMSF, root mean square fluctuation.

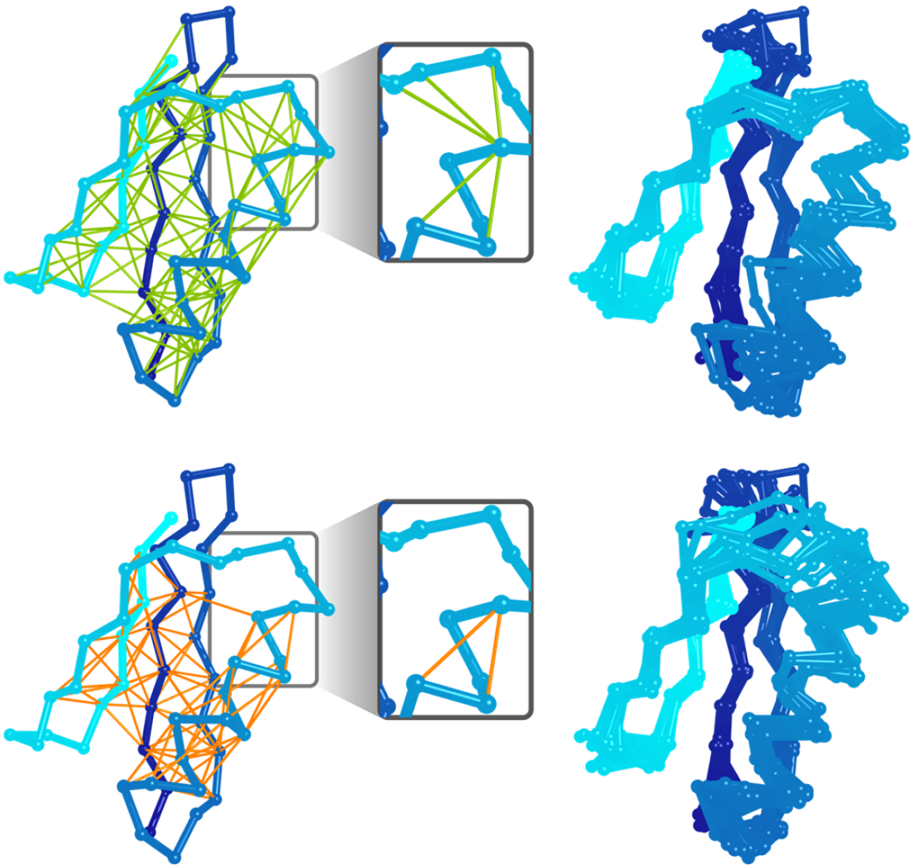

FIGURE 2 Distance restraints imposed on the protein (PDB ID: 2gb1) in the CABS-flex default modes SS1 (upper panel) and SS2 (lower panel). The left panels show the complete network of restraints shown as green and orange sticks, respectively for SS1 and SS2. In the middle, the restraints for a single residue are shown. The right panels show 10 protein models representing clusters of structures collected during one CABS-flex run with given default parameters. One can observe here a difference in the movement of the part depicted in the middle panel resulting from a different number of restraints.

enough to simulate the dynamics of the protein and to obtain an analysis of the fluctuations and their visualization. The server's capabilities have been significantly enhanced following a major upgrade to the 2.0 version available at http://biocomp.chem.uw.edu.pl/CABSflex2. In response to user feedback, we have expanded the restrictions on protein structure input from 400 to 2000 amino acids and, from single-chain proteins to those comprising up to 10 chains.

This progress was facilitated by increased computing resources. Additionally, we enhanced our result analysis by offering protein contact maps for the simulation trajectory that display the frequency of residue–residue contacts during the simulation process and for 10 selected protein models (Figure 1).

Furthermore, a customizable panel has been implemented, offering a range of simulation parameters and options to provide better control over the simulation process. This includes user-defined alterations to distance restraints. Through this approach, it is possible to incorporate experimental data into the simulation, immobilize specific segments of the model, or eliminate selected restraints. Within the web server, users have the option to choose from various restraint modes. Each "Mode" directs the algorithm in generating restraints based on the assignment of secondary structures to the residues. Specifically, if set to "SS1," the algorithm generates restraints when at least one of the residues is designated with a regular secondary structure, such as a helix or sheet. The "SS2" mode requires both residues to be assigned a regular secondary structure for restraints to be generated (Figure 2 presents a comparison of SS1 and SS2 restrain schemes for an examplary small protein). Selecting "All" instructs the algorithm to produce restraints for every residue, whereas choosing "None" results in no restraint generation. By default, the system is set to "SS2." Beyond the "Mode" setting, a collection of restraints is automatically produced based on additional parameters. These parameters determine the minimum distance along the protein chain required for two residues to be connected by a restraint, as well as the length and strength of these restraints (Kuriata et al., 2018).

2.2 | CABS-flex standalone application

Compared to the web server version, the CABS-flex standalone tool (Kurcinski et al., 2019) provides additional capabilities for modeling larger systems and customizing protocols. These enhancements enable full command-line access to various settings, such as the CABS coarse-grained model parameters, geometrical restraints utilized in CABS simulations, and result analysis. There are several options for analyzing the results, including customized output parameters, optional all-atom reconstruction (Badaczewska-Dawid et al., 2020a) with MODELLER (Webb and Sali, 2016) (including hydrogen atoms), personalized clustering and filtering, as well as contact map calculation and visualization, enabling users to analyze residue-to-residue contact frequencies with user-defined cutoffs.

As an additional feature, users can apply custom distance restraints to simulated models based on experimental data. It is also possible to manually adjust the flexibility of specific regions within a protein. Users can adjust flexibility factors for targeted residues at the local level, or they can opt for global modifications to the restraint generation system to modify the protein's overall flexibility across its structure.

With the standalone version of CABS-flex being written in Python, it can be seamlessly integrated into bioinformatics workflows and connected to additional libraries. This allows users to incorporate their data and analyses into more complex systems. The software is compatible with Linux, macOS, and Windows, and requires Python 2.7 and GFortran as the only essential additional software. While Python 2.7 is currently a limitation due to its obsolescence, an in-house version utilizing Python 3.12 is under development, indicating future updates to the standalone version. Installing MODELLER is recommended for full functionality. Importantly, CABS-flex can be operated on a standard up-to-date home computer without any special hardware requirements, making it both practical and accessible. Moreover, CABS-flex is free and open-source, ensuring wide accessibility. For more information, including an overview of all CABS-flex options, examples of use, and installation instructions, visit https://bitbucket.org/lcbio/cabsflex.

2.3 | Selecting optimal inputs for CABS-flex protein simulations

The primary and only required input for CABS-flex is the protein structure. When predicting the flexibility pattern of a protein with multiple conformations available, the following approach is recommended: First, employ an ensemble approach by analyzing multiple conformations, which is crucial for proteins with substantial structural diversity to ensure a comprehensive representation of their flexibility. Second, prioritize structures obtained under physiological conditions comparable to those of the biological study, enhancing the simulation's relevance. Third, utilize databases such as CoDNaS (Monzon et al., 2016) and PDB (Burley et al., 2023) to select a variety of conformations that demonstrate different functional states, enriching the foundation for predictive modeling.

The variability in flexibility patterns that arise from using different starting structures underscores the dynamic nature of proteins. Each conformation can distinctively affect simulation outcomes, identifying unique flexibility hotspots influenced by both structural and environmental factors.

Regarding simulation settings, the default parameters were set to mirror fluctuation patterns similar to those seen in 10-ns MD simulations using various force fields,

as referenced in earlier studies (Jamroz et al., 2013a). These settings are particularly calibrated for proteins up to approximately 200 amino acids, where the fluctuation profiles tend to be qualitatively consistent, with minimal differences due to the stochastic nature of the simulation algorithm. For proteins with high conformational diversity or larger sizes, it is advisable to increase the number of simulation cycles, potentially by tenfold, to capture more accurate dynamics (the advanced simulation options are discussed in the CABS-flex 2.0 documentation) (Kuriata et al., 2018).

3 | APPLICATIONS—CASE STUDIES

This section presents case studies showing the capabilities of CABS-flex in modeling protein systems. These examples are intended to inspire readers to explore CABS-flex applications in structure–function–flexibility relationship studies.

3.1 | Modeling the flexibility of globular protein structures

CABS-flex was initially employed to simulate the flexibility of folded globular proteins, demonstrating results in line with MD simulations across nanosecond timescales (Jamroz et al., 2013a, 2013b) and consistent with fluctuations observed in NMR ensembles (Jamroz et al., 2014). This section presents two case studies that compare the flexibility profiles predicted by CABS-flex simulations to those measured by the random coil index (RCI), which is derived from backbone chemical shifts (Fowler et al., 2020).

The RCI serves as a reliable measure of local protein rigidity (Berjanskii and Wishart, 2005, 2008), and is frequently referenced for assessing the structural accuracy of both NMR-derived and AlphaFold 2 (AF2)-predicted models (Fowler et al., 2021; Fowler and Williamson, 2022). Notably, it has been observed (Fowler and Williamson, 2022) that AF2-generated structures generally surpass NMR-derived structures in accuracy. However, the application of RCI is contingent upon the availability of backbone chemical shift assignments. In the absence of such data, AF2-predicted structures offer a provisional model for the protein's tertiary structure, whose flexibility can then be swiftly approximated using CABS-flex.

