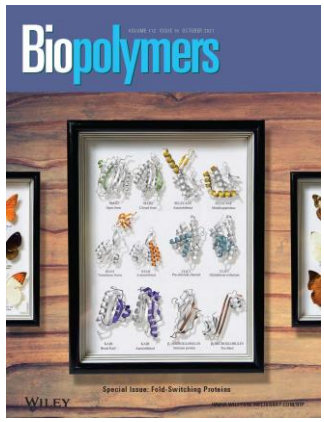
Biopolymers

Special Issue: Fold-Switching Proteins

Guest Editors: Andy LiWang, Lauren L. Porter and Lee-Ping Wang



EDITORIAL

Fold-Switching Proteins
Andy LiWang, Lauren L. Porter, Lee-Ping Wang, *Biopolymers* 2021, doi: 10.1002/bip.23478

REVIEW

Identification and characterization of metamorphic proteins: Current and future perspectives
Madhurima Das, Nanhao Chen, Andy LiWang, Lee-Ping Wang, *Biopolymers* 2021, doi: 10.1002/bip.23473

ARTICLES

Specific binding-induced modulation of the XCL1 metamorphic equilibrium
Acacia F. Dishman, Francis C. Peterson, Brian F. Volkman, *Biopolymers* 2021, doi: 10.1002/bip.23402

Dynamic and conformational switching in proteins
H. A. Scheraga, S. Rackovsky, *Biopolymers* 2021, doi: 10.1002/bip.23411

The C-terminal domain of transcription factor RfaH: Folding, fold switching and energy landscape
Bahman Seifi, Stefan Wallin, *Biopolymers* 2021, doi: 10.1002/bip.23420

Inducible fold-switching as a mechanism to fibrillate pro-apoptotic BCL-2 proteins
Daniel L. Morris, Nico Tjandra, *Biopolymers* 2021, doi: 10.1002/bip.23424

A high-throughput predictive method for sequence-similar fold switchers
Allen K. Kim, Loren L. Looger, Lauren L. Porter, *Biopolymers* 2021, doi: 10.1002/bip.23416

A sequence-based method for predicting extant fold switchers that undergo α-helix ↔ β-strand transitions
Soumya Mishra, Loren L. Looger, Lauren L. Porter, *Biopolymers* 2021, doi: 10.1002/bip.23471

Received: 31 December 2020 | Revised: 9 August 2021 | Accepted: 10 August 2021

DOI: 10.1002/bip.23473

REVIEW

Biopolymers WILEY

Identification and characterization of metamorphic proteins: Current and future perspectives

Madhurima Das[1] | Nanhao Chen[2] | Andy LiWang[1,3,4,5,6] | Lee-Ping Wang[2]

[1]School of Natural Sciences, University of California, Merced, California, USA

[2]Department of Chemistry, University of California, Davis, California, USA

[3]Department of Chemistry and Biochemistry, University of California, Merced, California, USA

[4]Center for Cellular and Biomolecular Machines, University of California, Merced, California, USA

[5]Health Sciences Research Institute, University of California, Merced, California, USA

[6]Center for Circadian Biology, University of California, San Diego, California, USA

Correspondence
Lee-Ping Wang, Department of Chemistry, University of California, Davis, CA 95616, USA.
Email: leeping@ucdavis.edu

Funding information
Life Sciences Division, Army Research Office; U.S. Army, Grant/Award Number: W911NF-17-1-0434

Abstract

Proteins that can reversibly alternate between distinctly different folds under native conditions are described as being metamorphic. The "metamorphome" is the collection of all metamorphic proteins in the proteome, but it remains unknown the extent to which the proteome is populated by this class of proteins. We propose that uncovering the metamorphome will require a synergy of computational screening of protein sequences to identify potential metamorphic behavior and validation through experimental techniques. This perspective discusses computational and experimental approaches that are currently used to predict and characterize metamorphic proteins as well as the need for developing improved methodologies. Since metamorphic proteins act as molecular switches, understanding their properties and behavior could lead to novel applications of these proteins as sensors in biological or environmental contexts.

KEYWORDS
fold switching, KaiB, lymphotactin, metamorphic protein, metamorphome

1 | INTRODUCTION

The classic paradigm "one sequence, one fold" proposed by Anfinsen postulates that the native structure of a protein is determined by its primary aminoacyl sequence.[1,2] In 1994, George Rose and Trevor Creamer put forth the "Paracelsus Challenge"—to transform a protein from its original fold to a different fold while retaining no less than 50% identity to the original aminoacyl sequence.[3] This challenge was solved in 1997 by Dalal *et al*. who designed a variant of a predominantly beta-sheet protein that instead adopted a helix bundle conformation.[4] At that time, it was unknown that some proteins can switch reversibly between two distinctly different folds under native conditions despite having just one aminoacyl sequence. The first protein that was found to definitively possess this property is thought to be XCL1/lymphotactin, identified by the lab of Brian Volkman in 2002.[5] These proteins have since come to be known as "metamorphic proteins," a term coined by Murzin in 2008.[6] Metamorphic proteins have close and sometimes partly overlapping relationships with other groups of proteins[7] including: intrinsically disordered proteins (IDPs), which lack ordered three-dimensional (3D) structure[8]; morpheeins, which possess multiple conformations to accommodate different oligomeric states[9]; moonlighting proteins, which exhibit multiple functions within a single aminoacyl sequence[10,11]; and prions[12,13] and amyloid proteins,[14] which are known for their ability to fold or change folds irreversibly.

Metamorphic proteins are of emerging interest because they evolved to adopt two different folds that interchange under native conditions. For example, XCL1/lymphotactin from the C family of chemokines[5,15] exists in a dynamic equilibrium between two distinctly different folds under physiological conditions each associated with a different function. The monomeric form binds XCL1/lymphotactin receptors while the dimeric form binds to heparin; as the two folds of XCL1/lymphotactin are mutually exclusive, so are its two distinct functions.[16] KaiB, the only known metamorphic protein in circadian clock systems,

This article was published on the 16th September 2021. On 5th October 2021 the article type was corrected to Review and biographies of the authors were added.

exists in an equilibrium between a unique fold that self-associates as a homotetramer and monomer with a thioredoxin-like fold.[17] Throughout the day, KaiB exists primarily in the tetrameric state but at night binds to its partner protein, KaiC, in its monomeric fold and this cycle repeats in synchrony with the Earth's rotation.[18,19]

In addition to their significance to the fields of evolution, biology, and protein folding, metamorphic proteins are also potential starting points for engineering new applications. For instance, biosensors have already been designed using proteins, and metamorphic proteins can potentially benefit this field, because they can change their inactive/active fold states (output) in response to a stimulus (input) and thus reversibly switch between "off" and "on" states.[20–22] As metamorphic proteins are an emerging class of proteins, it is important to advance methodology with which to better characterize the population of metamorphic proteins, called the "metamorphome."

Only a small number of proteins have been experimentally found to be metamorphic largely through serendipity because methods for identifying and characterizing metamorphic proteins using directed searches are in their infancy. A major challenge in the experimental identification of metamorphic proteins is that fold switching is not always observable as a conformational equilibrium under typical experimental conditions. Rather, fold switching often depends on an unknown environmental trigger (e.g., temperature, pH, concentration) that stabilizes the alternate fold, and without preexisting knowledge of the conditions that trigger the fold switch, only one of the conformations may be observable.[23] Therefore, if a conformational equilibrium is not initially observed and the trigger is unknown, then significant experimental efforts are needed to identify the fold switching behavior for even one protein sequence. A more efficient approach to identifying metamorphic protein sequences could be by using computational approaches for high-throughput screening, followed by experimentation for confirmation.

There is also a gap in knowledge on how metamorphic proteins convert between their alternate folds, impeding a comprehensive understanding of protein behaviors and developing protein-based biotechnologies. After the metamorphic behavior in a protein sequence has been *identified*, more detailed experimental and computational approaches should be used to *characterize* the fold switching behavior, which could include elucidating the mechanism of fold switching, identifying conformational intermediates along the fold switching pathway, and engineering the fold switching protein toward desired applications.

Recently, there have been new computational approaches developed to identify potentially metamorphic proteins from their aminoacyl sequences. For example, Looger and Porter[24] identified 192 extant structures in the Protein Data Bank (PDB) (96 structure pairs) that were similar or identical in sequence but had different 3D folds, and estimated that 0.5% to 4% of proteins in the PDB are capable of switching folds. The list was revised by Chen *et al.*[25] resulting in a dataset consisting of 201 potentially metamorphic proteins. These recent advances have improved our knowledge of the metamorphome, but perhaps more importantly also revealed the need for more powerful computational and experimental methodology to study this exciting family of proteins. This perspective is inspired by the possibility that the handful of labs working on metamorphic proteins today will attract numerous other researchers to create a vibrant community of metamorphome researchers. It summarizes recent advances in experimental and computational methodologies (see Table 2) with a future outlook on how to use these tools more effectively.

