Design and discovery of metamorphic proteins

Acacia F. Dishman1,2, Brian F. Volkman1
1Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI, USA
2Medical Scientist Training Program, Medical College of Wisconsin, Milwaukee, WI, USA

Abstract

Metamorphic proteins are single amino acid sequences that reversibly interconvert between multiple, dramatically different native structures, often with distinct functions. Since the discovery of the first metamorphic proteins in the early 2000s, several additional metamorphic proteins have been identified, and it was suggested that up to 4% of proteins in the PDB may switch folds. Metamorphic proteins have been found to share common features such as marginal thermostability and inconsistencies in predicted secondary structures. Outstanding challenges in the field include the search for more metamorphic proteins and the design of new proteins that switch folds. Identification of novel metamorphic proteins in nature will improve therapeutic targeting of fold-switching proteins involved in human pathology and will enhance the design of protein-based therapies. Designed fold switching proteins have applications as biosensors, molecular switches, molecular machines, and self-assembling systems.

Introduction

In the 70 years since Pauling, Corey and Branson predicted the α-helix and β-sheet, and the 50 years since Anfinsen’s thermodynamic hypothesis, our understanding of protein structure and folding coalesced around two main categories: globular proteins that find their native states spontaneously (either unaided or with help from chaperones) and fibrous proteins (multichain polymers typically assembled in the endoplasmic reticulum and the extracellular space). Then, in the late 1990s, intrinsically unfolded proteins and domains, which contain too few apolar sidechains to form a stable hydrophobic core, were recognized to comprise a significant fraction of the eukaryotic proteome [1].

The 2002 discovery that XCL1 (previously known as lymphotactin) simultaneously adopts two unrelated native state structures again challenged the one sequence-one fold paradigm [2]. With the subsequent identification of several other examples, metamorphic proteins are now established as a distinct category with special properties and functions that remain largely unexplored (Figure 1a) [3]. Our most recent work used ancestral sequence reconstruction to trace the evolutionary history of XCL1 and identified the first ancestor
that interconverted between two distinct conformations [4]. Dramatic shifts in the XCL1 metamorphic equilibrium from one ancestral node to the next suggest that the 1:1 ratio of β-sheet and chemokine conformations observed for human XCL1 is an optimized product of evolutionary selection (Figure 1b). Metamorphic folding was thus established as a favorable attribute that can improve protein fitness. In this article, we clarify the essential features and properties of a metamorphic protein, examine the search for more metamorphic proteins in nature, and assess the prospects for creation of metamorphic proteins as nano-scale tools or medical treatments.

**Observable characteristics of metamorphic proteins**

Anfinsen’s thermodynamic hypothesis (the native structure is the conformation with the lowest free energy) and Levinthal’s paradox (rapid folding cannot be achieved via random search of possible configurations) launched the ‘protein folding problem’ and the search for stepwise folding pathways and intermediate states. As the database of solved structures grew, the native structure was widely presumed to be unique — that is, each protein sequence would encode only one lowest-energy configuration. However, as others have noted [5] this restrictive definition of the native state appears to be a misreading of Anfinsen. This concept is widely depicted as a ‘folding funnel’ which relates the number of conformational states (width) to the change in Gibbs free energy (depth) for the transition from unfolded (top) to native (bottom) conformation (ΔG of folding or ΔGfold) (Figure 1c).

Moreover, by the time Dill and Chan popularized the folding funnel concept in 1997 it was already clear that “if landscapes are smooth, then native proteins should have small fluctuations, but if landscapes are rugged, then native proteins could fluctuate to very different conformations. Small changes in energy could lead to large changes in structure” (Figure 1d,e). In theory, a polypeptide with two very different folded configurations of equal free energy would adopt both in equal proportions and fold-switching transitions would occur at a rate governed by the transition state energy. This would require them to have a folding funnel with multiple similarly shallow wells (Figure 1e) to permit interconversion via partial or complete unfolding, as was demonstrated for XCL1 [6]. Within a few years the first examples of proteins exhibiting these unusual features were described and eventually designated ‘metamorphic’ proteins by Alexey Murzin [3]. Intrinsically disordered proteins (IDPs) also fit naturally into this paradigm as featureless energy landscapes (Figure 1f).

