

REVIEW ARTICLE

Functional, pathogenic, and pharmacological roles of protein folding intermediates

Emiliano Biasini¹ | Pietro Faccioli^{2,3} 

¹Department of Cellular Computational and Integrative Biology (CIBIO), University of Trento, Trento, Italy

²Department of Physics, University of Trento, Trento, Italy

³Trento Institute for Fundamental Physics and Applications, Italian Institute for Nuclear Physics, Trento, Italy

Correspondence

Pietro Faccioli, Physics Department, University of Trento, Trento, Italy.

Email: pietro.faccioli@unitn.it

Funding information

Fondazione Telethon; American CJD Foundation; Italian Association for Prion Encephalopathies

Abstract

Protein expression and function in eukaryotic cells are tightly harmonized processes modulated by the combination of different layers of regulation, including transcription, processing, stability, and translation of messenger RNA, as well as assembly, maturation, sorting, recycling, and degradation of polypeptides. Integrating all these pathways and the protein quality control machinery, deputed to avoid the production and accumulation of aberrantly folded proteins, determines protein homeostasis. Over the last decade, the combined development of accurate time-resolved experimental techniques and efficient computer simulations has opened the possibility of investigating biological mechanisms at atomic resolution with physics-based models. A meaningful example is the reconstruction of protein folding pathways at atomic resolution, which has enabled the characterization of the folding kinetics of biologically relevant globular proteins consisting of a few hundred amino acids. Combining these innovative computational technologies with rigorous experimental approaches reveals the existence of non-native metastable states transiently appearing along the folding process of such proteins. Here, we review the primary evidence indicating that these protein folding intermediates could play roles in disparate biological processes, from the posttranslational regulation of protein expression to disease-relevant protein misfolding mechanisms. Finally, we discuss how the information encoded into protein folding pathways could be exploited to design an entirely new generation of pharmacological agents capable of promoting the selective degradation of protein targets.

KEYWORDS

drug discovery, misfolding, molecular simulations, protein folding, protein homeostasis

1 | INTRODUCTION

In its classical definition, the native state of a protein refers to the three-dimensional conformation of the polypeptide that determines its biological activity. Such a conformation is reached at the end of a complex series of structural transitions that each chain of amino acids

undergoes soon after exiting the ribosomal channel, collectively referred to as the protein folding process. A correct folding pathway for newly synthesized proteins is fundamental for maintaining biological activities, influencing the rate and efficiency of protein expression and functions. This process is, in fact, closely monitored by multiple cellular mechanisms, which intervene to correct aberrantly folded

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Proteins: Structure, Function, and Bioinformatics* published by Wiley Periodicals LLC.

species or to reroute them to degradation in the proteasome or the lysosomes. These mechanisms collectively guarantee protein homeostasis, also called proteostasis.

Theoretical and experimental evidence gathered over the last 30 years have established that the thermodynamics and kinetics of the folding of single-domain globular proteins are determined by the polypeptide chain's multidimensional free-energy landscape, which is funneled toward the native state.¹ The critical hypothesis underlying the so-called energy landscape theory is that natural evolution has selected sequences of amino acids that minimize the intrinsic frustration associated with ordinary heteropolymers. The result is a highly cooperative transition that enables the chain to effectively reach its minimum energy structure by avoiding the random exploration of the whole space of configurations. In the extreme limit of a completely smooth energy landscape, proteins can roll down their folding funnel from any unfolded configuration and reach the native state via virtually infinite different folding pathways. An important prediction of this idealized model is that the folding transition should obey simple two-state kinetics, a feature confirmed by physics-based microscopic simulations² and by experiments carried out on small globular proteins.³

In contrast, understanding the folding of larger or multi-domain proteins has proven much more challenging, both from theoretical and experimental points of view. In fact, the relevant time scales of protein folding can extend to seconds, minutes, or even longer. Moreover, the folding kinetics is often characterized by the population of partially folded intermediates,⁴ which can be prone to misfolding and aggregation.^{5,6} The emerging picture is one in which the energy landscape of biologically relevant proteins is much more rugged than that of small globular proteins. This feature enables the emergence of predominant folding pathways. In particular, the folding of large proteins seems to occur through the rapid and cooperative formation of a few small structural building blocks (foldons), which then dock together to form the structure.⁷ All these concepts imply that the landscape of conformations that a protein may explore during the folding process includes a variety of dynamically interconverting metastable species, different from the native state, whose abundance and thermodynamic principles govern kinetics. Such a perspective suggests that if the docking of foldons occurs through relatively consistent steps, intermediate conformations appearing along the process could play fundamental roles by affecting both the rate and efficiency of protein folding. While the relevance of the whole folding process in modulating proteostasis is conceptually established, there is so far insufficient evidence that non-native protein states appearing along the folding pathways might hold relevant biological information and be directly implicated in functional or regulatory events.

In the last three decades, computer simulations have played a growing role in establishing a physics-rooted understanding of protein folding. In particular, molecular dynamics (MD) simulations can provide an accurate atomically detailed representation of the folding transition, thus elucidating the energy landscape's structure and identifying the prominent folding mechanisms. Unfortunately, even relying on the most powerful special-purpose supercomputer, all-atom MD simulations are extremely computationally expensive and can only be used to study the folding of small polypeptide chains with folding

times up to the millisecond scale.² Thus, while plain MD simulations can offer valuable insights into the folding kinetics of very small chains or the cooperative formation of protein domains or foldons, they do not provide a viable tool to investigate the kinetics involving the interconversion between partially folded structures. The development of computational enhanced-sampling techniques has been partially able to overcome those limitations, allowing for interesting mechanistic insights into the folding process of biologically relevant proteins.⁸

The combination of these innovative computational technologies with rigorous experimental approaches has recently provided evidence indicating that protein folding intermediates could play roles in disparate physiological and pathological processes. We refer to these protein conformers as functional protein folding intermediates to show precise protein structures appearing along the folding process and distinct from the native state that serves a specific function in the cell. Importantly, these intermediates are not considered misfolded but rather transiently populated in a dynamic equilibrium with the native state. This manuscript will review relevant literature regarding functional protein folding intermediates and their involvement in the posttranslational regulation of protein expression and protein misfolding disorders. We will also discuss new and unexpected pharmacological paradigms arising from the knowledge of functional protein folding intermediates and their unique biology.