For our analysis, we have chosen two examples from Fowler and Williamson (2022) to explore the correlation between flexibility profiles from CABS-flex simulations and those derived from RCI, which stands as a credible experimental benchmark. The first case, PDB ID 2ct5, illustrates an instance where the AF2-predicted structure is more accurate than the highest-scoring NMR structure. Conversely, PDB ID 2y4q represents a case where the NMR structure is more precise. The CABS-flex simulations were conducted using standard parameters, and the root mean square fluctuation (RMSF) was normalized to a unit scale to facilitate a direct comparison with the RCI-measured flexibility. These findings are illustrated in Figure 3. For PDB ID 2ct5, the flexibility profiles generated by CABS-flex for both the NMR and AF2 models accurately pinpoint the locations of flexible regions. Specifically, the segment between residues 18 and 28 and the terminal regions are identified as overly flexible in the NMR model predictions, with a notably better quantitative alignment observed with the AF2 structure's predictions. In the case of PDB ID 2y4q, CABS-flex accurately identifies the flexible regions as well. The AF2 model lacks a short beta-sheet found in the NMR model (Fowler and Williamson, 2022), which contributes to the AF2 model's diminished accuracy for this protein. Intriguingly, CABS-flex predicts the region encompassing residues 760–770 to exhibit excessive flexibility in both models. A further distinct difference is observed in the terminal region, where the AF2 model features a longer alpha-helix compared to the NMR model.

In summary, our evaluation indicates that CABS-flex serves as an effective tool for estimating backbone flexibility at a modest computational expense, offering accuracy on par with the RCI method for proteins lacking backbone chemical shift data. For such instances, employing AF2-derived structures—or x-ray structures when accessible—as the basis for analysis is expected to provide the most accurate estimates of flexibility. This approach underscores the utility of CABS-flex in complementing other flexibility measurement methods, particularly in scenarios where direct experimental data may be limited or unavailable.

3.2 | Structure–function analysis of protein complexes: Interactions and mutational dynamics

In this section, we have compiled several examples demonstrating applications of the CABS-flex method to explore the critical interactions and structural dynamics of the SARS-CoV-2 spike (S) protein. The S glycoprotein is a multidomain protein that binds to the angiotensin-converting enzyme 2 (ACE2) (Huang et al., 2020). Comprising 1273 residues, the S protein consists of an amino (N)-terminal S1 subunit and a carboxyl (C)-terminal S2 subunit. The S1 domain, which is responsible for interactions with the host receptor, includes

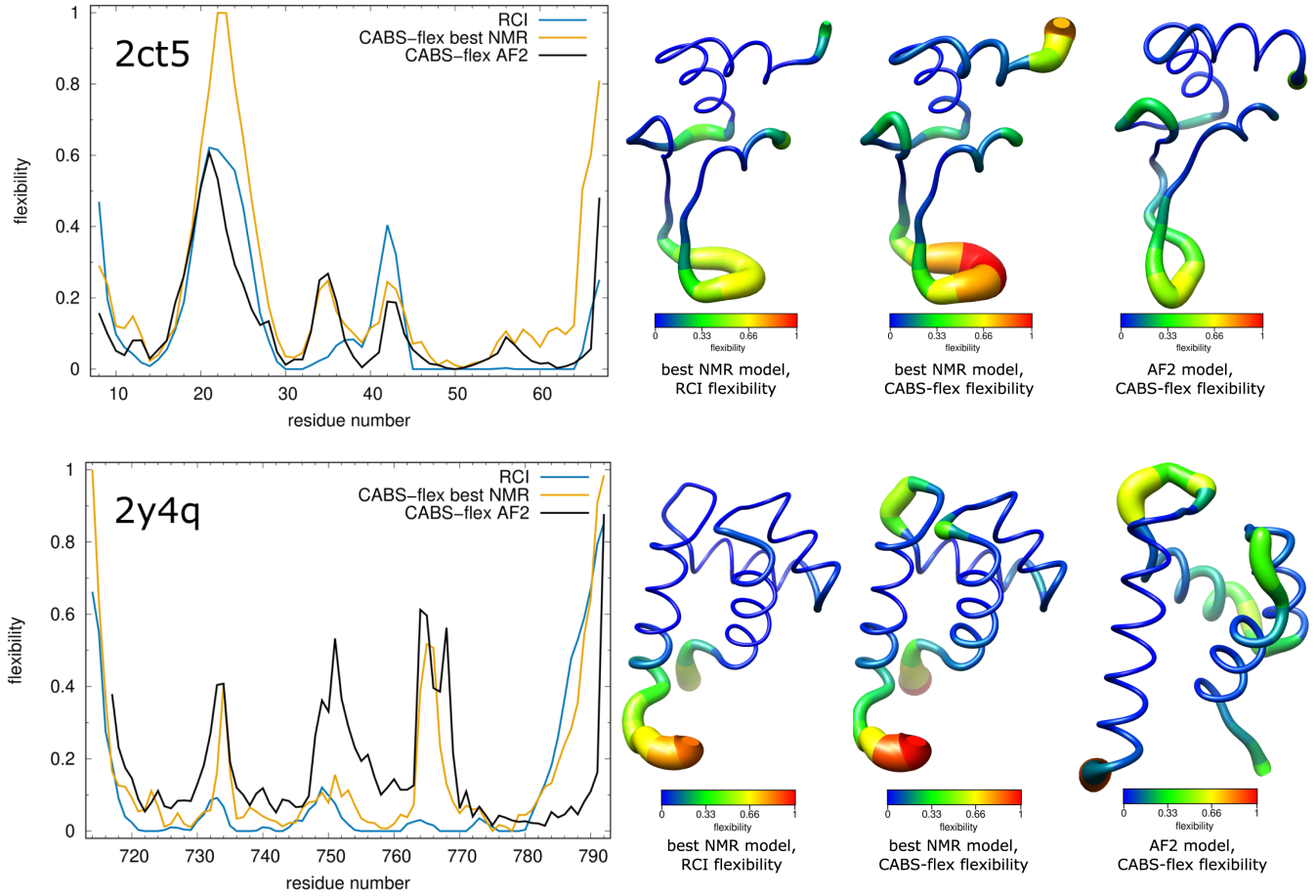

FIGURE 3 Comparison of flexibility profiles computed from the RCI and obtained from CABS-flex simulations. Top: PDB ID 2ct5 NMR model 3 and AF2 model (UniProt: O96006). Bottom: PDB ID 2y4q NMR model 3 and AF2 model (UniProt: Q13563). In the right panels, the color and the thickness of the tube representation are proportional to the flexibility of each residue. RCI, random coil index.

an N-terminal domain (NTD), a receptor binding domain (RBD), and a C-terminal domain (CTD) with two structurally conserved subdomains: CTD1 and CTD2. The S protein forms a homotrimer and the virus infection mechanism involves conformational transitions. Specifically, the RBDs alternate between a receptor-accessible "up" state (Figure 4a) and a receptor-inaccessible "down" state (Figure 4b) (Wrapp et al., 2020). Figure 4c highlights the residues involved in the binding of the RBD with ACE2 (using contact map MAPIYA web server [Badaczewska-Dawid et al., 2022a] with a distance cutoff of 8 Å). The mutations in the RBD domain significantly influence the virus's evolution.

CABS-flex can be particularly valuable for investigating the dynamic aspects of protein interactions that are not evident from static structural data. Although RMSF derived from multiple crystallized structures offers important insights into observed protein flexibility, CABS-flex provides additional value by modeling dynamics and interactions that may not be fully resolved in crystal structures. For example, in the study of the SARS-CoV-2 spike protein (Sanyal et al., 2022), CABS-flex highlighted significant flexibility in the distal ends of the ACE2 binding motif—a region often unresolved in crystallographic data. This modeling was crucial for understanding the enhanced binding affinity of the Omicron variant's spike protein to the ACE2 receptor (Ortega et al., 2022), which is speculated to contribute to its higher transmissibility. The insights gained from CABS-flex about the dynamic behavior of these critical regions complement static RMSF analyses, offering a more complete understanding of how mutations may influence virus–host interactions.

Beyond exploring the dynamics of entire regions of protein–protein interactions, CABS-flex has also been extensively used to study the effects of mutations on protein dynamics. It is important to clarify that CABS-flex does not inherently modify amino acid sequences or generate coordinates for newly substituted amino acids. Instead, users must prepare mutated structures prior to simulation, often employing tools such as Modeller (Webb and Sali, 2016) or FoldX (Buß et al., 2018). This crucial preparatory step ensures that the input structures

