

Chapter 6

Searching for Protein Folding Mechanisms: On the Insoluble Contrast Between Thermodynamic and Kinetic Explanatory Approaches



Gabriel Vallejos-Bacelliere and Davide Vecchi

Abstract The protein folding problem is one of the foundational problems of biochemistry and it is still considered unsolved. It basically consists of two main questions: what are the factors determining the stability of the protein's native structure and how does the protein acquire it starting from an unfolded state. Since its first formulation, two main explanatory approaches have dominated the field of protein folding research: a thermodynamic approach focused on energetic features and a kinetic approach focused on the temporal development of protein chains and structural considerations. Although these two approaches are tightly intertwined in biochemical practice and largely agree on which are the parts and activities in which the phenomenon under study should be decomposed to, there nevertheless exist important contrasts that have had repercussions on the development of the field and still engender vigorous debate. We shall analyse the historical development of the field and crucial aspects of current scientific debates. On this basis, we argue that the main sources of disagreement centre on the causal interpretation of thermodynamic and kinetic explanations, on the explanatory relevance assigned to different features of the phenomena under study and on the status of the ontological assumptions concerning the entities under study.

Keywords Protein folding problem · Thermodynamic hypothesis · Levinthal's paradox · Energy landscape · Brownian motion · Foldons · Equilibrium explanation

G. Vallejos-Bacelliere (✉)

Laboratorio de Bioquímica y Biología Molecular, Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

e-mail: gvallejos@ug.uchile.cl

D. Vecchi

Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_6

109

6.1 Introduction: What Is the Protein Folding Problem

Proteins are central components in the functioning of all known organisms, carrying out functions such as catalysis, regulation of cell processes, transport, movement (from the subcellular to the organismal level), signalling, body construction etc. Without much exaggeration, proteins are what make life as we know it possible.

Proteins are linear polymers composed of amino acids linked together by peptide bonds.¹ Polypeptide chains are synthesized in the cell by the ribosomes, which catalyse the formation of the peptide bonds between its different amino acids, which are thus arranged in a polypeptide in accordance with the order of the nucleotide sequence of a messenger RNA (mRNA). The sequences of both polymers are related through the “genetic code”, which maps each amino acid (carried by a tRNA molecule with a specific “anticodon”) to a specific triplet (“codon”) of nucleotides of the mRNA sequence.² In an analogous way, the sequence of the mRNA is generated, in accordance with the principle of complementarity, by an RNA polymerase according to the order of the sequence of nucleotides in a DNA segment, i.e., a gene. Polypeptide chains are synthesized as random coils. However, in order to realize their function and being soluble,³ a protein must get folded into a specific compact 3D structure (or some restricted set of 3D conformations), i.e., what is called its “native state” or “native structure”,⁴ which is stabilized by different kinds of interactions between its components, like hydrogen bonds, van der Waals (VDW) interactions, electrostatic effects, hydrophobic effects, etc.⁵ The developmental process by which a protein undergoes a series of compositional and structural changes to acquire this final structure is called “folding”.

The protein folding problem (PFP) is one of the foundational problems of biochemistry. Since its formulation in the early 50s, it has spurred a substantial amount

¹A peptide bond is a covalent bond formed between the amino group of an amino acid and the carboxylic group of another. Because this bonded structure forms the backbone of the protein, proteins are also called polypeptides or polypeptide chains.

²The ‘genetic code’ is not strictly universal (Krebs et al., 2018).

³Usually, an unfolded protein is insoluble. The accumulation of insoluble components inside a cell beyond a certain threshold would be deleterious, causing stress or even tissue damage (Austin, 2009).

⁴It is usual for biochemists to use the term “3D-structure” to refer to specific conformations with low degrees of freedom that are stabilized by non-covalent interactions between their components. The native state is one of them. In this article, we centre our attention on globular proteins, i.e., proteins whose native state is a compact and water-soluble spherical-like conformation. This focus is justified by the fact that globular proteins are the most studied in protein folding research. Thus, research on their folding can be considered a research field on its own.

⁵This stabilization process might also involve environmental or non-intrinsic factors, that is, extrinsic vis-à-vis the intrinsic properties of the components of the developing protein, e.g., water, protons, ions, cofactors, prosthetic groups, ligands or other proteins (which can be other polypeptide chains of the same type in the case of homooligomers). Entropic factors may also play relevant causal roles, like the increase in solvent entropy caused by the burial of the hydrophobic moieties of the macromolecule. See Santos et al. (2020) for an analysis of the causal role of extrinsic factors in protein folding.

of theoretical and experimental research. Basically, the problem consists in answering the questions concerning what factors determine the stability of the native structure and how polypeptide chains reach their final native structure in a given medium.⁶ The problem is far from being trivial. Proteins are compositionally and structurally very complex entities. Given the high number of degrees of freedom of a polypeptide chain, which depend on the vast number of possible 3D arrangements of the component parts, a protein can, in principle, acquire an enormous number of possible conformations. However, despite this complexity, proteins fold rapidly, which means that from all the possible conformations, only a very restricted set is selected, leading to stable, soluble and functional 3D structures that are generally acquired in less than a minute (or less than a second in the case of smaller proteins).⁷

The protein folding problem has important consequences not only for basic biochemical research, but also in applied fields such as biomedicine and industry. Indeed, the aetiology of many degenerative diseases, such as Alzheimer and Creutzfeldt-Jakob disease, are related to the misfolding of proteins and the formation of insoluble protein aggregates that seem to destroy neuronal tissues (Liu et al., 2019). Moreover, the possibility of predicting the native state of proteins – given knowledge of the polypeptide chain – would provide significant understanding concerning the causal role of each gene in the case of any kind of organism. This predictive accomplishment would also imply the identification of new possible pharmacological targets for the cure of different kinds of medical conditions. For industry, knowing the factors that stabilize proteins would allow to develop technological applications, for example, the production of more stable enzymes.

⁶It is important to distinguish the protein folding problem, which is addressed in this article, from the protein structure prediction problem. The latter corresponds to the aim of predicting the native structure of a protein given its amino acid sequence. In other words, the first is concerned with explanatory aims, while the second with predictive ones. In the beginning of protein folding research, both explanatory and predictive aims were indissociable. However, with the emergence of structural databases, it became possible to make predictions based only on observed structural patterns irrespective of the physical basis of its relative frequencies (i.e., prediction became independent from explanation, see Vallejos-Baccelliere, 2022). In this latter context, revolutionary advances have emerged thanks to the application of powerful computational methods and artificial intelligence, being softwares like AlphaFold2 and RoseTTAFold prominent examples of this field. In this article, when referring to predictive aims, it will only be in the context of its association to the explanatory aims of protein folding research.

⁷It is debatable whether intrinsically disordered proteins (IDPs) are an exception in this sense. On the one hand, while IDPs are characterised as a class of proteins that do not get folded to perform their function (i.e., they do not possess an identifiable native structure), many IDPs acquire some kind of “native” structure when performing their function (Gomes & Faísca, 2019), e.g., acquiring some form of “order” in a specific region as, for example, when interacting with some ligand. On the other hand, it is currently accepted that almost all proteins have regions with some degree of “disorder” (Medina et al., 2021).

6.2 Brief Historical Overview of Folding Research

For many years, the problem of how proteins acquire their native structure remained elusive. Because all proteins were extracted already folded from living cells, it was thought that proteins must get necessarily folded by some part of the cell machinery acting as a structural template. Given the high variety of protein types in a cell, this machinery was believed to have templates for each one (Tanford & Reynolds, 2003).⁸ Due to their ubiquity in organisms and their importance in almost all cell functions, an obvious proposal was to consider this machinery as constituted by proteins. However, this proposal generates a conundrum: if the templates necessary for folding any protein are other proteins, then how are such templates folded? Any answer to this puzzle seems to lead to an infinite regress. Later, in the context of protein synthesis research, it was proposed that ribosomes must carry out this template function. Nonetheless, all ribosomes in a cell turned out to have very similar structure, so the problem of how it is possible to have a template for all the variety of proteins remained unsolved (Tanford & Reynolds, 2003).

The problem was completely reformulated at the start of the 1960s thanks to the work of Christian Anfinsen and his collaborators (Anfinsen et al., 1961; Anfinsen, 1973).⁹ In a series of experiments, they managed to show that a purified protein (i.e., bovine pancreatic ribonuclease A), after being unfolded¹⁰ (i.e., losing its folded structure without the breaking of its peptide bonds) with urea and reducing agents, could be reversibly refolded once the denaturants were extracted from the medium, thus recovering its biological activity in the absence of any other cellular component. Based on these results, Anfinsen proposed the so called “thermodynamic hypothesis”: “... *the three-dimensional structure of a native protein in its normal physiological milieu is the one in which the Gibbs free energy of the whole system is lowest; that is, the native conformation is determined by the totality of the interatomic interactions and hence by the amino acid sequence, in a given environment*” (Anfinsen, 1973, p. 223). The hypothesis seemingly makes two different kinds of claim: the first is that the native structure of a protein corresponds to the conformation with minimal free energy; the second, of implicit causal nature, is that

⁸The original postulation of templates was speculative, referring to an entity that actively folds proteins by performing the role of master mould to be copied. This hypothesis was promoted, among others, by the defenders of the ‘one gene, one enzyme’ hypothesis (Tanford & Reynolds, 2003). This putative template role is different from that of entities either accelerating folding or restricting the possible conformations acquired by the polypeptide. For instance, chaperones, instead of acting as templates, perform other functions, such as the acceleration of the folding process, the isolation of the folding protein from the external environment (i.e., “Anfinsen’s cage”) or the partial restriction of the possible alternative conformations (Sorokina et al., 2022).