The manuscript is organized as follows. In Section 2, we briefly review some enhanced sampling schemes and approximations that have enabled, for the first time, the all-atom simulations of large proteins and the structural characterization of folding intermediates. Sections 2 and 3 discuss specific study cases, highlighting how folding intermediates can play an essential role in protein function, (mis-)regulation, and disease-relevant aggregation. Section 4 presents the hypothesis that folding intermediates can provide an unappreciated role in regulating protein homeostasis via posttranslational modifications (PTMs). In Section 5, we discuss how these studies have led to the development of an entirely new paradigm for drug discovery based on reducing the expression level of proteins by targeting folding intermediates. Conclusions and outlooks are summarized in Section 6.

2 | SIMULATIONS OF PROTEIN FOLDING, MISFOLDING, AND AGGREGATION WITH ALL-ATOM PHYSICS-BASED MODELS

All-atom MD simulations can aptly complement experiments investigating the kinetics of complex protein transitions, including in principle folding, misfolding and aggregation. The advantage of MD simulations is that they rely on physics-based theoretical modeling to yield the atomically resolved reconstruction of structural rearrangements. Unfortunately, in practice, even depending on the most powerful special-purpose supercomputer (Anton, developed by D.E.S. Research), MD simulations can only cover time intervals of up to a few milliseconds for polypeptide chains consisting of nearly 100 amino acids,² which are considerably smaller than most proteins of biological interest. While the next generations of the Anton supercomputer may be able to increase the time windows accessible to

MD, it is unlikely that a brute-force MD scheme will enable the simulation of many critical biomolecular processes that can occur at the time scales of seconds or longer and can involve several hundreds of amino acids.

Given the computational limitations of MD, more sophisticated alternative algorithms have been developed and then applied to investigate rare biological transitions.⁹ Most of these schemes typically require complementing MD simulations with additional information, such as identifying the slowest collective variables (CVs) in the system or choosing a good reaction coordinate. To the best of our knowledge, to date, the only enhanced sampling technique that has been systematically applied to perform all-atom simulations of the folding, misfolding, and aggregation of proteins consisting of several hundreds of amino acids are the so-called bias functional (BF) approach¹⁰ and its evolution called self-consistent path sampling (SCPS).¹¹ In the BF scheme, a specific type of biased dynamics called ratchet-and-pawl MD (rMD) is employed to efficiently generate a statistically significant number of productive trajectories. In rMD, an auxiliary history-dependent potential depending on a suitable CV is introduced to prevent the chain from backtracking toward the initial state.¹² Conversely, the biasing force is inactive when the system spontaneously progresses toward the final state. It has been shown that, in the ideal case in which the CV is an ideal reaction coordinate (the so-called committor function), rMD simulations yield the correct Boltzmann sampling in the region explored by the transition path ensemble.^{13,14} On the other hand, in practical applications, rMD must rely on a proxy of the ideal reaction coordinate, so it yields only an approximate sampling of the equilibrium distribution. The BF approach provides a scheme to keep the systematic errors of rMD to a minimum. In this approach, a variational principle derived from Langevin dynamics is used for scoring the reactive trajectories generated by rMD to identify those with the highest probability of occurring in the absence of any biasing force.¹⁰

SCPS exploits a different strategy to mitigate the systematic errors of plain rMD simulations: It is based on an iterative procedure through which the reactive rMD trajectories obtained in the previous iteration are analyzed to obtain a better proxy of the reaction coordinate, which is then used in the next iteration of rMD to generate improved reactive trajectories. It has been shown that the folding mechanism obtained by BF and SCPS schemes is statistically indistinguishable from that predicted by the plain MD simulations performed on the Anton Supercomputer.¹¹ The same enhanced sampling schemes have also been extensively validated against the result of different biophysical experiments.^{15–17} SCPS and BF have then been applied to several case studies of biological interest to be discussed below, to yield a complete structural characterization of protein folding intermediates.

3 | FUNCTIONAL ROLE OF FOLDING INTERMEDIATE IN PROTEIN AGGREGATION, FUNCTION, AND REGULATION, CASE I: PRIONS

Proteins evolve under two kinds of evolutionary pressure, one selecting their biological function, the other challenging their ability to remain

soluble under physiological conditions. However, in several pathologies, specific proteins lose their native fold and acquire a different conformation, ultimately clustering into aberrant aggregates.¹⁸ The phenomenon, known as protein misfolding, lies at the root of a wide variety of human diseases, such as neurodegenerative disorders, in which protein aggregation occurs in the brain. Prion diseases, also known as transmissible spongiform encephalopathies, represent a meaningful example. These disorders have the peculiarity of manifesting in a sporadic, inherited, or transmissible fashion. All forms of prion diseases, including Creutzfeldt-Jakob disease (CJD), fatal familial insomnia, and Gerstmann-Sträussler-Scheinker syndrome, are associated with the conformational conversion of the cellular prion protein (PrP), a cell surface glycoprotein mainly expressed in the central nervous system, into a misfolded isoform (PrP^{Sc}) that accumulates in the brain of affected individuals.¹⁹ PrP^{Sc} is a proteinaceous infectious particle (prion) capable of multiplying by directly recruiting other PrP molecules and causing their conformational rearrangement into growing PrP^{Sc} polymers.^{20,21} Despite the recent elucidation by cryogenic electron microscopy of the structure of PrP^{Sc} as a parallel-in-register- β -sheet fiber^{22,23} and decades of research from multiple laboratories, the definition of the mechanism underlying prion replication at atomic resolution has proven to be an unprecedented experimental challenge, mainly due to the lack of techniques combining deep spatial and temporal resolution.²⁴