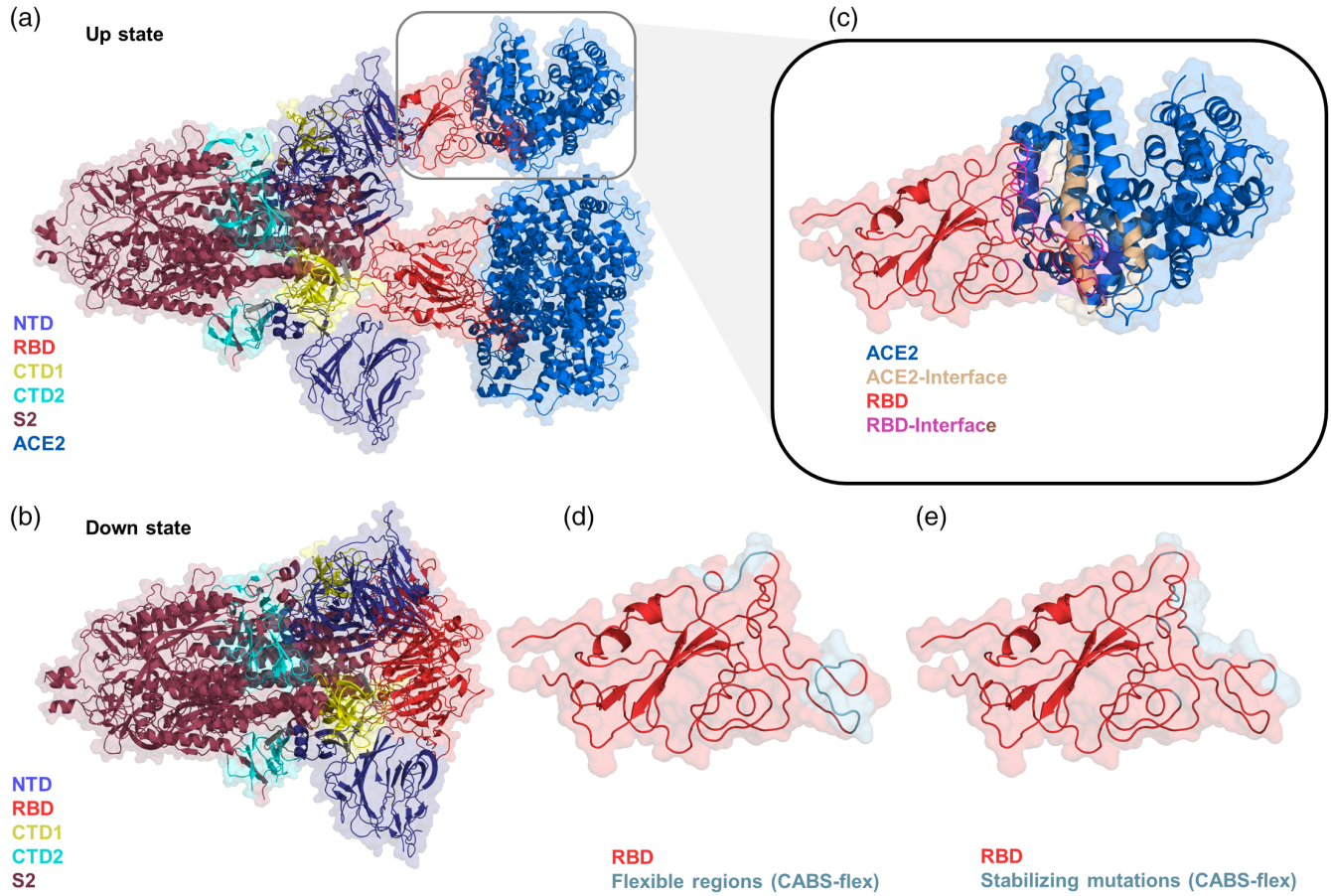

FIGURE 4 The SARS-CoV-2 spike (S) protein in (a) "up" conformation (PDB ID: 7a98) bound to angiotensin-converting enzyme 2 (ACE2) and in (b) "down" conformation (PDB ID: 7a98) which is without binding to the receptor. The N-terminal S1 subunit consists of NTD (residues 14–306) in blue, RBD (residues 331–528) in tv_red, CTD1 (residues 529–591) in yellow, and CTD2 (residues 592–686) in cyan. The remaining residues form the S2 subunit and are shown in raspberry. (c) The interaction between S and ACE2 happens through the residues shown in light magenta on the RBD and wheat on the ACE2 receptor. The residues identified from CABS-flex trajectories as (d) highly flexible and mutations could have significant effects on the evolution of the virus (Sanyal et al., 2022) and (e) have a significant stabilization effect on S-ACE2 complex (Mishra et al., 2022) are highlighted in sky blue. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

accurately reflect the physical characteristics of the mutated proteins, allowing for reliable simulations of their dynamics and interactions. This detail underscores the practical utility of CABS-flex in mutation analysis, demonstrating its capability to predict the functional consequences of genetic variations.

In combination with other analyses, CABS-flex has been employed to study the effects of mutations in the RBD of the spike protein. Sanyal et al. (2022) used CABS-flex trajectories to explore and identify the regions in the RBD that can have significant impacts on the stabilization of the RBD domain in its interactions with ACE2 reception. This study identified the residue stretches 441–445 and 477–484 (Figure 4d) as highly flexible and mutations in this region will have a significant effect on the evolution of SARS-CoV-2. Notably, the residues 477, 478, and 484 were mutated in the Omicron variant (Karim and Karim, 2021). In another study, Mishra et al. (2022) used CABS-flex to study the impact of mutations in S proteins on the stabilization or destabilization of the S-ACE2 complex. This study pinpointed 36 mutations involving 12 residues that could enhance the stability of the S-ACE2 complex (Figure 4e). The V503D mutation exhibited the highest stabilization of the S-ACE2 complex and is known to resist neutralization (Tortorici et al., 2021). Moreover, mutations K417Y, E484A, N501Y, D614G, and P681H identified in this study align with those observed in the Omicron variant and its sub-lineages (Cao et al., 2022; He et al., 2021; Zhang et al., 2022). These two studies used CABS-flex trajectories to identify the residues that could play a significant role in virus evolutions. Multiple residues and substitutions identified by these studies were also observed in the SARS-CoV-2 variants.

The CABS-flex was also used to investigate the impact of mutations on the ability of the S protein to bind to the ACE2 receptor. For example, the D614G substitution

introduces higher infectivity (Korber et al., 2020) and to understand the possible reasons for this, Verkhivker et al. (2022) used trajectories from 1000 independent CABS-flex simulations. The authors performed a principal component analysis (PCA) of the CABS-flex trajectories and used elastic network model analysis to determine the dynamics profiles from these trajectories. The D614G mutation introduces significant conformational adaptability to the "up" states, without compromising the folding stability or integrity of the S proteins (Cai et al., 2021). Based on the dynamic profiles from CABS-flex trajectories, the study concluded that the D614G mutation allows the S protein to exist in a dynamic equilibrium between closed and open states, with the RBD favoring the "up" conformation before ACE2 binding. In another study (Verkhivker, 2022a) exploring the effects of amino acid substitutions in SARS-CoV-2 variants, trajectories from 100 independent CABS-flex simulations were used to identify the diversity of "up" conformations. The study concluded that both A570D and D614G substitutions act as potential regulatory switches and play a role in modulating the dynamics of the S protein and its interaction with host receptors. Moreover, the study postulates that the two residues play an important role in the transitions between the "down" and "up" conformations of the RBD (Verkhivker, 2022a). Together, these two studies exemplify the use of CABS-flex trajectories to understand the implications of amino acid substitution introducing changes in the adaptability and evolution of the virus.

3.3 | Exploring allosteric mechanisms through integrative computational modeling

CABS-flex simulations have been utilized as a tool for investigating dynamics and elucidating allosteric regulation mechanisms. Although CABS-flex does not directly model ligand interactions, it effectively pinpoints regions of significant conformational flexibility within proteins that suggest potential allosteric sites. For instance, Verkhivker (2022b) applied CABS-flex in conjunction with atomistic simulations to explore the dynamic conformational landscapes of the Hsp90 chaperone complexes. This integrated approach, which demonstrated the compatibility of CABS-flex with MD simulations, helped identify critical dynamic areas essential for understanding allosteric regulation and client protein integration in Hsp90.

Similarly, Amusengeri et al. (2019) employed CABS-flex to analyze the intrinsic dynamics of heat shock proteins (Hsp70s) from *Plasmodium falciparum*, highlighting regions of conformational flexibility indicative of potential allosteric sites. Their study effectively bridges the resolution gap between coarse-grained simulations and all-atom MD simulations, confirming the consistency of allosteric sites identified by CABS-flex with those observed in more detailed simulations. Furthermore, Amusengeri et al. (2019) verified the interactions of ligands with the studied proteins through additional all-atom MD simulations. These simulations, focusing on both apo (ligand-free) and holo (ligand-bound) states, provided a detailed view of how ligands influence protein conformation and function.

Another significant contribution comes from a study focusing on the Hsp90 chaperone interactions with cochaperones and client proteins (Verkhivker, 2022c), which provided detailed atomistic insights from integrative modeling (using CABS-flex simulations followed by MODELLER-based all-atom reconstruction) and network analysis of conformational landscapes. This study reveals that dynamic-based allostery, driven by the interplay of conformational states and interactions within the chaperone complex, is crucial for the efficient recruitment and integration of client proteins.

CABS-flex was also adeptly employed to model the S protein's ability to switch between functional states, highlighting its utility in mapping dynamic regulatory switches within the protein structure (Verkhivker and Di Paola, 2021a, 2021b). Simulation trajectories from CABS-flex were validated by comparing them with those from all-atom MD simulations and experimental data, ensuring reliability in modeling the spike protein's conformational dynamics. These were further analyzed alongside atomistic reconstructions and PCA to identify hinge regions and characterize collective motions within the SARS-CoV-2 spike protein complexes, providing insights into allosteric communication pathways.

Ultimately, this collection of studies highlights the utility of CABS-flex in identifying potential allosteric sites and demonstrates that integrative simulation strategies can provide crucial insights into the complex mechanisms of allostery in protein dynamics, essential for advancing targeted drug development.