Random sequence polypeptides do not fold spontaneously, and methods for computational design have only recently gained the ability to design stable structures with a deep folding funnel. If stable proteins are unusual, encoding more than one folded structure in a single amino acid sequence is a daunting prospect. Furthermore, facile switching between structures would require a particular type of conformational energy landscape ensuring that metamorphic proteins are relatively rare and exhibit specific biophysical characteristics (Box 1). Metamorphic proteins are more likely to exhibit marginal thermodynamic stability than singlefold proteins (Figure 1g) [7].

Because discerning the presence of multiple conformational states is a challenge for most biophysical methods, naturally occurring metamorphic proteins can go undetected.
and are probably underrepresented in the protein data bank (PDB). Metamorphic proteins switch between conformations at rates that prevent the chromatographic separation of each native structure and may be detrimental to crystallization. As nuclear magnetic resonance (NMR) spectroscopy is uniquely suited to the detection of conformational heterogeneity, this method has been widely employed in the characterization of metamorphic proteins, starting with XCL1 (Figure 2a) [2]. If the rate of interconversion between metamorphic structures is limited by partial or complete unfolding (i.e., slower than $\sim 10 \text{ s}^{-1}$), a 2D heteronuclear single quantum coherence (HSQC) spectrum will contain distinct signals for each conformation with intensities proportional to their relative abundance in solution. Metamorphic behavior of well-studied metamorphic proteins RfaH (Figure 2b) and KaiB (Figure 2c) was also revealed by NMR analysis [8,9].

In a number of known metamorphic proteins, each fold encodes a distinct function [10,11]. For example, XCL1’s chemokine fold activates a G-protein couple receptor, while its $\beta$-sheet fold binds glycosaminoglycans (Figure 2a) [12]. One RfaH structure binds an $\alpha_{\text{ops}}$ DNA site, while the other structure binds a ribosomal subunit, meaning that fold-switching in RfaH couples bacterial transcription and translation (Figure 2b) [8]. The two folds of KaiB play different roles in a cyanobacterial circadian clock (Figure 2b) [13]. Other recent reviews catalogue known metamorphic proteins and their structure-function relationships in further detail [10,11,14,15]. Because identifying and separately characterizing the multiple conformations of metamorphic proteins is challenging, metamorphic protein functions may remain obscure long after the proteins are discovered. For example, XCL1 function was first described in 1994 [16], but it was not until 2008 that the two separate functions of XCL1’s two folds were disentangled [12]. Likewise, known proteins with single solved structures may have undetected alternate folds with alternate functions. Moonlighting proteins, which are known to have multiple functions [17], may thus be a particularly interesting group to search for previously undetected alternate folds.

**Searching for other natural metamorphic proteins**

When Murzin coined the term “metamorphic protein” in 2008, he cited XCL1 and Mad2 as two prime examples. Since then, despite the popular notion that metamorphic proteins were likely exceedingly rare, several other metamorphic proteins have been serendipitously discovered [10,11,14,15], including SARS-CoV-2 protein ORF9b [18]. Experimentally determined structures describe 17% of the amino acid residues in human protein sequences [19]. Protein structure prediction has taken a quantum leap with AlphaFold [20], which was recently used to expand structural coverage, in the form of high confidence structure predictions, to 58% of all residues in the human proteome [21]. Different estimates suggest that between 15 and 30% of residues in the proteome are disordered [22,23]. This still leaves at least one tenth of all residues without a high confidence prediction and unlikely to be unstructured. Additionally, even proteins with high confidence structure predictions may have alternate folds that remain uncaptured.

A systematic search by Porter and Looger identified 96 fold-switching proteins in the PDB in 2018, and estimated that as much as 4% of proteins in the PDB switch folds [24]. Fold-switching proteins may be difficult to identify, however. In the traditional paradigm,
one protein has one fold with one major function, and therefore once a single structure for a given protein is solved, additional structures are neither expected nor sought out. Additionally, metamorphic proteins are likely to be difficult to work with at the bench, particularly in the setting of structure determination by crystallography, due to their dynamic nature and low stability. Thus, metamorphic proteins are likely underrepresented in the PDB. It may also be difficult to generate high confidence structure predictions for metamorphic proteins due to their relative instability, as well as the fact that there may be a low-confidence prediction for each of their structures. Even predicting secondary structures for metamorphic proteins has been shown to generate inconsistencies [24–26].

