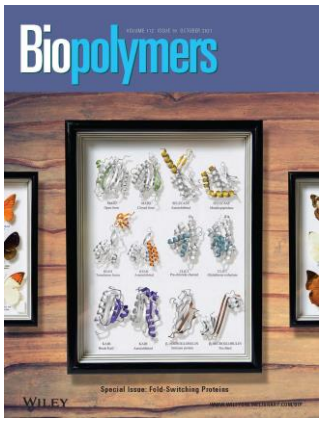
Biopolymers

Special Issue: Fold-Switching Proteins

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



EDITORIAL

REVIEW

ARTICLES

DOI: 10.1002/bip.23478

EDITORIAL

Biopolymers WILEY

Fold-switching proteins

1 | WHERE THE FIELD OF FOLD-SWITCHING PROTEINS STANDS TODAY

This special issue of *Biopolymers* is dedicated to the curious proteins classified as fold-switching. We have long been fascinated by how the wondrous structures of proteins are encoded within their amino acid sequences. Indeed, a Holy Grail of biochemistry is to be able to deduce the fold of a protein from sequence alone. Just witness the headline-grabbing breakthrough of AlphaFold2's performance in the recently completed critical assessment of protein structure prediction competition.[1]

In the early years of structural biology, our understanding of the relationship between protein sequence, structure, and function blossomed as X-ray crystallography provided unprecedented insights.[2] It became common knowledge soon thereafter that proteins, even those with well-defined folds, are not static and undergo changes in structure that often involve repositioning motifs and domains. Even so, the overall three-dimensional arrangement of secondary structures (i.e., architecture) and the path taken by the polypeptide chain through the structure (i.e., topology),[3] collectively defined as the fold of a protein, remains more or less the same. This one-sequence one-fold paradigm is largely attributed to Christian Anfinsen who was awarded the 1972 Nobel Prize in Chemistry for his work "concerning the connection between amino acid sequence and the biologically active conformation".[4] As is often the case with most aspects of the natural world though, the more we understood about proteins the more we realized their complexity. For example, it became apparent that a short sequence of aminoacyl residues that forms, say, an α-helix in one protein can form a β-strand in the context of a different protein.[5] Later it was discovered that an entire polypeptide chain can adopt more than one fold. Dramatically, prion proteins can switch folds where the fold with the higher content of β-strands aggregates irreversibly with devastating consequences.[6]

However, the discovery of a protein that can *reversibly* switch between two different folds was not reported until 2002.[7] Proteins with this type of fold-switching behavior are now classified as metamorphic[8] to distinguish them from single-fold (i.e., monomorphic) proteins. Metamorphic proteins can be thought of as providing a two-for-one deal because they switch reversibly between distinctly different folds and functions. They challenge our understanding of the rules governing protein structure and as such move the Holy Grail goalposts. In response to this new challenge, there is interest to estimate how many metamorphic proteins are masquerading in the RCSB Protein Data Bank as monomorphic. Carrying out such an estimation is formidable, however. Indeed, thus far metamorphic behavior has only been discovered serendipitously, explaining why just 20 proteins have been experimentally verified as metamorphic. Directed methods and tools to predict and validate reversible fold switching in proteins are now beginning to emerge and have much room to grow. Accordingly, as the field matures we expect the size of the metamorphome, and our understanding of protein structure and function, to blossom as well.

2 | IN THIS ISSUE

The papers in this issue present original research on identifying and characterizing fold switching proteins using both experimental and computational approaches, as well as giving an overview of the field as it currently stands.

The issue begins with Das and coworkers providing a perspective on the currently available experimental and computational methods for identifying and characterizing metamorphic proteins. They note (as mentioned) the metamorphic proteins identified thus far is largely a result of serendipitous discovery rather than a targeted search. They also highlight that the development of new experimental and computational tools is needed to identify additional metamorphic proteins and the environmental conditions that induce fold switching. The piece describes the known metamorphic proteins and their properties, as well as providing suggestions for the development of novel methods that could populate the "metamorphome" in the next few years.