3.4 | Structure-based prediction of aggregation-prone regions (APRs)

CABS-flex, as a tool for simulating near-native flexibility, can be easily used to introduce the aspect of protein flexibility to structure-based studies of static conformations of globular proteins. This capability was used in the Aggrescan3D (A3D) (Kuriata et al., 2019a) method for structure-based prediction of protein aggregation properties. By default, A3D runs in "static mode" where it evaluates aggregation tendencies of a static protein structure. However, A3D

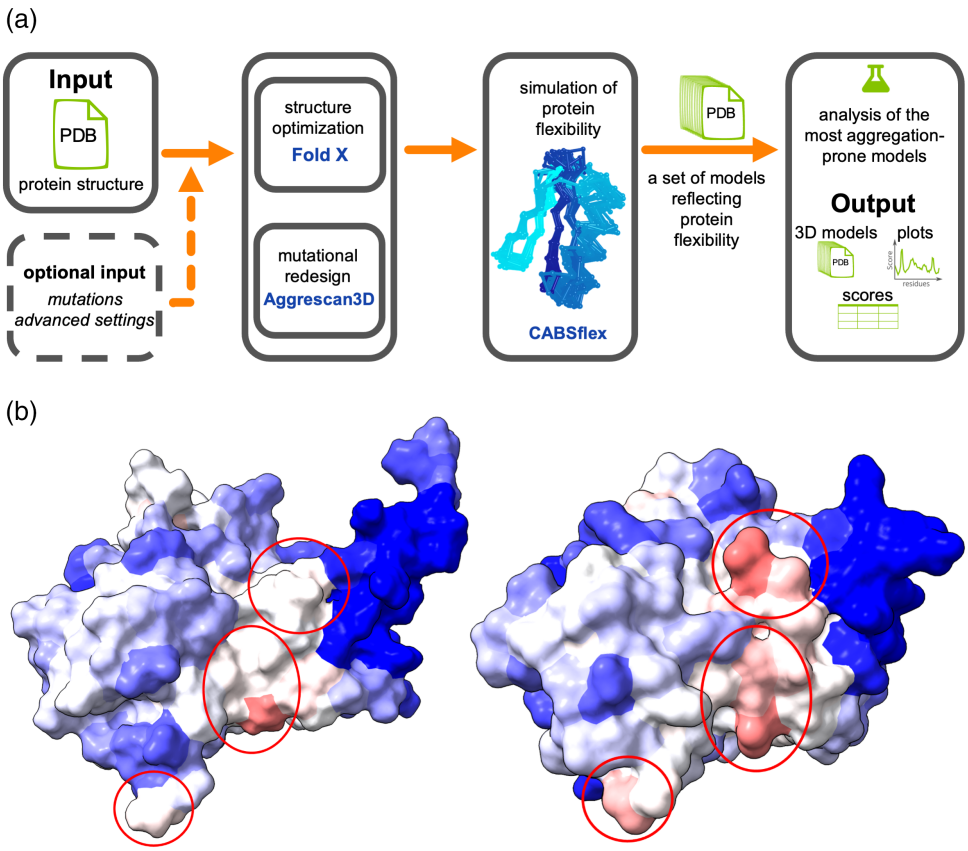

FIGURE 5 Aggrescan3D "dynamic mode." (a) A3D pipeline: the input structure's flexibility is simulated using CABS-flex to generate a set of predicted models that reflect the varying conformations of the structure; these models are then analyzed and scored based on their aggregation propensity, and the model with the highest aggregation propensity is presented as the final prediction. (b) Human p53 DNA-binding domain (PDB ID: 2fej) aggregation propensity predictions using A3D "static mode" (left side) and "dynamic mode" (right side). Residues are colored in A3D scale (blue—solubility; red—aggregation). Aggregation prone regions are encircled. A3D, Aggrescan3D.

calculations can be also performed using "dynamic mode" (Figure 5a) which harnesses CABS-flex to simulate protein fluctuations. In this mode, an initial static structure serves as a starting point for a CABS-flex simulation, and aggregation assessments are performed on CABS-flex output models, which reflect protein flexibility. The final structure is determined as the model with the highest aggregation score, offering a more comprehensive and dynamic view of protein aggregation tendencies. This dynamic approach is particularly valuable as in solution protein aggregation may arise not only from global unfolding but also as a result of local structural changes and fluctuations from the protein's native state (Chiti and Dobson, 2009). To emphasize the utility of this approach we can consider the illustrative example of the human p53 DNA-binding domain (PDB ID: 2fej). Calculations performed using A3D "static mode" (Figure 5b, left panel) do not reveal all potential APRs. However, after harnessing A3D "dynamic mode" (Figure 5b, right panel), CABS-flex unveils APRs that mirror experimental observations (Navarro and Ventura, 2019). In addition, A3D provides robust tools for introducing and analyzing mutations within protein sequences. Users can utilize the "Mutate residues" feature to manually input specific sequence changes or employ the automatic mutations functionality. This automatic tool identifies APRs and suggests strategic point mutations to charged residues, aiming to enhance protein solubility. Such modifications are crucial for exploring how changes in the flexibility of flanking regions can influence protein stability and aggregation propensity.

A3D method is available both as a web server (Kuriata et al., 2019a) and as a standalone tool (Kuriata et al., 2019b), offering a wide range of applications including automatic redesign toward soluble variants. Recently, A3D has been applied to predict regions responsible for aggregation across entire proteomes, a feat made possible through the use of structural predictions generated by AlphaFold2 (AF2) (Varadi et al., 2022). A3D databases have been created to seamlessly integrate A3D's analytical infrastructure with structural predictions for human (Badaczewska-Dawid et al., 2022b), yeast (Garcia-Pardo et al., 2023), and other proteomes (Badaczewska-Dawid et al., 2024a). Finally, the A3D method has recently undergone significant enhancements, notably including the ability to consider environmental pH. These advancements are now accessible through the A4D web service (Barcenas et al. 2024).

3.5 | Structure-based prediction of S-nitrosylation sites

In Section 3.4, we discussed the utilization of CABS-flex for revealing APRs on protein surfaces. A similar approach, integrating CABS-flex into structure-based

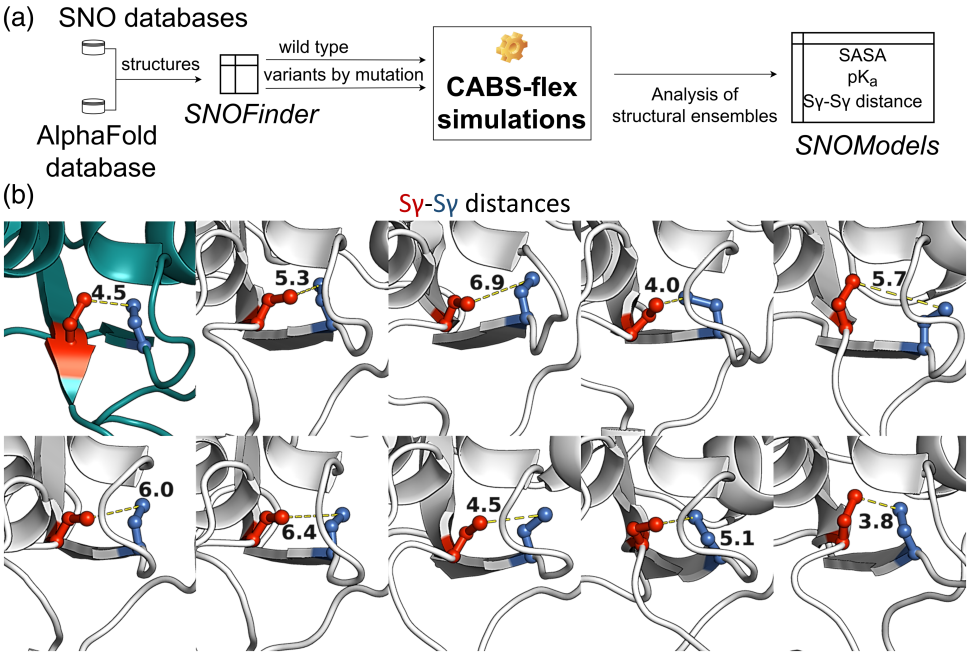

FIGURE 6 CABS-flex ensembles provided valuable insights into the structural dynamics of S-nitrosylated proteins. (a) The workflow utilizes SNOfinder to identify proteins susceptible to S-nitrosylation, generates a structural ensemble of 20 models per protein with CABS-flex to highlight their flexibility, and analyzes these models with SNOmodel to calculate key structural and biochemical parameters, thereby enhancing understanding of S-nitrosylation mechanisms. (b) The cartoon representations show the models from the CABS-flex ensemble (white) of the mitochondrial chaperone TRAP1 (deep teal). The stick and ball representations highlight the conformational dynamics of SNO cysteine (SNO, red), proximal cysteine (proximal, marine) of the ensembles, and the Sγ-Sγ distances between them.

analyses, was also employed by Papaleo et al. (2023) for identifying S-nitrosylation sites. S-nitrosylation—a reversible modification of cysteine residues to form S-nitrosothiols—is a key posttranslational modification (PTM) affecting cellular functions and is implicated in various diseases, making it a target for therapeutic interventions. The method involves identifying cysteines close in three-dimensional space, not just sequence, highlighting the importance of protein flexibility for exposing potential S-nitrosylation sites and influencing protein behavior and interactions within cellular pathways. For instance, in the mitochondrial chaperone TRAP1, SNO at Cys501 induces conformational changes at the distal site in the protein structure which results in altering the open and closing motions necessary for its chaperone function. This modification reduces its ATPase activity, and with its proximity to Cys527 suggests potential disulfide bridge formation (Faienza et al., 2020; Rizza et al., 2016).