Previous results collected using stopped-flow and force spectroscopy techniques suggested that a relatively stable folding intermediate is present along the folding pathway of PrP.^{25–27} However, they failed to provide high-resolution structural information about it. Interestingly, three more recent studies further supported the existence of a PrP folding intermediate. In the first one,²⁸ an intermediate folding conformer of a disease-relevant PrP mutant (T183A) was identified using NMR spectroscopy, biophysics, and metadynamics simulations. Strikingly, this conformer was nearly identical to a PrP folding intermediate identified in coarse-grained²⁹ and in all-atom simulations,³⁰ the latter based on the BF approach described in Section 2. Interestingly, computational and biochemical experiments revealed that such a PrP conformer might also spontaneously appear during the structural fluctuation of the protein starting from its native state, implying that its appearance in a cell may not be strictly limited to the subcellular compartment where the protein folds (i.e., the endoplasmic reticulum) but also occur on the cell surface, where the mature protein typically resides.^{28,30} Collectively, these data suggest that in addition to its role in the normal folding pathway of the protein, a folding intermediate could be involved in other biological aspects of PrP, including prion conversion, by acting as a direct substrate for PrP^{Sc} replication.

4 | FUNCTIONAL ROLE OF FOLDING INTERMEDIATE IN PROTEIN AGGREGATION, FUNCTION AND REGULATION, CASE II: SERPINS

Serpins (serine protease inhibitors) provide another notable example of proteins that are associated with a family of misfolding pathologies

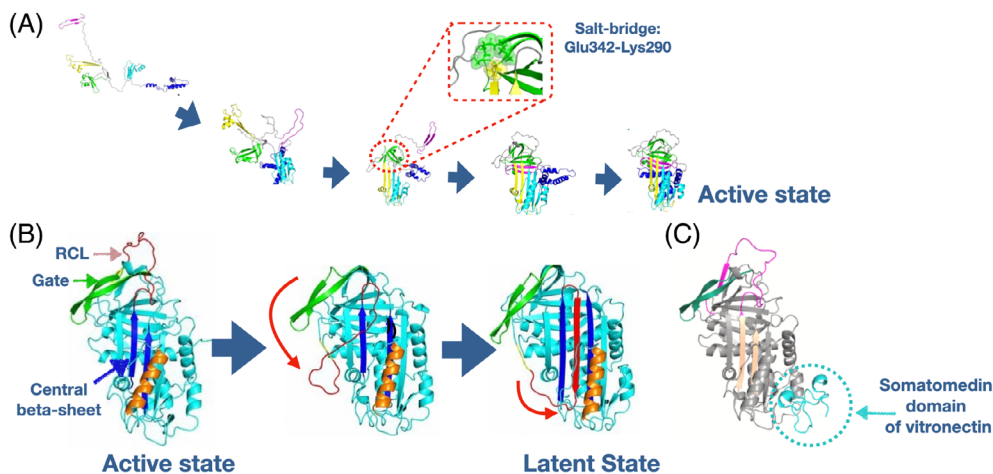


FIGURE 1 All-atom simulations serpin folding and latency transition. Panel (A): Schematic representation of a typical folding pathway for A1AT serpin observed in bias functional (BF) simulation. The red-dotted circle highlights the formation of the salt-bridge that is affected by the Z- and S- mutation. Panel (B): Schematic representation of the spontaneous latency transition of plasminogen activator 1 (PAI-1) serpin, displaying the existence of a meta-stable pre-latent intermediate. (C) The binding of Vitronectin to PAI-1 slows the latency transition by allosterically rigidifying the reactive center loop (RCL) and gate mobility.

(serpinopathies) that are induced mainly by the pathogenic dysregulation of their target proteases.³¹ The most common serpinopathies are induced by mutations in the canonical secretory serpin α 1-antitrypsin (A1AT) responsible for regulating leukocyte serine proteases. For example, the relatively benign S-mutation (Glu264Val) promotes the population of a state of misfolded conformations structurally and kinetically close to the native state. The resulting lung diseases can be slowed (although not wholly halted) by A1AT augmentation therapy.³² On the other hand, the highly pathogenic Z-mutation (Glu342Lys) is associated with an irreversible and aggregation-prone misfolded conformation. The resulting shallow levels of circulating active A1AT lead to severe liver diseases, for which the only known effective treatment is liver transplants.³³

Using the BF algorithm described in Section 2, it has become possible to simulate many folding and misfolding events with realistic all-atom models for different A1AT variants, including the wild-type, the S-mutant, and the Z-mutant.³⁴ These theoretical predictions were first validated using the available biophysical experiments to refold wild-type serpins and later applied to elucidate the chemical-physical mechanism responsible for the misfolding of S and Z mutants. In agreement with experimental observation, the Z mutation was found in simulations to disrupt the folding efficiency. In particular, an in-depth analysis of the folding trajectories revealed that the detouring from the correct folding pathway occurs at a relatively early stage of the transition (Figure 1A). A1AT misfolding is driven by the fact that the Z mutation affects an electrostatic interaction between two residues, changing it from attractive (Glu-Lys) into repulsive (Lys-Lys). This repulsion prevents the two residues from coming in contact at an early folding stage. If this contact does not occur early in the reaction, the chain cannot reach the native conformation. On the other hand, in the BF simulations for the S-mutant, a much more significant fraction of successful folding events was observed. The predicted misfolded

conformations were much more structurally close to the native state and less aggregation-prone. The comparative analysis of the successful and unsuccessful folding trajectories revealed that misfolding is associated with an alteration of the same electrostatic interaction involved in the misfolding of the Z-mutant. However, the S mutation does not turn the attractive Glu-Lys interaction into a repulsive one (which in Z-mutant is Lys-Lys) because the charged residue (Glu) is replaced by a neutral one (Val), so these two residues are not as strongly impeded to come close at the beginning of the folding process. As a result, the detour from the correct folding pathway occurs much more sporadically and at a later stage of the transition. The physicochemical mechanism underpinning the misfolding of A1AT mutants was validated by comparing experiments and simulations of several A1AT variants characterized by suitable rescue mutations. For example, the disruptive effect of the Z mutation could be rescued by replacing the Lys residue involved in the repulsive electrostatic interaction with a Glu. This way, the Glu-Lys pair is turned into a Lys-Glu, and the attractive electrostatic interaction is preserved. Proteins with such double mutation were predicted and observed to have a folding efficiency comparable to that of the wild-type.