2 | EXPERIMENTAL APPROACHES TO METAMORPHIC PROTEIN CHARACTERIZATION

Presently, only 20 proteins have been experimentally annotated and verified as being metamorphic (XCL1,[5,16] KaiB,[18] IscU,[26]

TABLE 1 Triggers and structure determination techniques for known metamorphic proteins

Metamorphic protein	Trigger	Technique	Metamorphic protein	Trigger	Technique
XCL1	Salt, temperature	NMR	XRCC1	Redox	NMR/X-ray
KaiB	Ligand binding	NMR	HIV-1 capsid maturation switches	Proteolysis	NMR/X-ray
IscU	Redox	NMR	HIV-1 reverse transcriptase	Unfolding	NMR/X-ray
Mad2	Ligand binding	NMR	P1 lysozyme	Redox	X-ray
OxyR	Redox	X-ray	Phytochrome photoreceptor	Light	X-ray
Caspase-6	Ligand binding	X-ray/HDX-MS/MD	Retinoic acid receptor	Ligand binding	X-ray
CLIC	Redox	X-ray	Selecase	Concentration	X-ray
RfaH	Ligand binding	NMR	Serpins	Proteolysis/domain swap	X-ray
Cytolysin A	Membrane insertion	X-ray	T7 RNA polymerase	Ligand binding	X-ray
Hemagglutinin	pH	X-ray	TCR ectodomain	Unknown	X-ray

Abbreviations: HDX-MS, hydrogen/deuterium-exchange mass spectrometry; HIV-1, human immunodeficiency viruses-1; MD, molecular dynamics; NMR, nuclear magnetic resonance.

TABLE 2 Summary of experimental and computational methodologies to identify and characterize known metamorphic proteins

Techniques	Applicability	Limitations	Throughput	Examples of metamorphic proteins
X-ray crystallography	Solving structure at very high atomic resolution	Traps structure at its lowest energy; low electron density for flexible regions	Low	OxyR,[28] CLIC,[30] Cytolysin A,[32] hemagglutinin,[33] XRCC1,[34,35] human immunodeficiency viruses-1 (HIV-1) capsid maturation switches,[36,37] HIV-1 reverse transcriptase,[38] P1 lysozyme,[39] phytochrome photoreceptor,[40] retinoic acid receptor,[41] Selecase,[42] Serpins,[43] T7 RNA polymerase,[44] TCR Ectodomain[45]
NMR	High structural resolution of protein ensembles; detection of sparsely populated intermediate states	Time consuming; size-limitation and requires high protein concentration	Med	XCL1,[16] KaiB,[18] IscU,[26] Mad2[27], RfaH,[31] XRCC1,[34,35] HIV-1 capsid maturation switches,[36,37] HIV-1 reverse transcriptase[38]
Cryo-electron microscopy	Atomic scale structural information; large macromolecular complexes can be investigated (>300 kDa)	Requires single crystals; poor tolerance to disordered or highly flexible regions	Low	
Circular dichroism (CD)	Changes in secondary structure; label-free; quick assay	Low sensitivity; formulation buffers can interfere	Low	Caspase-6,[29] Selecase[42]
Electron Paramagnetic Resonance (EPR)	Structural and dynamic information; <1.5 nm distance measurement in solution; membrane-embedded proteins	Needs paramagnetic labels; residue-level structural/ dynamic information	Low to Med	KaiB[18]
Small-angle X-ray Scattering (SAXS)	Information on shape, assembly, aggregation, flexibility; no size limitation	Low resolution	Variable	KaiB,[18] XRCC1,[34,35] HIV-1 reverse transcriptase[38] phytochrome photoreceptor,[40] Selecase[42]
Mass spectrometry (HDX-MS)	Structural, thermodynamic and kinetic information on molecular assemblies; high sensitivity; automated; suitable for coupling with other methods	Expensive; requires pure protein samples; low protein identification rate; manual work is required; no direct structural information	High	KaiB,[18] Caspase-6,[29] RfaH,[78] CLIC,[77] Selecase[42]
Single-molecule Förster Resonance Energy Transfer (smFRET) and Fluorescence Spectroscopy	Resolves inter- and intra-molecular distances (<5 nm) and long-range transitions; can be used to establish conformational changes in the two different folds of metamorphic protein	Requires multiple fluorescent labels to be in exact location; sensitive to inner filter effect, auto-fluorescence and photobleaching	High	XCL1,[16] KaiB,[18]Caspase-6,[29] XRCC1,[34,35] retinoic acid receptor[41]
Differential scanning calorimetry/fluorimetry (DSC/DSF)	Differentiates between folds and/or oligomerization states by tracking reversible changes in heat capacity	No direct structural information	Med	hemagglutinin[33]
Surface Plasmon Resonance (SPR)	Label-free; real-time kinetic measurements	Surface immobilization can interfere with binding event	Med	
Fourier-transform infrared spectroscopy (FTIR)/Raman spectroscopy	Conformation and orientation of membrane-associate macromolecules; detection of protein misfolding and aggregation	Difficulties in determination of secondary structure and orientation	Med	

(Continues)

TABLE 2 (Continued)

Techniques	Applicability		Limitations	Throughput	Examples of metamorphic proteins
Electron microscopy (EM)	Information on shape and assembly state of macromolecular complexes		Negative-staining might give rise to artifacts; cumbersome sample preparation and screening	Low	Mad2,[27] Cytolysin A,[32] HIV-1 capsid maturation switches[36,37]
Sequence-based approach	Porter *et al.* codes https://github.com/llporter/FoldSwitch.git	Determines independent cooperatively folding regions; discrepancies in secondary structure prediction programs	Misclassifications as reflected in false positives; small dataset of known metamorphic proteins	High	
	Chen *et al.* codes https://github.com/BeyondCNH/DiversityIndex	Predicts metamorphic tendency by quantifying uncertainty in secondary structure prediction programs	Misclassifications as reflected in false positives; small dataset of known metamorphic proteins	High	
Physics-based molecular simulation methods	Molecular dynamics	Simulates protein structural ensembles using classical dynamics; can determine fold switching mechanism and intermediates	Metamorphic time scale may exceed computational power; may require artificial perturbations to observe desired behavior	Med	Caspase-6,[29] phytochrome photoreceptor[40]

Mad2,[27] OxyR,[28] Caspase-6,[29] CLIC,[30] RfaH,[31] Cytolysin A,[32] hemagglutinin,[33] XRCC1,[34,35] human immunodeficiency viruses-1 (HIV-1) capsid maturation switches,[36,37] HIV-1 reverse transcriptase,[38] P1 lysozyme,[39] phytochrome photoreceptor,[40] retinoic acid receptor,[41] Selecase,[42] Serpins,[43] T7 RNA polymerase,[44] TCR ectodomain[45]). A list of known triggers (mostly adapted from Kulkarni *et al.*[46]) and techniques that were primarily used to recognize their metamorphic properties are presented in Table 1. Of all these proteins, structures of 11 of these proteins were solved by X-ray crystallography (e.g., see Figure 1), but their metamorphic behavior was serendipitously discovered while using other biophysical techniques. For example, crystal structures of both KaiB[47,48] and RfaH[49] were well-characterized, but their metamorphic properties could only be identified by nuclear magnetic resonance (NMR). For three proteins, NMR in combination with X-ray (XRCC1,[34,35] HIV-1 capsid maturation switches,[36,37] HIV-1 reverse transcriptase[38]) were used to discover their metamorphic properties, while H/DX-MS combined with molecular dynamics (MD) simulations were used to probe the distinct conformational dynamics of Caspase-6.[29] For the rest, NMR was primarily used in combination with other biophysical techniques to discover their metamorphic properties. The challenges of experimental techniques to identify metamorphic proteins are briefly discussed in the following sections.

2.1 | X-ray crystallography

X-ray crystallography is a powerful experimental technique for protein structure determination because of its potential to provide high spatial resolution over a broad molecular weight range. As of February 2020, approximately 89% of the structures in PDB were solved by X-ray crystallography. However, a drawback is that it traps proteins in a free energy minimum mostly defined by a single structure within a crystallographic lattice, whereas in solution proteins exist as an ensemble of thermodynamically accessible states. For example, the X-ray crystal structure of the metamorphic protein KaiB revealed that it adopts a unique fold as a homotetramer,[47,48] and its mode of interaction with KaiC remained speculative until the monomeric fold-switched structure of KaiB was solved by NMR.[18] Separately crystallizing a metamorphic protein in its different folds is challenging unless conditions favoring one fold over the other are known in advance.

