Metamorphic protein folding as evolutionary adaptation

Acacia F. Dishman[1,2], **Brian F. Volkman**[1,3]

[1]Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI 53226

[2]Medical Scientist Training Program, Medical College of Wisconsin, Milwaukee, WI 53226

[3]Program in Chemical Biology, Medical College of Wisconsin, Milwaukee, WI 53226

Abstract

Metamorphic proteins switch reversibly between multiple distinct, stable structures, often with different functions. It was previously hypothesized that metamorphic proteins arose as intermediates in the evolution of a new fold, rare and transient exceptions to the 'one sequence, one fold' paradigm. However, as described herein, mounting evidence suggests that metamorphic folding is an adaptive feature, preserved and optimized over evolutionary time as exemplified by the NusG family and the chemokine XCL1. Analysis of extant protein families and resurrected protein ancestors demonstrates that large regions of sequence space are compatible with metamorphic folding. As a category that enhances biological fitness, metamorphic proteins likely employ fold-switching to perform important biological functions and may be more common than previously thought.

Keywords

fold-switching protein; ancestral sequence reconstruction; adaptive trait; protein fitness; thermodynamic hypothesis

Are metamorphic proteins real?

It's been calculated that life on earth has sampled a tiny fraction of all possible protein sequences ($\sim 10^{130}$) [1], and that polypeptides randomly selected from sequence space rarely fold [2]. Of the relatively few ($\sim 10^{12}$) protein sequences found in nature that do fold, as well as the increasingly robust products of de novo protein design, almost all follow Anfinsen's **thermodynamic hypothesis** (see Glossary) [3, 4] [5]. Indeed, decades of experimental evidence have confirmed that globular proteins fold spontaneously to the lowest free energy conformation, and have fostered a "one sequence, one fold" dogma [6]. As evidence of conformational flexibility has accumulated, we now understand that a protein sequence can adopt different configurations of similar free energy. Thus, nothing in the thermodynamic

hypothesis restricts the native state to a single fold, and sequences encoding multiple folds can reasonably be expected to inhabit the protein universe.

A **metamorphic protein** or domain interconverts reversibly between two native, folded structures sharing few or no common tertiary contacts, usually with different functions [7–10] (Fig 1). Initially defined as reversible switches by Murzin in 2008 [7], metamorphic proteins are a separate category from fold-switching proteins, which includes irreversible switching, consistent with the definition coined by Kim and Porter [11]. Metamorphic proteins clearly break the "one sequence, one fold" rule. As scientists mining the universe of genome sequences discover new metamorphic proteins, there is a growing appreciation of their abundance and contribution to evolutionary fitness; however, the designation of a distinct category still rests on a relatively small set of well-characterized examples (described below). Because most experimental methods select for 'well-behaved' proteins that adopt a single, stable fold, metamorphic proteins are almost certainly underrepresented in the Protein Data Bank (PDB)[i] and other databases of solved and predicted structures.

As with any new and unexpected observations, questions arose. If the presence of two unrelated folded states for a single protein was not just an experimental artifact, could it represent an evolutionary bridge? In other words, metamorphic proteins might be exceedingly rare transitional snapshots in the evolutionary emergence of a new fold, a hypothesis aligned with a "one sequence, one fold" view of the protein universe (Fig 1). However, what if easy access to more than one native state structure instead conferred a functional advantage? Metamorphic folding could reasonably be an **adaptive trait**, i.e., the result of selective pressure, in which case metamorphic proteins might not be as scarce as they appear.

Below, we present recent evidence that metamorphic proteins are common across a large protein family and can be encoded by relatively large regions of sequence space, suggesting that they are not rare [12] (Fig 1). Multiple studies of a protein's evolutionary trajectory show that a single amino acid change can introduce a new function simply by altering conformational dynamics in either subtle or dramatic ways [13–15]. The existence of metamorphic proteins as more than a biological accident or artifact is also supported by an evolutionary analysis in which an alternative fold emerges as a minor species and eventually reaches an equilibrium state with the original fold [16].

Other reviews have elegantly covered the design of **fold-switching proteins** [8], the functional and regulatory roles of fold-switching proteins [11], and the evolution of protein multimerization and dynamics [17]. Here we focus on recent evidence for the evolutionary advantage conferred by metamorphic folding and the likelihood that metamorphic proteins are more abundant than previously suspected. There are multiple approaches to understanding emergence and prevalence of metamorphic folding: studies comparing structural features of modern day proteins (horizontal) and analysis of historical sequences inferred from patterns of amino acid conservation (vertical). For example, recent horizontal studies show that many members of the NusG protein family regulate gene

---

[i] https://www.rcsb.org/

expression using a broadly conserved metamorphic interaction domain [12, 18]. Vertical studies of the hemoglobin [15], guanylate kinase [13, 14], and chemokine [16] protein families using an approach called **ancestral sequence reconstruction** (ASR) show how single or few mutations lead to changes in conformational dynamics and allostery which confer new functions. These complementary horizontal and vertical studies show how new features, including metamorphic protein folding, are found to confer evolutionary advantages and are preserved over time.