Another exciting feature of serpin proteins is that their physiological function and regulation rely on meta-stable chain conformations. Indeed, the interaction with the target protease typically occurs in the so-called reactive center loop (RCL) and induces a large “mouse-trap” conformational transition to an inactive stable state, in which the RCL is inserted in a central beta-sheet (Figure 1B). To reach the central beta-sheet, the RCL has to overcome an exposed beta-sheet referred to as the gate. After this irreversible conformational change, the serpin becomes biologically inert. In most cases, the latency transition is initiated by the cleavage of the RCL, which releases a significant amount of mechanical energy. However, in some members of the serpin family, most notably plasminogen activator 1 serpin (PAI-1), the

transition can also occur spontaneously, albeit on a much longer time scale (in the range of hours). Hence, from a kinetic and thermodynamic point of view, the active state of PA1-1 may be regarded as a folding intermediate.

PAI-1 negatively regulates blood clot clearance by mechanically inhibiting several serine proteases, including tissue-type plasminogen activator and urokinase-type plasminogen activator.³⁵ The spontaneous latency transition, therefore, provides a facile way to regulate PAI-1 activity. In physiological conditions, it is achieved by binding to the cell adhesion factor vitronectin, which leads to a significant increase in the active state half-life. Since high levels of active PAI-1 are associated both with poor cancer prognosis and cardiovascular diseases, PAI-1 inhibitors that accelerate the latency transition are under development.^{7,35,36} For example, incubation of active PAI-1 with the antibody H4B3 has been shown to accelerate the latency transition in a concentration-dependent manner. Unfortunately, the lack of detailed molecular mechanisms for PAI-1 conformational changes has limited drug design efforts. For example, it has been noted that the epitope recognized by H4B3 is not accessible in PAI-1 active state, thus raising the question of how this antibody can accelerate the transition. Simulations performed in a native-centric coarse-grained model have confirmed that direct folding to the latent structure is disfavored by the presence of a large, entropy-dominated free-energy barrier.³⁷ On the other hand, simulations performed using the BF scheme have made it possible for the first time to obtain a fully atomistic reconstruction of the PAI-1 latency transition and provide a viable Physicochemical mechanism of action of H4B3.³⁸ Indeed, these simulations showed that the latency transition involves an on-pathway long-lived intermediate state in which the RCL is only partially inserted in the central loop (Figure 1C), where the interaction with H4B3 can take place. This binding prevents the protein from backtracking to the active state, thus promoting the latency transition's reaction rate.

Another puzzling feature of the PAI-1 latency transition concerns how the interaction with the cell adhesion factor vitronectin can slow the latency transition. This question was recently addressed by comparing BF simulations of the latency simulation with and without vitronectin.³⁹ An analysis of the productive trajectories revealed that vitronectin binding reduces RCL and gate mobility by allosterically rigidifying structural elements over 40 Å away from the binding site, thus blocking the transition to the latent conformation. The same simulations also revealed a cryptic drug-binding area (identified by crystallography) populated only in the vitronectin-bound state.

5 | FUNCTIONAL FOLDING INTERMEDIATES AND PTMs

PTMs are chemical changes that may occur to a protein after it is translated from its corresponding RNA molecule. These modifications include adding or removing chemical groups or entire molecules by specific enzymes and can occur on the protein's amino acid side chains or backbone. Known types of PTMs include, among others,

proteolytic cleavage, glycosylation, phosphorylation, acetylation, methylation, sumoylation, and ubiquitination. PTMs can have a wide range of effects on a protein's structure, function, and stability. For example, adding a phosphate group can change a protein's activity or strength, while adding a ubiquitin moiety may reroute the protein to degradation. Notably, the specific combination of PTMs on a given protein is highly regulated and can change in response to different cellular stimuli. Overall, PTMs play a critical role in regulating protein activity, stability, and localization and can also significantly impact a wide range of cellular processes, such as signaling pathways, enzyme activity, and protein-protein interactions.

PTMs are generally thought to occur on native states of proteins. However, PTMs could theoretically also target long-lived protein folding intermediates. There is ample evidence, for example, that glycosyl- and glycosyl-phosphatidylinositol chains occur co-translationally as soon as a polypeptide starts exiting the ribosomal channel.⁴⁰ In principle, any PTM targeting a protein folding intermediate could alter the folding pathway, thus impacting the expression rate of the protein. Regulating the homeostasis of protein folding intermediates, rather than just acting on transcription, translation, or clearance of mature protein forms, could theoretically be energetically favorable and allow the cell to respond rapidly to stimuli. Interesting evidence supporting this hypothesis has recently emerged from the analysis of the mtcPTM database, which contains information about phosphorylation sites identified through mass spectrometry-based experiments.⁴¹ A significant proportion (~15%) of all known phosphorylation sites exhibits solvent accessibility scores of their side chains lower than 10% in the unmodified form of the protein and under physiological conditions. Such a high proportion of hidden phosphosites is striking and unexpected, as an inaccessible positioning would not only dramatically hamper the ability of kinases to reach their target sites but also generate an energetically unfavorable environment for hosting the steric and electrostatic properties of a phosphate group. A plausible explanation for such observation is that the target structures of these phosphorylation events on hidden residues are non-native, functional folding intermediate conformers.