⁹Anfinsen received the chemistry Nobel Prize in 1972 for this work.

¹⁰Denatured and unfolded are different concepts. In the present context, the first is functional, referring to loss of biochemical activity (e.g., the loss of the catalytic capacity of RNase A). The second concept is structural, referring to the loss of 3D structure of a protein.

the native structure is determined by the amino acid sequence. Let us analyse each claim in turn and, particularly, their relation.¹¹

The first kind of claim has been interpreted as meaning that, of all the possible conformations a protein might acquire, the native structure is the most stable in the appropriate “physiological conditions”. This assertion is both independent of any consideration concerning the temporal development of a polypeptide chain from the unfolded to the native state and also independent of the possible regulation of the folding process *in-vivo*. The corollary of this view is that the folding process is “spontaneous”¹² in the specific sense that it is merely thermodynamically driven. In microscopic terms, this hypothesis is currently represented by a conformational energy landscape with a funnel-like shape characterised by a single global minimum.¹³ The total free energy of each possible conformation would be determined by the contribution of the totality of interatomic interactions that are established in each token case. Therefore, these interactions are what determine the shape of a protein’s energy landscape.

The second claim is of a causal nature and asserts that the native structure of a protein in a given medium is “determined” by its amino acid sequence. This claim accounts for the observed refolding capacities of the, by supposition, completely unfolded protein¹⁴ when denaturants are extracted from the medium in the absence of other cellular components. The spontaneity of folding is then explained by the intrinsic properties and potentialities possessed by the amino acid components of the polypeptide chain, which are immutable and unaffected by any causal interaction of the developing protein with extrinsic factors (Santos et al., 2020). Then, the acquisition of the native structure by a protein, either during translation or when folding/refolding, is due to the “activation” (or “manifestation”) or “suppression” of these potentialities. The role of the environment, which includes the interaction of the developing protein with extrinsic factors and regulatory processes – that, *in vivo*, include the causal role of cellular components (ribosomes, other proteins, chaperones, ligands, etc.) coupled to other energetic processes (e.g., ATP hydrolysis) –, would then only be that of activating or suppressing some of the immutable

¹¹ For a discussion of the dual nature of the thermodynamic hypothesis, see Santos et al., 2020.

¹² In thermodynamics, a process is defined as spontaneous when the initial state has a higher free energy than the final one. This definition is independent of how the transformation between states occurs and, hence, it is also independent of the time that this transformation would take.

¹³ Revisions of the thermodynamic hypothesis taking into consideration a plurality of local equilibria have of course been proposed (Dill & Chan, 1997). Such revisions are necessary to account for the existence of, for instance, conformational changes, folding intermediaries, misfolding and amyloid fibril structures (whereby the latter are more stable than the native one). This revision requires the postulation of shallower minima of the energy landscape. However, any pluralistic move seems to imply a weakening of Anfinsen’s original hypothesis; this is because the hypothesis that folding is a merely thermodynamically driven process (instead of being directed or regulated) becomes increasingly questionable the more complex the shapes of the energy landscape are.

¹⁴ In biochemistry, a frequently discussed issue is to characterize what a denatured state really is and whether there really are “completely unfolded” proteins (see Sorokina et al., 2022 for a good synthesis).

potentialities, a process that, in a physiological medium, will lead to the formation of the native state. Under this interpretation, phenomena like misfolding or the denaturation, aggregation and degradation through time (which occur in the natural cellular milieu and in-vitro) is attributable to "...secondary effects [that] are claimed to shift the equilibrium towards the unfolded state, preventing thermodynamically-driven folding" (Sorokina et al., 2022, p. 7), i.e., preventing the manifestation of the aforementioned intrinsic potentialities.

The thermodynamic hypothesis opened a whole new field of research whose aim was to predict the native state of each protein with known sequence by seeking its minimal energy conformation among the totality of possible ones. This research programme relied on already accepted knowledge about the physical properties common to all molecules, which would allow to obtain each conformation of a macromolecule just by tinkering with it.¹⁵ This body of knowledge concerned, for instance, the nature of the covalent bond and molecular geometries, including the length of the covalent bonds, the possible angles between bonds, the rotatability of two moieties separated by a single bond, the planarity of a double bond etc. Given this knowledge, it would become in principle possible to describe all the possible steric restrictions (e.g., two atoms cannot occupy the same place, two covalent bonds cannot go through each other) and all the possible interactions (e.g., attractive or repulsive) between each component of the protein in each possible conformation, thus being able to calculate the free energy associated to each physically plausible conformation and eventually find the one with the lowest value. This approach spurred the expectation that, basically, the folding problem had already been solved in thermodynamic terms: in principle, finding the native state would just require exploring all the conformations (or enough of them) and calculating their free energy until the minimum is found. The protein folding problem was basically framed as a computational one.¹⁶ However, things turned out to be much more complex.

Consider a simplified protein with 100 amino acids (which is considered a relatively short length), in which every amino acid can only assume two different conformations. This protein has approximately 10^{30} potential conformations, with the native state corresponding, according to the thermodynamic hypothesis, to just one of them (or, more appropriately and realistically, to a very restricted set of these possible conformations). Moreover, if each conformational change occurred in just 1 picosecond, the folding process would take more time than that of the entire age of the universe if it were a totally random search (see Gomes & Fafsa, 2019 p. 26). However, proteins get folded in the order of milliseconds to seconds. What we are describing is a mental experiment known, in honour to its formulator, "Levinthal's

¹⁵Of course, assuming we possess an ideal model to tinker with (See Francoeur, 2001, 2002 for a historical revision of the use of models to study protein conformations).

¹⁶This computational approach is different from current computational models for predicting protein structures, such as AlphaFold (see note 6). In the case of the thermodynamic hypothesis, what is computed is the free energy of any possible 3D conformation that might be potentially acquired by a polypeptide chain based on knowledge concerning the physical basis of the interatomic interactions between its constituent amino acids. The only similarity between the two programmes is the neglect of kinetic considerations.

paradox” (Levinthal, 1968, 1969). Its morale is straightforward: the folding process cannot be a random process, otherwise it would take too long; the alternative is that there must exist some pathway guiding or biasing the conformational search from the unfolded to the native state, otherwise phenomena like the cooperative nature of the process (i.e., the seemingly all-or-none character of the transition between states) would remain unexplained.¹⁷ The most important implication of the thought experiment is that the folding process cannot be merely thermodynamically driven, as kinetic factors must play an essential role. In this respect, Levinthal’s postulation of folding pathways had the implication of reframing the protein folding problem by focusing on kinetic considerations. As Levinthal (1968, p. 44) argued: “... a pathway of folding means that there exists a well-defined sequence of events which follow one another so as to carry the protein from the unfolded random coil to a uniquely folded metastable state.” Another major consequence is that the native state does not necessarily correspond to a global energy minimum in the conformational energy landscape, but rather to the conformation that is most rapidly reachable (or slowest to exit from) from the unfolded protein. More stable conformations could exist, but as it is slower to get to them, it would be highly improbable for those to be reached. The native state might thus be characterised as a local minimum in the conformational energy landscape, a “metastable state”, i.e., a folded set of state(s) separated from the unfolded one by lower energetic barriers. Therefore, the mere search for energy minima would be rather irrelevant for explaining the folding process and to predict the native structure. The kinetic approach involved a change in the question guiding protein folding research (PFR): to explain the folding phenomenon and predict native structure, just computing the free energies of the possible conformations is insufficient as it is also necessary to describe the actual conformational changes (including the formation and breaking of the molecular interactions between different protein parts) that characterise each temporal stage during folding, starting from the unfolded state up to the native state. These theoretical developments gave rise to a new research agenda aimed to describe the folding pathways by characterizing the stages of the process, which were conceived as discrete and structurally characterizable intermediaries and transient states.¹⁸ This

¹⁷We return to the issue of cooperativity in Sect. 6.4.2.

¹⁸Let us clarify the concepts of intermediary, transient and transition state. An intermediary state refers to a discrete metastable conformation that can in principle be characterized thermodynamically. In kinetic terms, it refers to a conformation that lasts long enough to be “detectable directly”. Transient (or transitory) states refer to a mere stage in a dynamic process. A good analogy is with a car travelling from a city to another. An intermediary would be the car making a stop at a service station, and a transient state would be the car passing through any part of the road. The concept of transient state is somehow problematic because its status as a proper state is dubious and there could be cases in which the difference between being an intermediary or a transient state gets blurred. In biochemical practice, it is common for transient states to be conceptualised as discrete states representing the structural elements that are already formed at some stage of the folding process (like a picture of a car passing a specific point on the road). In what follows, we will assume this interpretation. A third concept is the transition state of a kinetic process, a theoretical concept of chemical kinetics referring to the transient state with highest energy in a reaction coordinate; this defines the so-called activation energy of a chemical reaction. In other words, in the

approach is currently labelled as the “classical view” of protein folding (Dill & Chan, 1997).

Anfinsen’s and Levinthal’s seminal contributions gave rise to two alternative sets of explanatory practices dealing with the protein folding phenomenon: a thermodynamic approach and a kinetic approach (or kinetic-dynamic approach). Both approaches are deeply intertwined in PFR but, as indicated above, there are important contrasts between them. Indeed, we would argue that these contrasts are profound enough to have divided the research field into different epistemic cultures. In the next section we shall analyse how both approaches account for the protein folding process.