Insights into metamorphic folding from horizontal comparisons within a large protein family

One approach to better understanding protein metamorphosis is to compare modern-day members of a protein family that includes metamorphic proteins. One such family is the NusG superfamily of transcription factors, which is known to contain some **monomorphic proteins** (*e.g., E. Coli* NusG) and was known to contain at least one metamorphic protein, *E. coli* RfaH (*Ec*RfaH), a specialized NusG [19]. *Ec*RfaH is a well-studied metamorphic protein whose C-terminal domain (~40 amino acids) interconverts between two folded structures: a beta roll that resembles the C-terminal domain of monomorphic NusGs, and an alpha-helical hairpin [19]. The two folds perform different functions, and interconversion is triggered by the presence of binding partners, allowing for stringent control of transcriptional regulation [11]. Recently, the C terminal domain of *V. cholera* RfaH (*Vc*RfaH) was also found to switch between a beta roll and helical hairpin structure, closely resembling those in *Ec*RfaH [18].

To better understand how pervasive metamorphic protein folding might truly be, the Porter group examined more than 15,000 sequences from the NusG protein superfamily [12]. Using methods previously described by their group, which rely on secondary structure predictions, they predicted that 25% of the NusG family would switch folds, and that metamorphic NusGs would be present in bacteria, eukaryotes, and archaea. They experimentally validated their predictions for 10 proteins (6 predicted to be fold-switching and 4 predicted to be single-fold) and were correct in 10/10 cases. This suggests that metamorphic folding is a conserved feature in the NusG family, rather than an experimental or evolutionary artifact found in a single member. One reason protein metamorphosis might be conserved in the NusG superfamily is that it allows for fast, stringent control of protein function, which would be beneficial for some specialized NusG family members [11, 12, 18]. Together, this work suggests that fold switching might similarly be a pervasive, conserved feature in other protein families as well.

This study also performed coevolutionary analysis for subsets of the NusG superfamily, finding that shallower sequence alignments reflect the properties of the metamorphic RfaH subfamily, while deeper alignments reflect properties of the predominantly monomorphic NusG superfamily. This suggests that coevolutionary analyses of protein families may not identify metamorphic protein folding if deep alignments overwhelm the signal from metamorphic family members, but shallower alignments may capture properties of metamorphic proteins. This notion is further supported by a recent preprint which combines

AlphaFold2 (AF2) and coevolutionary data to predict the two structures of fold-switching proteins [20] (Box 1).

Interestingly, the six fold-switching proteins that were experimentally validated in the Porter group's study differed from one another at 68% of positions, with a mean sequence identity of only 21.0% [12]. In contrast, the mean sequence identity of the four experimentally validated monomorphic NusG proteins was significantly higher (43.2%). Likewise, the Knauer group showed that *Ec*RfaH and *Vc*RfaH both switch folds despite sharing only 43.6% sequence identity in the full length protein and 35.8% in the metamorphic domain [18]. These findings suggests that metamorphic protein folding can be encoded by highly divergent sequences (Fig 1); more disparate, even, than those of monomorphic proteins in the same family.

Vertical evolutionary analysis of protein conformational dynamics

New proteins arise through gene duplication, divergence, and recombination [21], and protein structure and function evolve as mutations accumulate over time. Horizontal comparisons between protein **orthologs** and **paralogs** are unlikely to reveal the history of sequence changes that gave rise to new structures and functions since each sequence evolved in a different species or on a different branch of the family tree. Without direct access to ancestral protein sequences, it is difficult to evaluate the evolutionary history of a protein. **Ancestral sequence reconstruction** (ASR), first envisioned by Linus Pauling [22], infers branch point amino acid sequences (nodes) from **phylogenetic analysis** of a database of extant sequences (Fig. 2), thereby allowing vertical analysis of protein evolution over time [23].

While ASR has been used mainly to trace functional evolution and map trends in thermostability [24], we discovered that resurrected proteins in the membrane associated guanylate kinase (MAGUK) family also reveal profound changes in conformational flexibility that introduce an entirely new function [14]. MAGUK proteins are present only in multicellular organisms and are essential for events that require spatial organization of subcellular components and organelles. Like the guanylate kinase enzyme, the GK module of MAGUK proteins contains a nucleotide binding domain, hinge, and lid domain. However, GK domains are catalytically inactive and instead function as protein interaction domains (PID) that bind phosphopeptide ligands in the substrate binding [25, 26].