The first primary research paper in this issue studies XCL1, a small cell signaling protein also known as lymphotactin. XCL1 is a metamorphic protein that has a chemokine fold that binds to a G-coupled protein receptor, and a dimeric all-beta fold that has antimicrobial activity. It has long been known that XCL1's metamorphic equilibrium can be shifted by various non-specific stimuli, but a specific means of controlling the equilibrium had not been found. In the issue, Dishman et al. show that the metamorphic equilibrium of XCL1 can be controlled by titration with an engineered XCL1 variant called CC5.[9] CC5 incorporates an artificial disulfide bond that stabilizes the all-beta fold of the monomer, shifting the XCL1 metamorphic equilibrium by the formation of CC5-XCL1 heterodimers. The work provides a new tool for studying protein metamorphosis and paves the way for therapeutic targeting of such proteins with roles in human health and disease.

Next, Scheraga and Rackovsky report a method to rapidly profile the dynamical behavior of proteins.[10] Starting from a large

https://doi.org/10.1002/bip.23478

experimental dataset of crystallographic structures and associated B-factors, they derive a set of averaged B-factors for each amino acid. For a given protein sequence, the sequence of amino acid-specific averaged B-factors (<B(X)>) is a profile of the dynamical properties of the protein, with higher values indicating greater mobility. Here, the authors show that their bioinformatic profiling approach to protein dynamics could provide insight into fold switching by demonstrating that among a class of GA/GB proteins, the average of <B(X)> over the protein sequence is highest for the fold-switching pair that differs by only one amino acid. Moreover, they show that differences in dynamical profiles (i.e., <B(X)>) between two sequences is correlated with differences in their 3D structure; the nature of this correlation varies across protein structural archetypes.

The C-terminal domain (CTD) of transcription factor RfaH adopts an α-helical hairpin structure as part of the RfaH protein, but switches folds to a β-barrel when expressed as a separate sequence. Seifi and Wallin present the results of computer simulations investigating this all-α to all-β fold switching transition. Their simulations used a model that combines physics-based energy functions with biasing potentials that favor the experimentally known folded structure.[11] They find the all-α structure is not a free energy minimum for RfaH-CTD in isolation, and that the unfolded protein contains an unusually high α-helical content. The authors propose that circular dichroism spectroscopy experiments could be used to test their predictions.

The final experimental work included in the issue concerns the human protein BAX. It regulates mitochondrial outer membrane permeabilization (MOMP), a process involved in apoptosis. Although native BAX assumes an all α-helical fold, previous work has shown that the natural peptide humanin sequesters BAX into β-sheet fibrils, preventing MOMP.[9] As with other known fold switchers, BAX fibrillation is environmentally sensitive. Here, Morris and Tjandra characterize the effects of different environmental factors, including temperature, pH and the presence of non-ionic detergent *n*-octyl-β-D-glucoside, on the fibrillation of BAX and its homolog BID, which also forms fibrils in the presence of humanin.

The issue ends with promising computational approaches to identify fold switching proteins. Sequence-similar fold switchers are pairs of sequences with distinct folds and differ by only a small number of amino acids in their sequence. Taking advantage of secondary structure (SS) prediction methods[10–12] Kim et al. identify sequence-similar fold switchers by generating distinct SS predictions for both input sequences. Using bioinformatics methods, they compiled a dataset of 19 sequence-similar fold switchers and 207 single-fold protein families. The JPred4 method was found to be the most effective discriminator between fold-switching and single-folding sequence pairs. Following on, in the final paper Mishra et al. use Jpred4 to identify fold switching regions (FSRs) within protein sequences by predicting different secondary structures for the FSR in isolation versus in the context of the full protein sequence. They applied the Jpred4 method to predict α-helix ↔ β-strand prediction discrepancies in a dataset of experimentally known fold switching proteins. Notably these discrepancies are very rare when the same method is applied to randomly selected regions of single-fold proteins.