Figure 6a presents a workflow devised by Papaleo et al. (2023), showcasing a two-step bioinformatic approach for studying S-nitrosylation in proteins. Initially, the workflow employs SNOfinder, a tool specifically designed for the identification of proteins susceptible to S-nitrosylation by leveraging databases of known PTMs and structural data. This identification process is critical for pinpointing proteins where S-nitrosylation plays a potential role in regulating protein function and signaling pathways. Following the identification, CABS-flex is utilized to generate a structural ensemble comprising 20 models for each identified protein. These models collectively represent the protein's structural flexibility and potential conformational states, taking into account the dynamic nature of protein structures in biological systems. The resulting ensemble is then analyzed using the SNOmodel pipeline, which is tailored to calculate specific structural and biochemical parameters, such as dihedral angles and solvent accessibility of cysteine residues involved in S-nitrosylation. This analysis is aimed at understanding the structural context of S-nitrosylation sites, including their proximity to other functional groups within the protein and potential interactions with cofactors or metal ions, thereby shedding light on the mechanistic underpinnings of S-nitrosylation in protein function and regulation.

For the mitochondrial chaperone TRAP1, the CABS-flex simulations focused on the cysteines C501 and C527. The Sγ-Sγ distance between the two sulfur groups of these cysteines is highlighted in Figure 6b. It showed that SNO at C501 could lead to a disulfide bridge with C527, a finding supported by experimental validation. This insight was important for understanding the redox regulation of TRAP1 and its role in diseases. The CABS-flex approach, a computationally inexpensive method, effectively

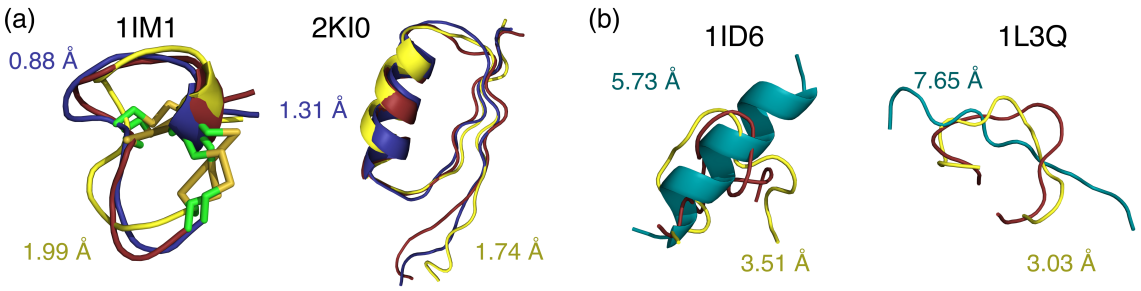

FIGURE 7 (a) Example CABS-flex predictions (best prediction out of 10,000 simulated structures—blue; best prediction out of top 10 structures ranked by CABS-flex—yellow) superimposed on the experimental PDB structures (red). (b) Comparison between CABS-flex predictions (yellow) and AlphaFold2 predictions (cyan) superimposed on the experimental PDB structures (red).

identified candidates with structural constraints that limit SNO or disulfide bridge formation. This study analyzed multiple SNO proteins, including the tumor protein TP53, focusing on its DNA-binding domain. For example, CABS-flex simulations were used to determine how mutations near the SNO and adjacent cysteine sites in TP53 influence its stability and DNA-binding capabilities. Key findings include the stabilizing N235H variant promoting S-nitrosylation and disulfide bridge formation, enhancing TP53's structure. On the other hand, the destabilizing variants could impair the structural mechanism.

3.6 | Structure prediction of linear and cyclic peptides

Recently, we introduced a new CABS-flex-based protocol that enables de novo structure prediction of linear and cyclic peptides (Badaczewska-Dawid et al., 2024b). The protocol has been introduced as an extension of the CABS-flex standalone application; it uses the CABS-flex methodology to perform an efficient exploration of a peptide's conformational space (Badaczewska-Dawid et al., 2024b). The protocol was extensively tested against other methods (AF2 [Jumper et al., 2021] implemented in the ColabFold [Mirdita et al., 2022], ESMFold [Lin et al., 2023], APPTEST [Timmons and Hewage, 2021], PEP-FOLD [Shen et al., 2014]) on a benchmark consisting of 159 linear and cyclic peptides ranging from short peptides of 9 residues to mini-proteins consisting of 49 residues. The two examples of CABS-flex predictions are highlighted in Figure 7. In most benchmarked cases, AlphaFold-based protocol outperformed CABS-flex, especially for longer peptides, due to its ability to capture complex long-range interactions. However, the diverse structural behaviors of short peptides, which range from disordered regions to well-defined secondary structures, pose a challenge for deep learning methods like AF2, as experimental structures sometimes inadequately represent this diversity. We have demonstrated that in these instances, coarse-grained methods such as CABS-flex remain competitive by effectively navigating this structural diversity through simplifications and reduced granularity. Figure 7b illustrates two examples in which CABS-flex was able to predict conformations that turned out to be closer to the experimental ones than those proposed by AF2.

In a standard CABS-flex simulation, the only required input is the initial structure provided in the form of a PDB code or PDB file. In the CABS-flex peptide structure prediction protocol, the only explicitly required input is an amino acid sequence. The starting structure is generated as a random coil based on that input sequence. There are other optional inputs, such as secondary structure in three-letter notation (H, helix; E, extended; C, coil). If it is not supplied, the secondary structure would be predicted by Psipred (Jones, 1999), though we recommend users to predict it using state-of-the-art deep-learning methods for enhanced accuracy (i.e., NetSurfP-3.0 [Høie et al., 2022] or SPOT-1D-LM [Singh et al., 2022]). The versatility of the presented CABS-flex modeling pipeline is a notable advantage, as it can be readily expanded by incorporating additional data or testing alternative configurations. Additional simulation data can be easily introduced as distance constraints between C-alpha atoms or side group centers of mass with appropriate confidence levels that may force a precisely known distance or provide a weak bias (Badaczewska-Dawid et al., 2024b). Potential pipeline extensions include the use of techniques like MD to investigate the peptide dynamics or explore the design of peptide-based drugs with modifications extending beyond natural amino acids, providing a broader scope for research and applications.

3.7 | Merging with high-accuracy all-atom reconstruction

CABS-flex employs the CABS coarse-grained model to create C-alpha protein models (refer to Section 2), which

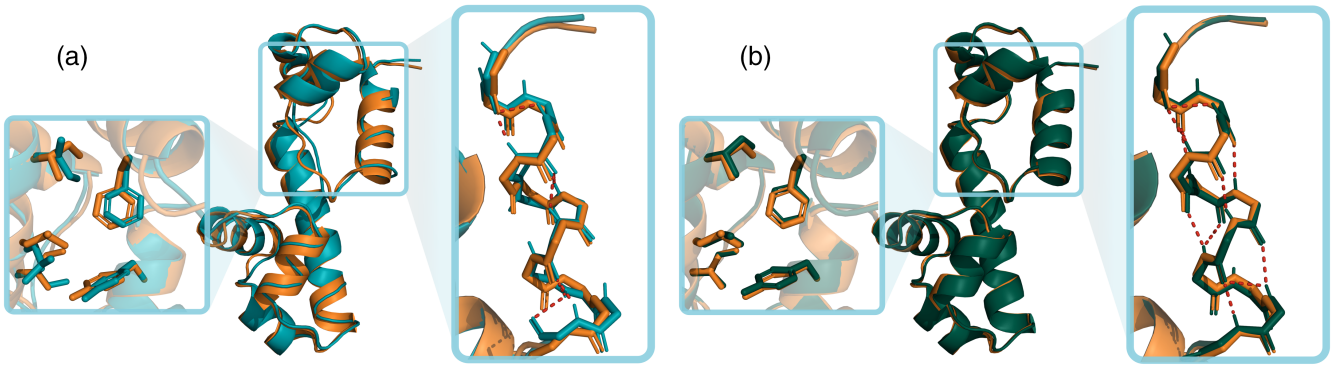

FIGURE 8 CABS-flex output structures rebuilt by two methods: (a) Modeller in cyan and (b) cg2all[72] in dark green, superimposed on the experimental PDB structure in orange (PDB ID: 3oou). We highlight the accuracy of side-chain reconstruction (left panels) and helical hydrogen bond networks (right panels) in the rebuilt structures.

are subsequently reconstructed into full-atom structures. Despite the advances in reconstruction methods, they are not without flaws, and as a result, suboptimal chain geometries can emerge in the final output (Badaczewska-Dawid et al., 2020b). Moreover, the process of reconstructing from coarse-grained models introduces additional challenges. Namely, the initial C-alpha trace may include unphysical distortions and different backbone reconstruction methods exhibit varying degrees of resilience to that challenge (Badaczewska-Dawid et al., 2020b). This can lead to missing portions of the rebuilt backbone or amplified distortions, which can substantially affect the quality of subsequent side chain reconstruction, all-atom energy minimization, and scoring (Badaczewska-Dawid et al., 2020b).

As of now, CABS-based tools are integrated with Modeller-based protocol (Webb and Sali, 2016), utilizing the coarse-grained models as templates and the automodel class from Modeller (Webb and Sali, 2016), to generate a comprehensive set of restraints for reconstructing all heavy atoms (Badaczewska-Dawid et al., 2020a). Unfortunately, the Modeller protocol often faces the reconstruction difficulties mentioned above. This prompted our exploration of a new reconstruction method. Recently, we obtained very promising results using the deep-learning-based method cg2all (Heo and Feig, 2024). The method, inspired by the success of AF2, employs a graph neural network and utilizes rigid-body blocks for 3D structure generation. It incorporates physical constraints and structural features, learned from known protein conformations during model training. Preliminary results indicate an improvement in C-alpha RMSDs, fewer clashes, and better MolProbity (Chen et al., 2010) scores of the rebuilt models by the cg2all method. This can be seen in Figure 8, where the quality of both the main chain and side chains rebuilt by Modeller (Figure 8a) is worse than in the structure rebuilt by cg2all (Figure 8b). The reconstructed structures demonstrate improved local quality and stability, assessed by using the reconstructed structures as a starting point for MD simulations. Furthermore, we observe a significant improvement (Figure 8b, right panel) in the previously observed issue of DSSP (Kabsch and Sander, 1983) being unable to identify secondary structure elements in the Modeller-reconstructed structures, due to local distortions (Figure 8a, right panel).