Prehoda and colleagues used ASR to infer the sequences that arose after a gene duplication of the guanylate kinase enzyme (GK_{dup}) in primordial unicellular eukaryotes [13]. The purpose was to trace the structural adaptation required for GK_{dup} to gain phosphopeptide binding activity, thereby establishing the primordial GK protein interaction domain (GK_{PID}). Unexpectedly, a single amino acid change (S35P) in the hinge region was sufficient to convert GK_{dup} into GK_{PID}, in the absence of any binding/active site remodeling [13]. We then used NMR measurements of backbone dynamics to show that introduction of a proline significantly slowed the rate of hinge closing, drastically reducing catalytic efficiency and leaving the nucleotide binding cleft open long enough to permit peptide binding [14]. ASR of the first GK PID demonstrated that a new protein family can arise from a single

mutation – unlikely by itself to create a new structure or active site – when it changes the internal motions of a dynamic folded state. Even single historical mutations that introduce small changes in dynamics can lead to large functional differences between generations of proteins. Metamorphic protein folding, an extreme of dynamics, can uniquely encode two functionally distinct conformers in the same amino acid sequence, yet can also be conferred by a single historical mutation, as described below.

Pillai and colleagues used ASR to study the evolution of another protein which exhibits conformational dynamics that are important to its function: Hemoglobin. Hemoglobin is responsible for delivery of oxygen to tissues and relies on cooperative oxygen binding and allosteric regulation by other effector molecules to do so effectively [15]. Hemoglobin has been well characterized in terms of multimerization, cooperativity, and allostery: modern day hemoglobin is a heterotetramer which cooperatively binds oxygen with moderate affinity and is regulated by allosteric effectors such as 2,3-diphosphoglycerate and inositol hexaphosphate (IHP). Pillai, *et al.* found that hemoglobin evolved from a common ancestor with myoglobin, a monomeric protein that binds oxygen non-cooperatively with high affinity and is not regulated allosterically by IHP (Fig 2B) [15]. A more recent ancestor, which they term Ancα/β, is a homodimer that also binds oxygen non-cooperatively with high affinity and is not regulated allosterically by IHP (Fig 2B) [15]. A gene duplication event then generated ancestral proteins Ancα and Ancβ, which come together to form a heterotetramer consisting of two subunits of each protein (Fig 2B) [15]. This tetramer binds oxygen cooperatively with moderate affinity and is allosterically regulated by IHP, similar to modern day hemoglobin [15]. These ancestral studies tease out the "missing link" through which the modern hemoglobin tetramer evolved, and show that tetramerization, cooperativity, and capacity for allosteric regulation evolved simultaneously (Fig 2B) [15]. Pillai and colleagues then distilled the historical mutational changes that occurred in hemoglobin's ancestors down to just two mutations that confer all three of these properties. Hemoglobin's evolution of new properties through few mutations was possible due to a preexisting linkage between the multimerization interface and the oxygen binding site in ancient hemoglobin predecessors. In other words, its predecessors were poised to evolve multimerization and cooperativity. As described below, metamorphic protein folding can also evolve via ancestors that are on the brink of becoming metamorphic, and proteins poised to become metamorphic through single or few mutations may be more common than expected.

As with the single amino acid change that alters global function and dynamics for GK, hemoglobin's evolutionary path shows that short mutational paths can lead to the evolution of new modes of multimerization and conformational dynamics and thereby confer new functional attributes [15]. Both studies indicate that simple genetic mechanisms can lead to the evolution of complex properties, especially when they build upon ancient features that predispose proteins to acquire new structures or functions (Fig 2B) [15, 17].

