HHS Public Access

Author manuscript

Curr Opin Struct Biol. Author manuscript; available in PMC 2024 June 01.

Published in final edited form as:

Curr Opin Struct Biol. 2024 June ; 86: 102807. doi:10.1016/j.sbi.2024.102807.

Metamorphic proteins and how to find them

Lauren L. Porter[1,2], **Irina Artsimovitch**[3], **César A. Ramírez-Sarmiento**[4,5]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

[2]Biochemistry and Biophysics Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

[3]Department of Microbiology and Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA

[4]Institute for Biological and Medical Engineering, Schools of Engineering, Medicine and Biological Sciences, Pontificia Universidad Católica de Chile, Santiago 7820436, Chile

[5]ANID, Millennium Science Initiative Program, Millennium Institute for Integrative Biology (iBio), Santiago 833150, Chile

Abstract

In the last two decades, our existing notion that most foldable proteins have a unique native state has been challenged by the discovery of metamorphic proteins, which reversibly interconvert between multiple, sometimes highly dissimilar, native states. As the number of known metamorphic proteins increases, several computational and experimental strategies have emerged for gaining insights about their refolding processes and identifying unknown metamorphic proteins amongst the known proteome. In this review, we describe the current advances in biophysically and functionally ascertaining the structural interconversions of metamorphic proteins and how coevolution can be harnessed to identify novel metamorphic proteins from sequence information. We also discuss the challenges and ongoing efforts in using artificial intelligence-based protein structure prediction methods to discover metamorphic proteins and predict their corresponding three-dimensional structures.

Keywords

Metamorphic proteins; Fold-switch; Protein biophysics; Coevolution; Protein structure prediction

Corresponding authors: Artsimovitch, Irina (artsimovitch.1@osu.edu); Ramírez-Sarmiento, César A (cesar.ramirez@uc.cl); Porter, Lauren L (porterll@nih.gov).

CrediT author statement

Lauren L. Porter: Writing – Original Draft, Writing – Review & Editing, Funding acquisition. **Irina Artsimovitch:** Writing – Original Draft, Writing – Review & Editing, Funding acquisition. **César A. Ramírez-Sarmiento:** Writing – Original Draft, Writing – Review & Editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Metamorphic proteins: a primer

Most experimentally characterized proteins fold into single and unique three-dimensional structures responsible for their biological function. However, metamorphic proteins discovered in the last two decades challenge this paradigm by encoding two or more native states in a single amino acid sequence. Once thought to be rare, almost 100 metamorphic proteins from all kingdoms of life have been discovered to date [1], after experimentally solving their disparate structures under different contexts (truncated proteins [2,3], fold-stabilizing mutants [4–6]) and physiological conditions [7,8] using X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryoEM).

Metamorphic proteins not only assume different native states but also can reversibly interconvert between these states to regulate protein function, as demonstrated by many targeted biophysical and functional assays. Figure 1 shows three examples: the bacterial transcription factor RfaH fold-switches to specifically bind RNA polymerases transcribing a small subset of genes to dramatically enhance their expression [3]; the cyanobacterial protein KaiB fold-switches to regulate the phosphorylation/dephosphorylation phases of the circadian clock KaiABC [4]; and the human CLIC1 transitions from a soluble monomeric form into a membrane chloride ion channel in response to changes in the oxidation state [7].

The existence of metamorphic proteins has been rationalized from an evolutionary point of view. Directed evolution experiments on a protein fragment from a natural monomeric protein into metamorphic oligomers suggested their importance in facilitating the emergence of novel folds [12], whereas recent ancestral sequence reconstructions [13] and analysis of metamorphic homologs on large protein families [9] provide evidence that fold-switching constitutes a functional adaptation that has been preserved during protein evolution as a novel regulatory mechanism [14].

Exploring the refolding landscapes of metamorphic proteins

While structural biology methods proved the existence of metamorphic proteins by experimentally solving the structures of their differently folded states, they did not reveal how the interconversions between these states take place. Different flavors of molecular dynamics rooted in protein folding principles and/or leveraging enhanced sampling methods can fill this gap. In this regard, dual-basin structure-based models (SBMs) [15,16] have been extensively used to simulate the refolding of metamorphic proteins.

SBMs replace commonplace physicochemistry-based force fields with simplified potential energy functions, which capture the essentials of protein folding by treating all residue pairs in contact in the native structure of a given protein (i.e. residues with any shared atom pair closer than a distance radius, typically <6 Å) as attractive, while also maintaining covalently bonded structure and torsional angles [15]. The significant reduction in the number of evaluated interactions makes SBM simulations computationally more tractable and efficient than molecular dynamics (MD) simulations using conventional force fields. Furthermore, their extension from single-basin to dual-basin models only requires to savvily merge the unique native contacts and torsions from two disparate structures into a single

potential energy function (Figure 2) [16,17]. Both all-atom and coarse-grained single-basin and dual-basin SBMs have been used to describe the refolding pathways of the following fold-switching proteins: human chemokine XCL1 [18], spindle checkpoint protein Mad2 [19], RfaH [20,21], cytolytic toxin ClyA [22], influenza hemagglutinin [23], KaiB [24] and SARS-CoV-2 spike protein [25]. Largely, these simulations reveal that fold-switching proceeds through intermediate states that retain a significant amount of native-like secondary and tertiary structure.