6.3 Two Explanatory Approaches in Protein Folding Research

To understand the explanatory aims in PFR, in this section we describe how an “ideal explanation” of the protein folding process may look like for both the thermodynamic and kinetic approach. We characterize the concept of “ideal explanation” in the protein folding case as that accounting for an explanandum in terms of a complete description of the folding process given one specific set of epistemic resources concerning the physical and chemical properties and interactions at the atomic and molecular levels.¹⁹ This explanatory basis is largely common to thermodynamic and kinetic approaches (see Sects. 6.3.2 and 6.5.1). However, there are distinctive epistemic resources to each approach since, as we have already stressed, the former is centred on energetic and thermodynamic considerations, while the second focuses on kinetic considerations and structurally characterizable steps.

6.3.1 *Thermodynamic and Kinetic Explanations*

There are two main explananda in PFR: the first concerns the factors that determine the stability of the native structure; the second concerns how the protein acquires its native 3D structure starting from an unfolded state. We will call the first the native state stability problem (NSSP) and the second the folding dynamic problem (FDP).

protein case, this would be the least stable conformation a protein must transit through to reach native state (or to transit from any metastable state, like an intermediary, to another state), which defines the limiting step of the process. It is by definition non-detectable and non-isolatable, so it is only modelled theoretically. The aim of transition state theory was to account for the thermodynamic properties of chemical processes in terms of energy barriers between states.

¹⁹In biochemical practice, research results are often presented in terms of tokens. This is a deliberate idealization used for mainly narrative purposes. Consequently, the concept of ideal explanation is hereby characterised by reference to tokens; however, this is an analytical choice to account for the epistemic aims of the two explanatory approaches that will be addressed. Nevertheless, as we will show, biochemical practice is based on protein types.

In the case of the NSSP, an ideal thermodynamic explanation (i.e., leaving aside Levinthal's problem and assuming it is possible to identify all the possible conformations of a protein) would require the calculation of the free energy of all possible conformations of a polypeptide chain. The free energy of each one will be the result of the contribution of each (attractive and repulsive) interaction exerted between the protein parts in virtue of their intrinsic properties. The lower the conformational energy, the higher its stability. As anticipated in Sect. 6.2, the graphical representation of possible conformational energy states is called the energy landscape of the protein (Dill & Chan, 1997; Onuchic et al., 1997). Usually, it is illustrated as a graph in which the vertical axis represents the internal free energy and the other axes represent the conformational space.²⁰ Besides the possibility of finding the native conformation, knowing the shape of the energy landscape would also allow finding local energy minima, indicating the existence of possible alternative metastable conformations. Moreover, it would also be possible to describe energy barriers between the different metastable conformations, that is, the energies of the conformations that the protein should overcome to transition from one state to another. The higher the energy values of those intermediate conformations, the lower the probability of crossing from one state to another. In this way, a thermodynamic approach can account for the FDP. Importantly, in a thermodynamic explanatory approach there is neither an explicit appeal to the temporal variable nor to the actual pathways that a token protein (or populations of token proteins) will transit through when going from one state to another.

Conversely, the kinetic approach explicitly takes into consideration the time variable. The ideal explanation in this case is the description of the temporal development of a polypeptide chain when transiting from an unfolded state to the native state. This explanatory aim originates directly from Levinthal's postulation of folding pathways. In this case, what is explanatorily central are not the energy differences between conformational states, but the structural changes in conformation manifested by the developing protein during the folding process. This ideal explanation involves the characterisation of the temporal order in which the interactions between the parts of the protein occur and the identification of the new structures emerging as the native state is reached (Fersht, 1995, 1998; Baldwin, 2008; Englander & Mayne, 2017a, b). It is thus straightforward to see how the kinetic approach accounts for the FDP. Regarding the NSSP, from a kinetic perspective the stability of the native state is accounted for in terms of its maintenance. Basically, when the rate of reaching one state from another is higher in comparison to the rate of abandoning it, this state will be dynamically maintained.

²⁰More specifically, the vertical axis represents the 'internal free energy' and the additional axes represent the conformational space defined in terms of the conformational coordinates accounting for the degrees of freedom of a protein. The total internal free energy depends on physical factors (e.g., the sum of the energy contributions of hydrogen bonds, ion-pairs, torsion angles, hydrophobic contacts and salvation free energies) and the environmental factors on which the former depend (e.g., temperature, solvent). For each possible conformation, different interactions between protein parts will occur (or not), and that is what will define the total internal free energy of each conformation.

Although these two epistemic endeavours can be clearly distinguished conceptually, they are tightly intertwined in biochemical practice, so that it is usual to interpret kinetic features in thermodynamic terms and vice versa. The thermodynamic approach might explain kinetic features. For example, as we mentioned above, it is possible to describe the speed of folding/unfolding in terms of the height of the energetic barriers between the two states (using transition state theory): the higher the barrier, the slower the transition will be (Fersht, 1998). Conversely, as has already been related, the kinetic approach might explain thermodynamic features in terms of maintenance of states. For example, the stability of the native state can be explained by rapid refolding in contrast to a slow unfolding kinetic. This dynamic can be interpreted, for example, in terms of the early formation of strong stable interactions that “guide” the chain to the native state and which are then later difficult to break. In other cases, explanations of some specific features mesh thermodynamic and kinetic considerations, blurring their distinction. For example, the topology of the native state is sometimes assumed to play an important role both in the folding process and native state stability (Plaxco et al., 1998). A native fold with a complex topology (e.g., with very high contact order) will be reached more slowly than a native fold with a simpler one. But, from a thermodynamic point of view, a complex native state topology also increases the stability of a protein by the generation of atomic and molecular interactions constraining the unfolding process. Knotted topologies are an example of this latter case (Gomes & Faísca, 2019).

In summary, in various contexts both explanatory approaches are indeed intertwined and biochemists make a complementary use of their respective epistemic resources to generate explanations. Biochemists largely agree on the issue of which are the salient parts and activities in which the phenomenon under study should be decomposed to. However, as we shall argue in the next section, biochemists advocating thermodynamic and kinetic approaches engage in theoretical debates regarding the causal nature of their explanations, the explanatory relevance assigned to different aspects of the phenomena under study as well as on the status of some underlying ontological assumptions concerning the nature of the entities under study. These debates produce genuine clashes concerning both the interpretation of experimental results and the appropriate way to seek explanations of protein folding phenomena.

6.3.2 Mechanistic Credentials of Thermodynamic and Kinetic Explanations of Folding

To analyse to what extent both kinds of ideal explanations are causal, we will address the problem by considering to what extent they can be characterised as mechanistic explanations. The mechanistic framework of analysis is justified because of the ubiquitous appeal to underlying causes accounting for the

phenomena under study in PFR. To do this, a working definition of mechanism is needed. A largely consensual minimal definition of mechanism characterises the notion in terms of “entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon” (Glennan et al., 2022, p. 145). This implies that, in the case of protein folding, it would be necessary to identify the phenomenon to be explained, the parts and activities that are responsible for it and the organisation between the parts.

As we pointed out in Sect. 6.3.1, in the case of folding, the phenomena to be explained are at least two: the stability of the native state (NSSP) and the process of its acquisition (FDP). The second explanandum seems to be clearly amenable to mechanistic analysis. In fact, many standard characterizations of mechanism refer to an organised start-to-finish causal sequence of operations/activities performed by parts/entities producing a phenomenon (Machamer et al., 2000; Bechtel, 2011). However, mechanistic explanations do not only encompass start-to-finish causal sequences. Even when the maintenance of a state – another dynamical process, e.g., homeostasis – is at issue, a mechanistic explanation can be legitimately sought (Glennan et al., 2022). Given that kinetic explanations of both folding dynamics (which are classic examples of input-output aetiological explanations, see Krickel, Chap. 2, this volume) and maintenance of the native state are straightforwardly causal and mechanistic in nature, the causal nature of thermodynamic explanations for native state stability and acquisition will be a major concern in Sects. 6.5.1 and 6.5.2.

Concerning the relevant ontology of entities or parts, as indicated at the beginning of Sect. 6.3, there are aspects that are common to both explanatory approaches. Independently of whether the ideal explanatory aim is kinetic or thermodynamic, there is widespread agreement between the advocates of the kinetic and thermodynamics approach that the components of the polypeptide chain must be considered relevant parts in any explanation of native state stability and folding dynamics. These relevant parts must include the different covalently bonded atoms that compose the polypeptide chain, which are organized in the backbone as well as in the different residues of each amino acid. Accordingly, the relevant activities would then be accounted for in terms of the interactions established between these different parts, like hydrogen bonding, electrostatic interactions, VDW interactions, etc. which occur, at least partially (discounting relational extrinsic properties, see Santos et al., 2020) in virtue of the parts’ intrinsic properties, such as net electric charge, polar or non-polar nature, aromaticity etc. Other activities might be associated to the nature of the bonds established between its parts, like torsions between bonds, proline isomerization, steric clashes, chain collapse etc. Moreover, if we consider the environment in which the protein is embedded, other activities like the interaction with extrinsic factors like water, salts, protons, etc. may be considered. In many ways, all these parts and activities are common to both kinds of explanations.

6.4 Clashes Between Thermodynamic and Kinetic Approaches

Despite the general agreement just highlighted, there exists a clear difference between both approaches regarding the appropriate decomposition of folding phenomena. In this section we will analyse two main sources of disagreement. One pivotal source of this contrast concerns the explanatory relevance of the microscopic features of the folding process (Sect. 6.4.1). Another concerns the ontological commitments related to the decomposition of the system at hand (Sect. 6.4.2).