Direct evidence for a functional role of protein folding intermediates in PTM-mediated regulation of protein expression emerged from the analysis of the regulatory mechanism of the androgen receptor (AR), a paradigmatic example of proteins undergoing folding-upon-binding transition. This AR is responsible for the hormone-dependent transcription of genes involved in cell proliferation and differentiation. Once produced in the Leydig cells of the testes, testosterone circulates in the bloodstream bound to serum sex hormone-binding globulin. The free form of this hormone can cross the cell membrane and may be converted into 5-dihydrotestosterone (DHT) in the cytoplasm. The AR's inactive form resides in the cytoplasm bound to heat shock proteins, which are then displaced upon binding of the AR to DHT, leading the receptor to reach its functional conformation, translocate into the nucleus, and activate transcription of target genes. Ligand binding to the AR stabilizes its conformation by inducing an intramolecular structural reorganization of helix 12 that caps the ligand-binding pocket.

Additionally, the AR could be regulated by several PTMs, including phosphorylation, ubiquitination, SUMOylation, acetylation, and methylation. In one of these cases, it has been shown that serine 791 (S791) within the ligand-binding domain of the AR could be phosphorylated by protein kinase B. This modification inhibits ligand interaction and promotes polyubiquitination and degradation of the AR by the proteasome. Interestingly, in natively folded, ligand-bound AR, the S791 residue is buried inside the protein core, suggesting that regulation of this phosphorylation site likely happens before the AR binds to its ligand. This was the possibility offered by computational analyses of the AR folding pathway that revealed the existence of complex folding kinetics, including a folding intermediate in which residue S791 is exposed to the solvent.⁴² Therefore, residue S791 appears as a regulatory site representing a crossroad between a nonphosphorylated, viable ligand-binding folding pathway of AR and a phosphorylated, nonviable folding pathway leading to AR degradation.

6 | PHARMACOLOGICAL VALUE OF TARGETING FOLDING INTERMEDIATES

Conventional rational drug discovery is primarily based on identifying small molecules that can hinder the biological activity of a target protein by binding to pockets in its native structure. Unfortunately, this is not a viable therapeutic strategy for many pathologies. In particular, neurodegenerative diseases rooted in protein misfolding and aggregation are paradigmatic examples of pathologies for which conventional therapeutic methods are largely unsuccessful.

The observation that a protein's physiological regulation can occur at the level of the folding pathway inspired us to test the possibility of pharmacologically modulating the expression of a given protein by targeting folding intermediates.^{30,42,43} Strategies to modulate protein folding kinetics using small peptides have been based on theoretical simulations and in-vitro biophysical kinetic experiments. In particular, Das et al. employed a variety of biophysical experiments and coarse-grained simulations to show that small peptides that provide a template secondary motif in the folding nucleus of a target protein can speed up its folding kinetics.⁴⁴ Broglia et al. used a lattice-based coarse-grained model to theoretize the design of small peptides capable of docking to specific regions of a polypeptide and inhibiting its folding.⁴⁵

The design of small molecules that suppress the expression of target proteins by acting on folding pathways was made possible by the development of a completely new paradigm for rational drug discovery named pharmacological protein inactivation by folding intermediate targeting (PPI-FIT; Figure 2). In contrast to the conventional rational drug discovery approach, PPI-FIT is based on reducing the cellular level of a target protein through small molecules that bind to specific folding intermediates, thus triggering protein degradation via the cell's protein folding quality control machinery. The rationale is that the cellular quality control machinery recognizes artificially stabilized intermediates as improperly folded species. In practice, the PPI-

FIT protocol is made possible by a novel generation of molecular simulations based on the enhanced sampling schemes discussed in Section 2, which enable the prediction of the folding process of the target protein with an atomic degree of resolution. Virtual screening campaigns are then run to identify small molecules that can bind to pockets in long-lived protein configurations that are transiently visited only during the folding process and are not present in the native state. The PPI-FIT can be applied in virtually all therapeutic areas and protein targets, including those considered undruggable with conventional methods.

Indeed, the first proof-of-concept application of PPI-FIT led to the identification of small molecules capable of lowering the levels of the cellular prion protein in a dose-dependent fashion, with no detected effect on a control protein with identical biogenesis. Biochemical assays performed with one of these small molecules (named SM875) demonstrated that the level of PrP in different cell types is reduced by promoting the lysosomal-dependent degradation pathway. Further essays on inducible cell cultures showed that the molecule only acts on nascent proteins, while it shows no effects on native forms of the protein. Remarkably, SM875 also suppresses prion replication in cell cultures.³⁰ The PPI-FIT methodology is now being industrially exploited to identifying drug candidates in different therapeutic areas.⁴⁶

7 | CONCLUSIONS

The elucidation of the sequence of structural rearrangements through which globular proteins spontaneously reach their active native state (often referred to as the "Part II of the protein folding problem") has been challenging theoretical and experimental biophysicists and biochemists for several decades. Until recently, such an intense research activity has mostly focused on addressing fundamental chemical-physics questions: how can these polypeptide chains avoid the exhaustive exploration of their huge configuration space to efficiently and rapidly reach their native state? Is the folding mechanism sequential and deterministic, or is it completely heterogeneous and stochastic? What is the rate-limiting step in folding, and what is the origin of the underlying free-energy barrier? From a biological perspective, biological interest in understanding folding has mainly concerned pathogenic implications of misfolding. In contrast, the details of productive folding transitions have mostly been regarded as having little physiological relevance, assuming that the function and regulation of proteins only involved interactions with the polypeptide's native structures.