2.2 | NMR spectroscopy

Solution-state NMR spectroscopy has been used to discover and characterize approximately half of experimentally validated metamorphic proteins (XCL1,[5,16] KaiB,[18] IscU,[26] Mad2,[27] RfaH,[31] XRCC1,[34,35] HIV-1 capsid maturation switches,[36,37] HIV-1 reverse transcriptase[38]), despite comprising only 8% of protein structures deposited in the RCSB PDB. For example, ^{15}N-^{1}H HSQC spectra of KaiB free and bound to a monomerized domain of KaiC had almost no

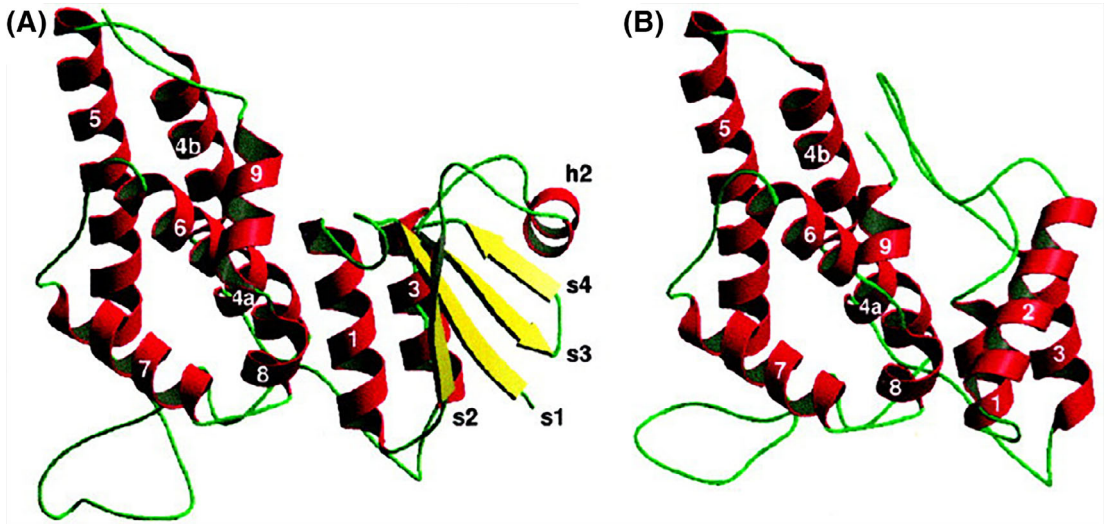


FIGURE 1 Reversible monomer-dimer structural transition in CL1C, identified in X-ray crystal structures. The major changes in the N-domain are primarily due to the transition of residues from the beta-sheet region of Ramachandran space (A) into a loop conformation (B). Reproduced from Ref. [30]

chemical shifts in common[50] and secondary backbone chemical shifts[51–53] (Figure 2) indicated they had different secondary structures.[18] During NMR sample optimization of XCL1/lymphotactin, two distinct sets of chemical shifts were observed depending on salt concentration and temperature,[5] as a result of an equilibrium between two unrelated folds.[16] XCL1/lymphotactin exemplifies the ability of NMR to observe multiple states of a solubilized protein simultaneously, which is why many metamorphic proteins were found serendipitously using this technique.

Inherent limitations associated with NMR spectroscopy, such as time-intensive sample preparation, data collection and analysis, and low-throughput, are likely prohibitive for screening proteins for metamorphic properties. More realistically, NMR could be used to test candidate proteins predicted to be metamorphic by computational approaches. Advanced NMR methods including Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion, paramagnetic relaxation enhancement, and chemical-exchange saturation transfer (CEST) spectrocopies[54–57] are capable of detecting sparsely populated states depending on fold-switching time scales and may be applicable for detecting alternate folds of metamorphic proteins. CEST based on longitudinal magnetization makes it amenable to study an invisible or sparsely populated conformational state of a protein, and was used to detect the sparsely-populated "open" conformations of RfaH in slow chemical exchange with its visible ground-state "closed" conformation.[58] This demonstrates the power of NMR as a primary method for studying structural plasticity of small to moderate size proteins.

2.3 | Cryo-electron microscopy

An increasingly popular and fast-emerging technique is single-particle cryo-electron microscopy (cryo-EM), which is now routinely being used to determine the structures of proteins to approximately 3 Å resolution or better.[59] In these experiments, an electron beam is used to image a flash-frozen protein sample, and single-particle analysis (SPA) or sub-tomogram averaging (STA) techniques are used to reconstruct 3D models from the microscope images.[60–62] Because samples are not crystalline, this technique can observe conformational differences between different single particles within a single protein preparation. Looger and Porter found that 11% of the fold-switching pairs in their dataset contained one structure that was solved by cryo-EM, suggesting that this method has outsize potential for identifying new metamorphic proteins.[24]

Although we do not know of any large-scale studies that leverage cryo-EM for identifying metamorphic proteins, we anticipate that this technique will be used successfully for these discoveries in the near future. A possible cryo-EM experiment to search for protein metamorphism could involve developing new data analysis techniques that identify protein particles in an image that do not fit the primary 3D model, and using these outlier particles to construct a model of the alternate fold. Additional data to improve the models could be obtained by tuning experimental conditions to favor the alternate fold.

2.4 | Alternative experimental approaches

Protein metamorphism can be characterized by a number of spectroscopic techniques long employed to study protein conformational changes. Circular dichroism (CD) is an example of such an alternate technique that uses the differential absorption in left- and right-circularly polarized light to estimate the percentage of secondary structure types in a protein (Figure 3); this technique was used by Dalal *et al*. to design a protein with a four-helix bundle fold with 50% sequence identity to a predominantly β-sheet protein, thereby

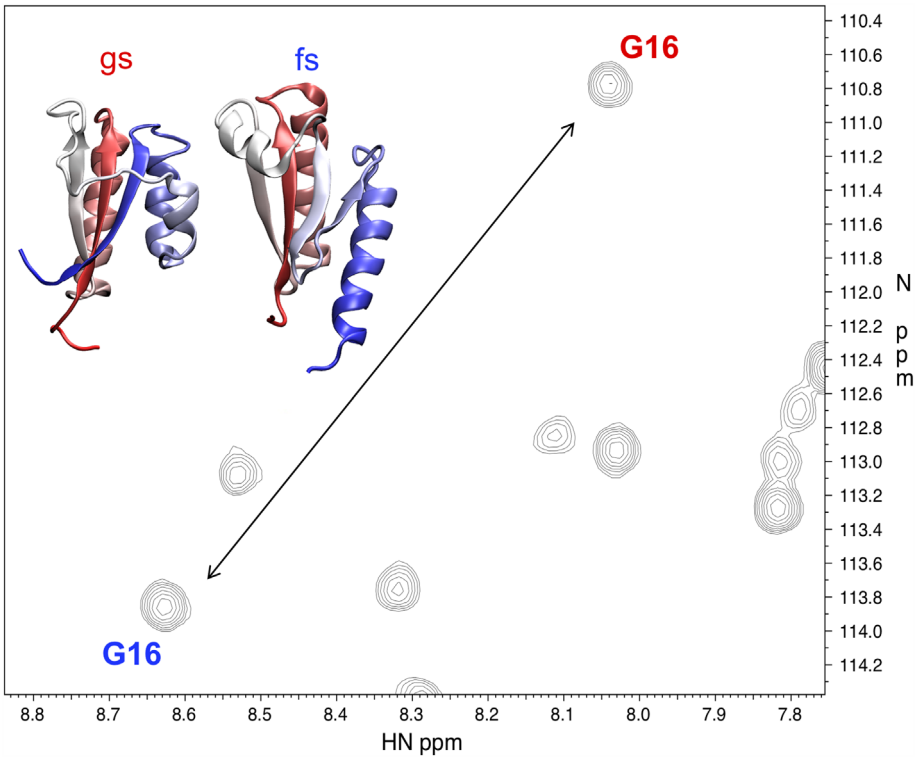


FIGURE 2 1H-15 N HSQC spectra of cyanobacterial KaiB that exists in a two-state conformational equilibrium between the fold-switch (fs) and ground state (gs). Arrows connect resonances for the same residue G16 in the two conformational states: (fs) and (gs)