6.4.1 *Micro Versus Macro Analyses*

As we argued in Sect. 6.3.1, one first difference between thermodynamic and kinetic approaches concerns the appeal to the time variable. To put it bluntly, without countenancing the temporal aspect, it is difficult to see how thermodynamic explanations can be counted as causal. The rationale of the thermodynamic hypothesis is that reaching native conformation is dependent on exploring enough possible backbone conformations whose formation in turn depend on the intrinsic properties of the residues and peptide bonds. Obviously, this search takes time. However, from an ideal thermodynamic perspective, this temporal aspect would be explanatorily irrelevant to make sense of the directionality of the folding process and the stability of the native state. What is relevant is to account for free energy differences between the native state and the other physically possible conformations that the protein could attain. What is required is thus an explanation in terms of energetics. The issue of directionality in the thermodynamic approach is solved by assuming that the search for native state is (significantly) thermodynamically driven. In order to explain folding speed and cooperativity (Sects. 6.2 and 6.4.2), what is relevant are the energetic biases, which are represented by the currently proposed funnelled shapes of proteins' energy landscapes (Fig. 6.1). In addition to this, the peculiar idea that the energy landscape "directs the folding protein into the native state without the need for a definite pathway" (Govindarajan & Goldstein, 1998 p. 5545) or, put differently, that "native structure is determined only by the final native conditions" (Dill & Chan, 1997, p. 10), as if it were an attractor, are added. Basically, when folding starts, the number of possible conformations the protein can explore (i.e., the internal entropy of the chain) is gradually reduced due to the energetic biases accounted for by the enthalpic factors (such as the formation of intermolecular interactions) and the increment in solvent entropy as hydrophobic moieties get buried.²¹ In this respect, it is postulated that, considering a given protein type, folding may start at many different locations of the chain in each token's case, with the

²¹Even among the defenders of thermodynamic approaches there is considerable debate about which are the main factors that account for the energy biases of the folding process (Dill, 1999;

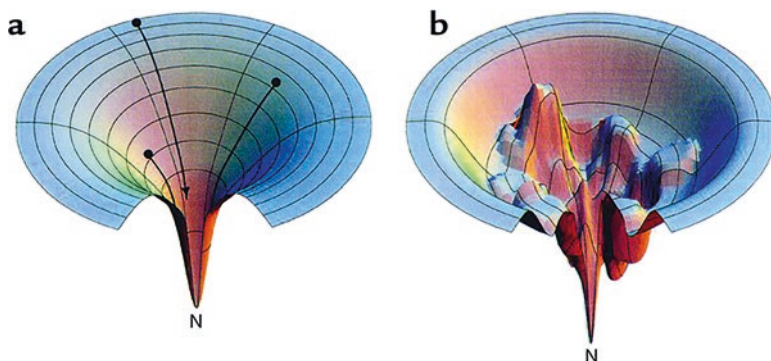


Fig. 6.1 Typical representations of energy landscapes of protein folding with funnelled shapes. The figure on the left corresponds to an idealized smooth funnel. Inside, three possible folding trajectories are marked as black lines starting from specific points located on the “denatured state ensemble”. The figure on the right corresponds to a rough and more realistic energy landscape with several local minima and energy barriers. (From Ken A. Dill, CC BY 4.0, via Wikimedia Commons)

further consequence that it will occur on many independent pathways (Fig. 6.1). Therefore, it becomes meaningless to postulate an order of folding events or the existence of single pathways to the native state.

This interpretation of folding dynamics is at odds with the notion of productive causal explanation because of the omission of the time variable and also because it focuses on energy differences instead of actual causal processes. This description thus leaves unexplained why, although all conformational potentialities can, in principle, be physically realised, some conformational potentialities are either not realised or transient (as suggested by experimental evidence); the thermodynamic explanation just assumes that it is because these conformations are energetically disadvantageous and unstable. Even at the microscopic level of description of the folding process (i.e., the level of the intrinsic properties of amino acids and peptide bonds, e.g., dihedral angles, side chain rotamers, etc.), which is the one explanatorily relevant for the thermodynamic account, the only concern is about differences in terms of stability between conformations, neglecting how the transition between conformations occurs. To solve this theoretical problem and make possible the explanation of aspects like the cooperativity of the folding process, the defenders of thermodynamic approaches resort to Brownian motion. Indeed, this dependence of folding on random processes makes folding analogous to a “parallel microscopic multi-pathway diffusion-like” process (Dill & Chan, 1997 p. 18), captured by the analogy between folding and the trickle of rainwater or skiers skying down a mountain, as is indicated in Fig. 6.1. This not only means that token proteins of the same type will inevitably fold differently, but that even the same token differently spatio-temporally localised (or even, at the extreme, the same spatio-temporally localised

Rose et al., 2006; Ben-Naim, 2015). Other driving forces in addition to the hydrophobic effect have been proposed, like the formation of hydrogen bonds, secondary structure propensities, etc.

token in case Brownian motion is an indeterministic process) will inevitably fold differently. This is why a central concept of the thermodynamic approach is the “denatured state”,²² which refers to an ensemble:

We can draw an analogy between the denatured ‘state’ and an ensemble of skiers distributed over a mountainside. When folding conditions are initiated, each skier proceeds down the funnel following his own private trajectory. Skiers skiing down funnels reach a global minimum (satisfying Anfinsen’s hypothesis) by many different routes (not a single microscopic pathway), yet they do so in a directed and rapid way (satisfying Levinthal’s concerns). Dill & Chan, 1997, p. 12.

From the thermodynamic perspective, the folding process is understood in terms of ensembles of different microscopic conformations. The concept of ensemble is difficult to characterize with precision due to its vagueness. In the case of protein folding, it can be conceptualised as the distribution of conformations that might be acquired by the token proteins of a population characterized by some macroscopic parameter (e.g., enthalpy differences, observable signals like fluorescence, etc.) and in a given environment (fixed temperature and pressure). This population is highly dynamic as each token protein is constantly fluctuating between different conformations: the broader the distribution of conformations, the higher the entropy and degrees of freedom of the population. The unfolded state would correspond to the ensemble with highest entropy. During folding, the entropy decreases until reaching native state, which corresponds to an ensemble with a very restricted conformational distribution. The distribution of conformations that define an ensemble is given by their stability differences, which, in their turn, are explained at the microscopic level by the interactions established between the protein components given their intrinsic properties.²³ The stages of protein folding are then conceived as ensembles with different conformational distributions in a protein population. Thus, assuming that the kinetic concept of pathway is only meaningful when referring to token and spatio-temporally localized polypeptide chains that actually start folding from one specific conformation, the thermodynamic approach denies the legitimacy of the kinetic (“classical view”) approach:

.... folding a protein does not involve starting from one specific conformation, A. The denatured state of a protein is not a single point on the landscape: it is all the points on the landscape, except for N. A pathway is too limited an idea to explain the flow from everywhere else, the denatured ensemble, to one point N. The concept of a pathway is useful for explaining the milestones we see in travels along a road or along a hiking trail, but not for describing how rain flows down a funnel. Dill & Chan, 1997 p. 12.

²²As stated in note 10, denatured is not the same as unfolded. In this case, the term “denatured” refers to any non-native state.

²³The energy landscape is thus said to “encode” the dynamic properties of the protein type. Encoding is due to the fact that the energy landscape represents the potency of a protein type, i.e., all possible conformations it may acquire given the intrinsic properties of the polypeptide chain’s components (in analogy with the fixed phase space of a dynamical system characterized in statistical mechanical terms). The energy landscape is therefore fixed from the outset (see Sect. 6.5.2).

Unlike the thermodynamic approach, kinetic approaches consider the time variable and aim to track the protein folding sequence of events through the identification of intermediate and transient states – structurally characterised – in the hope of uncovering the pathway leading to the native state. Folding dynamics are not random (otherwise, as Levinthal argued, the folding process would be too slow); they are rather constrained by processes (not necessarily thermodynamically driven)²⁴ that, despite being elusive, are open to experimental investigation. The discovery of such processes or principles of folding is the basic aim of kinetic approaches and what grounds their mechanistic ethos. Moreover, kinetic approaches do not deny that the folding pathways of different tokens of the same type of polypeptide chain might vary to some degree (if only because of Brownian motion). However, at some level of analysis, such pathways might share significant features, such as the generation of similar biochemically relevant intermediate and transient states, which can be described in structural terms by generalizing over the average behaviour of the same type of system (e.g., a population of tokens of the same protein type). A classic way to assess this is by treating the folding process as a chemical reaction going from the unfolded state (U) to the native state (N) and applying transition state theory to interpret kinetic experimental data. This permits modeling general features of the *transition state* of the process, which represents the highest energy state through which the protein must go through to transition from the unfolded to the native state. Using a previously determined structure of the native state (usually by X-ray crystallography) as a guide, and performing destabilizing site-directed mutations in the protein, it is possible to map the interactions which are already formed in this highest energy state, thus constructing a structural characterization of the limiting steps of the process (Fersht, 1995), which correspond to a global feature of a type representing the average behaviour of a protein population. This has allowed to propose different kinds of possible global mechanisms for protein folding, depending on which are considered the main events of the process (Fig. 6.2), e.g., formation and collision of secondary structure elements, hydrophobic collapse, nucleation-propagation, etc.²⁵

Advocates of the thermodynamic approach reject this macro-level kind of analysis as illegitimate for two reasons. First of all, at the atomistic level that is relevant for the thermodynamic approach, it is impossible that the pathways of two tokens can ever be identical (Eaton & Wolynes, 2017), if only because they are affected by thermal agitation. Secondly, the actual and extremely varied dynamics of folding tokens cannot be decomposed in terms of biochemically significant and structurally characterizable intermediate or transient states; the folding pathways for proteins of the same type uncovered by kinetic approaches will inevitably be too coarse-grained

²⁴This does not mean that they are in principle unexplainable in thermodynamic terms. What is important is that those processes are not merely driven (or accountable) by differences in stability and biased conformational search, but by sequences of causally related events.