Over the last decade, however, several theoretical and experimental studies have challenged this view, suggesting that protein folding intermediates may be involved in many critical biological processes. In this perspective, A1AT serpin provides an extreme paradigmatic example of a biologically active protein in a metastable state on the folding pathway. Furthermore, in many cases, folding intermediates may be the only states where a polypeptide chain can undergo PTMs. In addition, many proteins are believed to be able to reach their

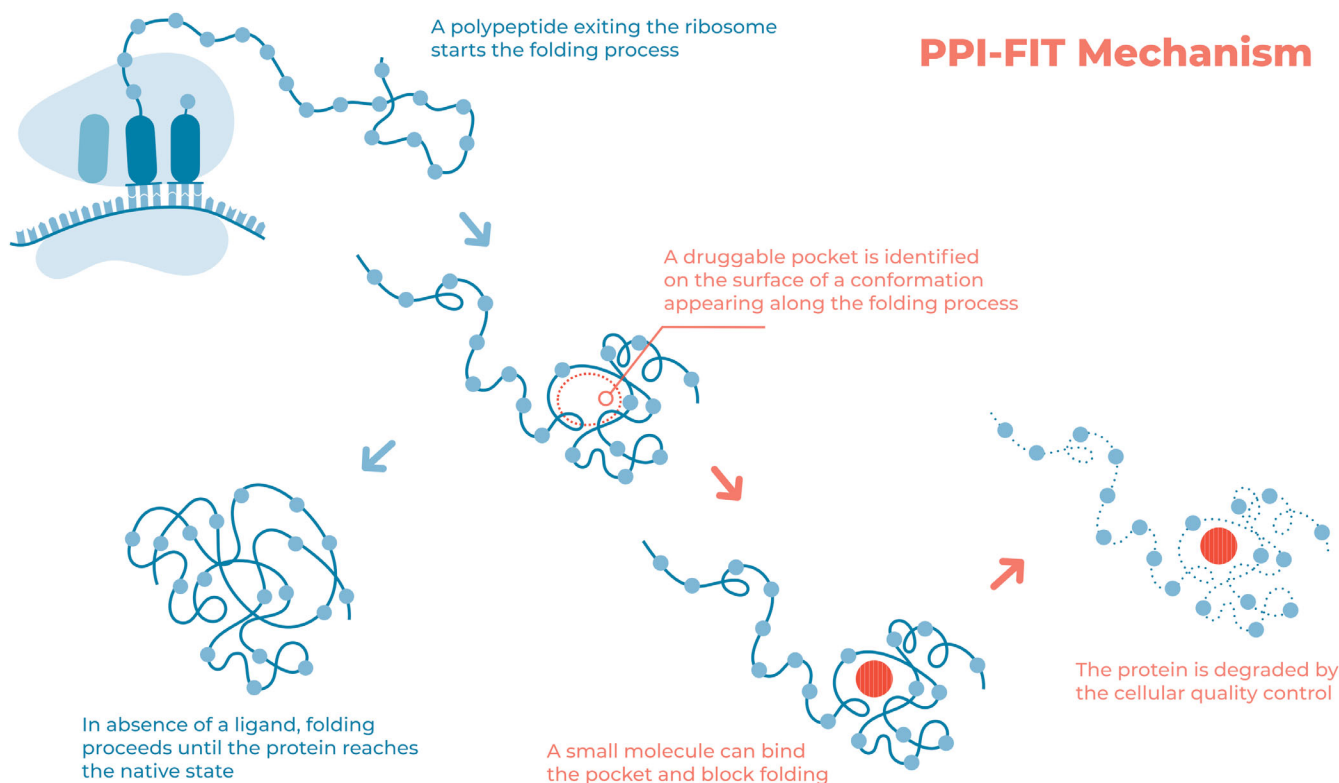


FIGURE 2 Rationale of targeting folding intermediates pharmacologically. The image illustrates the rationale underlying the pharmacological protein inactivation by folding intermediate targeting (PPI-FIT) paradigm. Polypeptides exiting the ribosomal channel start the folding process. Folding intermediates transiently appearing along this process could be targeted with small molecules binding to unique pockets on their surface. Such a binding event ultimately interferes with the protein folding process, prompting the quality control machinery of the cells to reroute the polypeptide to degradation pathways.

stable state only upon binding to specific targets, suggesting the existence of complex feedback-based regulatory mechanisms.

Reconstructing folding pathways at high resolution also provides an entirely new class of pharmacological targets, potentially paving the way for treating diseases for which conventional rational drug discovery schemes are inefficient or inapplicable. Examples involve proteins which do not display druggable pockets in their native structure or proteins responsible for misfolding diseases.

Several factors have driven this paradigmatic change of perspective on the role of protein folding intermediates. On the one hand, the development of sophisticated time-resolved experimental techniques has revealed that the folding kinetics of biologically relevant proteins is far more complex than that of small globular domains, as it typically involves many time scales and rate-limiting steps.⁴ On the other hand, the development of powerful algorithms and accurate models has made it possible to reconstruct these folding events at various degrees of resolution, ranging from relatively coarse-grained to fully atomistic. While the initial seminal theoretical activity of the previous decades focused on so-called native-centric models that emphasize the role of the native structure in shaping the folding pathways, several models or approaches developed in the last 15 years have made it possible for the first time to predict folding pathways while retaining the information about the chemical composition of the chain. This

progress paves the way to investigating the impact of mutations on folding kinetics and folding efficiency and the role of non-native interactions in shaping the folding pathway. Finally, various methods are currently being developed that coherently integrate experimental information into numerical simulations, for example, Bayesian statistics approaches.^{47,48} These schemes may correct for possible inaccuracies of the underlying atomistic models. At the same time, powerful machine learning and information-theory-based techniques are being exploited to obtain reliable multiscale and coarse-grained models (for a recent review, see Reference 49) to consistently analyze vast amounts of data generated in simulations⁵⁰ and to enhance sampling.⁵¹

In the next decade, the combination and integration of all these experimental and theoretical developments will likely have a great impact by generating a breakthrough in the appreciation of biological aspects connected with the folding mechanisms, potentially unveiling additional layers of complexity in the fundamental dogma of molecular biology.

ACKNOWLEDGMENTS

The authors thank Giovanni Spagnoli and Lidia Pieri from Sibylla Biotech for helping with the design of Figure 2. This work was supported by grants from Fondazione Telethon (GGP20043), the American CJD

Foundation, and the Italian Association for Prion Encephalopathies (Aienp) to E.B. The authors also thank Mrs Michela Acquaviva for a generous donation to the E.B. laboratory at University of Trento. Open Access Funding provided by Università degli Studi di Trento within the CRUI-CARE Agreement.