Recently, these inconsistencies in secondary structure prediction have been harnessed to predict protein metamorphism. Of the 96 fold switching proteins identified by Porter and Looger, each represented by two structures, 85 pairs had at least one structure with substantial discrepancies between predicted and experimentally determined secondary structure [24]. Since then, secondary structure predictions for fold switching protein regions have been found to be consistently less accurate than secondary structure predictions for randomly selected, equally long regions of non-fold-switching proteins [26]. Alpha-beta discrepancies detected by secondary structure prediction algorithm JPred4 can also be used to predict whether protein families with similar sequences will have different folds, i.e., whether they will behave as “sequence-similar fold switchers” [25]. A recent preprint describes the use of such discrepancies in predicted secondary structure to identify fold-switching members of the >15,000-member NusG superfamily, which contains prototypical metamorphic protein RfaH [27]. The study predicted that ~25% of the superfamily would switch folds. A sparse survey of 10 sequence-diverse variants by circular dichroism (CD) and NMR supported this prediction. Additionally, this study identified fold switching behaviors in variants with as little as 32% sequence identity, suggesting that diverse sequences can encode fold switching behavior, and supporting the idea that there may be more metamorphic proteins in nature than would otherwise be expected.

Identifying more fold-switching proteins in nature will allow fold-switching proteins to be therapeutically targeted. Fold switching can be involved in pathology, as in the case of the most common cancer-associated mutation (D83V) in the protein MEF2B, a major source of somatic mutations in non-Hodgkin lymphoma. This mutation causes an α-helical region to switch to a β-strand [28]. This indicates that a single amino acid change can trigger a fold switch implicated in cancer pathogenesis. Knowing how to identify fold-switching proteins, and sequence similar fold-switchers that switch folds in response to a few mutations, will enhance targeting of protein structure-function relationships in disease. Fold-switching can also be important for certain proteins’ physiologic roles in fighting disease. XCL1 switches between two dramatically different folds, one of which binds and activates its cognate GPCR on dendritic cell surfaces, and the other of which binds glycosaminoglycans and directly kills microbial pathogens via membrane disruption [12,29,30]. XCL1 variants locked in each of the two folds have been engineered by the Volkman lab [31,32]. Matsuo and coworkers sought to develop a cancer vaccine using XCL1 as an adjuvant but had limited success until they used an XCL1 variant locked in the GPCR-binding fold by an additional a disulfide bond [33]. In mice, at the injection site and draining lymph nodes, this locked variant then achieved the authors’ goal of inducing accumulation of dendritic cells capable of antigen
cross-presentation to CD8+ T cells, triggering production of long-term antigen-specific memory CD8+ T cells [33]. This strategy would not have been possible without knowing that XCL1 has two folds with two functions. These examples highlight the importance of searching for more fold-switching proteins. If structural biologists stop after solving one structure for each protein, an entire universe of other protein folds may never be discovered.

**Design and engineering of fold switching proteins**

Design of amino acid sequences not found in nature that (1) stably adopt *multiple, distinct* well-defined structures and (2) *reversibly* interconvert between them remains a challenge. To date, proteins have been designed that undergo subtle, designed structural fluctuations reversibly, such as proteins termed DANCERs in which a reporter tryptophan residue flips from a buried to a solvent exposed conformation [34]. Proteins have also been designed that adopt distinct structures but do not reversibly interchange between them, such as Hori and Suigura’s Ant-F [35] and Wei and colleague’s XAA_GVDQ (Figure 3b) [36]. Sequences which differ by one amino acid and adopt different folds have been designed as well (Figure 3a) [37]. Additionally, a recent Bryan lab preprint reports the design of identical amino acid sequences that adopt distinct structures based on the presence or absence of additional residues preceding or following the sequence [38].

The Baker lab has designed a biologically active protein switch system called co-LOCKR (Colocalization-dependent Latching Orthogonal Cage—Key pRotein) using Rosetta, which consists of two proteins [39,40] (Figure 3d). The first protein switches from a monomeric, helical bundle, closed or “latched” conformation to a heterodimeric, open state in response to the addition of the second protein, called the “key,” which binds at the same interface as the “latch” domain [39,40]. The switch in this system is not reversible and the open state is not structurally well defined [39,40].