²⁵Interestingly, an ongoing debate between the advocates of the kinetic approach concerns which kinds of mechanisms are most relevant and frequent to explain folding phenomena. See Gomes & Fafsa, 2019 p. 27–29 for an overview.

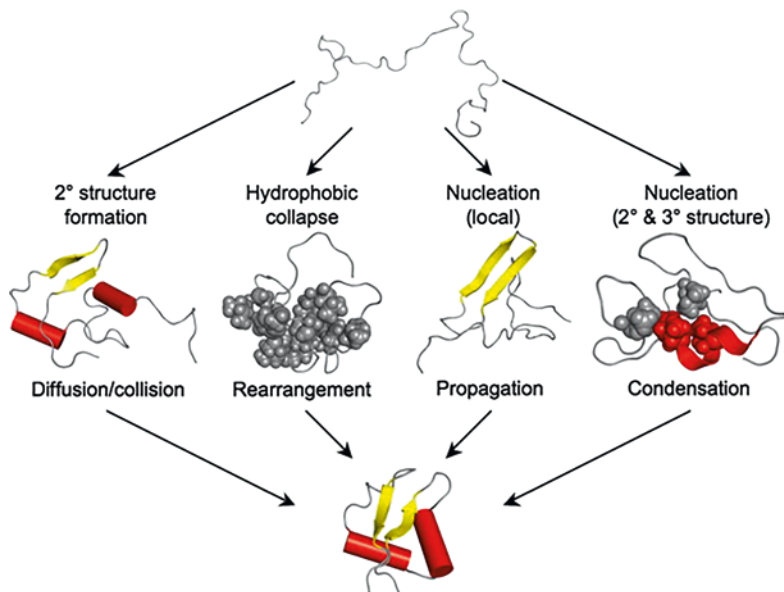


Fig. 6.2 Main kinds of folding routes that have been described for different proteins based on the structural characterization of the most relevant stages defining the folding process. (From Nickson and Clarke (2010), CC BY 3.0)

to ground significant generalizations. Ultimately, the thermodynamic approach denies the value of the structural characterization of the stages the protein transits through on the path to native state:

What is notable about the transition states of folding ... is not that they are specific structures, but that they are ensembles. The classical [kinetic] view focuses on specific structure (which experiments see), whereas the new view [thermodynamic]²⁶ is an ensemble perspective that recognizes the importance of disorder and that random processes and wrong steps are also major contributors to folding speed. Dill & Chan, 1997, p. 15.²⁷

This quotation, which is representative of the contrast characterizing current debates about folding, reveals a central issue. The characterization of the specific structures of intermediate and transient states makes theoretical sense in the context of models used to account for empirical data. In this context, for example, it is meaningful to treat the unfolded, intermediate and transient states as discrete populations. Meanwhile, random processes occurring at the microscopic level in ensembles of molecules can only be accounted for through theoretical representations such as ensembles and energy landscapes. This state of affairs illustrates the crucial point that one of the main clashes between thermodynamic and kinetic approaches

²⁶With “new view” these authors refer to the energy landscape theory.

²⁷The current form of the thermodynamic approach is commonly labeled as the “new view” by its advocates because of the novelty of the energy landscape theory and its application to protein folding. In this sense, it stands in opposition to the so-called “classical view” (see end of Sect. 6.2) based on the search for structurally characterizable stages in the folding process.

concerns the relative explanatory relevance that is attributed to empirical and theoretical considerations as well as to different theoretical models (pathway/sequential vs. landscape/parallel) and methods of analysis (structural, traditionally mechanistic, chemical kinetic vs. thermodynamic, statistical mechanical, chemical thermodynamics).

To summarise, the ideal thermodynamic explanation of folding dynamics aims to describe the whole space of possible conformations or potentials of any protein of a given type and the relative stability of each of them in terms of their free energy. The assumptions of the thermodynamic approach are that, while conformational search is dependent on micro-level causal factors, it is solely “constrained” thermodynamically (which also means that it is not totally random). In this sense, what is important in a thermodynamic explanation are microscopic level features, based on which protein’s type ensembles are defined. However, ensembles are not structurally, but thermodynamically characterized. Ultimately, the thermodynamic approach needs to explain in what specific causal sense thermodynamic “constraining” accounts for native state stability and folding dynamics. Section 6.5 shall delve on this issue, specifically on whether thermodynamic explanations of protein folding might be considered mechanistic or even causal. On the other hand, ideal kinetic explanations aim to uncover the temporal development of a polypeptide chain when transiting from the unfolded to the native state. The assumption of the kinetic approach is that this trajectory can be accounted for by describing the intermediate and transient states in structural terms. Ultimately, the kinetic approach needs to discover whether significant structural principles of folding exist notwithstanding variation in folding dynamics. The explanatorily relevant features are, then, macroscopic level properties corresponding to the average behaviours of polypeptide chains of the same type when transiting from the unfolded to the native state.

6.4.2 *The Issue of Decomposition*

Another general issue that emerges from the previous section is whether different kinds of analytic decompositions are possible. Despite the agreement between advocates of both approaches regarding the parts and activities composing the folding phenomena articulated in Sect. 6.3.2, the answer to this question is positive and shall be illustrated with one particular example: foldon kinetics (Englander & Mayne, 2017a, b).

Foldons might be defined as structural elements of a protein type acting as distinguishable cooperative units during the folding process (Fig. 6.3). Cooperativity means that the folding of one foldon influences the folding of the others, resulting in a stepwise folding process in which foldons acquire their native structure sequentially (i.e., when one foldon gets folded, this event triggers the folding of the next one and so on; conversely, a foldon cannot fold until the previous in the sequence gets folded first; more generally, a foldon is not stable enough on its own to last long enough unless the next foldon gets folded). Generally speaking, two elements of this different analytic decomposition are relevant. First, foldons – as relevant causal

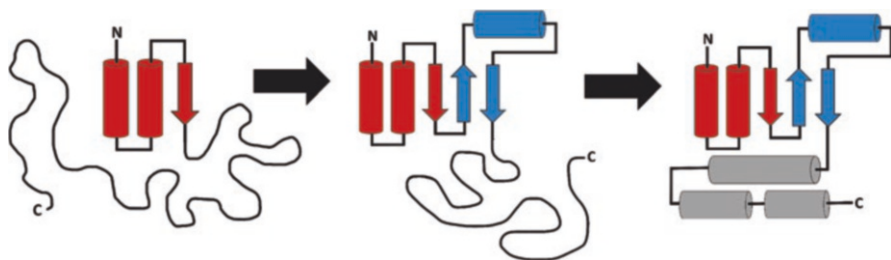


Fig. 6.3 Schematic representation of a sequential folding pathway containing three foldons (red, blue and grey). The scheme corresponds to a topological diagram in which α -helices are represented as cylinders and β -strands as arrows. Each stage corresponds to a transient state in which the folding of each foldon leads to the folding of the next one. The scheme represents the average behaviour of a protein type and not (necessarily) the actual pathway of a token protein. In this scheme the three foldons are concatenated in the protein sequence, but more complex cases (e.g., re-entrant topologies) are also possible. Additionally, this scheme takes foldons to be clusters of secondary structure elements, but other kinds of structural organizations (nucleations, hydrophobic centres, etc.) are not ruled out

parts with characteristic activities – cannot be identified at the initial stages of folding; they rather emerge as significant parts with specific causal roles during the transition from the unfolded to the native state. Thus, foldons provide a vivid example of the behaviour of dynamical systems with no fixed parts Levy and Bechtel (2016) refer to. Secondly, this example shows that the analytic decomposition into parts at the atomistic level is a commitment that is only strictly necessary for the thermodynamic approach. In fact, the denial on the part of the advocates of the thermodynamic approach that there are folding pathways is grounded on the assumption that conformational search occurs at the microscopic level: “*The multipathway idea stems from the early presumption that structure formation must occur through microscopic amino acid-level searching*” (Englander & Mayne, 2017b p. E9761). As we showed in the previous section, when looked at from this perspective, it becomes difficult to believe in significant folding pathways. Indeed, as Eaton and Wolynes (2017, p. E9759) admit: “*At an atomistic level, no two trajectories from the unfolded state to the folded state can possibly be identical, so there is an unimaginably large number of detailed pathways for folding a protein.*” The issue at this juncture is more significantly about the interpretation of the experimental evidence (gathered both *in vivo* and *in vitro*): the existence of pathway variation at the microscopic level is not under dispute, but its explanatory significance is at stake.