CONFLICT OF INTEREST

The authors are co-founders and share holders of Sibylla Biotech SPA, a drug discovery company.

ORCID

Pietro Faccioli  <https://orcid.org/0000-0002-5546-5054>

REFERENCES

- Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem.* 1997;48:545-600. doi:10.1146/annurev.physchem.48.1.545
- Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. *Science.* 2011;334:517-520. doi:10.1126/science.1208351
- Eaton WA. Modern kinetics and mechanism of protein folding: a retrospective. *J Phys Chem B.* 2021;125:3452-3467. doi:10.1021/acs.jpcc.1c00206
- Walters BT, Mayne L, Hinshaw JR, Sosnick TR, Englander SW. Folding of a large protein at high structural resolution. *Proc Natl Acad Sci U S A.* 2013;110:18898-18903. doi:10.1073/pnas.1319482110
- Yu H, Liu X, Neupane K, et al. Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc Natl Acad Sci U S A.* 2012;109:5283-5288. doi:10.1073/pnas.1107736109
- Pirchi M, Ziv G, Riven I, et al. Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat Commun.* 2011;2:493. doi:10.1038/ncomms1504
- Lin Z, Jensen JK, Hong Z, et al. Structural insight into inactivation of plasminogen activator inhibitor-1 by a small-molecule antagonist. *Chem Biol.* 2013;20:253-261. doi:10.1016/j.chembiol.2013.01.002
- Gershenson A, Gosavi S, Faccioli P, Wintrode PL. Successes and challenges in simulating the folding of large proteins. *J Biol Chem.* 2020;295:15-33. doi:10.1074/jbc.REV119.006794
- Yang YI, Shao Q, Zhang J, Yang L, Gao YQ. Enhanced sampling in molecular dynamics. *J Chem Phys.* 2019;151:070902. doi:10.1063/1.5109531
- Beccara SA, Fant L, Faccioli P. Variational scheme to compute protein reaction pathways using atomistic force fields with explicit solvent. *Phys Rev Lett.* 2015;114:098103. doi:10.1103/PhysRevLett.114.098103
- Orioli S, Beccara SA, Faccioli P. Self-consistent calculation of protein folding pathways. *J Chem Phys.* 2017;147:064108. doi:10.1063/1.4997197
- Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J Mol Biol.* 1999;288:441-459. doi:10.1006/jmbi.1999.2670
- Bartolucci G, Orioli S, Faccioli P. Transition path theory from biased simulations. *J Chem Phys.* 2018;149:072336. doi:10.1063/1.5027253
- Vanden Eijnden S, Martinovici D. Neonatal diabetes due to neonatal hemochromatosis. *J Pediatr Endocrinol Metab.* 2010;23:1207; author reply 1208.
- Wang F, Cazzolli G, Wintrode P, Faccioli P. Folding mechanism of proteins Im7 and Im9: insight from all-atom simulations in implicit and explicit solvent. *J Phys Chem B.* 2016;120:9297-9307. doi:10.1021/acs.jpcc.6b05819
- Ianeselli A, Orioli S, Spagnoli G, et al. Atomic detail of protein folding revealed by an ab initio reappraisal of circular dichroism. *J Am Chem Soc.* 2018;140:3674-3682. doi:10.1021/jacs.7b12399
- Dingfelder F, Macocco I, Benke S, Nettels D, Faccioli P, Schuler B. Slow escape from a helical misfolded state of the pore-forming toxin cytolysin A. *JACS Au.* 2021;1:1217-1230. doi:10.1021/jacsau.1c00175
- Chiti F, Dobson CM. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu Rev Biochem.* 2017;86:27-68. doi:10.1146/annurev-biochem-061516-045115
- Prusiner SB. Prions. *Proc Natl Acad Sci U S A.* 1998;95:13363-13383. doi:10.1073/pnas.95.23.13363
- Prusiner SB. Novel proteinaceous infectious particles cause scrapie. *Science.* 1982;216:136-144. doi:10.1126/science.6801762
- Spagnoli G, Rigoli M, Orioli S, et al. Full atomistic model of prion structure and conversion. *PLoS Pathog.* 2019;15:e1007864. doi:10.1371/journal.ppat.1007864
- Kraus A, Hoyt F, Schwartz CL, et al. High-resolution structure and strain comparison of infectious mammalian prions. *Mol Cell.* 2021;81:4540-4551 e4546. doi:10.1016/j.molcel.2021.08.011
- Manka SW, Zhang W, Wenborn A, et al. 2.7 Å cryo-EM structure of ex vivo RML prion fibrils. *Nat Commun.* 2022;13:4004. doi:10.1038/s41467-022-30457-7
- Spagnoli G, Requena JR, Biasini E. Understanding prion structure and conversion. *Prog Mol Biol Transl Sci.* 2020;175:19-30. doi:10.1016/bs.pmbts.2020.07.005
- Apetri AC, Maki K, Roder H, Surewicz WK. Early intermediate in human prion protein folding as evidenced by ultrarapid mixing experiments. *J Am Chem Soc.* 2006;128:11673-11678. doi:10.1021/ja063880b
- Apetri AC, Surewicz K, Surewicz WK. The effect of disease-associated mutations on the folding pathway of human prion protein. *J Biol Chem.* 2004;279:18008-18014. doi:10.1074/jbc.M313581200
- Apetri AC, Surewicz WK. Kinetic intermediate in the folding of human prion protein. *J Biol Chem.* 2002;277:44589-44592. doi:10.1074/jbc.C200507200
- Sanz-Hernandez M, Barritt JD, Sobek J, Hornemann S, Aguzzi A, De Simone A. Mechanism of misfolding of the human prion protein revealed by a pathological mutation. *Proc Natl Acad Sci U S A.* 2021;118:e2019631118. doi:10.1073/pnas.2019631118
- Mondal B, Reddy G. A transient intermediate populated in prion folding leads to domain swapping. *Biochemistry.* 2020;59:114-124. doi:10.1021/acs.biochem.9b00621
- Spagnoli G, Massignan T, Astolfi A, et al. Pharmacological inactivation of the prion protein by targeting a folding intermediate. *Commun Biol.* 2021;4:62. doi:10.1038/s42003-020-01585-x
- Gooptu B, Lomas DA. Conformational pathology of the serpins: themes, variations, and therapeutic strategies. *Annu Rev Biochem.* 2009;78:147-176. doi:10.1146/annurev.biochem.78.082107.133320
- Chapman KR, Burdon JG, Piitulainen E, et al. Intravenous augmentation treatment and lung density in severe alpha1 antitrypsin deficiency (RAPID): a randomised, double-blind, placebo-controlled trial. *Lancet.* 2015;386:360-368. doi:10.1016/S0140-6736(15)60860-1
- Teckman JH, Jain A. Advances in alpha-1-antitrypsin deficiency liver disease. *Curr Gastroenterol Rep.* 2014;16:367. doi:10.1007/s11894-013-0367-8
- Wang F, Orioli S, Ianeselli A, et al. All-atom simulations reveal how single-point mutations promote serpin misfolding. *Biophys J.* 2018;114:2083-2094. doi:10.1016/j.bpj.2018.03.027
- Van De Craen B, Declerck PJ, Gils A. The biochemistry, physiology and pathological roles of PAI-1 and the requirements for PAI-1 inhibition in vivo. *Thromb Res.* 2012;130:576-585. doi:10.1016/j.thromres.2012.06.023
- Fjellstrom O, Deinum J, Sjögren T, et al. Characterization of a small molecule inhibitor of plasminogen activator inhibitor type 1 that accelerates the transition into the latent conformation. *J Biol Chem.* 2013;288:873-885. doi:10.1074/jbc.M112.371732