Insights into metamorphic protein folding from Ancestral Sequence Reconstruction

Akin to the above studies, we used ASR to study the evolution of fold switching in the well characterized, model metamorphic protein XCL1 [16]. XCL1 is a human chemokine that orchestrates antigen cross presentation by dendritic cells to T Cells [27]. Like nearly all chemokines, XCL1 must bind and activate its cognate GPCR and bind to extracellular matrix glycosaminoglycans. XCL1 switches between two distinct folded structures, each of which performs one of these functions [28]. We hypothesized that ancestral XCL1 sequences would document the emergence of metamorphic folding. We found that XCL1 evolved from a monomorphic (single-fold) ancestor which gained the ability to switch folds about 150 million years ago (Fig 2C). Resurrection of several nodes along the evolutionary trajectory from its common ancestor with the CCL20 family (Anc.0) to extant XCL1 showed that loss of a conserved disulfide bond in the interval between Anc.1 and Anc.2 did not enable metamorphism. However, both Anc.3 and Anc.4 interconverted reversibly between two distinct folds despite sharing only 60 and 68% sequence identity with human XCL1, respectively [16]. The first metamorphic XCL1 ancestor, Anc.3, adopted the chemokine fold ~90% and the alternate fold ~10% at equilibrium under near physiologic solution conditions. Under the same conditions, the more recent XCL1 ancestor, Anc.4, preferred the alternate fold, adopting the chemokine fold ~10% and the alternate fold ~90%. Modern day XCL1 adopts each fold in equal proportion [16, 28]. As noted above, some have hypothesized that metamorphic proteins may represent evolutionary bridges: snapshots of proteins evolving from one fold to another [29]. However, if XCL1 were indeed evolving from an ancient fold to a new one, one would expect that modern XCL1 would occupy the alternate fold 100%. Instead, XCL1's evolutionary trajectory "swung back" to equal occupancy of each fold, supporting our opinion that metamorphic folding is an adaptive trait that was selected for in XCL1 (Fig 1). The evolutionary bridge and adaptive trait hypotheses are not mutually exclusive in principle. In XCL1's case, for example, some progeny of the first metamorphic ancestor could have gone on to adopt exclusively the new fold, while others could have evolved to remain metamorphic – in that case, the metamorphic ancestor would have acted as a bridge for some species but not for others.

Just as ancestral hemoglobin proteins were poised to evolve cooperativity, ancestral XCL1 sequences were also poised to evolve metamorphic folding (Fig 2D). All modern chemokines contain two or more disulfide bonds except XCL1, which contains only one. The presence of both disulfide bonds prevents metamorphic interconversion: restoring the missing disulfide in XCL1 renders it monomorphic [30]. As alluded to above, ASR shows that XCL1 ancestors Anc.0 and Anc.1 contain both disulfide bonds. While the second disulfide bond was lost, Anc.2 remained monomorphic, but was poised to evolve metamorphic folding (Fig 2D). In the interval between Anc.2 and Anc.3 (Fig 2B), we identified a single amino acid substitution that confers metamorphic folding. This shows that metamorphic folding is another complex trait that can evolve along short mutational paths, similar to the evolution of hinge dynamics in GK, and allostery and tetramerization in hemoglobin [2]. The examples presented here illustrate that small numbers of mutations conferring either large or small changes in conformational dynamics contribute to functional

expansion. More examples are required to properly judge the relative importance of metamorphic folding in the evolution of new protein functions.

Concluding Remarks

One might expect metamorphic protein folding to be a rare exception to the "one sequence, one fold" rule - only possible for a handful of highly specific sequences and unlikely to arise and persist over the course of evolution (see Outstanding Questions). Metamorphic proteins, in this hypothesis, would be unlikely to accumulate over evolutionary time if fold-switching is useful only as a transient evolutionary bridge and not as a functional adaptation. However, numerous fold-switching proteins are known to parse two distinct functions between their two folds, using switching between folds as a unique regulatory mechanism [11]. Fold-switching is an evolutionarily conserved feature in the NusG superfamily and may also be conserved in other families [12]. Moreover, XCL1's evolutionary trajectory suggests that metamorphic folding was selected for in XCL1, and could be selected for in other proteins. We believe that metamorphic folding therefore confers a functional advantage and is evolutionarily selected for and conserved (Fig 1).

Multiple examples of profound changes in structure and function arising from one or a few evolutionary mutations have been documented. In addition to the examples described herein, a recent ASR study of a large family of response regulator proteins shows that a new fold can arise as a result of only two mutations [31]. Collectively, these studies demonstrate that complex traits like metamorphosis, oligomerization, cooperativity, and allostery, can evolve via short mutational paths.

Because metamorphic folding requires an amino acid sequence that encodes two distinct, stable sets of intramolecular interactions corresponding to two folded structures, one might expect that it can only be accomplished by few, highly specific sequences. However, we found that while both modern day XCL1 and the first metamorphic XCL1 ancestor switch folds, they differ at 40% of sequence positions [16]. Likewise, the Knauer group structurally characterized two RfaH variants that share only 43.6% sequence identity yet both switch folds [18]. Further, the Porter group found that six fold-switching RfaH variants are even more different from one another, with a median sequence identity of only 21%. Together, this shows that an unexpectedly wide range of sequences can switch between the same two folds.

We believe if many sequences can encode multiple structures, and metamorphic folding can evolve via short mutational paths, then the natural abundance of metamorphic proteins may be greater than anticipated (Fig 1). In support of this, some primary sequence-based searches for more metamorphic proteins have harnessed discrepancies in secondary structure predictions with moderate success [32, 33]. Search methods have seen recent breakthroughs with the incorporation of coevolutionary signatures that are missed by other structure prediction algorithms such as AF2 [20]. Future work will search for new metamorphic proteins, uncover their functions, and determine relevance to human health and disease (see Outstanding Questions).