In 2006, Ambroggio and Kuhlman designed fold-switching protein Sw2, marking the first use of computational protein design to optimize a single sequence for two different structures (Figure 3c) [41]. Notably, one of Sw2’s two folds had numerous exposed hydrophobic residues and thus aggregated unless suspended in buffer containing 200 mM guanidine. Similarly, Cerasoli and colleagues designed a 27-residue peptide that switches between two folds reversibly, using sequence comparison and visual pattern recognition rather than energy based calculations [42]. Since these examples, to our knowledge, a *de novo* designed protein that reversibly exchanges between two distinct structures has not been designed.

The challenge of consistently designing single sequences which interconvert reversibly between two distinct, arbitrary structures, particularly structures not found in nature, remains unsolved. Current protein design methods search for amino acid sequences that fold into a chosen structure with the lowest free energy, a strategy which has been successful in designing highly stable, monomorphic proteins, but poses challenges for the design of bistable proteins. Current protein design methods also struggle to predict folding energies of designed proteins at high accuracy, making it difficult to design a protein with two folds that have highly similar energies of folding, as is required for fold-switching. Even if two
folds with similar energies are designed, the energy barrier between the two must not be so high as to prevent switching, which is currently difficult to predict using available protein design software. To circumvent this challenge, one can imagine designing proteins with two relatively unstable folds; however, the folds must then not be so unstable that the designed protein is intrinsically disordered or adopts a heterogeneous mixture of conformations. In all, the design of bi-stable proteins is a formidable challenge.

New methods for bistable protein design, such as the Rosetta-based multistate design approach called RECON from the Meiler lab [43], show promise for designing sequences with multiple characteristics, e.g., multi-specific influenza antibodies [44]. Additionally, recent advances in structure prediction methodologies such as AlphaFold2 [20] will likely improve our ability to predict which designed proteins will be capable of stably adopting two different folds. Predictive methods such as a Baker lab deep learning approach have even been shown to be capable of detecting alternative minima in energy landscapes [45]. In a recent preprint, Jendrusch and colleagues embed AlphaFold in a novel protein design framework and use this framework to design sequences that are predicted to switch folds upon oligomerization with high confidence. While these designs remain untested in the laboratory, this framework demonstrates the way AlphaFold, or other structure prediction techniques, can advance design of fold switching proteins [46]. Together, these recent advances in multistate design and structure prediction provide hope for the future design of fold switching proteins.

De novo designed fold-switching proteins have diverse applications in healthcare and industry, and knowledge gained from the design process will have fundamental implications for understanding protein folding and dynamics [47]. For example, fold-switching proteins could be designed to act as molecular switches, for use as therapeutics, sensors, or components of cellular computers. Molecular switch systems have already been built using de novo designed proteins, such as the co-LOCKR system described above, or a system of de novo designed proteins which switch from monomer to trimer assemblies based on pH [39,40,48]. Co-LOCKR variants have been engineered to recognize certain combinations of surface antigens specific for cancer cells — e.g., Her2 and EGFR on the surface of breast cancer cells — switching “on” only in the presence of both antigens [49]. The goal is to use this system to target therapies specifically to cancer cells while sparing non-cancerous host cells [49].

Protein switch systems designed to date largely undergo hinge motions rather than the more substantial changes in intramolecular contacts and even secondary structure undergone by metamorphic proteins (Figure 3e,f). Metamorphic switches provide the opportunity for more complete rearrangement of the primary sequence in space, allowing multiple distinct spatial combinations of amino acids to be encoded in the same sequence. Moreover, metamorphic proteins that interconvert via complete or partial unfolding are likely to switch more slowly (e.g. seconds) than hinge-based (millisecond) conformational rearrangements. Slower switches would provide longer access to each fold-switched state, better enabling slower binding events. Each type of switch may be optimal for regulating different biological events. Control of molecular switch systems by fold switching proteins would provide an advantage over many current designs: reversible switching with the potential for tunable
kinetics. Not only could fold switching proteins serve a unique role as components of molecular sensors, switches, and machines, but they could provide key information on the fundamental processes of protein folding and dynamics. Efforts to create fold switching proteins will test and refine the design principles for sequences that encode multiple 3D structures. Protein engineering successes should in turn facilitate the search for other fold switching proteins in nature.