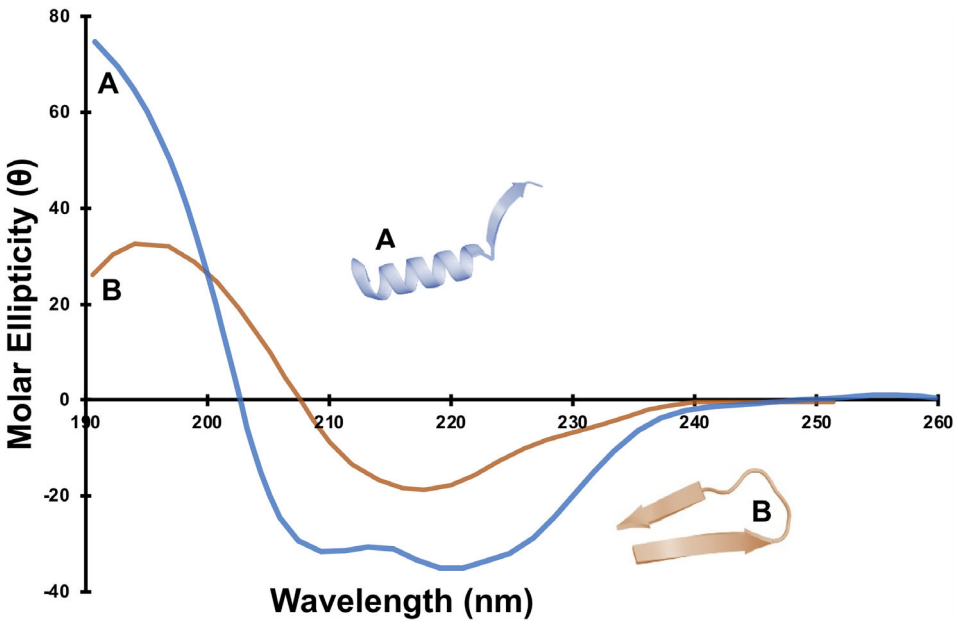


FIGURE 3 Illustration of how circular dichroism (CD) spectra can be sensitive to protein secondary structure changes. The helical structure (A) is expected to generate a stronger CD signal than the beta hairpin (B)

winning the "Paracelsus Challenge."[4] The amide absorption band in Fourier transform infrared (FT-IR) spectra can be interpreted for secondary structure content, and this technique has been applied to study reversible secondary structure transitions in transmembrane domains.[63] Raman spectroscopy of amide absorption bands[64] has the advantage of being applicable in dilute solutions, and has been used to capture conformational ensembles of natively unfolded proteins.[65] Stopped-flow fluorescence experiments, which induce a rapid change in solution conditions followed by measuring tryptophan fluorescence after a controlled time delay, have been employed to

study the fold-switching mechanism; it was found that Mad2 interconverts via metastable intermediates,[66] while XCL1/lymphotactin interconversion requires a large-scale unfolding process.[67] Single-molecule Förster resonance energy transfer (FRET) experiments have been used to monitor the fold switching equilibrium in a metamorphic protein[68]; while this is a powerful technique to investigate dynamics at the single-molecule level, it requires the addition of fluorescent probes that restricts the method to characterizing individual systems where metamorphism is already known. Surface plasmon resonance (SPR) has been used to measure conformational changes of proteins immobilized at the interface of a metal and aqueous solution, and has the advantage of being label-free with very small sample size requirements (on the order of a few nano grams).[69]

Compared to X-ray crystallography and cryo-EM, these types of spectroscopy could make it easier to perturb the conditions of the sample (e.g., pH, concentration, temperature) and measure the response in the signal, thereby facilitating the search for triggers of metamorphism. High-throughput protein expression and purification methods[70,71] combined with recently developed high-throughput IR spectroscopy methods[72,73] are currently capable of characterizing the secondary structures of hundreds of proteins in parallel, and we think that if such an experimental approach could be used to measure how structures depend on changes in sample conditions, it could serve as a powerful experimental screen for metamorphic protein candidates.

Mass spectrometry techniques can also be leveraged to study protein metamorphism, for example: (a) ion-mobility mass spectrometry with electrospray ionization provides rotationally averaged cross-sections of proteins, which can be used to infer the existence of multiple folds[74,75]; and (b) hydrogen/deuterium-exchange mass spectrometry (HDX-MS) measures the deuterium uptake of proteins dissolved in a deuterated solvent, thereby quantifying the solvent-accessible surface area, which may take on multiple values for different folds.[76,77] HDX-MS has been successfully employed to explore the conformational dynamics and study the local stability of metamorphic proteins like RfaH,[78] KaiB,[18] Caspase-6,[29] Selecase,[42] and CL1C.[77] A newly developed approach called limited proteolysis-coupled mass spectrometry (LiP-MS) involves subjecting proteins to a structure-dependent enzymatic digestion procedure, resulting in peptide fragments with masses that depend on the initial protein conformation.[79] Because LiP-MS was developed to detect protein structural changes in the entire proteome taken from a biological extract, it has the potential capability to detect metamorphism in a mixture of many different proteins, thereby removing the need to carry out experiments on individually purified proteins.

In conclusion, we think that experimental approaches with the potential to detect environmentally induced structural changes in a high-throughput manner could be a powerful screening tool for identifying metamorphic proteins. In contrast, methods that are not high throughput and require more intensive sample preparation or experimental setup are better suited for investigating the thermodynamics, kinetics, and/or mechanism in systems where fold switching has already been identified. Computational methods are another promising route for screening the vast sequence space for metamorphic candidates, which we describe in the next section.

3 | COMPUTATIONAL APPROACHES TO METAMORPHIC PROTEIN PREDICTION

Currently, of the approximately 162 thousand protein structures in RCSB PDB, only 20 of them are experimentally annotated as metamorphic proteins, around one in 10 000. On the other hand, recent estimates of the percentage of metamorphic proteins in the PDB are as high as 0.5% to 4%,[24] which implies that hundreds to thousands of experimentally known protein structures possess as-yet-undiscovered alternate folds. However, it is currently intractable to experimentally search for alternate folds of all proteins in the PDB, even for high-throughput experimental approaches. This has consequently lead to the development of computational approaches to identify and characterize metamorphic proteins that build on existing structural biology tools (see Figure 1). An advantage of computational approaches is that they are relatively high throughput compared to experimentation, but due to the inherently approximate nature of predictions, confirmation by targeted experimental studies is required.

In addition to the relatively tiny number of known metamorphic proteins, there is precious little experimental knowledge on how metamorphic proteins switch folds. Experimentally capturing a metamorphic protein in the process of switching folds is highly difficult unless intermediates along the pathway can be stabilized. Therefore, there is also much interest in computational methods to elucidate the mechanisms by which metamorphic proteins switch folds. These studies could provide valuable predictions such as intermediate states along the pathway that could be experimentally tested using mutagenesis or other approaches.

Computational approaches can be subdivided into *knowledge-based approaches* that make predictions based on existing experimental data, and *physics-based approaches* that involve computer simulations based on physical principles. These categories of methods are complementary because drawing on existing knowledge can greatly simplify the complexity of a calculation, whereas physics-based simulations can generate more accurate predictions of properties difficult to experimentally access. There exists a broad spectrum of computational methods where knowledge-based and physics-based methods blend into one another; for example, protein design and MD software both incorporate knowledge-based and physics-based elements.

3.1 | Sequence-based methods for predicting metamorphism

These knowledge-based methods are based mainly on the sequence and secondary structures of preferably single-domain proteins. These methods do not explicitly involve computation on 3D protein structures, and thus are among the least expensive computational methods. The PDB provides secondary structure annotations of its

deposited structures, and this wealth of data has led to the development of data-driven secondary structure prediction tools over the past three decades.[80–83]

Porter *et al*. proposed that fold switching proteins can be placed into two categories: (a) Extant fold switchers are individual proteins that remodel their secondary structure in response to cellular stimuli to form alternative folds and, (b) Evolved fold switchers are pairs of proteins with highly similar sequences (>98% sequence identity) and yet adopt different stable folds.[84] They report two distinguishing features of extant fold-switching proteins: (a) cooperatively folding regions that are likely to unfold/refold independently in response to environmental triggers and, (b) discrepancies between predicted and experimentally determined secondary structure as reflected in secondary structure prediction programs.[24,85] This inconsistency in secondary structure prediction accuracy was used to discriminate 21 known evolved metamorphic proteins from known monomorphic proteins (built from CATH domains with <35% identity from single-fold families).[84]

A desirable goal for computational methods is to reliably predict whether a protein is likely to be metamorphic using only the aminoacyl sequence as input. A sequence-based classifier model was recently described by Chen *et al*.[25] where protein metamorphism (or its absence) is predicted based on the so-called diversity index, which is a measure of uncertainty in secondary structure prediction. The diversity index is computed from the probabilities of secondary structure classifications generated by a secondary structure prediction program for each aminoacyl residue in an input sequence. A protein is predicted to be metamorphic if the computed diversity index exceeds a threshold value for a sufficiently long contiguous portion of the sequence. The two parameters of the classifier were trained using a revised version of Porter's metamorphic dataset and a novel dataset consisting of 200 manually annotated monomorphic (single-fold proteins). The performance of the classifier is characterized by a Matthews correlation coefficient of approximately 0.36 and true positive/true negative rates of 65% to 80%. The performance varies depending on which secondary structure prediction program is used, with SPIDER2[86,87] providing the overall highest accuracy in terms of accuracy and robustness. The paper also describes "misclassifications" as reflected in false positives (monomorphic proteins mis-classified as metamorphic) and false negatives (metamorphic proteins mis-classified as monomorphic), which revealed some limitations to the purely secondary structure-based approach, but also some irregularities in the input monomorphic and metamorphic data sets. In the near future, we hope that improved sequence-based computational methods will facilitate discovery and experimental characterization of new metamorphic proteins.