There are additional advantages to the use of dual-basin SBMs to explore the refolding of metamorphic proteins. On the one hand, these simplified models can be combined with more sophisticated physicochemical force fields, enabling the use of the changes in native contacts between the two folds as a reaction coordinate for enhanced sampling. Such combinations were employed to explore how the breakage and reformation of secondary structure during XCL1 fold-switching relies on hydrogen bond networks that facilitate the refolding process [27], and how the fold-switch of the C-terminal domain (CTD) of RfaH from an α-helical hairpin to a β-roll structure traverses through partially unfolded intermediates that retain some native-like secondary structures [28,29]. On the other hand, these SBMs have proven successful in following structural interconversions in very complex biomolecular systems, including very large influenza hemagglutinin [23] and SARS-CoV-2 spike proteins [25]; metamorphic proteins whose fold-switching is coupled to changes in quaternary structure, such as the tetramer-to-monomer transformation of KaiB [4]; and fold-switching processes that are triggered by interactions with large macromolecular entities, such as the refolding of RfaH upon binding to a transcription elongation complex [21].

More recently, dual-basin SBMs have been employed to simulate the effect of macromolecular crowders, which mimic the intracellular environment, on fold-switching processes [30]. These simulations suggested that crowding might preferentially stabilize one fold over the other, thus causing a population shift [30], and were validated by NMR experiments, which showed that crowding agents shift the equilibrium of KaiB towards its ground-state tetramer and of XCL1 towards its monomeric chemokine fold [31].

From MD simulations to savvy experimental biophysics

Experiments are required both to test the findings derived from these MD simulations and to measure important biophysical quantities they cannot yet estimate, such as timescales of structural interconversions. Several experimental biophysics strategies have been adopted to explore the structural changes and timescales of these challenging fold-switching processes.

NMR spectroscopy has been extensively used to characterize structural interconversions of metamorphic proteins and their timescales by following the chemical shifts of specific isotopically labeled residues. For example, the [^{1}H,^{13}C]-labeling of the methyl groups of isoleucine, leucine, and valine in RfaH enabled the exploration of RfaH refolding in the presence of its target, a paused transcription elongation complex, by NMR [32]. These experiments demonstrated that RfaH switches from the all-α to the all-β state only upon recruitment to the complete transcription complex and refolds back into the all-α state after release. Likewise, the use of hydrophobicity-sensitive ^{19}F-labeled residues, located

in regions that undergo changes in solvent exposure upon fold-switching, revealed that the PimA mannosyltransferase from tuberculosis-associated mycobacteria refolds on the seconds timescale, with its two native states coexisting under equilibrium conditions, and that the addition of its substrate accelerates PimA fold switching [33]; a similar approach was recently used to characterize RfaH unfolding [34]. 15N ZZ-exchange NMR experiments on XCL1 determined that it also switches folds on the seconds timescale [35]. In stark contrast, ^{1}H chemical shift measurements of Mad2 revealed that its structural interconversion occurs in the timescale of several hours [36], similar to what was shown for KaiB using heteronuclear single quantum coherence (HSQC) experiments [31]. A striking conclusion from these NMR experiments is that the refolding timescales of metamorphic proteins can vary from seconds to hours.

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) is another promising strategy to interrogate local structure changes during metamorphic refolding. Recently, HDX-MS was used to demonstrate that the ends of the α-helical hairpin in RfaH CTD are less stable than the rest of this structure [37], in agreement with MD simulations [20,28] suggesting that RfaH refolding starts from the ends of the α-helical hairpin [37]. HDX-MS has also been used to study changes in the local structural stability of RfaH upon binding to transcription elongation complexes [38]; it also showed that KaiB mutants that alter the cyanobacterial circadian clock periodicity resemble the fold-switched KaiB state [24]. Lastly, HDX-MS revealed distinct trimeric conformations of the SARS-CoV-2 spike protein, including a novel open-trimer conformation in reversible interconversion with its prefusion state [39].

When a fold switch is accompanied by large changes in sensitivity to proteases, a time-tested low-cost limited proteolytic digestion could prove instrumental – for example, while the all-α state of RfaH is resistant to chymotrypsin, the CTD release (and subsequent refolding) exposes the hydrophobic residues and sensitizes RfaH to digestion [40].

Predicting metamorphic proteins using evolutionary information as guide

Several recent studies of fold evolution suggest that metamorphic proteins typically (1) evolve from single-folding ancestors, (2) are conserved within protein subfamilies, and (3) serve specialized cellular functions. The metamorphic transcriptional regulator, RfaH [41], illustrates these three points. RfaH has a structurally conserved N-terminal domain (NTD) that binds RNA polymerase and a CTD that reversibly transitions between α-helical bundle and β-roll folds [32]. Recent computational predictions–supported by experiments–suggest that clades of RfaH homologs with similar sequences often have the same fold-switching activity [9]. Further, phylogenetic analysis suggests that RfaH likely evolved from its single-folding homolog, NusG, whose CTD assumes the β-roll fold but not the α-helical hairpin [42]. This specialized α-helical fold limits the activity of RfaH to a specific DNA sequence known as the operon polarity suppressor (*ops*), while the β-roll form interacts directly with the ribosome, fostering efficient translation. Similarly, the human protein XCL1 reversibly interconverts between two folds: a chemokine fold that binds the G-protein coupled receptor XCR1 and a dimeric β-sheet fold that binds fungal pathogens [13]. Phylogenetic analysis suggests that XCL1 evolved from single-folding chemokine ancestors. Further, its ability

to bind fungal pathogens is likely a specialized function of its alternative fold [13]. As expected, there are exceptions to every rule: fold switching may be a universally conserved trait in some protein families, such as the cell-division-related proteins MinE (bacteria [43]) and Mad2 (eukaryotes, [44]). Nevertheless, these rules of thumb indicate that phylogenetic order is often an important factor when predicting metamorphic proteins.