Therefore, one of the shortcomings of using coarse-grained models, including the CABS-flex method, is the difficulty in reconstructing physically meaningful full-atom structures. We believe that this problem can be addressed by employing new deep learning networks, which would allow for the utilization of extensive knowledge of known structures in the task of reconstruction based on the coarse-grained data.

Another common challenge in reconstructing protein structures is the incorporation of nonstandard amino acids or PTMs. While CABS-flex is primarily designed to work with natural amino acids and does not directly support the integration of nonstandard amino acids or PTMs, users can approximate the effects of such modifications. For instance, users can substitute phosphoserine with aspartic acid or acetyllysine with lysine to mimic the effects of these modifications. For a more precise analysis, researchers may start with a CABS-flex generated ensemble of structures and subsequently refine these models using more detailed MD simulations. Tools like CHARMM-GUI, which has recently incorporated the PDB Manipulator for easier handling of nonstandard amino acids and PTMs (Park et al., 2023), offer an advanced platform for such detailed studies. Additionally, the PyMOL plugin pyTM can be used to visualize and introduce common PTMs into models (Warnecke et al., 2014). This layered approach leverages the rapid simulation capabilities of CABS-flex alongside the detailed analysis possible with MD,

providing a framework for studying the structural and functional impacts of PTMs.

3.8 | AlphaFold's pLDDT: Bridging predictive confidence and protein flexibility

The pLDDT score, developed by AlphaFold, serves as a confidence metric for the predicted positions of amino acids within a protein structure. Regions with high pLDDT scores generally indicate high structural confidence and presumed rigidity, whereas lower scores often suggest potential flexibility where structural predictions are uncertain. Importantly, pLDDT scores can indicate flexibility even in regions with high scores if the predictive model, like AlphaFold, has incorporated multiple conformations in its training dataset.

This underscores the complexity of interpreting pLDDT scores, necessitating a nuanced approach to assessing protein flexibility. AF2 adaptations for modeling multiple protein states enhance the structural heterogeneity of predicted ensembles (Sala et al., 2023). These adaptations allow for the simulation of various conformational states beyond a single rigid configuration, which is essential for proteins undergoing significant structural changes. The pLDDT metric reliably assesses the structural confidence of these predicted models, assigning higher values to more accurately predicted conformations.

Here, we analyze two example proteins from the ATLAS database (Vander Meersche et al., 2024), 6o2v_A and 1d3y_B, which were subjected to three 100 ns simulations using GROMACS v2019.4 with the CHARMM36m force field. By scaling the RMSF values yielded by CABS-flex alongside those from MD simulations, we maintained comparability and shape consistency of fluctuation values. Spearman's correlation was employed as a metric to measure the similarity between these two sets of fluctuation data.

For 6o2v_A, the correlation between CABS-flex and ATLAS's MD simulations was notably high at 0.86, indicating strong predictive accuracy by CABS-flex (Figure 9a). Interestingly, the shape of the pLDDT curve inversely mirrored the RMSF curves, leading to further analysis where RMSF and pLDDT values from MD were scaled together. This analysis resulted in a Spearman correlation of −0.80, suggesting a strong negative interdependence, with most fluctuations aligning with protein loops that are prone to movement.

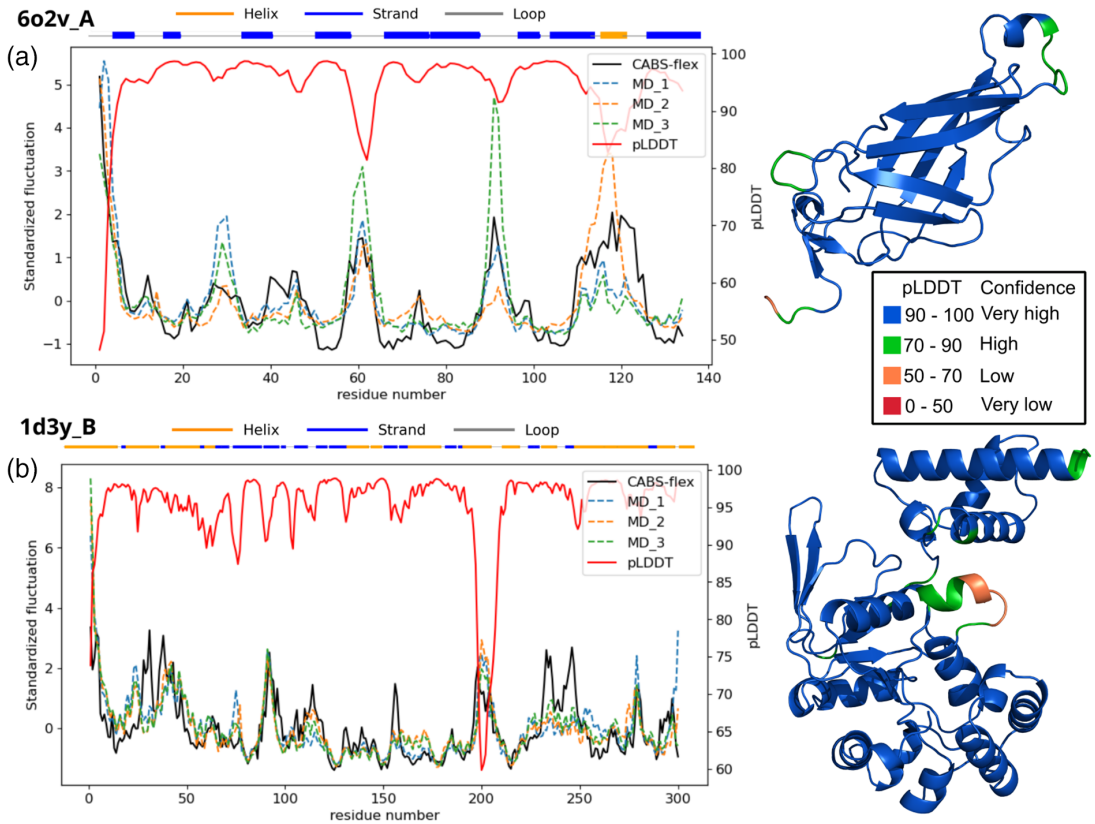

FIGURE 9 Examples of protein structure fluctuations from CABS-flex and molecular dynamics (MD) runs, and pLDDT confidence scores for proteins: (a) 6o2v_A and (b) 1d3y_B. MD runs were retrieved from the ATLAS database (Vander Meersche et al., 2024). On the right, the figure shows pLDDT confidence scores visualized on the protein structures. pLDDT, predicted local distance difference test.

Conversely, 1d3y_B, which is rich in secondary structure elements and has fewer loops, showed significant fluctuations loosely correlated with pLDDT scores (Figure 9b). Particularly in the region 195–205, a significant drop in pLDDT scores was observed. This region was not originally present in the PDB data and was computationally modeled by AF2 for inclusion in the ATLAS database (Vander Meersche et al., 2024), explaining the lower pLDDT scores. Such scores often suggest structural uncertainty, which, in this case, is a direct result of the absence of empirical data for this segment. Despite the low pLDDT, this region exhibits a level of flexibility comparable to other regions with much higher pLDDT scores, emphasizing the importance of source data in interpreting both pLDDT scores and flexibility.

These findings underscore the complex and context-dependent nature of using pLDDT scores to infer flexibility in protein structures. They highlight the importance of integrating these scores with dynamic simulation data from CABS-flex to obtain a comprehensive understanding of protein behavior, facilitating more informed predictions about protein functionality under diverse physiological conditions.

4 | CONCLUSIONS

CABS-flex has established itself as an efficient and cost-effective tool for simulating the backbone flexibility of proteins across diverse biological contexts. It has been particularly useful for applications in the analysis of globular proteins, mechanisms of viral adaptation, allostery, prediction of aggregation-prone and S-nitrosylation sites, as well as in structure prediction for linear and cyclic peptides, thereby enhancing our understanding of protein dynamics.

In this work, we have shown how CABS-flex effectively integrates into research workflows, complementing traditional experimental techniques such as x-ray crystallography, NMR spectroscopy, and cryo-electron microscopy (cryo-EM). This integration proves especially valuable in scenarios where experimental data are sparse, or where traditional methods like crystallization fail due to the dynamic nature of the target proteins. Additionally, CABS-flex has demonstrated its utility in studying the effects of mutations, further enhancing its versatility in addressing various biological questions. In the various workflows presented, CABS-flex is primarily used to simulate the flexibility of globular proteins at a low computational cost, offering a practical and less resource-intensive alternative to the more computationally demanding all-atom methods. This capability allows for the exploration of a wide range of protein conformations and provides insights into dynamics that are often inaccessible through conventional experimental approaches alone, making CABS-flex a useful tool in computational biology for studying protein dynamics with reduced computational overhead.