3.2 | Knowledge-based protein structure modeling

This widely used class of computational methods uses experimental knowledge of protein structures to predict the 3D structures of proteins, providing a means to fill experimental knowledge gaps or design entirely new proteins from scratch. Examples of these methods now widely used in protein design applications include Rosetta, SWISS-MODEL, I-TASSER, RaptorX, AlphaFold and so on.[83,88–92] A commonly employed strategy of these methods is to construct a model structure for an input sequence by joining together 3D structural "fragments" (each <10 amino acids in length) derived from experimental databases. Multiple structural models are generated algorithmically, and a physically motivated energy model can be used to optimize and rank the structures. The 3D structure of a template sequence can also be provided to guide the modeling.[93] Because protein modeling methods always generate multiple candidate structures, they could be used to predict alternate folds of potentially metamorphic proteins. Rosetta has recently been used to design a pair of protein sequences that are very similar (over 94% identity) but fold into distinct 3D structures, a short and thick helical bundle in one case, and a long and thin bundle in the other.[94]

A prerequisite for protein modeling software to effectively identify metamorphic proteins is the ability to predict the existing folds of known metamorphic proteins. This is a more difficult requirement than simply predicting whether metamorphism exists in a sequence, because a full 3D model involves predicting much more than a binary classification. The use of template structures in structure prediction is both an advantage and a drawback, because the "correct" template could guide the software to correctly predicting the native or alternate fold, whereas other templates could lead to incorrect structural predictions or wasted computational effort. The sequence-based methods in the previous section could be used together with 3D structure prediction methods described here by first identifying which proteins are most likely to be metamorphic, then using sequence alignment to search for suitable template structures to guide the 3-D modeling. The closeness in the scores/energy models of the alternate folds may be a useful predictor of whether both structures are thermodynamically accessible, with or without an environmental trigger.

3.3 | Physics-based molecular simulation methods

In contrast to knowledge-based approaches, physics-based simulation methods employ physical principles to make predictions of protein structure and behavior. Classical MD is perhaps the most extensively used physics-based computer simulation method for proteins, and is characterized by its use of empirical potential energy functions, also called "force fields." Force fields are used to evaluate the potential energy and classical forces on the atoms of an input protein conformation, and the atomic positions are propagated forward in time using Newton's equations of motion. The experimental knowledge in these simulations is mainly contained in the empirical parameters of the force field, which include terms that describe, for example, the strength of intermolecular interactions between aminoacyl side chains, and the energy barriers associated with torsion angles along the protein backbone. MD is well-known for its ability to simulate the folding processes of small proteins starting from an unfolded polypeptide chain and ending at a structure approximating the experimentally

determined conformation, without being guided by explicit knowledge of the experimental structure.[95–97] MD is increasingly being used to simulate the pathways of conformational change, such as in G-protein coupled receptors and Src kinases, both important targets for drug development.[98–100]

In principle, it is possible to observe the pathways of protein metamorphosis simply by observing a sufficiently long MD simulation trajectory (Figure 4). However, this is prohibitive in terms of computational cost as modern computers are only capable of simulating a few μs of physical time per day of computer time. A specialized supercomputer designed for MD is capable of achieving a few milliseconds of physical time,[101,102] which is still insufficient if the goal is to observe a rare event that occurs on timescales of seconds or more. For example, the two states of IscU typically interconvert on millisecond timescales,[103] Mad2 folded states interconvert slowly with a lifetime of 9.4 hours, and XCL1 interconversion time scale is approximately 1 second under physiologic conditions.[67] In order to overcome limitations of time scale, MD simulations of protein metamorphosis require the researcher to provide mechanistic hypotheses that effectively bias the simulation trajectory in the desired direction of fold switching, while still allowing sufficient freedom for the underlying force field and equations of motion to produce a meaningful prediction. Accelerated sampling methods such as targeted molecular dynamics, metadynamics, milestoning, and replica exchange are all approaches that could speed up the conformational space being explored by MD simulations.[104–106] For example, a short 25 ns implicit solvent replica exchange MD (REMD) simulation of the helix-turn-helix motif in RfaH that exhibits metamorphosis showed an α-to-β conversion in all 16 replicas, with one replica even attaining the conformation closest to the experimentally observed RfaH-CTD β-structure.[107] All of these methods require defining a progress variable or method of applying the bias, which could affect the pathways that are found; therefore, it is important to verify the theoretical predictions with experiments, which could involve using EPR, NMR, or FRET probes to measure changes in inter-residue distances, or mutagenesis to stabilize intermediates along the folding pathway.

3.4 | Evolution-based methods

Meyerguz *et al*. developed a computational network model of protein structure space, showing how point mutations are able to cause protein sequences to switch from one 3D fold to another.[108] Sfriso *et al*. predicted alternate conformers of proteins using a co-evolutionary approach that identifies residue pairs in an experimental structure that are correlated in their evolution (based on multiple sequence alignments) but are not in close contact.[109] These residue pairs are then brought close together using MD simulations with an added energy restraint, resulting in prediction of the alternate fold. These examples suggest that evolution-based approaches could be effective for identifying potentially metamorphic proteins.

3.5 | Hybrid experiment/simulation approaches

In recent years, there have been many growing opportunities to combine computational and experimental approaches for predicting conformational dynamics of macromolecular disordered protein structures at high resolution.[110–113] Some of the successful applications of this approach has been made by Loquet *et al*.[114] who used an integrated combination of solid-state NMR, electron microscopy and Rosetta modeling methods to solve the structure of type III secretion system needle, and Lasker *et al*.[115] in the same year used restraints from EM, X-ray crystallography, chemical crosslinking, and

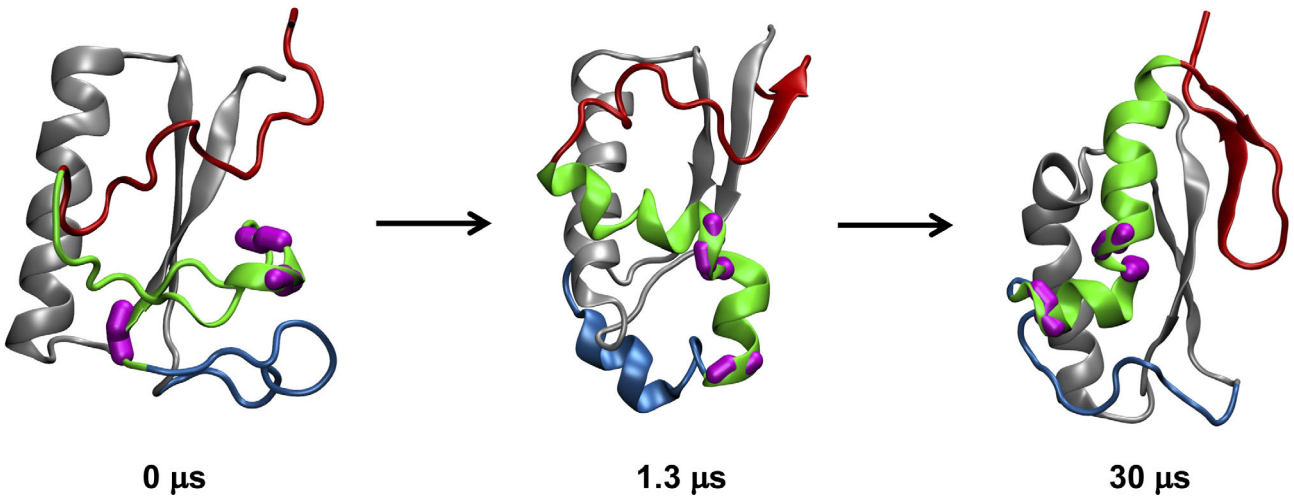


FIGURE 4 A plausible fold-switching pathway of KaiB metamorphosis as revealed by MD simulation. The starting initial structure is a fsKaiB analog with P63, P70, and P72 forced in trans conformation along with a denatured C-terminal helix (red) and β-hairpin (green). After 1.3 μs, an intermediate structure is formed where the C-terminal helix attains a β-sheet conformation (red) and the β-hairpin is converted into an α-helix (green). Finally, after 30 μs a structure very similar to gsKaiB, is obtained, which is stable for more than 10 μs

proteomics to build a model of 26S proteasome. Hybrid approaches were used to solve the structures of the yeast spindle pole body,[116] nuclear pore complex,[117] and in identifying two structural states of a multidomain transmembrane protein, PhoQ, by integrating MD simulations with experimental cross-linking data.[118,119] Although molecular dynamics is a physics-based method, there are many ways to integrate experimental data to guide the simulations. By way of context, the determination of experimental NMR structures is actually a molecular dynamics-based method, with a large number of energy restraints for distances, angles, and orientations derived from NMR measurements of various types.[120–122] Structural studies of intrinsically disordered proteins are also known to often require a combination of experimental data input from NMR and small-angle X-ray scattering (SAXS) which is then quantitatively interpreted by construction of conformational ensembles using restrained MD simulations or Bayesian statistics.[123,124]