Leveraging these observations, we recently found evidence that evolution has selected for both conformations of 56/56 fold-switching proteins from diverse families and organisms [45]. Early studies of protein structure recognized that homologous protein sequences can have covarying amino acid pairs that tend to be in direct contact [46–48], greatly constraining the conformational possibilities that computational methods must sample to predict a protein's fold [49]. Applying these methods over thousands of genomic and metagenomic RfaH sequences enabled correct predictions of several interacting residue pairs unique to both folds, and SBM simulations where native contacts were replaced by these sequence-derived contacts allowed RfaH to be folded into both native states [50]. However, standard coevolutionary analysis uncovers only a small number of contacts unique to alternative folds [48]. By applying two coevolutionary inference methods [51,52] to superfamily and subfamily-specific sets of sequences homologous to metamorphic proteins, contacts uniquely corresponding to these alternative folds were enhanced greatly [45]. The high statistical significance of these unique contacts suggests that both folds likely confer a selective advantage. Further, these dual-fold contacts suggested a pipeline to blindly predict fold-switching proteins in which the predicted contacts can be used to cross-validate differing AlphaFold2 (AF2) and ESMFold structure predictions [45].

Limitations of artificial intelligence (AI)-based predictions of metamorphic proteins

Though AI-based methods have revolutionized protein structure prediction, they have so far proven to be insufficient predictors of metamorphic proteins, even those likely in AF2's training set. An 18-month-old study benchmarking AF2's ability to predict fold-switching proteins indicated that it correctly predicted both conformations in only 8/93 fold switchers [53], and a preprint benchmarking the most recent implementations of AF2 yielded a 23% prediction success rate from extensive sampling of ~300,000 structures [54]. For context, a BLASTsearch of the PDB would have had a 100% success rate. Further, the predictive success of fold switchers outside the training set was very poor, with 1/7 successes [54].

AF2's lackluster performance likely results from its limitations as a deep learning (DL) model. Like all other DL models, AF2 is limited by its training set. AF2's impressive accuracy in predicting single-fold proteins is directly related to its training on >100,000 experimentally determined structures [55,56] followed by an additional round of training on ~300,000 subsequent predictions, known as the self-distillation set [57]. By contrast, there are only ~100 experimentally characterized fold-switching proteins [1]. In such a data-poor area, it is no wonder that AF2 struggles to predict fold switching. Further, a recent benchmarking study indicates that AF2 has not learned the underlying physics of protein folding [54]. First, it found that AF2's main measure of prediction confidence–the

predicted local distance difference test (plDDT)–selects against alternative conformations of fold-switching proteins with statistical significance. Second, though AF2's structure module accurately ranked good and poor models of single-folding proteins [58], it failed to discriminate between low and high-energy conformations of metamorphic proteins [54]. All proteins are governed by the same physical principles; until those principles are recognized, understood, and modeled effectively, predicting metamorphic proteins will likely remain a challenge.

To move forward, AI-based methodologies for predicting metamorphic proteins must overcome several challenges. For instance, AF2 was recently exploited to predict metamorphic proteins from three distinct families and several organismal sources [59]. The authors attributed the success of their method – AF-cluster – to successful coevolutionary inference of very shallow sequence clusters. Although AF-cluster predicted the dominant conformations of KaiB proteins, all ≥47% identical to their closest experimentally characterized homologs, it systematically failed on most RfaH homologs [54,60] despite clear coevolution of both folds [45]. Likewise, AF-cluster failed to predict both folds for selecase, XCL1 and CLIC1 [59], which also have coevolutionary signals unique to each of their experimentally characterized folds [45]. Furthermore, the claim that AF-cluster harnesses coevolutionary information has been recently challenged by additional calculations using random sequence sampling and single sequence predictions on AF2 [60,61]. These contradictory findings highlight the difficulty of interpreting the DL networks used to predict protein structure and underscore the need for a community-wide effort to explore the limits of current AI-based structure prediction methods and their usefulness for predicting metamorphic proteins.

Experimental screening of fold switchers

Because the training set of metamorphic proteins is so small, robust experimental assays are needed to identify and characterize new fold-switching proteins, whose structures could be then determined and utilized to properly train a meaningful DL model for generalized fold-switching prediction. These assays could be also employed to experimentally validate the potential metamorphic proteins predicted by novel coevolution-based methods such as ACE [45] and AF2-based methods such as AF-cluster [59]. Ultimately, these methods could open the door to design fold-switching proteins, a task that nowadays requires a significant amount of computational work [62]. Several biophysical approaches to achieve this have been discussed previously [63,64] and are summarized as follows (Figure 3):

- HDX-MS can be used to unveil local structural changes and identify distinct trimeric conformations of metamorphic proteins, as in the case of RfaH [37], KaiB [24], and SARS-CoV-2 spike protein [39].
- Circular dichroism can be used to detect large changes in secondary structure or differences in secondary structure between different protein variants, such as RfaH, whose 50-residue CTD completely transforms from α-helix to β-sheet [9].

- Distinct HSQC spectra collected from the same protein sample under different conditions indicated that XCL1 and KaiB switch folds [4,65]. Other proteins that switch folds on a slow timescale may also be identified using their HSQCs.
- Förster Resonance Energy Transfer (FRET)-based assays have been used to observe large conformational changes or differences in end-to-end distances, such as the Rop dimer, which can assume different oligomeric states with distinct hydrophobic packing [66].
- Fold switching can alter functional properties that can be assayed in a purified *in vitro* system or in a cellular context. For example, the β-roll form of RfaH simultaneously binds RNA polymerase and ribosome and is fully active in activating transcription and translation [32] but loses the ability to discriminate between cognate *ops* targets and nonspecific sites [67,68]. This fortuitous gain of function turns on “wrong” genes and is lethal because RfaH excludes its essential paralog NusG from RNA polymerase [69] but is instrumental in assessing the RfaH state by monitoring its activity on *ops*-less reporters [67,68]. While this analysis requires a great deal of pre-existing knowledge, it could greatly facilitate studies of other specialized NusG paralogs that have very divergent sequences but are thought to share metamorphic and activation properties of RfaH. Similar designs of experiments for other metamorphic proteins, such as the real-time monitoring of KaiB-dependent circadian clock oscillations *in vitro* by fluorescence anisotropy [70], could enable the high-throughput screening of hundreds of homologs to unveil their potential metamorphic behaviors.