Despite its advantages, the coarse-grained nature of CABS-flex, which relies on a knowledge-based force field, might not capture all detailed molecular interactions. This limitation is particularly evident in scenarios involving transient interactions, highly dynamic regions, or intricate regulatory mechanisms within proteins. Such interactions often require a detailed atomic-level understanding that coarse-grained models may oversimplify. This gap highlights the importance of supplementing CABS-flex with higher-resolution methods, such as all-atom MD simulations, to validate and refine computational predictions. Integrating these methods ensures a more comprehensive understanding of protein dynamics, particularly in understanding the subtle nuances of allosteric regulations and the complexity of protein–protein interaction interfaces.

From our experience, the adept application of distance restraints in CABS-flex simulations is crucial yet frequently overlooked by users. Often, a default set of distance restraints and simulation parameters is employed, optimized for the short-term dynamics of folded globular proteins as derived from all-atom MD simulations. Adjustments to these restraints are necessary for different protein systems and timescales to accommodate diverse functional states and conformations. Modifying these restraints presents challenges, requiring expertise and sometimes extensive customization. An interesting and important challenge is the design of distance restraints based on low-resolution experimental data, which can include techniques such as small-angle x-ray scattering (SAXS), cryo-EM, or cross-linking mass spectrometry (XL-MS). These methods provide valuable structural information that can guide the placement and adjustment of distance restraints. Recent significant advances in AI-based structural prediction have further expanded the possibilities for integrative modeling pipelines (Rout and Sali, 2019), which combine modeling methods with experimental data. For instance, AI-predicted models can now be integrated with data from MS (Allison et al., 2022) or cryo-electron tomography (Mosalaganti et al., 2022), setting the stage for simulating large-scale dynamics. Simplifying the process of adjusting restraints and their weights remains a key area for future development to enhance user accessibility and flexibility in simulation setups.

Additionally, exploring the integration of CABS-flex with advanced AI-based structural predictions, such as those from AF2, could be promising. The conceptual use

of AlphaFold's pLDDT scores to inform distance restraints in CABS-flex simulations might lead to more nuanced depictions of protein flexibility, particularly in regions where the predictive model's confidence varies. This approach could potentially enhance CABS-flex's capability to model protein dynamics more accurately, especially in areas that are critical for understanding functional mechanisms. Such integration could improve the utility of simulations, providing insights into protein behavior that are complementary to traditional experimental approaches.

As we advance, refining the CABS-flex algorithm and simplifying its user interface will be crucial for facilitating more precise simulations of large biological systems. Continued development will ensure that CABS-flex remains an useful tool in computational biology, aiding researchers in effectively exploring the dynamic nature of proteins and their complexes.

AUTHOR CONTRIBUTIONS

Chandran Nithin: Writing – original draft; writing – review and editing; visualization. **Rocco Peter Fornari:** Conceptualization; investigation; formal analysis; visualization; writing – original draft; writing – review and editing. **Smita P. Pilla:** Conceptualization; investigation; formal analysis; writing – original draft; writing – review and editing; visualization. **Karol Wroblewski:** Writing – original draft; investigation; formal analysis; visualization. **Mateusz Zalewski:** Investigation; visualization; writing – original draft. **Rafał Madaj:** Writing – original draft; investigation; visualization. **Andrzej Kolinski:** Writing – review and editing. **Joanna M. Macnar:** Writing – original draft; writing – review and editing; visualization. **Sebastian Kmiecik:** Conceptualization; writing – review and editing; writing – original draft; funding acquisition.

ACKNOWLEDGMENTS

S.K., R.P.F., S.P.P., C.N., and K.W. acknowledge funding from the National Science Centre, Poland [OPUS 2020/39/B/NZ2/01301]; S.K., J.M.M., and M.Z. acknowledge funding from the National Science Centre, Poland [2021/40/Q/NZ2/00078]; C.N. and S.P.P. acknowledge Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support [computational grant no. PLG/2024/016931].

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Chandran Nithin https://orcid.org/0000-0001-8212-6093
Rocco Peter Fornari https://orcid.org/0000-0002-3792-8262
Smita P. Pilla https://orcid.org/0000-0002-4520-3210
Karol Wroblewski https://orcid.org/0000-0001-5182-1376
Mateusz Zalewski https://orcid.org/0000-0002-3436-8829
Rafał Madaj https://orcid.org/0000-0002-6330-6344
Andrzej Kolinski https://orcid.org/0000-0002-8830-2315
Joanna M. Macnar https://orcid.org/0000-0003-1760-2668
Sebastian Kmiecik https://orcid.org/0000-0001-7623-0935

REFERENCES

Allison TM, Degiacomi MT, Marklund EG, Jovine L, Elofsson A, Benesch JLP, et al. Complementing machine learning-based structure predictions with native mass spectrometry. Protein Sci. 2022;31:e4333.

Amusengeri A, Astl L, Lobb K, Verkhivker GM, Tastan Bishop Ö. Establishing computational approaches towards identifying malarial allosteric modulators: a case study of *Plasmodium falciparum* Hsp70s. Int J Mol Sci. 2019;20:5574.

Badaczewska-Dawid A, Wróblewski K, Kurcinski M, Kmiecik S. Structure prediction of linear and cyclic peptides using CABS-flex. Brief Bioinform. 2024b;25:bbae003.

Badaczewska-Dawid AE, Garcia-Pardo J, Kuriata A, Pujols J, Ventura S, Kmiecik S. A3D database: structure-based predictions of protein aggregation for the human proteome. Bioinformatics. 2022b;38:3121–3.

Badaczewska-Dawid AE, Khramushin A, Kolinski A, Schueler-Furman O, Kmiecik S. Protocols for all-atom reconstruction and high-resolution refinement of protein–peptide complex structures. In: Kihara D, editor. Protein structure prediction. Methods in molecular biology. New York, NY: Springer US; 2020a. p. 273–87. https://doi.org/10.1007/978-1-0716-0708-4_16

Badaczewska-Dawid AE, Kolinski A, Kmiecik S. Computational reconstruction of atomistic protein structures from coarse-grained models. Comput Struct Biotechnol J. 2020b;18:162–76.

Badaczewska-Dawid AE, Kuriata A, Pintado-Grima C, Garcia-Pardo J, Burdukiewicz M, Iglesias V, et al. A3D model organism database (A3D-MODB): a database for proteome aggregation predictions in model organisms. Nucleic Acids Res. 2024a;52:D360–7.

Badaczewska-Dawid AE, Nithin C, Wroblewski K, Kurcinski M, Kmiecik S. MAPIYA contact map server for identification and visualization of molecular interactions in proteins and biological complexes. Nucleic Acids Res. 2022a;50:W474–82.

Bárcenas, O., Kuriata, A., Zalewski, M., Iglesias, V., Pintado-Grima, C., Firlik, G., Burdukiewicz, M., Kmiecik, S., & Ventura, S. Aggrescan4D: structure-informed analysis of pH-dependent protein aggregation. Nucleic Acids Research. 2024;52:W170–5. https://doi.org/10.1093/nar/gkae382

Berjanskii MV, Wishart DS. A simple method to predict protein flexibility using secondary chemical shifts. J Am Chem Soc. 2005;127:14970–1.

Berjanskii MV, Wishart DS. Application of the random coil index to studying protein flexibility. J Biomol NMR. 2008;40:31–48.

Burley SK, Bhikadiya C, Bi C, Bittrich S, Chao H, Chen L, et al. RCSB Protein Data Bank (RCSB.Org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. Nucleic Acids Res. 2023;51:D488–508.

Buß O, Rudat J, Ochsenreither K. FoldX as protein engineering tool: better than random based approaches? Comput Struct Biotechnol J. 2018;16:25–33.

Cai Y, Zhang J, Xiao T, Lavine CL, Rawson S, Peng H, et al. Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants. Science. 2021;373:642–8.

Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. Nature. 2022;602:657–63.

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010;66:12–21.

Chiti F, Dobson CM. Amyloid formation by globular proteins under native conditions. Nat Chem Biol. 2009;5:15–22.

Ciemny MP, Debinski A, Paczkowska M, Kolinski A, Kurcinski M, Kmiecik S. Protein-peptide molecular docking with large-scale conformational changes: the p53-MDM2 interaction. Sci Rep. 2016;6:37532.

Faienza F, Lambrughi M, Rizza S, Pecorari C, Giglio P, Salamanca Viloria J, et al. S-nitrosylation affects TRAP1 structure and ATPase activity and modulates cell response to apoptotic stimuli. Biochem Pharmacol. 2020;176:113869.

Fowler NJ, Sljoka A, Williamson MP. A method for validating the accuracy of NMR protein structures. Nat Commun. 2020;11:6321.

Fowler NJ, Sljoka A, Williamson MP. The accuracy of NMR protein structures in the Protein Data Bank. Structure. 2021;29:1430–1439.e2.

Fowler NJ, Williamson MP. The accuracy of protein structures in solution determined by AlphaFold and NMR. Structure. 2022;30:925–933.e2.

Garcia-Pardo J, Badaczewska-Dawid AE, Pintado-Grima C, Iglesias V, Kuriata A, Kmiecik S, et al. A3DyDB: exploring structural aggregation propensities in the yeast proteome. Microb Cell Factories. 2023;22:186.

Guo H-B, Perminov A, Bekele S, Kedziora G, Farajollahi S, Varaljay V, et al. AlphaFold2 models indicate that protein sequence determines both structure and dynamics. Sci Rep. 2022;12:10696.

He X, Hong W, Pan X, Lu G, Wei X. SARS-CoV-2 omicron variant: characteristics and prevention. MedComm. 2021;2:838–45.

Heo L, Feig M. One bead per residue can describe all-atom protein structures. Structure. 2024;32:97–111.e6.