In the near future, hybrid methods incorporating both experiment and simulation could be highly useful for modeling the alternative folds of metamorphic protein candidates. Broadly speaking, these could take the form of simulations that start from the native structure and incorporate experimental data to bias the ensemble toward alternate folds. For example, the MELD software[123] treats physics-based MD simulations as a prior probability distribution, then adds experimental data as a bias to steer the simulations toward physically realistic structures that agree with experimental data. This approach is flexible with respect to the type of experimental data, and the authors showed its ability to build models from diverse types of data including solid-state NMR measurements, spin-label EPR measurements, and/or evolutionarily inferred contacts. A future experiment for modeling the alternate fold of a metamorphic protein, based purely on experimental data that measures only shifts in CD or Fourier-transform infrared spectroscopy (FTIR) spectra, could involve this approach to add a biasing potential function designed to reproduce the observed spectral shift. The biasing potential could drive the native structure toward the experimentally observed changes in secondary structure content or other observables, thereby building a plausible 3D model of the alternate fold, and perhaps even generating some insight into the fold switching mechanism.

4 | CONCLUSION

In 1984, Kabsch and Sander made an observation that aminoacyl stretches five residues in length can adopt different secondary structures in different proteins.[125] Today, there are more dramatic examples as demonstrated by metamorphic proteins, which defy the classical "one sequence-one fold" paradigm, although how common they are await the development of innovative approaches to discover them. The better we can identify and characterize metamorphic proteins, the better we will understand the rules governing protein folding (the holy grail of biochemistry), and the role of metamorphic proteins in the evolutionary adaptation of viruses, archaea, bacteria, and eukaryotes. To increase our rate of progress, which is rather slow at the moment, powerful new approaches will need to be developed through interdisciplinary collaborations that integrate knowledge and expertise across the biological and physical sciences. As we described earlier in this perspective, computational methods based on secondary structure prediction can be used to screen the vast sequence space for likely fold switchers. Computational protein design tools can be used to predict the alternate fold starting from candidate fold switching sequences and template structures. With the increasing availability of high-throughput experimental methods, they can be applied to measure how secondary (and higher) structure responds to changes in environmental conditions for batches of proteins, thereby increasing the capacity to verify the fold switching candidates identified computationally. After the initial experimental detection of metamorphism, a variety of more targeted experimental and simulation methods and hybrid methods could then be used to solve the alternate fold and investigate the fold switching mechanism. One way to stimulate the development of new computational methodologies could be for the CASP competition to assess protein structure prediction methods, or other "blind challenge" type competitions, to use metamorphic proteins as targets; this could motivate future computational methods to incorporate uncertainty or multiplicity in the final result as a design goal, rather than as a byproduct needing to be minimized. New knowledge on protein metamorphosis will have translational potential by facilitating for example the design of customized biosensors that switch according to specific changes in the local environment.

ACKNOWLEDGEMENTS

We thank Supratim Dey, Archana Chavan, Joel Heisler, and Ning Zhang for valuable discussions. This work was supported by the U.S. Army Research Office Award W911NF-17-1-0434.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Lee-Ping Wang https://orcid.org/0000-0003-3072-9946

REFERENCES

[1] C. B. Anfinsen, E. Haber, *J Biol Chem* **1961**, *236*, 1361.
[2] C. B. Anfinsen, E. Haber, M. Sela, F. H. White, *Proc. Natl. Acad. Sci. U. S. A.* **1961**, *47*, 1309.
[3] G. D. Rose, T. P. Creamer, *Proteins* **1994**, *19*, 1.
[4] S. Dalal, S. Balasubramanian, L. Regan, *Nat. Struct. Biol.* **1997**, *4*, 548.
[5] E. S. Kuloğlu, D. R. McCaslin, J. L. Markley, B. F. Volkman, *J Biol Chem* **2002**, *277*, 17863.
[6] A. G. Murzin, *Science* **2008**, *320*, 1725.
[7] A. F. Dishman, B. F. Volkman, *ACS Chem. Biol.* **2018**, *13*, 1438.
[8] R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, P. E. Wright, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 11504.
[9] E. K. Jaffe, *Trends Biochem. Sci.* **2005**, *30*, 490.