Acknowledgements

This work was supported in part by NIH grants R56 AI155881 (to B.F.V.) and F30 CA236182 (to A.F.D). A.F.D. is a member of the NIH supported (T32 GM080202) Medical Scientist Training Program at the Medical College of Wisconsin.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

1. Wright PE, Dyson HJ: Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. J Mol Biol 1999, 293:321–331. [PubMed: 10550212]

2. Kuloglu ES, McCaslin DR, Markley JL, Volkman BF: Structural rearrangement of human lymphotactin, a C chemokine, under physiological solution conditions. J Biol Chem 2002, 277: 17863–17870. [PubMed: 11889129]

3. Murzin AG: Biochemistry. Metamorphic proteins. Science 2008, 320:1725–1726. [PubMed: 18583598]

4. Dishman AF, Tyler RC, Fox JC, Kleist AB, Babu MM, Peterson FC, Volkman BF: Evolution of fold switching in a metamorphic protein. Science 2021, 371:86–90. [PubMed: 3384377] *Science: Authors resurrected ancient XCL1 sequences and studied them by NMR, finding that metamorphic folding is likely an adaptive trait for XCL1, and identifying principles for encoding two folds in one sequence.

5. Kulkarni P, Solomon TL, He Y, Chen Y, Bryan PN, Orban J: Structural metamorphism and polymorphism in proteins on the brink of thermodynamic stability. Protein Sci 2018, 27: 1557–1567. [PubMed: 30144197]

6. Tyler RC, Murray NJ, Peterson FC, Volkman BF: Native-state interconversion of a metamorphic protein requires global unfolding. Biochemistry 2011, 50:7077–7079. [PubMed: 21776971]

7. Vila JA: Metamorphic proteins in light of Anfinsen’s Dogma. J Phys Chem Lett 2020, 11:4998–4999. [PubMed: 32496777]

8. Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, Landick R, Artsimovitch I, Rosch P: An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell 2012, 150: 291–303. [PubMed: 22817892]

9. Tseng R, Goularte NF, Chavan A, Luu J, Cohen SE, Chang YG, Heisler J, Li S, Michael AK, Tripathi S, et al.: Structural basis of the day-night transition in a bacterial circadian clock. Science 2017, 355:1174–1180. [PubMed: 28302851]

10. Dishman AF, Volkman BF: Unfolding the mysteries of protein metamorphosis. ACS Chem Biol 2018, 13:1438–1446. [PubMed: 29787234]

11. Kim AK, Porter LL: Functional and regulatory roles of fold-switching proteins. Structure 2021, 29:6–14. [PubMed: 33176159]

12. Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, Volkman BF: Interconversion between two unrelated protein folds in the lymphotactin native state. Proc Natl Acad Sci U S A 2008, 105:5057–5062. [PubMed: 18364395]
13. Chang YG, Cohen SE, Phong C, Myers WK, Kim YI, Tseng R, Lin J, Zhang L, Boyd JS, Lee Y, et al.: Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. Science 2015, 349: 324–328. [PubMed: 26113641]

14. Das M, Chen N, Li Wang A, Wang LP: Identification and characterization of metamorphic proteins: current and future perspectives. Biopolymers 2021, 112, e23473. [PubMed: 34528703]

15. London RE: HIV-1 reverse transcriptase: a metamorphic protein with three stable states. Structure 2019, 27:420–426. [PubMed: 30639227]

16. Kelner GS, Kennedy J, Bacon KB, Kleyensteuber S, Largaespada DA, Jenkins NA, Copeland NG, Bazan JF, Moore KW, Schall TJ, et al.: Lymphotactin: a cytokine that represents a new class of chemokine. Science 1994, 266: 1395–1399. [PubMed: 7973732]

17. Chen C, Liu H, Zabad S, Rivera N, Rowin E, Hassan M, Gomez De Jesus SM, Llinas Santos PS, Kravchenko K, Mikhova M, et al.: MoonProt 3.0: an update of the moonlighting proteins database. Nucleic Acids Res 2021, 49:D368–D372. [PubMed: 33245761]

18. Porter LL: Predictable fold switching by the SARS-CoV-2 protein ORF9b. Protein Sci 2021, 30:1723–1729. [PubMed: 33934422]

19. Brogden KA, Johnson GK, Vincent SD, Abbasi T, Vali S: Oral inflammation, a role for antimicrobial peptide modulation of cytokine and chemokine responses. Expert Rev Anti Infect Ther 2013, 11:1097–1113. [PubMed: 24124799]

20. Jumper J, Evans R, Prizel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, et al.: Highly accurate protein structure prediction with AlphaFold. Nature 2021, 596:583–589. [PubMed: 34265844]

21. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zieliinski M, Zidek A, Bridgland A, Cowie A, Meyer C, Laydon A, et al.: Highly accurate protein structure prediction for the human proteome. Nature 2021, 596:590–596. [PubMed: 34293799]

22. Piovesan D, Necci M, Escobedo N, Monzon AM, Monos A, Micetic I, Quaglia F, Paladin L, Ramasamy P, Dosztanyi Z, et al.: MobiDB: intrinsically disordered proteins in 2021. Nucleic Acids Res 2021, 49:D361–D367. [PubMed: 33237329]

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

24. Porter LL, Looger LL: Extant fold-switching proteins are wide spread. Proc Natl Acad Sci U S A 2018, 115: 5968–5973. [PubMed: 29784778] *PNAS: Porter and Looger searched the PDB for proteins with multiple folds, identifying 96 fold-switching proteins whose different folds have different functions, and predicting that as much as 4% of the PDB will switch folds.

25. Kim AK, Looger LL, Porter LL: A high-throughput predictive method for sequence-similar fold switchers. Biopolymers 2021, 112, e23416. [PubMed: 33462801]

26. Mishra S, Looger LL, Porter LL: Inaccurate secondary structure predictions often indicate protein fold switching. Protein Sci 2019, 28:1487–1493. [PubMed: 31148305]

27. Porter LL, Kim AK, Looger LL, Majumdar A, Starich M: Pervasive fold switching in ubiquitous protein superfamilies. bioRxiv 2021. *bioRxiv: Authors predict which members of the >15,000-member NusG superfamily, which contains prototypical metamorphic protein RfaH, will switch folds, and show that their predictions are accurate through a sparse survey of 10 sequence-diverse variants by CD and NMR.

28. Lei X, Kou Y, Fu Y, Rajashekar N, Shi H, Wu F, Xu J, Luo Y, Chen L: The cancer mutation D83V induces an alpha-helix to beta-strand conformation switch in MEF2B. J Mol Biol 2018, 430:1157–1172. [PubMed: 29477338]

29. Nevins AM, Subramanian A, Tapia JL, Delgado DP, Tyler RC, Jensen DR, Ouellette AJ, Volkman BF: A requirement for metamorphic interconversion in the antimicrobial activity of chemokine XCL1. Biochemistry 2016, 55:3784–3793. [PubMed: 27305837]

30. Dishman AF, Lee MW, de Anda J, Lee EY, He J, Huppler AR, Wong GCL, Volkman BF: Switchable membrane remodeling and antifungal defense by metamorphic chemokine XCL1. ACS Infect Dis 2020, 6.

31. Tuinstra RL, Peterson FC, Elgin ES, Pelzek AJ, Volkman BF: An engineered second disulfide bond restricts lymphotactin/XCL1 to a chemokine-like conformation with XCR1 agonist activity. Biochemistry 2007, 46:2564–2573. [PubMed: 17302442]
32. Fox JC, Tyler RC, Guzzo C, Tuinstra RL, Peterson FC, Lusso P, Volkman BF: Engineering metamorphic chemokine lymphotactin/XCL1 into the GAG-binding, HIV-inhibitory dimer conformation. ACS Chem Biol 2015, 10:2580–2588. [PubMed: 26302421]

33. Matsuo K, Kitahata K, Kawabata F, Kamei M, Hara Y, Takamura S, Oiso N, Kawada A, Yoshie O, Nakayama T: A highly active form of XCL1/lymphotactin functions as an effective adjuvant to recruit cross-presenting dendritic cells for induction of effector and memory CD8(+) T cells. Front Immunol 2018, 9:2775. [PubMed: 30542351]

34. Davey JA, Damry AM, Goto NK, Chica RA: Rational design of proteins that exchange on functional timescales. Nat Chem Biol 2017, 13:1280–1285. [PubMed: 29058725]

35. Hori Y, Sugiyama Y: Conversion of antennapedia homeodomain to zinc finger-like domain: Zn(II)-induced change in protein conformation and DNA binding. J Am Chem Soc 2002, 124: 9362–9363. [PubMed: 12167017]

36. Wei KY, Moschidi D, Bick MJ, Nerli S, McShan AC, Carter LP, Huang PS, Fletcher DA, Sgourakis NG, Boyken SE, et al.: Computational design of closely related proteins that adopt two well-defined but structurally divergent folds. Proc Natl Acad Sci U S A 2020, 117:7208–7215. [PubMed: 32188784] *PNAS: Authors designed a protein inspired by hemagglutinin which adopts one fold in a structure solved by crystallography and a dramatically different fold in a structure solved by NMR.