Funding

This work is supported by the National Institutes of Health (GM067153 to I.A.), the Intramural Research Program of the National Library of Medicine from the National Institutes of Health (LM202011 to L.L.P.), and the National Agency for Research and Development (ANID) through Millennium Science Initiative Program (ICN17_022 to C.A.R–S.) and Fondo de Desarrollo Científico y Tecnológico (FONDECYT 1201684 to C.A.R.-S).

Data availability

No data were used for the research described in the article.

References

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

* of special interest

* * of outstanding interest

1* *. Porter LL, Looger LL: Extant fold-switching proteins are widespread. Proc Natl Acad Sci U S A 2018, 115:5968–5973. [PubMed: 29784778] In this work, the PDB was systematically searched for fold-switching proteins with likely biological relevance. Nearly 100 examples from all kingdoms of life were identified, and their biological triggers and functions were detailed.

2. Chen J, Skehel JJ, Wiley DC: N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. Proc Natl Acad Sci U S A 1999, 96: 8967–8972. [PubMed: 10430879]

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

4. Chang Y-G, Cohen SE, Phong C, Myers WK, Kim Y-I, 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]

5. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, Graham BS, McLellan JS: Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 2020, 367:1260–1263. [PubMed: 32075877]

6. Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM Jr, Rawson S, Rits-Volloch S, Chen B: Distinct conformational states of SARS-CoV-2 spike protein. Science 2020, 369: 1586–1592. [PubMed: 32694201]

7. Littler DR, Harrop SJ, Fairlie WD, Brown LJ, Pankhurst GJ, Pankhurst S, DeMaere MZ, Campbell TJ, Bauskin AR, Tonini R, et al. : The intracellular chloride ion channel protein CLIC1 undergoes a redox-controlled structural transition. J Biol Chem 2004, 279:9298–9305. [PubMed: 14613939]

8. López-Pelegrín M, Cerdà-Costa N, Cintas-Pedrola A, Herranz-Trillo F, Bernadó P, Peinado JR, Arolas JL, Gomis-Rüth FX: Multiple stable conformations account for reversible concentration-dependent oligomerization and autoinhibition of a metamorphic metallopeptidase. Angew Chem Int Ed Engl 2014, 53:10624–10630. [PubMed: 25159620]

9*. Porter LL, Kim AK, Rimal S, Looger LL, Majumdar A, Mensh BD, Starich MR, Strub M-P: Many dissimilar NusG protein domains switch between α-helix and β-sheet folds. Nat Commun 2022, 13:1–12. [PubMed: 34983933] The first work to successfully predict fold switching throughout an entire protein family. Differences in JPred4-based secondary structure predictions were used to successfully distinguish between fold-switching and single-folding NusG proteins in 10/10 experimentally confirmed cases. These predictions suggest that nearly 25% of sequence-diverse NusG proteins switch folds.

10. Chen EA, Porter LL: SSDraw: software for generating comparative protein secondary structure diagrams. Protein Sci 2023, 32, e4836. [PubMed: 37953705]

11. Schrödinger LLC: The PyMOL molecular graphics system, version 2.5. 2022.

12. Yadid I, Kirshenbaum N, Sharon M, Dym O, Tawfik DS: Metamorphic proteins mediate evolutionary transitions of structure. Proc Natl Acad Sci U S A 2010, 107:7287–7292. [PubMed: 20368465]

13* *. Dishman AF, Tyler RC, Fox JC, Kleist AB, Prehoda KE, Babu MM, Peterson FC, Volkman BF: Evolution of fold switching in a metamorphic protein. Science 2021, 371:86–90. [PubMed: 33384377] In this work, ancestral sequence reconstruction was performed on the family of metamorphic protein, XCL1. Distant ancestors of XCL1 assumed one fold, while closer ancestors were metamorphic, indicating that XCL1's fold-switching activity became fixed over evolutionary time. This was the first work to solidly suggest that evolution selects for fold-switching proteins; it also uses NMR experiments to uncover the biophysical basis of XCL1's metamorphic behavior.

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

15. Noel JK, Onuchic JN: The many faces of structure-based potentials: from protein folding landscapes to structural characterization of complex biomolecules. In Computational modeling of biological systems: from molecules to pathways. Edited by Dokholyan NV, Springer US; 2012:31–54.

16. Retamal-Farfán I, González-Higueras J, Galaz-Davison P, Rivera M, Ramírez-Sarmiento CA: Exploring the structural acrobatics of fold-switching proteins using simplified structure-based models. Biophys Rev 2023, 15:787–799. [PubMed: 37681096]

17. Best RB, Chen Y-G, Hummer G: Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. Structure 2005, 13: 1755–1763. [PubMed: 16338404]

18. Camilloni C, Sutto L: Lymphotactin: how a protein can adopt two folds. J Chem Phys 2009, 131, 245105. [PubMed: 20059117]

19. Ghosh C, Jana B: Curious case of MAD2 protein: diverse folding intermediates leading to alternate native states. J Phys Chem B 2022, 126:1904–1916. [PubMed: 35230837]

20. Ramírez-Sarmiento CA, Noel JK, Valenzuela SL, Artsimovitch I: Interdomain contacts control native state switching of RfaH on a dual-funneled landscape. PLoS Comput Biol 2015, 11, e1004379. [PubMed: 26230837]

21*. Zuber PK, Said N, Hilal T, Wang B, Loll B, González-Higueras J, Ramírez-Sarmiento CA, Belogurov GA, Artsimovitch I, Wahl MC, et al. : Concerted transformation of a hyper-paused transcription complex and its reinforcing protein. bioRxiv 2023, 10.1101/2023.06.26.546588.This study mapped the pathway of RfaH recruitment to the transcription complex, using cryoEM to obtain structures of key intermediates and all-atom dual-basin SBM to trace RfaH transformation from the autoinhibited to the active state.