[10] J. Piatigorsky, W. E. O'Brien, B. L. Norman, K. Kalumuck, G. J. Wistow, T. Borras, J. M. Nickerson, E. F. Wawrousek, *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 3479.
[11] C. J. Jeffery, *Trends Biochem. Sci.* **1999**, *24*, 8.
[12] R. Zahn, A. Liu, T. Lührs, R. Riek, C. von Schroetter, F. López García, M. Billeter, L. Calzolai, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 145.
[13] R. Tycko, R. Savtchenko, V. G. Ostapchenko, N. Makarava, I. V. Baskakov, *Biochemistry* **2010**, *49*, 9488.
[14] C. P. J. Maury, *J Intern Med* **2009**, *265*, 329.
[15] E. S. Kuloğlu, D. R. McCaslin, M. Kitabwalla, C. D. Pauza, J. L. Markley, B. F. Volkman, *Biochemistry* **2001**, *40*, 12486.
[16] R. L. Tuinstra, F. C. Peterson, S. Kutlesa, E. S. Elgin, M. A. Kron, B. F. Volkman, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5057.
[17] M. Ishiura, S. Kutsuna, S. Aoki, H. Iwasaki, C. R. Andersson, A. Tanabe, S. S. Golden, C. H. Johnson, T. Kondo, *Science* **1998**, *281*, 1519.
[18] Y.-G. Chang, S. E. Cohen, C. Phong, W. K. Myers, Y.-I. Kim, R. Tseng, J. Lin, L. Zhang, J. S. Boyd, Y. Lee, S. Kang, D. Lee, S. Li, R. D. Britt, M. J. Rust, S. S. Golden, A. LiWang, *Science* **2015**, *349*, 324.
[19] R. Tseng, N. F. Goularte, A. Chavan, J. Luu, S. E. Cohen, Y.-G. Chang, J. Heisler, S. Li, A. K. Michael, S. Tripathi, S. S. Golden, A. LiWang, C. L. Partch, *Science* **2017**, *355*, 1174.
[20] A. J. DeGrave, J.-H. Ha, S. N. Loh, L. T. Chong, *Nat. Commun.* **2018**, *9*, 1013.
[21] J.-H. Ha, S. N. Loh, *Chem. – Eur. J.* **2012**, *18*, 7984.
[22] M. M. Stratton, S. N. Loh, *Protein Sci.* **2011**, *20*, 19.
[23] P. K. Zuber, I. Artsimovitch, M. NandyMazumdar, Z. Liu, Y. Nedialkov, K. Schweimer, P. Rösch, S. H. Knauer, *elife* **2018**, *7*, e36349.
[24] L. L. Porter, L. L. Looger, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 5968.
[25] N. Chen, M. Das, A. LiWang, L.-P. Wang, *Biophys. J.* **2020**, *119*, 1380.
[26] J. L. Markley, J. H. Kim, Z. Dai, J. R. Bothe, K. Cai, R. O. Frederick, M. Tonelli, *FEBS Lett.* **2013**, *587*, 1172.
[27] X. Luo, Z. Tang, G. Xia, K. Wassmann, T. Matsumoto, J. Rizo, H. Yu, *Nat. Struct. Mol. Biol.* **2004**, *11*, 338.
[28] H. Choi, S. Kim, P. Mukhopadhyay, S. Cho, J. Woo, G. Storz, S. E. Ryu, *Cell* **2001**, *105*, 103.
[29] K. B. Dagbay, N. Bolik-Coulon, S. N. Savinov, J. A. Hardy, *J Biol Chem* **2017**, *292*, 4885.
[30] D. R. Littler, S. J. Harrop, W. D. Fairlie, L. J. Brown, G. J. Pankhurst, S. Pankhurst, M. Z. DeMaere, T. J. Campbell, A. R. Bauskin, R. Tonini, M. Mazzanti, S. N. Breit, P. M. G. Curmi, *J Biol Chem* **2004**, *279*, 9298.
[31] B. M. Burmann, S. H. Knauer, A. Sevostyanova, K. Schweimer, R. A. Mooney, R. Landick, I. Artsimovitch, P. Rösch, *Cell* **2012**, *150*, 291.
[32] M. Mueller, U. Grauschopf, T. Maier, R. Glockshuber, N. Ban, *Nature* **2009**, *459*, 726.
[33] P. A. Bullough, F. M. Hughson, J. J. Skehel, D. C. Wiley, *Nature* **1994**, *371*, 37.
[34] M. J. Cuneo, R. E. London, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 6805.
[35] M. R. Gryk, A. Marintchev, M. W. Maciejewski, A. Robertson, S. H. Wilson, G. P. Mullen, *Structure* **2002**, *10*, 1709.
[36] J. M. Wagner, K. K. Zadrozny, J. Chrustowicz, M. D. Purdy, M. Yeager, B. K. Ganser-Pornillos, O. Pornillos, *elife* **2016**, *5*, e17063.
[37] J. L. Newman, E. W. Butcher, D. T. Patel, Y. Mikhaylenko, M. F. Summers, *Protein Sci.* **2004**, *13*, 2101.
[38] X. Zheng, L. C. Pedersen, S. A. Gabel, G. A. Mueller, M. J. Cuneo, E. F. DeRose, J. M. Krahn, R. E. London, *Nucleic Acids Res.* **2014**, *42*, 5361.
[39] M. Xu, A. Arulandu, D. K. Struck, S. Swanson, J. C. Sacchettini, R. Young, *Science* **2005**, *307*, 113.
[40] H. Takala, A. Björling, O. Berntsson, H. Lehtivuori, S. Niebling, M. Hoernke, I. Kosheleva, R. Henning, A. Menzel, J. A. Ihalainen, S. Westenhoff, *Nature* **2014**, *509*, 245.
[41] A. le Maire, C. Teyssier, C. Erb, M. Grimaldi, S. Alvarez, A. R. de Lera, P. Balaguer, H. Gronemeyer, C. A. Royer, P. Germain, W. Bourguet, *Nat. Struct. Mol. Biol.* **2010**, *17*, 801.
[42] M. López-Pelegrín, N. Cerdà-Costa, A. Cintas-Pedrola, F. Herranz-Trillo, P. Bernadó, J. R. Peinado, J. L. Arolas, F. X. Gomis-Rüth, *Angew. Chem., Int. Ed.* **2014**, *53*, 10624.
[43] H. Nar, M. Bauer, J. M. Stassen, D. Lang, A. Gils, P. J. Declerck, *J. Mol. Biol.* **2000**, *297*, 683.
[44] T. H. Tahirov, D. Temiakov, M. Anikin, V. Patlan, W. T. McAllister, D. G. Vassylyev, S. Yokoyama, *Nature* **2002**, *420*, 43.
[45] G. I. van Boxel, S. Holmes, L. Fugger, E. Y. Jones, *J. Mol. Biol.* **2010**, *400*, 828.
[46] P. Kulkarni, T. L. Solomon, Y. He, Y. Chen, P. N. Bryan, J. Orban, *PROTEINSCIENCE* **2018**, *27*, 1557.
[47] K. Hitomi, T. Oyama, S. Han, A. S. Arvai, E. D. Getzoff, *J Biol Chem* **2005**, *280*, 19127.
[48] R. Iwase, K. Imada, F. Hayashi, T. Uzumaki, M. Morishita, K. Onai, Y. Furukawa, K. Namba, M. Ishiura, *J Biol Chem* **2005**, *280*, 43141.
[49] G. A. Belogurov, M. N. Vassylyeva, V. Svetlov, S. Klyuyev, N. V. Grishin, D. G. Vassylyev, I. Artsimovitch, *Mol. Cell* **2007**, *26*, 117.
[50] Y.-G. Chang, R. Tseng, N.-W. Kuo, A. LiWang, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16847.
[51] S. Spera, A. Bax, *J. Am. Chem. Soc.* **1991**, *113*, 5490.
[52] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213.
[53] D. S. Wishart, B. D. Sykes, F. M. Richards, *Biochemistry* **1992**, *31*, 1647.
[54] G. M. Clore, *Protein Sci.* **2011**, *20*, 229.
[55] A. Bax, G. M. Clore, *J. Magn. Reson.* **2019**, *306*, 187.
[56] A. J. Baldwin, L. E. Kay, *Nat. Chem. Biol.* **2009**, *5*, 808.
[57] A. Sekhar, L. E. Kay, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 12867.
[58] P. K. Zuber, K. Schweimer, P. Rösch, I. Artsimovitch, S. H. Knauer, *Nat. Commun.* **2019**, *10*, 702.
[59] M. Wu, G. C. Lander, *Biophys. J.* **2020**, *119*, 1281.
[60] T. A. M. Bharat, C. J. Russo, J. Löwe, L. A. Passmore, S. H. W. Scheres, *Structure* **2015**, *23*, 1743.
[61] E. Callaway, *Nature* **2015**, *525*, 172.
[62] E. Callaway, *Nature* **2020**, *578*, 201.
[63] W. Yassine, N. Taib, S. Federman, A. Milochau, S. Castano, W. Sbi, C. Manigand, M. Laguerre, B. Desbat, R. Oda, J. Lang, *Biochim. Biophys. Acta* **2009**, *1788*, 1722.
[64] J. L. Lippert, D. Tyminski, P. J. Desmeules, *J. Am. Chem. Soc.* **1976**, *98*, 7075.
[65] N. C. Maiti, M. M. Apetri, M. G. Zagorski, P. R. Carey, V. E. Anderson, *J. Am. Chem. Soc.* **2004**, *126*, 2399.
[66] Y. Zhao, L. Li, C. Wu, X. Jiang, B. Ge, H. Ren, F. Huang, *Protein Eng., Des. Sel.* **2016**, *29*, 23.
[67] R. C. Tyler, N. J. Murray, F. C. Peterson, B. F. Volkman, *Biochemistry* **2011**, *50*, 7077.
[68] Y. Gambin, A. Schug, E. A. Lemke, J. J. Lavinder, A. C. M. Ferreon, T. J. Magliery, J. N. Onuchic, A. A. Deniz, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 10153.
[69] D. Dell'Orco, K.-W. Koch, *ACS Chem. Biol.* **2016**, *11*, 2390.
[70] B. Jia, C. O. Jeon, *Open Biol* **2016**, *6*, 160196.
[71] R. Vincentelli, A. Cimino, A. Geerlof, A. Kubo, Y. Satou, C. Cambillau, *Methods* **2011**, *55*, 65.
[72] L. L. Liu, L. Wang, J. Zonderman, J. C. Rouse, H.-Y. Kim, *J. Pharm. Sci.* **2020**, *109*, 3223.
[73] J. De Meutter, E. Goormaghtigh, *Anal. Chem.* **2021**, *93*, 3733.
[74] R. Beveridge, Q. Chappuis, C. Macphee, P. Barran, *Analyst* **2013**, *138*, 32.
[75] D. Stuchfield, A. P. France, L. G. Migas, A. Thalhammer, A. Bremer, B. Bellina, P. E. Barran, *Meth Enzymol* **2018**, *611*, 459.
[76] I. Oganesyan, C. Lento, D. J. Wilson, *Methods* **2018**, *144*, 27.
[77] S. H. Stoychev, C. Nathaniel, S. Fanucchi, M. Brock, S. Li, K. Asmus, V. L. Woods, H. W. Dirr, *Biochemistry* **2009**, *48*, 8413.