Høie MH, Kiehl EN, Petersen B, Nielsen M, Winther O, Nielsen H, et al. NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. Nucleic Acids Res. 2022;50:W510–5.

Huang Y, Yang C, Xu X, Xu W, Liu S. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin. 2020;41:1141–9.

Jamroz M, Kolinski A, Kmiecik S. CABS-flex: server for fast simulation of protein structure fluctuations. Nucleic Acids Res. 2013b;41:W427–31.

Jamroz M, Kolinski A, Kmiecik S. CABS-flex predictions of protein flexibility compared with NMR ensembles. Bioinformatics. 2014;30:2150–4.

Jamroz M, Orozco M, Kolinski A, Kmiecik S. Consistent view of protein fluctuations from all-atom molecular dynamics and coarse-grained dynamics with knowledge-based force-field. J Chem Theory Comput. 2013a;9:119–25.

Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292:195–202.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.

Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22:2577–637.

Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. Lancet. 2021;398:2126–8.

Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-grained protein models and their applications. Chem Rev. 2016;116:7898–936.

Kmiecik S, Gront D, Kouza M, Kolinski A. From coarse-grained to atomic-level characterization of protein dynamics: transition state for the folding of B domain of protein A. J Phys Chem B. 2012;116:7026–32.

Kmiecik S, Kolinski A. Characterization of protein-folding pathways by reduced-space modeling. Proc Natl Acad Sci. 2007;104:12330–5.

Kmiecik S, Kolinski A. Folding pathway of the B1 domain of protein G explored by multiscale modeling. Biophys J. 2008;94:726–36.

Kolinski A. Protein modeling and structure prediction with a reduced representation. Acta Biochim Pol. 2004;51:349–71.

Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020;182:812–827.e19.

Kurcinski M, Badaczewska-Dawid A, Kolinski M, Kolinski A, Kmiecik S. Flexible docking of peptides to proteins using CABS-dock. Protein Sci. 2020;29:211–22.

Kurcinski M, Kolinski A, Kmiecik S. Mechanism of folding and binding of an intrinsically disordered protein as revealed by ab initio simulations. J Chem Theory Comput. 2014;10:2224–31.

Kurcinski M, Oleniecki T, Ciemny MP, Kuriata A, Kolinski A, Kmiecik S. CABS-flex standalone: a simulation environment for fast modeling of protein flexibility. Bioinformatics. 2019;35:694–5.

Kuriata A, Gierut AM, Oleniecki T, Ciemny MP, Kolinski A, Kurcinski M, et al. CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures. Nucleic Acids Res. 2018;46:W338–43.

Kuriata A, Iglesias V, Kurcinski M, Ventura S, Kmiecik S. Aggrescan3D standalone package for structure-based prediction of protein aggregation properties. Bioinformatics. 2019b;35:3834–5.

Kuriata A, Iglesias V, Pujols J, Kurcinski M, Kmiecik S, Ventura S. Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility. Nucleic Acids Res. 2019a;47:W300–7.

Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023;379:1123–30.

Ma P, Li D-W, Brüschweiler R. Predicting protein flexibility with AlphaFold. Proteins. 2023;91:847–55.

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022;19:679–82.

Mishra PM, Anjum F, Uversky VN, Nandi CK. SARS-CoV-2 spike mutations modify the interaction between virus spike and human ACE2 receptors. Biochem Biophys Res Commun. 2022; 620:8–14.

Monzon AM, Rohr CO, Fornasari MS, Parisi G. CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state. Database. 2016;2016:baw038.

Mosalaganti S, Obarska-Kosinska A, Siggel M, Taniguchi R, Turoňová B, Zimmerli CE, et al. AI-based structure prediction empowers integrative structural analysis of human nuclear pores. Science. 2022;376:eabm9506.

Navarro S, Ventura S. Computational re-design of protein structures to improve solubility. Expert Opin Drug Discov. 2019;14:1077–88.

Ortega JT, Jastrzebska B, Rangel HR. Omicron SARS-CoV-2 variant spike protein shows an increased affinity to the human ACE2 receptor: an in silico analysis. Pathogens. 2022;11:45.

Papaleo E, Tiberti M, Arnaudi M, Pecorari C, Faienza F, Cantwell L, et al. TRAP1 S-nitrosylation as a model of population-shift mechanism to study the effects of nitric oxide on redox-sensitive oncoproteins. Cell Death Dis. 2023;14:284.

Park S-J, Kern N, Brown T, Lee J, Im W. CHARMM-GUI PDB manipulator: various PDB structural modifications for biomolecular modeling and simulation. J Mol Biol. 2023;435:167995.

Puławski W, Koliński A, Koliński M. Integrative modeling of diverse protein-peptide systems using CABS-dock. PLoS Comput Biol. 2023;19:e1011275.

Rizza S, Montagna C, Cardaci S, Maiani E, Di Giacomo G, Sanchez-Quiles V, et al. *S*-nitrosylation of the mitochondrial chaperone TRAP1 sensitizes hepatocellular carcinoma cells to inhibitors of succinate dehydrogenase. Cancer Res. 2016;76:4170–82.

Rout MP, Sali A. Principles for integrative structural biology studies. Cell. 2019;177:1384–403.

Ruff KM, Pappu RV. AlphaFold and implications for intrinsically disordered proteins. J Mol Biol. 2021;433:167208.

Sala D, Engelberger F, Mchaourab HS, Meiler J. Modeling conformational states of proteins with AlphaFold. Curr Opin Struct Biol. 2023;81:102645.

Saldaño T, Escobedo N, Marchetti J, Zea DJ, Mac Donagh J, Velez Rueda AJ, et al. Impact of protein conformational diversity on AlphaFold predictions. Bioinformatics. 2022;38:2742–8.

Sanyal D, Banerjee S, Bej A, Chowdhury VR, Uversky VN, Chowdhury S, et al. An integrated understanding of the evolutionary and structural features of the SARS-CoV-2 spike receptor binding domain (RBD). Int J Biol Macromol. 2022;217: 492–505.

Schwarz D, Georges G, Kelm S, Shi J, Vangone A, Deane CM. Co-evolutionary distance predictions contain flexibility information. Bioinformatics. 2021;38:65–72.

Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. J Chem Theory Comput. 2014;10:4745–58.

Singh J, Paliwal K, Litfin T, Singh J, Zhou Y. Reaching alignment-profile-based accuracy in predicting protein secondary and tertiary structural properties without alignment. Sci Rep. 2022;12:7607.

Timmons PB, Hewage CM. APPTEST is a novel protocol for the automatic prediction of peptide tertiary structures. Brief Bioinform. 2021;22:bbab308.

Tortorici MA, Czudnochowski N, Starr TN, Marzi R, Walls AC, Zatta F, et al. Broad sarbecovirus neutralization by a human monoclonal antibody. Nature. 2021;597:103–8.

Vander Meersche Y, Cretin G, Gheeraert A, Gelly J-C, Galochkina T. ATLAS: protein flexibility description from atomistic molecular dynamics simulations. Nucleic Acids Res. 2024;52:D384–92.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50:D439–44.

Verkhivker GM. Conformational flexibility and local frustration in the functional states of the SARS-CoV-2 spike B.1.1.7 and B.1.351 variants: mutation-induced allosteric modulation mechanism of functional dynamics and protein stability. Int J Mol Sci. 2022a;23:1646.

Verkhivker GM. Conformational dynamics and mechanisms of client protein integration into the Hsp90 chaperone controlled by allosteric interactions of regulatory switches: perturbation-based network approach for mutational profiling of the Hsp90 binding and allostery. J Phys Chem B. 2022b;126:5421–42.

Verkhivker GM. Exploring mechanisms of allosteric regulation and communication switching in the multiprotein regulatory complexes of the Hsp90 chaperone with cochaperones and client proteins: atomistic insights from integrative biophysical modeling and network analysis of conformational landscapes. J Mol Biol. 2022c;434:167506.

Verkhivker GM, Agajanian S, Kassab R, Krishnan K. Landscape-based protein stability analysis and network modeling of multiple conformational states of the SARS-CoV-2 spike D614G mutant: conformational plasticity and frustration-induced allostery as energetic drivers of highly transmissible spike variants. J Chem Inf Model. 2022;62:1956–78.

Verkhivker GM, Di Paola L. Dynamic network modeling of allosteric interactions and communication pathways in the SARS-CoV-2 spike trimer mutants: differential modulation of conformational landscapes and signal transmission via cascades of regulatory switches. J Phys Chem B. 2021a;125:850–73.

Verkhivker GM, Di Paola L. Integrated biophysical modeling of the SARS-CoV-2 spike protein binding and allosteric interactions with antibodies. J Phys Chem B. 2021b;125:4596–619.

Warnecke A, Sandalova T, Achour A, Harris RA. PyTMs: a useful PyMOL plugin for modeling common post-translational modifications. BMC Bioinformatics. 2014;15:370.

Webb B, Sali A. Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics. 2016;54:5.6.1–5.6.37.

Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367:1260–3.

Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S. AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. Nucleic Acids Res. 2015;43:W306–13.

Zhang N-N, Zhang R-R, Zhang Y-F, Ji K, Xiong X-C, Qin Q-S, et al. Rapid development of an updated mRNA vaccine against the SARS-CoV-2 omicron variant. Cell Res. 2022;32:401–3.

How to cite this article: Nithin C, Fornari RP, Pilla SP, Wroblewski K, Zalewski M, Madaj R, et al. Exploring protein functions from structural flexibility using CABS-flex modeling. Protein Science. 2024;33(9):e5090. https://doi.org/10.1002/pro.5090