Given the realization that metamorphic sequences are widespread yet can be separated from monomorphic sequences by only one or a few mutations, design of metamorphic proteins should be within reach. Our understanding of principles which enable metamorphic folding is growing [8, 18]. For example, some protein folds may be more conducive to metamorphosis than others based on intrinsic stabilities of different topologies and proportion of disordered content [8, 18, 34]. Indeed, two of the first engineered metamorphic proteins were recently reported [35, 36], increasing confidence that methods for protein design will gain the ability to encode metamorphic properties. It is our belief that future efforts to design new fold-switching proteins will have applications as components of molecular motors, switchable therapeutics, and biosensors [8].

Glossary

Adaptive trait

A genetic variation that enhances reproductive success or fitness of the organism and is consequently selected for during evolution.

Ancestral Sequence Reconstruction

A technique that uses multiple sequences from families of modern-day proteins to predict the most likely amino acid sequences of ancient proteins going back through evolutionary time. These predicted sequences can then be "resurrected": expressed, purified, and studied in the laboratory.

Fold-switching protein

A protein that remodels all or part of its secondary structure in response to a stimulus. Unlike metamorphic proteins, fold-switching proteins need not switch reversibly, and may only contain small fold-switching regions within their structures. The family of fold-switching proteins thus encompasses metamorphic proteins, fibril-forming proteins, and others.

Intrinsically disordered protein (IDP)

Proteins, or regions of proteins, that lack stable native folded structures owing to a low content of bulky apolar amino acids. IDPs may adopt a folded structure when bound to a partner.

Metamorphic protein

A protein that interconverts reversibly between two native folded structures sharing few or no common tertiary contacts, usually with distinct functions.

Monomorphic protein

A protein with a single, stable folded structure.

Paralogs

A set of related genes, often but not necessarily within the same species, that have arisen via gene duplication.

Phylogenetic analysis

Elucidation of the evolutionary relationship between both extant (modern-day) and extinct genes, or organisms, or species. Nodes (i.e., branch points) in a phylogenetic tree (also called a phylogeny) represent biological entities, such as genes or organisms, and the lines between nodes represent evolutionary relationships.

Protein folding problem

Longstanding question in biology: If the amino acid sequence of a protein is sufficient to specify its folded native state, what is the code relating sequence to 3D structure?

Orthologs

A set of related genes that have arisen via speciation, *i.e.,* vertical descent.

Thermodynamic hypothesis

Anfinsen's discovery that a protein's "native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment." The native conformation corresponds to the lowest free energy state.

Literature cited

1. Koonin EV and Galperin MY (2003) In Sequence - Evolution - Function: Computational Approaches in Comparative Genomics.

2. Baker D (2019) What has de novo protein design taught us about protein folding and biophysics? Protein Sci 28 (4), 678–683. [PubMed: 30746840]

3. Haber E and Anfinsen CB (1962) Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. J Biol Chem 237, 1839–44. [PubMed: 13903380]

4. Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181 (96), 223–30. [PubMed: 4124164]

5. Huang PS et al. (2016) The coming of age of de novo protein design. Nature 537 (7620), 320–7. [PubMed: 27629638]

6. Nelson DM (2021) Lehninger Principles of Biochemistry, 8 edn., W. H. Freeman.

7. Murzin AG (2008) Biochemistry. Metamorphic proteins. Science 320 (5884), 1725–6. [PubMed: 18583598]

8. Dishman AF and Volkman BF (2022) Design and discovery of metamorphic proteins. Curr Opin Struct Biol 74, 102380. [PubMed: 35561475]

9. Dishman AF and Volkman BF (2018) Unfolding the Mysteries of Protein Metamorphosis. ACS Chem Biol 13 (6), 1438–1446. [PubMed: 29787234]

10. Stryer L (2019) Biochemistry, 9 edn., W. H. Freeman.

11. Kim AK and Porter LL (2021) Functional and Regulatory Roles of Fold-Switching Proteins. Structure 29 (1), 6–14. [PubMed: 33176159]

12. Porter LL et al. (2022) Many dissimilar NusG protein domains switch between alpha-helix and beta-sheet folds. Nat Commun 13 (1), 3802. [PubMed: 35778397]

13. Anderson DP et al. (2016) Evolution of an ancient protein function involved in organized multicellularity in animals. Elife 5, e10147. [PubMed: 26740169]

14. Whitney DS et al. (2016) Evolution of a Protein Interaction Domain Family by Tuning Conformational Flexibility. J Am Chem Soc 138 (46), 15150–15156. [PubMed: 27502157]

15. Pillai AS et al. (2020) Origin of complexity in haemoglobin evolution. Nature 581 (7809), 480–485. [PubMed: 32461643]

16. Dishman AF et al. (2021) Evolution of fold switching in a metamorphic protein. Science 371 (6524), 86–90. [PubMed: 33384377]

17. Pillai AS et al. (2022) Simple mechanisms for the evolution of protein complexity. Protein Sci 31 (11), e4449. [PubMed: 36107026]

18. Zuber PK et al. (2022) Structural and thermodynamic analyses of the beta-to-alpha transformation in RfaH reveal principles of fold-switching proteins. Elife 11.