22. Giri Rao VVH, Desikan R, Ayappa KG, Gosavi S: Capturing the membrane-triggered conformational transition of an α-helical pore-forming toxin. J Phys Chem B 2016, 120:12064–12078. [PubMed: 27797514]

23. Lin X, Eddy NR, Noel JK, Whitford PC, Wang Q, Ma J, Onuchic JN: Order and disorder control the functional rearrangement of influenza hemagglutinin. Proc Natl Acad Sci U S A 2014, 111:12049–12054. [PubMed: 25082896]

24. Rivera M, Galaz-Davison P, Retamal-Farfán I, Komives EA, Ramírez-Sarmiento CA: Dimer dissociation is a key energetic event in the fold-switch pathway of KaiB. Biophys J 2022, 121: 943–955. [PubMed: 35151633]

25*. Dodero-Rojas E, Onuchic JN, Whitford PC: Sterically confined rearrangements of SARS-CoV-2 Spike protein control cell invasion. Elife 2021, 10, e70362. [PubMed: 34463614] The conformational rearrangements of the SARS-CoV-2 spike protein from the prefusion to the postfusion state were explored using SBMs, showing that glycans stabilize an intermediate that is crucial for facilitating host membrane capture during infection.

26. Noel JK, Levi M, Raghunathan M, Lammert H, Hayes RL, Onuchic JN, Whitford PC: SMOG 2: a versatile software package for generating structure-based models. PLoS Comput Biol 2016, 12, e1004794. [PubMed: 26963394]

27. Khatua P, Ray AJ, Hansmann UHE: Bifurcated hydrogen bonds and the fold switching of lymphotactin. J Phys Chem B 2020, 124:6555–6564. [PubMed: 32609521]

28. Seifi B, Wallin S: The C-terminal domain of transcription factor RfaH: folding, fold switching and energy landscape. Bio-polymers 2021, 112, e23420.

29. Galaz-Davison P, Román EA, Ramírez-Sarmiento CA: The N-terminal domain of RfaH plays an active role in protein fold-switching. PLoS Comput Biol 2021, 17, e1008882. [PubMed: 34478435]

30. Bazmi S, Seifi B, Wallin S: Simulations of a protein fold switch reveal crowding-induced population shifts driven by disordered regions. Commun Chem 2023, 6:191. [PubMed: 37689829]

31*. Zhang N, Guan W, Cui S, Ai N: Crowded environments tune the fold-switching in metamorphic proteins. Commun Chem 2023, 6:117. [PubMed: 37291449] This study employed ZZ-exchange and HSQC NMR experiments in solutions containing Ficoll, PEG and BSA concentrations that mimic macromolecular crowding conditions to demonstrate that crowding shifts the equilibrium towards one of the two native states of KaiB and XCL1.

32* *. Zuber PK, Schweimer K, Rösch P, Artsimovitch I, Knauer SH: Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. Nat Commun 2019, 10:1–13. [PubMed: 30602773] This study used NMR to show that the RfaH fold-switch requires a complete ops-paused transcription complex and that, upon release from RNAP, its CTD transforms back into the autoinhibitory α-state, resetting the cycle.

33. Liebau J, Tersa M, Trastoy B, Patrick J, Rodrigo-Unzueta A, Corzana F, Sparrman T, Guerin ME, Mäler L: Unveiling the activation dynamics of a fold-switch bacterial glycosyl-transferase by 19F NMR. J Biol Chem 2020, 295:9868–9878. [PubMed: 32434931]

34. Bhuvaneshwari RA, Shivamani A, Sengupta I: Line shape analysis of 19F NMR-monitored chemical denaturation of a fold-switching protein RfaH reveals its slow folding dynamics. J Phys Chem B 2024, 128:465–471. [PubMed: 37991741]

35. Tyler RC, Wieting JC, Peterson FC, Volkman BF: Electrostatic optimization of the conformational energy landscape in a metamorphic protein. Biochemistry 2012, 51:9067–9075. [PubMed: 23102260]

36. Luo X, Tang Z, Xia G, Wassmann K, Matsumoto T, Rizo J, Yu H: The Mad2 spindle checkpoint protein has two distinct natively folded states. Nat Struct Mol Biol 2004, 11:338–345. [PubMed: 15024386]

37*. Galaz-Davison P, Molina JA, Silletti S, Komives EA, Knauer SH, Artsimovitch I, Ramírez-Sarmiento CA: Differential local stability governs the metamorphic fold switch of bacterial virulence factor RfaH. Biophys J 2020, 118:96–104. [PubMed: 31810657] Through a combination of HDX-MS and MD simulations, this work identified that the ends of the α-helical hairpin of RfaH CTD are the least stable elements of the autoinhibited state of RfaH, and the first ones to dissociate from the NTD to enable the fold switch of this protein.

38. Molina JA, Galaz-Davison P, Komives EA, Artsimovitch I, Ramírez-Sarmiento CA: Allosteric couplings upon binding of RfaH to transcription elongation complexes. Nucleic Acids Res 2022, 50:6384–6397. [PubMed: 35670666]

39*. Costello SM, Shoemaker SR, Hobbs HT, Nguyen AW, Hsieh C-L, Maynard JA, McLellan JS, Pak JE, Marqusee S: The SARS-CoV-2 spike reversibly samples an open-trimer conformation exposing novel epitopes. Nat Struct Mol Biol 2022, 29:229–238. [PubMed: 35236990] This work provides a beautiful example of how HDX-MS experiments can be used to identify alternative protein conformations in slow exchange.