37. Alexander PA, He Y, Chen Y, Orban J, Bryan PN: A minimal sequence code for switching protein structure and function. Proc Natl Acad Sci U S A 2009, 106:21149–21154. [PubMed: 19923431]

38. Chen Y, He Y, Ruan B, Choi EJ, Chen Y, Motabar D, Solomon T, Simmons R, Kauffman T, Gallagher DT, et al.: Rules for designing protein fold switches and their implications for the folding bioRxiv 2021. 2021.2005.2018.444643.

39. Langan RA, Boyken SE, Ng AH, Samson JA, Dods G, Westbrook AM, Nguyen TH, Lajoie MJ, Chen Z, Berger S, et al.: De novo design of bioactive protein switches. Nature 2019, 572:205–210. [PubMed: 31341284]

40. Ng AH, Nguyen TH, Gomez-Schiavon M, Dods G, Langan RA, Boyken SE, Samson JA, Waldburger LM, Dueber JE, Baker D, et al.: Modular and tunable biological feedback control using a de novo protein switch. Nature 2019, 572:265–269. [PubMed: 31341280]

41. Ambroggio XI, Kuhlman B: Computational design of a single amino acid sequence that can switch between two distinct protein folds. J Am Chem Soc 2006, 128:1154–1161. [PubMed: 16433531]

42. Cerasoli E, Sharpe BK, Woolfson DN: ZiCo: a peptide designed to switch folded state upon binding zinc. J Am Chem Soc 2005, 127:15008–15009. [PubMed: 16248623]

43. Sevy AM, Jacobs TM, Crowe JE Jr, Meiler J: Design of protein multi-specificity using an independent sequence search reduces the barrier to low energy sequences. PLoS Comput Biol 2015, 11, e1004300. [PubMed: 26147100]

44. Sevy AM, Wu NC, Gilchuk IM, Parrish EH, Burger S, Youcis D, Nagel MBM, Schey KL, Wilson IA, Crowe JE Jr, et al.: Multistate design of influenza antibodies improves affinity and breadth against seasonal viruses. Proc Natl Acad Sci U S A 2019, 116: 1597–1602. [PubMed: 30642961]

45. Norn C, Wicky BIM, Juergens D, Liu S, Kim D, Tischer D, Koepnick B, Anishchenko I, Foldit P, Baker D, et al.: Protein sequence design by conformational landscape optimization. Proc Natl Acad Sci U S A 2021, 118.

46. Jendrusch M, Korbel JO, Sadiq SK: AlphaDesign: a de novo protein design framework based on AlphaFold. bioRxiv 2021.

47. Alberstein RG, Guo AB, Kortemme T: Design principles of protein switches. Curr Opin Struct Biol 2021, 72:71–78. [PubMed: 34537489]

48. Boyken SE, Benhaim MA, Busch F, Jia M, Bick MJ, Choi H, Klima JC, Chen Z, Walkey C, Mileant A, et al.: De novo design of tunable, pH-driven conformational changes. Science 2019, 364:658–664. [PubMed: 31097662]

49. Lajoie MJ, Boyken SE, Saltar AI, Bruffey J, Rajan A, Langan RA, Olshesky A, Muhunthan V, Bick MJ, Gewe M, et al.: Designed protein logic to target cells with precise combinations of surface antigens. Science 2020, 369:1637–1643. [PubMed: 32820060] *Science: Using the Co-LOCKR system previously designed by the Baker lab, authors engineer a molecular switch
system that becomes activated in response to specific sets of antigens on cell surfaces. Authors show that this system can direct chimeric antigen receptor T (CAR-T) cells to specific tumor cells.
What makes a protein metamorphic?

Proteins fold spontaneously into the lowest energy 3D configuration as originally shown by Anfinsen for RNase and illustrated by Wolynes as an energy landscape or ‘folding funnel.’ 3D structures solved (mainly by X-ray crystallography) through the 1970s, 80s and 90s cemented a particular interpretation of the thermodynamic hypothesis, namely that the amino sequence of a globular protein contains the information required to specify exactly one native state conformation, i.e., ‘one sequence, one fold’.