[78] P. Galaz-Davison, J. A. Molina, S. Silletti, E. A. Komives, S. H. Knauer, I. Artsimovitch, C. A. Ramírez-Sarmiento, *Biophys. J.* **2020**, *118*, 96.
[79] S. Schopper, A. Kahraman, P. Leuenberger, Y. Feng, I. Piazza, O. Müller, P. J. Boersema, P. Picotti, *Nat. Protoc.* **2017**, *12*, 2391.
[80] D. T. Jones, *J. Mol. Biol.* **1999**, *292*, 195.
[81] Y. Yang, J. Gao, J. Wang, R. Heffernan, J. Hanson, K. Paliwal, Y. Zhou, *Brief Bioinformatics*, Vol. 19, Oxford University Press, Oxford, England, UK **2016**, p. 482.
[82] W. Wardah, M. G. M. Khan, A. Sharma, M. A. Rashid, *Comput. Biol. Chem.* **2019**, *81*, 1.
[83] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583.
[84] A. K. Kim, L. L. Looger, L. L. Porter, *BioRxiv* **2020**, 956805.
[85] S. Mishra, L. L. Looger, L. L. Porter, *Protein Sci.* **2019**, *28*, 1487.
[86] R. Heffernan, K. Paliwal, J. Lyons, A. Dehzangi, A. Sharma, J. Wang, A. Sattar, Y. Yang, Y. Zhou, *Sci. Rep.* **2015**, *5*, 11476.
[87] Y. Yang, R. Heffernan, K. Paliwal, J. Lyons, A. Dehzangi, A. Sharma, J. Wang, A. Sattar, Y. Zhou, *Methods Mol. Biol.* **2017**, *1484*, 55.
[88] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, *Nat. Methods* **2015**, *12*, 7.
[89] K. W. Kaufmann, G. H. Lemmon, S. L. Deluca, J. H. Sheehan, J. Meiler, *Biochemistry* **2010**, *49*, 2987.
[90] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, *Nucleic Acids Res.* **2018**, *46*, W296.
[91] G. Xu, Q. Wang, J. Ma, *J. Chem. Theory Comput.* **2020**, *16*, 3970.
[92] J. Peng, J. Xu, *Proteins* **2011**, *79*, 161.
[93] A. Fiser, *Methods Mol. Biol.* **2010**, *673*, 73.
[94] V. K. Kumirov, E. M. Dykstra, B. M. Hall, W. J. Anderson, T. N. Szyszka, M. H. J. Cordes, *Protein Sci.* **2018**, *27*, 1767.
[95] D. Shukla, C. X. Hernández, J. K. Weber, V. S. Pande, *Acc. Chem. Res.* **2015**, *48*, 414.
[96] G. R. Bowman, V. A. Voelz, V. S. Pande, *Curr. Opin. Struct. Biol.* **2011**, *21*, 4.
[97] V. A. Voelz, G. R. Bowman, K. Beauchamp, V. S. Pande, *J. Am. Chem. Soc.* **2010**, *132*, 1526.
[98] K. J. Kohlhoff, D. Shukla, M. Lawrenz, G. R. Bowman, D. E. Konerding, D. Belov, R. B. Altman, V. S. Pande, *Nat. Chem.* **2014**, *6*, 15.
[99] B. Yi, A. Jahangir, A. K. Evans, D. Briggs, K. Ravina, J. Ernest, A. B. Farimani, W. Sun, J. Rajadas, M. Green, E. N. Feinberg, V. S. Pande, M. Shamloo, *PLoS One* **2017**, *12*, e0180319.
[100] S. Singh, X. Sun, K. J. Blumer, G. Bowman, *Biophys J* **2020**, *118*, 320a.
[101] D. E. Shaw, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, R. O. Dror, S. Piana, Y. Shan, B. Towles, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, Bank, J. A, C. Young, B. Batson, *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09*, ACM Press, New York, New York, USA **2009**, p. 1.
[102] D. E. Shaw, J. P. Grossman, Bank, J. A, B. Batson, J. A. Butts, J. C. Chao, M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp, C. R. Ho, D. J. Ierardi, L. Iserovich, J. S. Kuskin, R. H. Larson, T. Layman, C. Young, *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* **2014**, 41.
[103] J. H. Kim, A. K. Füzéry, M. Tonelli, D. T. Ta, W. M. Westler, L. E. Vickery, J. L. Markley, *Biochemistry* **2009**, *48*, 6062.
[104] R. Elber, *Annu. Rev. Biophys.* **2020**, *49*, 69.
[105] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141.
[106] A. Barducci, M. Bonomi, M. Parrinello, *WIREs Comput. Mol. Sci.* **2011**, *1*, 826.
[107] J. B. Gc, Y. R. Bhandari, B. S. Gerstman, P. P. Chapagain, *J Phys Chem B* **2014**, *118*, 5101.
[108] L. Meyerguz, J. Kleinberg, R. Elber, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11627.
[109] P. Sfriso, M. Duran-Frigola, R. Mosca, A. Emperador, P. Aloy, M. Orozco, *Structure* **2016**, *24*, 116.
[110] A. B. Ward, A. Sali, I. A. Wilson, *Science* **2013**, *339*, 913.
[111] H. van den Bedem, J. S. Fraser, *Nat. Methods* **2015**, *12*, 307.
[112] M. P. Rout, A. Sali, *Cell* **2019**, *177*, 1384.
[113] A. Srivastava, S. P. Tiwari, O. Miyashita, F. Tama, *J. Mol. Biol.* **2020**, *432*, 2846.
[114] A. Loquet, N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker, A. Lange, *Nature* **2012**, *486*, 276.
[115] K. Lasker, F. Förster, S. Bohn, T. Walzthoeni, E. Villa, P. Unverdorben, F. Beck, R. Aebersold, A. Sali, W. Baumeister, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1380.
[116] S. Viswanath, M. Bonomi, S. J. Kim, V. A. Klenchin, K. C. Taylor, K. C. Yabut, N. T. Umbreit, H. A. Van Epps, J. Meehl, M. H. Jones, D. Russel, J. A. Velazquez-Muriel, M. Winey, I. Rayment, T. N. Davis, A. Sali, E. G. Muller, *Mol. Biol. Cell* **2017**, *28*, 3298.
[117] S. J. Kim, J. Fernandez-Martinez, I. Nudelman, Y. Shi, W. Zhang, B. Raveh, T. Herricks, B. D. Slaughter, J. A. Hogan, P. Upla, I. E. Chemmama, R. Pellarin, I. Echeverria, M. Shivaraju, A. S. Chaudhury, J. Wang, R. Williams, J. R. Unruh, C. H. Greenberg, E. Y. Jacobs, M. P. Rout, *Nature* **2018**, *555*, 475.
[118] T. Lemmin, C. S. Soto, G. Clinthorne, W. F. DeGrado, M. Dal Peraro, *PLoS Comput. Biol.* **2013**, *9*, e1002878.
[119] K. S. Molnar, M. Bonomi, R. Pellarin, G. D. Clinthorne, G. Gonzalez, S. D. Goldberg, M. Goulian, A. Sali, W. F. DeGrado, *Structure* **2014**, *22*, 1239.
[120] P. Robustelli, K. Kohlhoff, A. Cavalli, M. Vendruscolo, *Structure* **2010**, *18*, 923.
[121] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *J. Magn. Reson.* **2003**, *160*, 65.
[122] C. D. Schwieters, G. A. Bermejo, G. M. Clore, *Protein Sci.* **2018**, *27*, 26.
[123] J. L. MacCallum, A. Perez, K. A. Dill, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 6985.
[124] T. N. Cordeiro, F. Herranz-Trillo, A. Urbanek, A. Estaña, J. Cortés, N. Sibille, P. Bernadó, *Curr. Opin. Struct. Biol.* **2017**, *42*, 15.
[125] W. Kabsch, C. Sander, *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 1075.

AUTHOR BIOGRAPHIES

Madhurima Das is a postdoctoral fellow at the University of California, Merced (USA). After receiving her PhD in Life Sciences (2012) from the Bose Institute (India), she joined the Bioinformatics Centre, University of Kansas (USA) as a postdoctoral fellow. In 2019, she joined the Andy LiWang group to study the metamorphic behavior of a cyanobacterial clock protein, KaiB, using biophysical techniques. Her research interests are directed at bridging experiments and

computation to provide structural and mechanistic insights into the functioning of metamorphic proteins.

ANDY LIWANG is a Professor in the Department of Chemistry & Biochemistry at the University of California, Merced (USA). His interest in metamorphic proteins was sparked when his lab serendipitously discovered that the circadian clock protein, KaiB, is metamorphic – adopting one fold needed during the day and a distinctly different fold needed at night. Today, Andy is interested in the extent to which the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank is populated with metamorphic proteins masquerading as monomorphic proteins.

LEE-PING WANG is an Associate Professor in the Department of Chemistry at the University of California, Davis (USA). His research in theoretical and computational chemistry includes developing first-principles methods for exploring reaction mechanisms, and improving force fields for simulating proteins and other biomolecular systems. He is also interested in how knowledge-based approaches in structural biology could be developed to uncover previously unknown metamorphic proteins.

How to cite this article: M. Das, N. Chen, A. LiWang, L.-P. Wang, *Biopolymers* **2021**, *112*(10), e23473. https://doi.org/10.1002/bip.23473