The foldon hypothesis stems from experimental approaches using new technologies (e.g., hydrogen exchange). Using these experimental approaches, foldon theorists have supposedly vindicated a series of kinetic hypotheses concerning the limited role (to the initial phases of folding) of random conformational search and, most prominently, the actual existence of biochemically relevant and structurally characterisable intermediates as well as the repeatable, stable, linear²⁸ and stepwise

²⁸This approach does not deny the possibility of parallel or ramified routes. Rather, the point is that these would consist, in their turn, of defined pathways with their respective structurally characterisable steps.

sequential nature of folding, at least in the case of some proteins (e.g., Rnase H): “... *Proteins fold by putting their structural elements into place over and over again in the same reproducible sequence*” (Englander & Mayne, 2017a, p. 8256). In particular, structurally characterizable transient states – i.e., “foldons” – pave the significant stepping stones of the folding process. This is the most important experimental finding in foldon research: foldons fold as units in sequential order (Englander & Mayne, 2017a, p. 8254). The foldon hypothesis suggests that conformational search is due to macro-level (i.e., not solely at the amino acid level) interactions between the components of foldons and between foldons, what Englander and Mayne (2017a) call cooperatively organized native-like intrafoldon and interfoldon interactions. In that sense, foldon research is centred on the structural characterization of the transient states that occur during the folding process, an endeavour that, to reiterate the point, is at odds with the thermodynamic approach to folding dynamics.

The foldon hypothesis gives rise to new research questions, such as: how do foldons and “foldon-based body plans” (Englander & Mayne, 2017a, p. 8256) evolve? What drives foldon assembly and interactions? For our present analytic purposes, the relevance of the foldon hypothesis is that, ultimately, the principles of folding are to be sought at the level of macro-entities such as foldons, cooperative units of amino acids, rather than atomistically. Probably the most significant ontological aspect of this view is that foldons in isolation are less stable than in the complex. This cooperativity between foldon components and between foldons suggests an anti-reductionist and relational view (see Santos, Chap. 12, this volume) of folding whereby the units of decomposition are macro-level structural or organizational units: a foldon cannot be characterised as just the sum of its components – i.e., amino acids – taken in isolation or, put differently, the behaviour of a foldon is not accountable in terms of the intrinsic properties of its components independently of their relational context. Thus, at least some version of the kinetic approach, such as the foldon hypothesis, should be contrasted to the thermodynamic approach in terms of their differing ontological commitments concerning the nature of the relevant entities and activities realizing the folding phenomena.

6.5 What Kind of Explanations Are Thermodynamic Explanations of Folding?

As we argued in Sect. 6.3.2, the kinetic approach aims to generate explanations of folding phenomena that are straightforwardly causal and mechanistic. Meanwhile, the causal nature of thermodynamic explanations remains dubious. In this section, we shall analyse how thermodynamic explanations of native state stability and folding dynamics may be interpreted.

6.5.1 *Thermodynamic Explanations of Native state Stability*

Our first attempt is to consider thermodynamic explanations of native state stability as instances of equilibrium explanations (Sober, 1983; Sperry-Taylor, 2019), in which the stable equilibrium condition that is maintained and to which the system returns when perturbed is, of course, the native state. According to Sober (1983), equilibrium explanations are not causal, as they do not refer to actual initial conditions or actual processes. However, Sober also argues that equilibrium explanations can be more informative than causal ones as they provide a particular kind of “understanding” related to situations whereby a system’s dynamical behaviour is governed by global equilibria, which are those to which a system reverts (or is “attracted to”) independently of initial conditions. In this case “...an event can be explained in the face of considerable ignorance of the actual forces and initial conditions that in fact caused the system to be in its equilibrium state. In this circumstance, we are, in one natural sense, ignorant of the event’s cause, but explanation is possible nonetheless” (Sober, 1983 p. 209).

However, this interpretation has some problems: when it is asked “why protein x reverts to putatively global equilibrium N ?”, are we not seeking a causal explanation? Consider this analogy with organismal homeostasis. For instance, internal temperature regulation is dependent on sensing external temperature; when the temperature increases, the organism responds (e.g., by producing some kind of metabolic change that, by assumption, is mechanistically accountable); thus, an explanation of heat regulation seems to be partially causal and mechanistic.

Note also that two explananda can be identified at this juncture: why a system reverts to equilibrium state and why the equilibrium state has that particular nature. What we are arguing here is that, in the case of homeostasis, we clearly rely on a causal and mechanistic explanans to explain equilibrium maintenance, even though in different organisms a different mechanism might act. The second explanandum (i.e., why internal temperature 37 °C is a global attractor or equilibrium) might have a different kind of explanation (e.g., evolutionary), which might nevertheless still be causal.

Analogous considerations, we surmise, pertain to native structure stability. In particular, just knowing the energy values of each possible conformation, which would give us the minimal energy value (or the maximally stable conformation, i.e., the native state), is neither enough to explain the dynamics of reversion to the native state nor to explain why the native state is the state of maximum stability. To achieve the first explanatory aim, it is necessary to appeal to the underlying mechanisms of stabilization in terms of the energy contributions of all the interactions formed between the parts of the protein, as well as the relations of the protein with the environment. Features like having a strong hydrophobic core, a great number of electrostatic interactions (e.g., salt bridges) on the surface, a high number of hydrogen bonds in certain configurations, etc. could explain why the native state of a protein is more stable than all (or most of) the other possible conformations. In the same way, the localization of a charged residue within the core, the existence of

hydrophobic patches at the surface, the existence of torsions, etc. could explain the instability of the native state of a given protein. In summary, to explain why the system reverts to the native state, it is necessary to appeal to the properties of the parts of the protein and the interaction between them and the environment. Dynamic reversion ($U \leftrightarrow N$) is grounded on structural properties (e.g., the existence and number of salt bridges, hydrogen and disulphide bonds), where proteins acquire such properties during folding in ways that are *prima facie* mechanistically accountable. Furthermore, given that the maintenance of the native state (including resistance to perturbation or return to equilibrium state after perturbation) is a dynamical process, kinetic considerations seem to be necessary (at least to complement ideal thermodynamic explanations) to account for proteins' behaviour.

To achieve the second explanatory aim – i.e., why the equilibrium state is that particular native state rather than another – an explanation might resort not only to principles of chemical stability but also to natural selection. Whether such principles are in principle mechanistically accountable is difficult to say.

Overall, the thermodynamic explanations of the two explananda related to native state stability (i.e., the reversion to equilibrium state and why to that particular equilibrium state) seem to us clearly amenable to be interpreted in causal and mechanistic terms, but only when complemented by either kinetic considerations (when the aim is to explain the maintenance of the native state) or chemical/evolutionary ones (when the aim is to explain the nature of the equilibrium state).

6.5.2 *Thermodynamic Explanations of Folding Dynamics*

When we consider thermodynamic explanations of folding dynamics, their mechanistic credentials are even more suspicious. First of all, thermodynamic explanations of folding dynamics seem intuitively more compatible with a formal deduction schema. They seem exemplars of the covering law model (by referring to a variety of thermodynamic generalisations and laws concerning Gibbs free energy, enthalpy and entropy) against which new mechanists have originally dedicated so much ink. However, as many authors have argued throughout history (see Santos, Chap. 12, this volume), even the purest mechanistic explanation must inevitably refer to some form of generalisation (a point more recently argued by Cartwright et al., 2020). We largely agree with this latter position and see no good reason to distinguish so sharply between mechanistic explanations and explanations referring to putative law-like generalisations. At the same time, advocates of the thermodynamic approach should clarify which thermodynamic generalizations are causal and, as such, proper explanantia for the phenomenon to be explained, that is, the acquisition of native structure.

Secondly, because of the absence of the temporal variable and the apparent omission of the causal details concerning how a protein reaches native state, it is difficult to make sense of the putative mechanistic (or even causal) nature of thermodynamic explanations. The lack of an explicit temporal and causal interpretation might

suggest that thermodynamic explanations are instances of constitutive mechanistic explanations. Indeed, since the energy landscape on which the ideal explanatory aims of this approach depends on is nothing more than the conformational possibility phase state of a protein type together with the constraints associated to the free energy of each conformation, it would be tempting to affirm that this description of the biochemical system constitutes (or is even identical to, see Craver et al., 2021) the phenomenon to be explained, that is, the thermodynamic behaviour of the protein. However, this constitutive interpretation would be at odds with the conception endorsed by practicing scientist advocating the thermodynamic approach, for whom the putative mechanisms underlying native state stability as well as those underlying folding dynamics cause rather than constitute the phenomena.

A third alternative interpretation is that explanations of folding dynamics are developmental explanations of a particular kind, that is, hybrids between constitutive and causal explanations. Following Ylikoski's (2013, p. 293) analysis, we might consider a developmental hybrid explanation of the following form: "Biochemical system S (i.e., the polypeptide chain type) has the causal capacity of folding to native state in a folding environment E due to S's components (i.e., the intrinsic properties of amino acids and peptide bonds) and their organization O (i.e., the arrangement of the amino acids in a linear polypeptide with a given sequence)." From a thermodynamic perspective, this explanation accounts for the fact that S generates different conformational distributions or ensembles at different stages of the process. However, thermodynamic approaches need to account for the directionality of the folding process in terms that are causally rich enough to make sense of its supposed "spontaneity", otherwise spontaneity is black boxed.²⁹ Whenever thermodynamic approaches provide causally rich details, for instance by appealing to the hydrophobic effect, they are implicitly committed to a causal account in terms of structural modifications, which is standardly mechanistic. This is indeed what kinetic approaches do. According to these, S as an unfolded polypeptide chain and S as a folded and functional protein differ in their organization, often in composition (e.g., if extrinsic components are integrated in the system) and, consequently, in their behaviour, as they acquire new properties as the folding process proceeds. The issue is thus not about constitution, but rather about whether thermodynamic approaches account for folding dynamics causally.