40. Shi D, Svetlov D, Abagyan R, Artsimovitch I: Flipping states: a few key residues decide the winning conformation of the only universally conserved transcription factor. Nucleic Acids Res 2017, 45:8835–8843. [PubMed: 28605514]

41. Artsimovitch I, Ramírez-Sarmiento CA: Metamorphic proteins under a computational microscope: lessons from a fold-switching RfaH protein. Comput Struct Biotechnol J 2022, 20: 5824–5837. [PubMed: 36382197]

42. Wang B, Gumerov VM, Andrianova EP, Zhulin IB, Artsimovitch I: Origins and molecular evolution of the NusG paralog RfaH. mBio 2020, 11:e02717–e02720. [PubMed: 33109766]

43. Cai M, Huang Y, Shen Y, Li M, Mizuuchi M, Ghirlando R, Mizuuchi K, Clore GM: Probing transient excited states of the bacterial cell division regulator MinE by relaxation dispersion NMR spectroscopy. Proc Natl Acad Sci U S A 2019, 116: 25446–25455. [PubMed: 31772021]

44. Jain S, Sekhar A: Transient excited states of the metamorphic protein Mad2 and their implications for function. Proteins 2024, 10.1002/prot.26667.

45*. Schafer JW, Porter LL: Evolutionary selection of proteins with two folds. Nat Commun 2023, 14:1–13. [PubMed: 36596776] Though fold switching was previously believed to be an evolutionary epiphenomenon, this work demonstrates that evolution has selected for both conformations of fold-switching proteins, indicating that they likely confer selective advantage. A pipeline for blind prediction of fold switchers was proposed employing the computational methodologies used to infer dual-fold evolutionary selection.

46. Altschuh D, Lesk AM, Bloomer AC, Klug A: Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. J Mol Biol 1987, 193:693–707. [PubMed: 3612789]

47. Yanofsky C, Horn V, Thorpe D: Protein structure relationships revealed by mutational analysis. Science 1964, 146: 1593–1594. [PubMed: 14224506]

48. Anishchenko I, Ovchinnikov S, Kamisetty H, Baker D: Origins of coevolution between residues distant in protein 3D structures. Proc Natl Acad Sci U S A 2017, 114:9122–9127. [PubMed: 28784799]

49. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D: Improved protein structure prediction using predicted inter-residue orientations. Proc Natl Acad Sci U S A 2020, 117: 1496–1503. [PubMed: 31896580]

50. Galaz-Davison P, Ferreiro DU, Ramírez-Sarmiento CA: Coevolution-derived native and non-native contacts determine the emergence of a novel fold in a universally conserved family of transcription factors. Protein Sci 2022, 31:e4337. [PubMed: 35634768]

51. Kamisetty H, Ovchinnikov S, Baker D: Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc Natl Acad Sci U S A 2013, 110:15674–15679. [PubMed: 24009338]
52. Rao RM, Liu J, Verkuil R, Meier J, Canny J, Abbeel P, Sercu T, Rives A: MSA transformer. In Proceedings of the 38th International conference on machine learning. Edited by Meila M, Zhang T, PMLR; 2021:8844–8856.
53. Chakravarty D, Porter LL: AlphaFold2 fails to predict protein fold switching. Protein Sci 2022, 31:e4353. [PubMed: 35634782]
54. Chakravarty D, Schafer JW, Chen EA, Thole JR, Porter LL: AlphaFold2 has more to learn about protein energy landscapes. bioRxiv 2023, 10.1101/2023.12.12.571380.
55. Chen S-J, Hassan M, Jernigan RL, Jia K, Kihara D, Kloczkowski A, Kotelnikov S, Kozakov D, Liang J, Liwo A, et al. : Opinion: protein folds vs. protein folding: differing questions, different challenges. Proc Natl Acad Sci U S A 2023, 120, e2214423119. [PubMed: 36580595]
56. Ahdritz G, Bouatta N, Floristean C, Kadyan S, Xia Q, Gerecke W, O'Donnell TJ, Berenberg D, Fisk I, Zanichelli N, et al. : OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. bioRxiv 2023, 10.1101/2022.11.20.517210.
57. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. : Highly accurate protein structure prediction with AlphaFold. Nature 2021, 596:583–589. [PubMed: 34265844]
58. Roney JP, Ovchinnikov S: State-of-the-art estimation of protein model accuracy using AlphaFold. Phys Rev Lett 2022, 129, 238101. [PubMed: 36563190]
59. Wayment-Steele HK, Ojoawo A, Otten R, Apitz JM, Pitsawong W, Hömberger M, Ovchinnikov S, Colwell L, Kern D: Predicting multiple conformations via sequence clustering and Alpha-Fold2. Nature 2023, 625:832–839. [PubMed: 37956700]
60. Porter LL, Chakravarty D, Schafer JW, Chen EA: ColabFold predicts alternative protein structures from single sequences, coevolution unnecessary for AF-cluster. bioRxiv 2023, 10.1101/2023.11.21.567977.
61. Schafer JW, Chakravarty D, Chen EA, Porter LL: Sequence clustering confounds AlphaFold2. bioRxiv 2024, 10.1101/2024.01.05.574434.
62. Ruan B, He Y, Chen Y, Choi EJ, Chen Y, Motabar D, Solomon T, Simmerman R, Kauffman T, Gallagher DT, et al. : Design and characterization of a protein fold switching network. Nat Commun 2023, 14:431. [PubMed: 36702827]
63. Porter LL: Fluid protein fold space and its implications. Bio-essays 2023, 45, e2300057.
64. Chakravarty D, Schafer JW, Porter LL: Distinguishing features of fold-switching proteins. Protein Sci 2023, 32, e4596. [PubMed: 36782353]
65. 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]
66. Gambin Y, Schug A, Lemke EA, Lavinder JJ, Ferreon ACM, Magliery TJ, Onuchic JN, Deniz AA: Direct single-molecule observation of a protein living in two opposed native structures. Proc Natl Acad Sci U S A 2009, 106:10153–10158. [PubMed: 19506258]
67. Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin NV, Vassylyev DG, Artsimovitch I: Structural basis for converting a general transcription factor into an operon-specific virulence regulator. Mol Cell 2007, 26:117–129. [PubMed: 17434131]
68. Zuber PK, Artsimovitch I, NandyMazumdar M, Liu Z, Nedialkov Y, Schweimer K, Rösch P, Knauer SH: The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. Elife 2018, 7, e36349. [PubMed: 29741479]
69. Belogurov GA, Mooney RA, Svetlov V, Landick R, Artsimovitch I: Functional specialization of transcription elongation factors. EMBO J 2009, 28:112–122. [PubMed: 19096362]
70. Heisler J, Chavan A, Chang Y-G, LiWang A: Real-time in vitro fluorescence anisotropy of the cyanobacterial circadian clock. Methods Protoc 2019, 2:42. [PubMed: 31164621]