3 | FUTURE DIRECTIONS FOR RESEARCH ON FOLD-SWITCHING PROTEINS

Previous work, along with articles published in this issue, suggest that fold-switching proteins regulate diverse biological processes[13] relevant to human diseases such as COVID-19,[14] cancer,[15] and autoimmune dysfunction.[16] Thus discovering more fold switchers could reveal novel regulatory processes associated with other human diseases. As stated previously, all known metamorphic proteins have been discovered by chance.[17] Moving forward, the field needs systematic methods, likely integrating computation and experiment, to uncover more.

Recently published computational methods for predicting fold switchers,[10,12] as well as those in this issue, suggest that at least some fold switchers can be identified from their sequences alone. These proofs-of-concept now require experimental validation. To that end, one method has been used to identify fold switchers in the universally conserved NusG family of transcriptional regulators.[18] Validation from circular dichroism and nuclear magnetic resonance (NMR) spectroscopy suggest that this approach accurately classified 10/10 of diverse sequences as fold switchers (6) and single folders (4). By contrast, current state-of-the-art methods, such as AlphaFold2,[19] did not correctly predict the ground-state helical hairpin fold in 5/6-fold-switching variants (the sixth was in the training set[18]). These findings indicate that accurate predictions of fold switchers are still in their infancy, and their biochemical features are yet to be discovered. Nevertheless, the success of existing predictive methods holds promise for further progress.

As computational methods advance, experimental methods to assay and characterize fold switchers will also require development. Currently, the mechanisms by which proteins switch folds are largely opaque, though some insights have been gained through NMR spectroscopy[20] and single-molecule experiments.[21] Cryo-EM is likely to elucidate more fold switchers as many of these proteins switch upon binding another macromolecule, such as the SARS-CoV-2 protein, Orf9b.[14] Additionally, because cryo-EM can be used to determine the structures of smaller and smaller proteins that may resist crystallization, it could reveal more fold-switching proteins. Given the rapid advances of both computational and experimental techniques, the future is bright for the field.

Andy LiWang[1]
Lauren L. Porter[2,3]
Lee-Ping Wang[4]

[1]*University of California, Merced, California, USA*
[2]*National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA*
[3]*National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA*
[4]*University of California, Davis, California, USA*

REFERENCES

[1] 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.

[2] M. Jaskolski, Z. Dauter, A. Wlodawer, *FEBS J.* **2014**, *281*, 3985.

[3] R. Kolodny, L. Pereyaslavets, A. O. Samson, M. Levitt, *Annu. Rev. Biophys.* **2013**, *42*, 559.

[4] The Nobel Prize in Chemistry, Christian Anfinsen, https://www.nobelprize.org/prizes/chemistry/1972/anfinsen/facts/ (accessed: 2021)

[5] W. Kabsch, C. Sander, *Proc. Natl. Acad. Sci. USA.* **1984**, *81*, 1075.

[6] S. B. Prusiner, *Science* **1997**, *278*, 245.

[7] E. S. Kuloğlu, D. R. McCaslin, J. L. Markley, B. F. Volkman, *J. Biol. Chem.* **2002**, *277*, 17863.

[8] A. G. Murzin, *Science.* **2008**, *320*, 1725.

[9] D. L. Morris, D. W. Kastner, S. Johnson, M.-P. Strub, Y. He, C. K. E. Bleck, D. Y. Lee, N. Tjandra, *J. Biol. Chem.* **2019**, *294*, 19055.

[10] S. Mishra, L. L. Looger, L. L. Porter, *Protein Sci.* **2019**, *28*, 1487.

[11] L. L. Porter, L. L. Looger, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 5968.

[12] N. Chen, M. Das, A. LiWang, L.-P. Wang, *Biophys. J.* **2020**, *119*, 1380.

[13] A. K. Kim, L. L. Porter, *Structure* **2021**, *29*, 6.