Nothing in Anfinsen’s description limits the native state of a polypeptide to a single structure. If the energy landscape contains two different energy minima of near-equal depth, each funnel represents a distinct 3D structure. If spontaneous folding results in both structures being populated and they are observed to interconvert, the protein may be metamorphic.

Metamorphic proteins are single protein sequences that reversibly switch between multiple, dramatically different native structures, often with different functions.

What are the criteria or common characteristics for defining a metamorphic protein?

- **Marginal thermostability** (i.e., $\Delta G_{\text{fold}} > -3 \text{ kcal/mol}$), in contrast to $\Delta G_{\text{fold}}$ values ranging from −15 to −5 kcal/mol for most globular proteins.

- **Spontaneous unfolding.** Marginal thermostability implies that >1% of protein molecules are unfolded at any given moment (Figure 1g), making alternative native-state structures readily accessible.

- **Structural dissimilarity.** Helix–sheet conversions or changes to $\beta$-sheet topology will reorganize a large fraction of tertiary contacts within the metamorphic protein or independently folding domain. These changes can be catalogued and quantified using Protein Contact Atlas (https://www.mrc-lmb.cam.ac.uk/rajini) or other bioinformatic tools.

What is not a metamorphic protein? A partial list:

- **Chameleon sequences**, which are identical sequence fragments encompassed by larger sequences that differ

- **Prions**, which undergo irreversible fold changes

- **Local/subtle conformational changes**, which are less dramatic than the changes undergone by metamorphic proteins

- **Coupled binding/folding of IDPs**, which converts an unfolded protein to a folded protein, as opposed to converting one folded state to another

- **Irreversible PTM-driven conformational changes** (e.g., cleavage).

However, certain PTM-induced changes may be viewed as metamorphic interconversion, particularly when a reversible modification (e.g., phosphorylation) could mimic a slow equilibrium process.
- Proteins that occupy different folds as a result of sequence modifications such as mutations, extensions, or cleavage
Figure 1.
Thermodynamic characteristics of different categories of proteins. (a) Metamorphic proteins reversibly interconvert between two distinct, incompatible structures encoded by a single sequence. (b) Percent occupancy of the two XCL1 folds (chemokine fold, orange; β-sheet fold, blue) over evolutionary time, arriving at the modern day ratio of ∼50/50 (c–f) Hypothetical energy landscapes (i.e., folding funnels) for a stable monomorphic protein (c), a protein that undergoes subtle conformational change (d), a metamorphic protein (e), and an intrinsically unfolded protein (f). (g) Unfolded population versus ΔG of folding as calculated using ΔG_{fold} = −RT ln(K_{eq}). Points from each energy landscape are plotted in their corresponding colors.
Figure 2.
Metamorphic protein folding. (a-c) Example metamorphic proteins XCL1, RfaH, and KaiB. The dashed line in the RfaH structures represents a region which was not resolved in the RfaH crystal structure. N and C termini are labeled, and their alpha carbons are shown as spheres.
Design and applications of fold switching proteins. (a–d) Fold-switching proteins designed to date. The alpha carbons of the N and C termini are shown as spheres and labeled. PDB IDs are shown below each structure. (a) Proteins $G_A^{98}$ and $G_B^{98}$ differ by one amino acid and adopt different folds [37] (b) XAA_GVDQ is a de novo designed protein inspired by hemagglutinin which adopts one fold in a structure solved by NMR, and a significantly different fold in a structure solved by crystallography but has not been demonstrated to interconvert reversibly between the two structures [36]. (c) Sw2 is a computationally designed 30-residue protein that reversibly exchanges between a trimeric, coiled-coil, 3-$\alpha$ fold to a zinc finger fold in response to the addition of zinc [41]. (d) Co-LOCKR (Colocalization-dependent Latching Orthogonal Cage–Key pRotein) is a de novo designed protein system in which a “key” peptide displaces a “latch” helix from a 5-helix bundle [39]. The structure of the protein bound to the “key” peptide has not been solved. (e) De novo designed fold switching protein systems to date, such as Co-LOCKR, undergo hinge motions without substantially rearranging secondary structure or intramolecular contacts in the subunits that hinge with respect to one another. (f) Metamorphic proteins undergo large-scale conformational rearrangements often involving changes in secondary structure and substantial reorganization of intramolecular contacts.