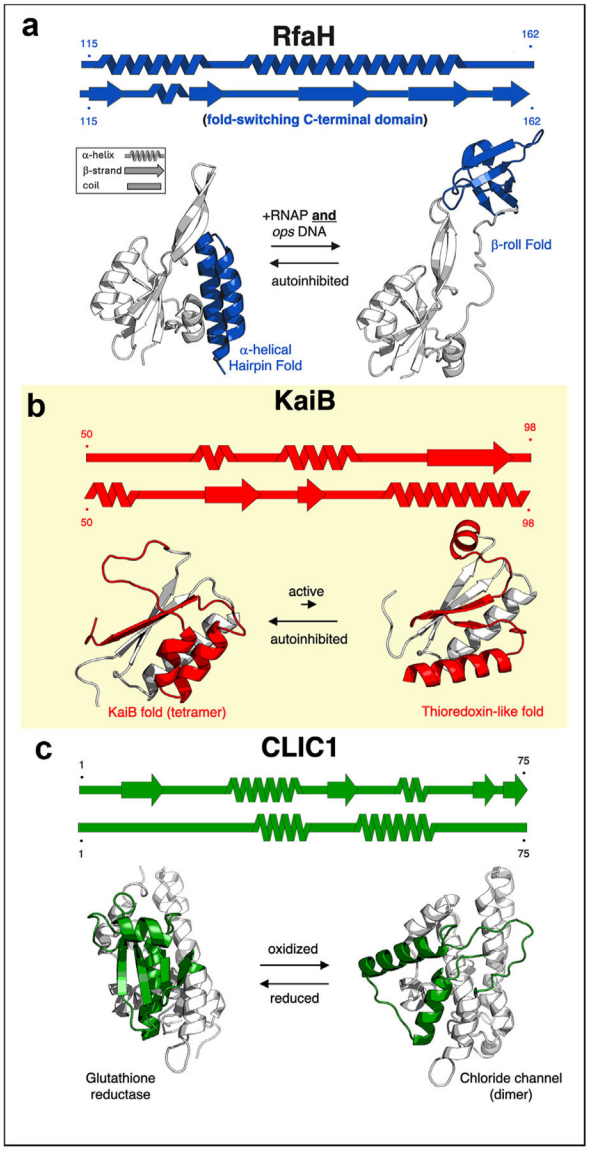

Figure 1.

Three examples of metamorphic proteins. (**a**) The C-terminal domain (CTD, blue) of *E. coli* RfaH reversibly switches from an α-helical to β-sheet fold upon binding RNA polymerase (RNAP) and *ops* DNA. Adapted from Ref. [9]. (**b**) The C-terminal subdomain of *S. elongatus* KaiB (red) switches secondary structures when its ground state tetrameric form dissociates into a monomer. (**c**) The N-terminal domain of *H. sapiens* CLIC1 (green) switches from a glutathione reductase fold to a helical fold that may foster chloride efflux in response to oxidative stress. Single-folding regions of all proteins are colored gray. Secondary structures diagrams were generated using SSDraw [10] and protein ribbon diagrams with PyMOL [11].

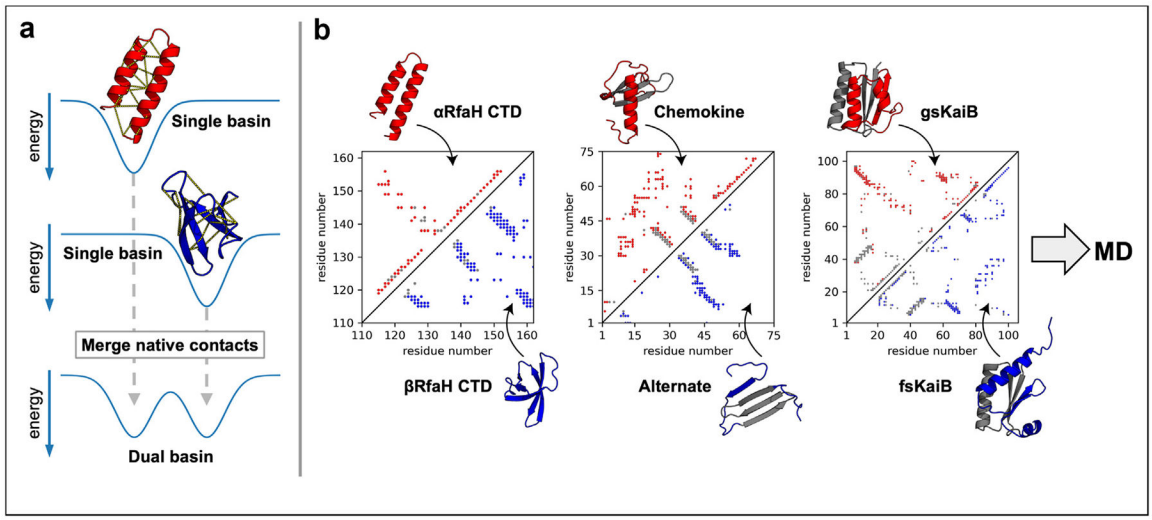

Figure 2.