Another alternative to make sense of thermodynamic explanations of folding dynamics in causal terms should be mentioned. Often, the claims of the advocates of thermodynamic approaches seem to reify the geometrical properties of graphical representations such as energy landscapes, attributing causal roles to them that are difficult to comprehend. For instance, the claim according to which the conformational search is dependent on the "... *bias toward native interactions intrinsic to a funneled landscape*" (Eaton & Wolynes, 2017, p. E9759) gives rise to this kind of interpretation. Sometimes these claims are accompanied by others concerning the

²⁹The supposed spontaneity of the folding process has recently been the subject of criticism (see for example Sorokina et al., 2022).

directedness imparted to the search or its “encoding” (see note 23) by the energy landscape:

However, how this propensity [i.e., the thermodynamic bias] might be encoded in the physical chemistry of protein structure has never been discovered. One simply asserts the general proposition that it is encoded in the shape of the landscape and to an ad hoc principle named minimal frustration imposed by natural evolution. Englander and Mayne 2017a p. 8256

As this quotation illustrates, the reification of the properties of the energy landscape and the attribution of causal capacities to them is accompanied by assuming the biochemical relevance of debatable folding principles (e.g., minimal frustration)³⁰ as well as by a seemingly teleological interpretation of folding:

For effective performance, folding proteins must “know” how to select native as opposed to nonnative interactions. This information is said to be contained in the shape of the energy landscape, but how it is implemented in the physical chemistry of any given protein, or proteins in general, is unknown. Englander and Mayne, 2017a p. 8253

Therefore, the appeal to energy landscapes raises suspicions about the causal character of thermodynamic explanations of folding dynamics. What exactly do energy landscapes represent? A possible interpretation is that, since an energy landscape represents all the possible conformations a protein could attain, it also represents all the possible pathways or potential causal sequences a protein could take during folding. This interpretation would again lead us back to the issue concerning the causal nature of the processes underlying folding.

Unless we discount the causal legitimacy of thermodynamic explanations of folding dynamics *tout court*, there must be a way to bridge thermodynamic and causal talk. Indeed, in our opinion the most appropriate interpretation of thermodynamic explanations of folding dynamics is by bridging mechanistic analysis and energetics. This connects to the recent proposal that mechanistic explanations bottom out in energetics (Bechtel & Bollhagen, 2021). Basically, activities would be grounded on constraints on “the flow of free energy”. Thermodynamic explanations of folding emphasising energetic considerations approximate this new breed of extended mechanistic analysis focused on identifying the sources of free energy necessary for the mechanism to perform work and being active. In thermodynamic approaches of protein folding, Brownian motion must be countenanced as a significant force. In a sense, it might be considered as one source of activity or free energy underlying folding. However, if Brownian motion solely regulated the folding process, we would be left wondering how native state can as a matter of fact be reached. Therefore, to answer the question of why some atomistic interactions tend to occur

³⁰Frustration refers, in this context, to a general property of any linear chain composed of monomers of different nature like, for example, a random peptide chain. In the case of a random peptide chain, the most common fate will be a collapse in many different 3D structures instead of a unique (or very restricted ensemble of) globular, stable and soluble 3D structures. Nevertheless, extant proteins do most often fold into a native state. This behaviour is called minimal frustration: the fact that extant proteins acquire native state would be the result of the selection of “minimally frustrated” peptide chains through evolution. This “minimal frustration principle” is at the core of current versions of the thermodynamic hypotheses.

with higher probability than others, what is needed is a more rigorous causal account of the conformational search in thermodynamic terms. One possible suggestion is that the differences in stability (i.e., accounted in terms of Gibbs free energy) between the different conformations determine which ones are transient and which ones are acquired during the process leading to native state. However, as Englander and Mayne (2017a, p. 8257) note, this way of interpreting the thermodynamic constraint does neither refer to nor identify any relevant molecular properties of proteins:

Atypically, the funneled landscape emblematic of energy landscape theory does not deal with molecular properties that would serve to guide interactions. It portrays some external thermodynamic constraints that are valid for the folding of proteins, RNA, or any other polymer. It contains in itself no molecular information or molecule-based constraints or predictions.

If the energetic considerations considered central to folding dynamics by advocates of the thermodynamic approach are so general as to pertain to any polymer, it is not surprising that a causal and mechanistic interpretation of the thermodynamics of folding dynamics is not easily forthcoming.

In a sense, this difficulty is not surprising, as thermodynamic approaches are grounded on disciplines such as chemical thermodynamics (whose primary focus is on the direction of chemical reactions independently of the underlying reaction mechanisms) and inspired by statistical mechanics. One source of inspiration is the analogy with ideal gases, that captures the important point that, despite continuous change at the micro-level (due to Brownian motion for instance), the ensemble during folding changes only in terms of the probability distribution of the fixed set of possible protein conformations – which are analogous to microstates (see Sect. 6.4.1 and note 23). This analogy is questionable in many senses. First, the uniformity assumption (i.e., with the idea that the molecules of an ideal gas are identical) is questionable in the protein case; in fact, proteins of a same type might vary in composition, for instance by acquiring new structural components from the environment or even by acquiring variations in composition that might occur *in-vivo* because of mistranslation. Secondly, as advocates of kinetic approaches stress, there is no reason to believe, experimentally and theoretically, that all possible conformations of a protein will as a matter of fact occur during folding. In the biochemical case, some conformations might never be realized. Thirdly, as advocates of kinetic approaches also stress, there is no experimental reason to deny that only some conformations are biochemically significant. To argue for the contrary position, as advocates of the thermodynamic approach do, is to borrow uncritically the analogy with ideal gases, where all microstates are assumed to be (“in the long term”) equiprobable (i.e., ergodic hypothesis of statistical thermodynamics). Fourthly, and most importantly, unlike phase space, the energy landscape might not be fixed (Sect. 6.4.1) at the outset by the intrinsic properties of the protein type, grounded on its characteristic amino acid composition. It is rather co-determined by the properties of the protein environment, including the varieties of molecules the developing protein interacts with during folding (see Sorokina et al., 2022). In a nutshell, unlike

phase space, the energy landscape might be dynamic and, consequently, cannot be reduced to a mere representation of the polypeptide chain's degrees of freedom considered as an isolated system (i.e., solely characterized in terms of the intrinsic properties of its components).³¹

At the same time, as already indicated in Sect. 6.5.1, thermodynamic approaches make implicit reference to causal processes such as the formation of salt bridges, hydrogen bonds or hydrophobic effects. In this sense, note that the hydrophobic effect can either be given an energetic interpretation (whereby, in an aqueous medium, the solvent entropy is higher when hydrophobic moieties are not being solvated, i.e., the free energy of the system is lower, so that they are grouped together) or a mechanistic interpretation (whereby hydrophobic amino acids tend to move inside the folding structure while hydrophilic ones tend to position themselves on the external part of the protein structure during the folding process, leading to the compaction of the polypeptide chain). Both interpretations have a sound rationale and might be considered, as it is often the case in biochemical practice, complementary. The contrast between thermodynamic and kinetic approaches might as a result be more properly diagnosed as a dispute concerning the appropriate analytic strategy to explain folding phenomena.

6.6 Conclusion

The protein folding problem is a philosophically fertile field that has received, as far as we know, limited attention in the philosophies of chemistry and biology despite its central importance in biochemistry and, more generally, biology. There are many aspects of this problem that remain to be addressed beyond those considered in this chapter. In this sense, it should be stressed that part of the contrast between thermodynamic and kinetic approaches is intimately related to the use of different experimental and modelling techniques. Advocates of the kinetic approach tend to base their arguments mostly on experimental results, while advocates of the thermodynamic approach tend to adopt mostly theoretical and computational practices, being computer simulations of simplified models a central one. At the same time, both approaches have common difficulties when attempting to explain why extant proteins fold as they do, independently of whether the answer is sought by asking the question of why a protein has a particular energy landscape or, alternatively, why a specific order of stages occurs. When faced with these deep questions, both approaches often resort to evolutionary biology. The most common answer, which

³¹Analysing the case of allostery, Neal (2021, p. 209) indeed argues that the analogy with ideal gases is broken because "The perturbation of the ensemble by the allosteric ligand remodels the energy landscape of the entire system because the energetic properties of the microstates themselves were differentially altered by the allosteric ligand." It remains unclear whether Neal is arguing that this is just a change in the probability distribution of microstates or whether a new set of microstates is formed in the process.

betrays an adaptationist bias (see the principle of minimal frustration, note 24), is that extant proteins are as they are and fold as they fold because they are the results of adaptive evolution. In other words, extant proteins are assumed to be adaptive traits that can reach certain conformations in their physiological contexts at a speed that allows them to perform their biological functions.