19. Burmann BM et al. (2012) An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell 150 (2), 291–303. [PubMed: 22817892]

20. Schafer JW and Porter LL (2023) Evolutionary selection of proteins with two folds. bioRxiv, 2023.01.18.524637.

21. Chothia C et al. (2003) Evolution of the protein repertoire. Science 300 (5626), 1701–3. [PubMed: 12805536]

22. Pauling L and Zuckerkandl E (1963) Chemical paleogenetics. Molecular "restoration studies" of extinct forms of life. Acta Chem Scand 17, S9–S16.

23. Harms MJ and Thornton JW (2010) Analyzing protein structure and function using ancestral gene reconstruction. Curr Opin Struct Biol 20 (3), 360–6. [PubMed: 20413295]

24. Wheeler LC et al. (2016) The thermostability and specificity of ancient proteins. Curr Opin Struct Biol 38, 37–43. [PubMed: 27288744]

25. Johnston CA et al. (2012) Structure of an enzyme-derived phosphoprotein recognition domain. PLoS One 7 (4), e36014. [PubMed: 22545154]

26. Olsen O and Bredt DS (2003) Functional analysis of the nucleotide binding domain of membrane-associated guanylate kinases. J Biol Chem 278 (9), 6873–8. [PubMed: 12482754]

27. Lei Y and Takahama Y (2012) XCL1 and XCR1 in the immune system. Microbes Infect 14 (3), 262–7. [PubMed: 22100876]

28. Tuinstra Robbyn L., F.C.P., Snjezana Kutlesa, Elgin E. Sonay, Kron Michael A., and Volkman Brian F. (2008) Interconversion between two unrelated protein folds in the lymphotactin native state. Proc Natl Acad Sci U S A 105 (13), 5057–5062. [PubMed: 18364395]

29. Yadid I et al. (2010) Metamorphic proteins mediate evolutionary transitions of structure. Proc Natl Acad Sci U S A 107 (16), 7287–92. [PubMed: 20368465]

30. Tuinstra RL et al. (2008) Interconversion between two unrelated protein folds in the lymphotactin native state. Proc Natl Acad Sci U S A 105 (13), 5057–62. [PubMed: 18364395]

31. Chakravarty D et al. (2022) Identification of a covert evolutionary pathway between two protein folds. bioRxiv, 2022.12.08.519646.

32. Kim AK et al. (2021) A high-throughput predictive method for sequence-similar fold switchers. Biopolymers 112 (10), e23416. [PubMed: 33462801]

33. Mishra S et al. (2019) Inaccurate secondary structure predictions often indicate protein fold switching. Protein Sci 28 (8), 1487–1493. [PubMed: 31148305]

34. Porter LL and Looger LL (2018) Extant fold-switching proteins are widespread. Proc Natl Acad Sci U S A 115 (23), 5968–5973. [PubMed: 29784778]

35. Solomon TL et al. (2023) Reversible switching between two common protein folds in a designed system using only temperature. Proc Natl Acad Sci U S A 120 (4), e2215418120. [PubMed: 36669114]

36. Ruan B et al. (2023) Design and characterization of a protein fold switching network. Nat Commun 14 (1), 431. [PubMed: 36702827]

37. Chen SJ et al. (2023) Opinion: Protein folds vs. protein folding: Differing questions, different challenges. Proc Natl Acad Sci U S A 120 (1), e2214423119. [PubMed: 36580595]

38. Chakravarty D and Porter LL (2022) AlphaFold2 fails to predict protein fold switching. Protein Sci 31 (6), e4353. [PubMed: 35634782]

Text Box 1:

Does AlphaFold solve the metamorphic protein folding problem?