Native contacts derived from experimentally solved structures of alternative states of metamorphic proteins can be combined to enable their use in MD simulations. The scheme in A shows how the unique native contacts (yellow lines) from two different structures (red, blue) of a metamorphic protein (in this case, RfaH CTD) can be merged to generate a dual-basin SBM that enables to simulate the transition between both states. In B, the native contact maps of the two native states of several metamorphic proteins that have been used for MD simulations of fold-switching are presented. While a significant number of native contacts are unique to each fold (red, blue), other contacts can be common to both folds (gray). The residue pair native contacts presented herein were obtained using the SMOG2 tool [26] on the experimentally solved structures of αRfaH CTD (PDB 2OUG) and βRfaH CTD (PDB 2LCL), the chemokine (PDB 2HDM) and alternate (PDB 2JP1) folds of XCL1, and the ground (gs, PDB 1VGL) and fold-switched (fs, PDB 5JYT) states of KaiB.

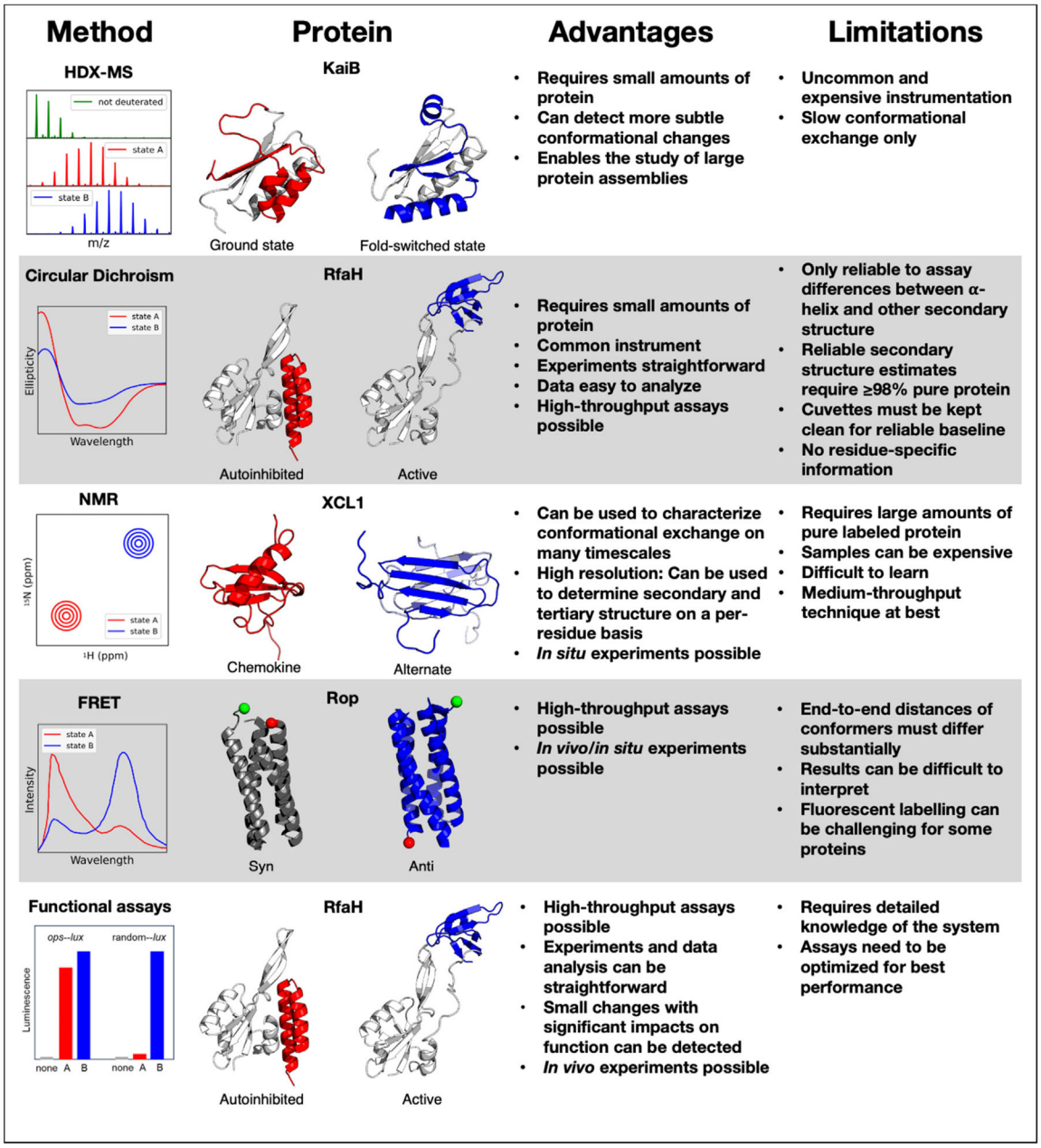

Figure 3.

Biophysical and functional approaches to interrogate the fold-switching of metamorphic proteins. Examples of a protein characterized, as well as advantages and limitations, are listed for each method.