It must also be noted that, despite its relative antiquity, the protein folding problem remains an unsolved problem in biochemical and biophysical research. Despite experimental and theoretical advances (including data-driven approaches such as AlphaFold, see note 6), new questions and new debates are continuously emerging. In this article we have assessed one major source of disagreement, rooted in the divergence between two major explanatory approaches to the protein folding problem that can be traced back to when the problem was firstly formulated: a thermodynamic approach focused on energetic features and a kinetic approach centred on temporal and structural ones. Despite the partial agreement between these approaches on various aspects of the folding process and their complementarity – evident in biochemical practice – in generating hybrid explanations, there remain significant contrasts between them. We have tried to uncover some aspects of such contrast related to the relevance assigned to different epistemic resources and the causal nature of the explanations proposed. We also identified their different ontological assumptions. While the thermodynamic approach localizes the relevant epistemic resources at the atomic level, thus aiming to define the properties of ensembles, the kinetic approach considers as central the average behaviour of populations of proteins, thus aiming to provide a structural description of the stages of the folding process. Thermodynamic approaches are based on a conceptualization of the folding process whereby pathways are unwarranted postulations. Conversely, kinetic approaches deny the relevance of the atomic level of analysis (by itself) because it is experimentally inaccessible. Concerning the causal nature of the explanations, kinetic explanations are straightforwardly causal and mechanistic, while the causal nature of thermodynamic ones is elusive and difficult to interpret. Here the incipient contrast between thermodynamic and structural approaches comes to the fore, which is a further expression of the difficulty of applying a structure-based mechanistic model of explanation to chemistry (Scerri, Chap. 8, this volume) and physics (Falkenburg, Chap. 10, this volume) alike.³² Finally, the thermodynamic approach conceives the folding process as a manifestation (or suppression) of the predetermined intrinsic properties of the protein components, while the kinetic approach – at least in some forms – seems compatible with an anti-reductionist and relational perspective (Santos, Chap. 12, this volume) whereby proteins, during development, diachronically acquire novel properties, some of which only characterizable macroscopically. What results from this state of affairs is a “hardly-to-integrate” pluralism (Bolinska, 2022 reaches similar conclusions in the case of protein structure determination). At first glance, these disagreements may be interpreted as a classical

³² Structural approaches in biochemistry become especially problematic when protein function is not dependent on structural change, as IDPs (see note 7) seem to show.

example of “relative significance” debate (Beatty, 1997). However, in this case both approaches have led to interpretations of natural phenomena that are difficult to reconcile. Indeed, from the thermodynamic perspective, it is always possible to argue that structural pathways are chimeras produced by centring attention on selected macroscopic features. Conversely, for the advocates of the kinetic approach, it is always possible to argue that multiple and parallel folding routes are irrelevant as a basis for generalizable explanations. These contrasts are in our opinion hardly resolvable.

Acknowledgements Gabriel Vallejos-Baccelliere acknowledges the financial support of Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT grant N° 3210758). Davide Vecchi acknowledges the financial support of the FCT—Fundação para a Ciência e a Tecnologia (DL57/2016/CP1479/CT0072; Grants N. UIDB/00678/2020 and UIDP/00678/2020). We would like to thank Gil Santos, Eric Scerri and two anonymous reviewers for stimulating criticisms and suggestions to improve the clarity of the manuscript. Additional thanks to Ariel Roffé, Santiago Ginnobili, Daniel Nicholson, Luis González-Flecha and the audience at the 17th CLMPST, held in Buenos Aires, Argentina. Gabriel Vallejos-Baccelliere also thanks the academic staff of Laboratorio de Bioquímica y Biología Molecular, Departamento de Biología, Facultad de Ciencias, Universidad de Chile.

References

- Austin, R. C. (2009). The unfolded protein response in health and disease. *Antioxidants & Redox Signaling*, 11(9), 2279–2287. <https://doi.org/10.1089/ars.2009.2686>
- Baldwin, R. L. (2008). The search for folding intermediates and the mechanism of protein folding. *Annual Review of Biophysics*, 37, 1–21. <https://doi.org/10.1146/annurev.biophys.37.032807.125948>
- Beatty, J. (1997). Why do biologists argue like they do? *Philosophy of Science*, 64, S432–S443. <http://www.jstor.org/stable/188423>
- Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science*, 78(4), 533–557.
- Bechtel, W., & Bollhagen, A. (2021). Active biological mechanisms: Transforming energy into motion in molecular motors. *Synthese*, 199(5–6), 12705–12729.
- Ben-Naim, A. (2015). *Myths and verities in protein folding theories*. WSPC.
- Bolinska, A. (2022). Monist proposal: Against integrative pluralism about protein structure. *Erkenn*. <https://doi.org/10.1007/s10670-022-00601-2>
- Cartwright, N., Pemberton, J., & Wieten, S. (2020). Mechanisms, laws and explanation. *European Journal of Philosophy of Science*, 10, 25. <https://doi.org/10.1007/s13194-020-00284-y>
- Craver, C. F., Glennan, S., & Povich, M. (2021). Constitutive relevance & mutual manipulability revisited. *Synthese*, 199(3–4), 8807–8828. <https://doi.org/10.1007/s11229-021-03183-8>
- Dill, K. A. (1999). Polymer principles and protein folding. *Protein Science: A Publication of the Protein Society*, 8(6), 1166–1180. <https://doi.org/10.1110/ps.8.6.1166>
- Dill, K. A., & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural Biology*, 4(1), 10–19.
- Eaton, W. A., & Wolynes, P. G. (2017). Theory, simulations, and experiments show that proteins fold by multiple pathways. *PNAS*, 114(46), E9759–E9760.
- Englander, S. W., & Mayne, L. (2017a). The case for defined protein folding pathways. *PNAS*, 114(31), 8253–8258.
- Englander, S. W., & Mayne, L. (2017b). Reply to Eaton and Wolynes: How do proteins fold? *PNAS*, 114(46), E9761–E9762.

- Fersht, A. R. (1995). Characterizing transition states in protein folding: An essential step in the puzzle. *Current Opinion in Structural Biology*, 5, 79–84.
- Fersht, A. R. (1998). *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. W. H. Freeman.
- Francoeur, E. (2001). Molecular models and the articulation of structural constraints in chemistry. In U. Klein (Ed.), *Tools and modes of representation in the laboratory sciences* (Boston studies in the philosophy and history of science) (Vol. 222). Springer. https://doi.org/10.1007/978-94-015-9737-1_7
- Francoeur, E. (2002). Cyrus Levinthal, the Kluge and the origins of interactive molecular graphics. *Endeavour*, 26(4), 127–131. [https://doi.org/10.1016/s0160-9327\(02\)01468-0](https://doi.org/10.1016/s0160-9327(02)01468-0)
- Glennan, S., Illari, P., & Weber, E. (2022). Six theses on mechanisms and mechanistic science. *Journal for General Philosophy of Science*, 53, 143–161.
- Gomes, C., & Faisca, P. (2019). *Protein folding: An introduction*. Springer.
- Govindarajan, S., & Goldstein, R. A. (1998). On the thermodynamic hypothesis of protein folding. *PNAS*, 95(10), 5545–5549.
- Krebs, J. E., Goldstein, E. S., & Kilpatrick, S. T. (2018). *Lewin's genes XII*. Jones & Bartlett Learning.
- Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique*, 65, 44–45.
- Levinthal, C. (1969). How to fold gracefully. Mossbauer spectroscopy in biological systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois: 22–24.
- Levy, A., & Bechtel, W. (2016). *Towards mechanism 2.0: Expanding the scope of mechanistic explanation*. <http://philsci-archiv.pitt.edu/12567/>
- Liu, P. P., Xie, Y., Meng, X. Y., & Kang, J. S. (2019). History and progress of hypotheses and clinical trials for Alzheimer's disease. *Signal Transduction and Targeted Therapy*, 4, 29. <https://doi.org/10.1038/s41392-019-0063-8>. Erratum in: *Signal Transduct Target Ther*. 2019 Sep 23;4:37. PMID: 31637009; PMCID: PMC6799833.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1–25.
- Medina, E., R Latham, D., & Sanabria, H. (2021). Unraveling protein's structural dynamics: From configurational dynamics to ensemble switching guides functional mesoscale assemblies. *Current Opinion in Structural Biology*, 66, 129–138. <https://doi.org/10.1016/j.sbi.2020.10.016>
- Neal, J. P. (2021). *Protein structure, dynamics, and function: A philosophical account of representation and explanation in structural biology*. Ph.D. Dissertation, The University of Pittsburgh, USA. <http://d-scholarship.pitt.edu/41715/13/NealJP%20ETD%20History%20%26%20Philosophy%20of%20Proteins.pdf>
- Nickson, A. A., & Clarke, J. (2010). What lessons can be learned from studying the folding of homologous proteins? *Methods (San Diego, Calif.)*, 52(1), 38–50. <https://doi.org/10.1016/j.ymeth.2010.06.003>
- Onuchic, J. N., Luthey-Schulten, Z., & Wolynes, P. G. (1997). Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry*, 48, 545–600. <https://doi.org/10.1146/annurev.physchem.48.1.545>
- Plaxco, K. W., Simons, K. T., & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology*, 277(4), 985–994. <https://doi.org/10.1006/jmbi.1998.1645>
- Rose, G. D., Fleming, P. J., Banavar, J. R., & Maritan, A. (2006). A backbone-based theory of protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 103(45), 16623–16633. <https://doi.org/10.1073/pnas.0606843103>
- Santos, G., Vallejos, G., & Vecchi, D. (2020). A relational-constructionist account of protein macrostructure and function. *Foundations of Chemistry*, 22(3), 363–382.
- Sober, E. (1983). Equilibrium explanation. *Philosophical Studies*, 43(2), 201–210.

- Sorokina, I., Mushegian, A. R., & Koonin, E. V. (2022). Is protein folding a thermodynamically unfavorable, active, energy-dependent process? *International Journal of Molecular Sciences*, 23, 521. <https://doi.org/10.3390/ijms23010521>
- Sperry-Taylor, A. T. (2019). Reassessing equilibrium explanations: When are they causal explanations? *Synthese*, 198(6), 5577–5598.
- Tanford, C., & Reynolds, J. A. (2003). *Nature's robots: A history of proteins*. Oxford University Press.
- Vallejos-Bacelliere, G. (2022). Problemas contemporáneos en la filosofía de la bioquímica. *Culturas Científicas*, 3(1), 45–77. <https://doi.org/10.35588/cc.v3i1.5584>
- Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis*, 78(2), 277–297.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