AlphaFold2 (AF2) is an artificial intelligence system that is widely recognized as the most accurate protein structure prediction algorithm, while falling short of a general solution to the **protein folding problem** [37]. A generational advance over previous methods, AF2 reliably predicts the folded structures of stable, monomeric proteins from their amino acid sequences with high confidence. There is also a significant proportion of the proteome for which AF2 predictions have low confidence values, which largely overlaps with regions of the proteome that are known or predicted to be intrinsically disordered. If AF2 accurately folds monomorphic proteins and flags **intrinsically disordered proteins** (**IDPs**), can it also recognize metamorphic proteins and predict their structures? Chakravarty and Porter assessed the ability of AF2 to predict the structures of fold-switching proteins [38], whose stabilities likely fall between those of stable monomers and those of IDPs [8, 9]. They examined AF2 predictions for 98 fold-switching proteins, finding that AF2 predicted one fold-switched structure but not the other in 94% of cases. In the remaining 4/98 cases, both conformations were sampled with moderate to good accuracy. Unlike for IDPs, AF2 confidences were moderate-to-high for 74% of fold-switching regions. Chakravarty and Porter concluded that AF2 fails to predict whether a protein will switch folds: it almost always misses the second structure and does so with moderate to high confidence. If one were to use AF2 to search the proteome for fold-switching proteins, one would fail to detect most of the proteins that switch folds. So far, no tool has been shown to reliably identify fold-switching proteins based on primary sequence.

Why does AF2, a highly accurate predictor of monomorphic protein structure, fail for proteins with more than one fold? One possible reason is that AF2 is trained on a dataset that is likely enriched with proteins that crystallize readily, such as stable, monomorphic proteins. Of course, it is also possible that few fold-switching proteins exist; however, recent work described elsewhere in this review suggests otherwise. As more fold-switching proteins are identified, better training sets will become available for machine learning-based prediction methods, and our ability to predict metamorphic protein folding will improve. Additionally, promising new prediction methods are under development, which use secondary structure prediction algorithms and coevolutionary data to predict whether a primary sequence encodes a protein that switches folds [11, 12, 33, 34].

Box:

Outstanding Questions

- **How common are metamorphic proteins?** A recent survey counted 96 naturally occurring fold-switching proteins, roughly 10 of which are well-characterized metamorphic proteins [1]. However, current estimates suggest that up to 4% of proteins in the PDB are metamorphic. Fold-switching may be overlooked under the expectation that proteins have a single structure, combined with the difficulty inherent in characterizing dynamic proteins that are relatively unstable compared to monomorphic proteins.

- **Is metamorphic protein folding selected for evolutionarily?** Recent work suggests that metamorphic folding is an adaptive trait that was evolutionarily selected for in the human chemokine XCL1 [2]. Is this the case for other metamorphic proteins? If so, metamorphic proteins may be more common than previously expected.

- **What are common sequence features of metamorphic proteins?** Identifying primary sequence features shared by metamorphic proteins but not monomorphic proteins could facilitate a search for metamorphic proteins in the proteome.

- **Which proteins are close, in sequence space, to becoming metamorphic?** Which proteins are on the brink of metamorphosis, such that one or a few mutations would make them start switching folds? Understanding features that predispose a protein to become metamorphic could enhance the ability to design proteins that switch folds. This could also improve our understanding of how disease-related mutations exert their effects.

Highlights

- Proteins are generally presumed to fold into a single native state structure.

- Some proteins clearly break the "one sequence, one fold" rule.

- A metamorphic protein switches between two native, folded structures that are significantly different from each other.

- Evolutionary aspects of metamorphic proteins can be studied in multiple ways, including the resurrection of extinct ancestral versions.

- Emerging evidence suggests that metamorphic protein folding is selected for during evolution.
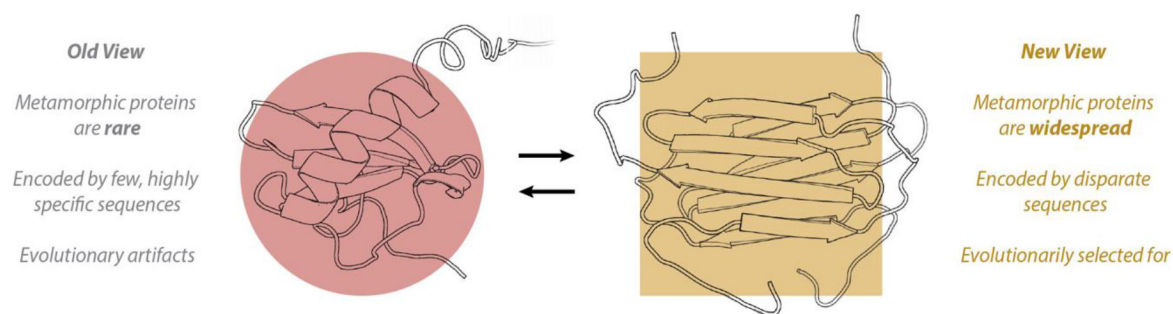
Old View