37. Giri Rao VVH, Gosavi S. On the folding of a structurally complex protein to its metastable active state. *Proc Natl Acad Sci U S A*. 2018;115:1998-2003. doi:[10.1073/pnas.1708173115](https://doi.org/10.1073/pnas.1708173115)
38. Cazzolli G, Wang F, Beccara S, Gershenson A, Faccioli P, Wintrode PL. Serpin latency transition at atomic resolution. *Proc Natl Acad Sci U S A*. 2014;111:15414-15419. doi:[10.1073/pnas.1407528111](https://doi.org/10.1073/pnas.1407528111)
39. Kihn K, Marchiori E, Spagnolli G, et al. Long-range allostery mediates the regulation of plasminogen activator inhibitor-1 by cell adhesion factor vitronectin. *J Biol Chem*. 2022;298:102652. doi:[10.1016/j.jbc.2022.102652](https://doi.org/10.1016/j.jbc.2022.102652)
40. Gemmer M, Forster F. A clearer picture of the ER translocon complex. *J Cell Sci*. 2020;133:jcs231340. doi:[10.1242/jcs.231340](https://doi.org/10.1242/jcs.231340)
41. Jimenez JL, Hegemann B, Hutchins JR, Peters JM, Durbin R. A systematic comparative and structural analysis of protein phosphorylation sites based on the mtcPTM database. *Genome Biol*. 2007;8:R90. doi:[10.1186/gb-2007-8-5-r90](https://doi.org/10.1186/gb-2007-8-5-r90)
42. Giovanni Spagnolli TM, Astolfi A, Biggi S, et al. Pharmacological protein inactivation by targeting folding intermediates. *bioRxiv*. 2020. doi:[10.1101/2020.03.31.018069](https://doi.org/10.1101/2020.03.31.018069)
43. Tania Massignan AB, Terruzzi L, Spagnolli G, et al. Antimalarial artefenomel inhibits human SARS-CoV-2 replication in cells while suppressing the receptor ACE2. *arXiv*. 2021. doi:[10.48550/arXiv.2004.13493](https://doi.org/10.48550/arXiv.2004.13493)
44. Das A, Yadav A, Gupta M, et al. Rational design of protein-specific folding modifiers. *J Am Chem Soc*. 2021;143:18766-18776. doi:[10.1021/jacs.1c09611](https://doi.org/10.1021/jacs.1c09611)
45. Broglia RA, Tiana G, Berera R. Resistance proof, folding-inhibitor drugs. *J Chem Phys*. 2003;118:4754-4758. doi:[10.1063/1.1545087](https://doi.org/10.1063/1.1545087)
46. Dolgin E. Turning transient structures into drug targets. *Nature*. 2021. doi:[10.1038/d41586-021-01668-7](https://doi.org/10.1038/d41586-021-01668-7)
47. Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K. Combining experiments and simulations using the maximum entropy principle. *PLoS Comput Biol*. 2014;10:e1003406. doi:[10.1371/journal.pcbi.1003406](https://doi.org/10.1371/journal.pcbi.1003406)
48. Cavalli A, Camilloni C, Vendruscolo M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J Chem Phys*. 2013;138:094112. doi:[10.1063/1.4793625](https://doi.org/10.1063/1.4793625)
49. Giulini M, Menichetti R, Shell MS, Potestio R. An information-theory-based approach for optimal model reduction of biomolecules. *J Chem Theory Comput*. 2020;16:6795-6813. doi:[10.1021/acs.jctc.0c00676](https://doi.org/10.1021/acs.jctc.0c00676)
50. Brandt S, Sittel F, Ernst M, Stock G. Machine learning of biomolecular reaction coordinates. *J Phys Chem Lett*. 2018;9:2144-2150. doi:[10.1021/acs.jpcclett.8b00759](https://doi.org/10.1021/acs.jpcclett.8b00759)
51. Noe F, Olsson S, Kohler J, Wu H. Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science*. 2019;365:eaaw1147. doi:[10.1126/science.aaw1147](https://doi.org/10.1126/science.aaw1147)

How to cite this article: Biasini E, Faccioli P. Functional, pathogenic, and pharmacological roles of protein folding intermediates. *Proteins*. 2023;1-9. doi:[10.1002/prot.26479](https://doi.org/10.1002/prot.26479)