*Metamorphic proteins
are rare*

*Encoded by few, highly
specific sequences*

Evolutionary artifacts

New View

*Metamorphic proteins
are widespread*

*Encoded by disparate
sequences*

Evolutionarily selected for

Figure 1. Metamorphic protein folding.
Left, a summary of previously held views about metamorphic proteins; right, newer views in the field. Center, the two structures of a model metamorphic protein, XCL1 (PDB IDs 1J9O, left, and 2JP1, right).
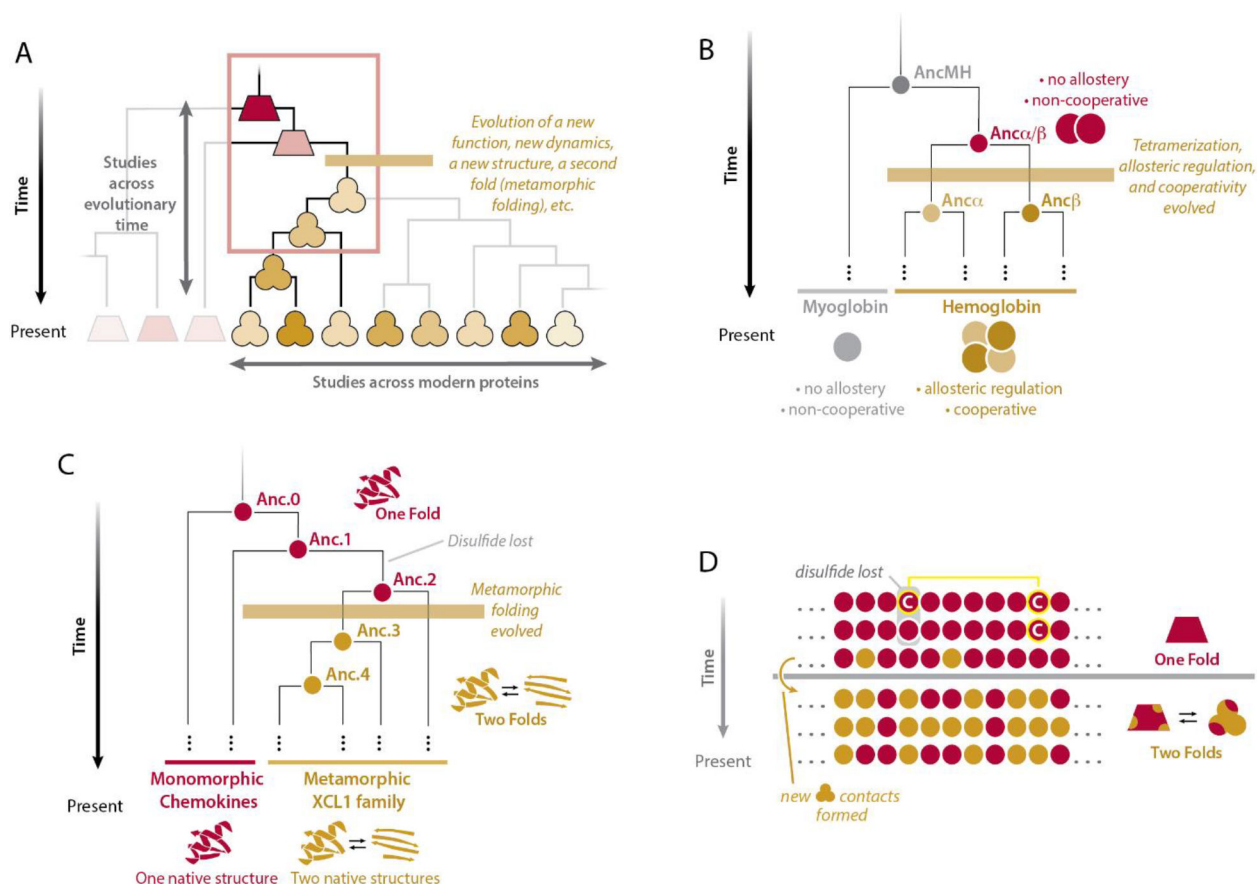
Figure 2. Evolution of protein structure and function.
(A) Comparisons of modern protein sequences (horizontal studies) can reveal patterns of conservation and variation. Comparisons of historical protein sequences inferred by ancestral sequence reconstruction (ASR) (vertical studies) can reveal the evolutionary emergence of new features. The change from a red trapezoid to a gold trefoil represents the emergence of a new feature, such as a new fold or function, a second fold, an altered oligomeric state, or a change in conformational dynamics. (B) Modern hemoglobin evolved from non-cooperative homodimers resurrected using ASR. Assembly into heterotetramers, allosteric regulation, and cooperative oxygen binding then emerged during the same evolutionary interval. (C) Emergence of metamorphic folding was detected in the evolutionary trajectory of XCL1 ancestors resurrected using ASR. Access to a second native state structure emerged after loss of a conserved disulfide and the accumulation of additional mutations that destabilized the chemokine fold or favored the alternative structure. (D) Historical sequence changes may predispose a protein to gain new features. Gray and gold highlighting illustrate sequence changes. These changes may predispose or introduce a new feature, respectively.