[14] D. E. Gordon, J. Hiatt, M. Bouhaddou, V. V. Rezelj, S. Ulferts, H. Braberg, D. E. Gordon, J. Hiatt, M. Bouhaddou, V. V. Rezelj, S. Ulferts, H. Braberg, A. S. Jureka, K. Obernier, J. Z. Guo, J. Batra, R. M. Kaake, A. R. Weckstein, T. W. Owens, M. Gupta, S. Pourmal, E. W. Titus, M. Cakir, M. Soucheray, M. McGregor, Z. Cakir, G. Jang, M. J. O'Meara, T. A. Tummino, Z. Zhang, H. Foussard, A. Rojc, Y. Zhou, D. Kuchenov, R. Hüttenhain, J. Xu, M. Eckhardt, D. L. Swaney, J. M. Fabius, M. Ummadi, B. Tutuncuoglu, U. Rathore, M. Modak, P. Haas, K. M. Haas, Z. Z. C. Naing, E. H. Pulido, Y. Shi, I. Barrio-Hernandez, D. Memon, E. Petsalaki, A. Dunham, M. C. Marrero, D. Burke, C. Koh, T. Vallet, J. A. Silvas, C. M. Azumaya, C. Billesbølle, A. F. Brilot, M. G. Campbell, A. Diallo, M. S. Dickinson, D. Diwanji, N. Herrera, N. Hoppe, H. T. Kratochvil, Y. Liu, G. E. Merz, M. Moritz, H. C. Nguyen, C. Nowotny, C. Puchades, A. N. Rizo, U. Schulze-Gahmen, A. M. Smith, M. Sun, I. D. Young, J. Zhao, D. Asarnow, J. Biel, A. Bowen, J. R. Braxton, J. Chen, C. M. Chio, U. S. Chio, I. Deshpande, L. Doan, B. Faust, S. Flores, M. Jin, K. Kim, V. L. Lam, F. Li, J. Li, Y.-L. Li, Y. Li, X. Liu, M. Lo, K. E. Lopez, A. A. Melo, F. R. Moss 3rd., P. Nguyen, J. Paulino, K. I. Pawar, J. K. Peters, T. H. Pospiech Jr.., M. Safari, S. Sangwan, K. Schaefer, P. V. Thomas, A. C. Thwin, R. Trenker, E. Tse, T. K. M. Tsui, F. Wang, N. Whitis, Z. Yu, K. Zhang, Y. Zhang, F. Zhou, D. Saltzberg, QCRG Structural Biology Consortium, A. J. Hodder, A. S. Shun-Shion, D. M. Williams, K. M. White, R. Rosales, T. Kehrer, L. Miorin, E. Moreno, A. H. Patel, S. Rihn, M. M. Khalid, A. Vallejo-Gracia, P. Fozouni, C. R. Simoneau, T. L. Roth, D. Wu, M. A. Karim, M. Ghoussaini, I. Dunham, F. Berardi, S. Weigang, M. Chazal, J. Park, J. Logue, M. McGrath, S. Weston, R. Haupt, C. J. Hastie, M. Elliott, F. Brown, K. A. Burness, E. Reid, M. Dorward, C. Johnson, S. G. Wilkinson, A. Geyer, D. M. Giesel, C. Baillie, S. Raggett, H. Leech, R. Toth, N. Goodman, K. C. Keough, A. L. Lind, Zoonomia Consortium, R. J. Klesh, K. R. Hemphill, J. Carlson-Stevermer, J. Oki, K. Holden, T. Maures, K. S. Pollard, A. Sali, D. A. Agard, Y. Cheng, J. S. Fraser, A. Frost, N. Jura, T. Kortemme, A. Manglik, D. R. Southworth, R. M. Stroud, D. R. Alessi, P. Davies, M. B. Frieman, T. Ideker, C. Abate, N. Jouvenet, G. Kochs, B. Shoichet, M. Ott, M. Palmarini, K. M. Shokat, A. García-Sastre, J. A. Rassen, R. Grosse, O. S. Rosenberg, K. A. Verba, C. F. Basler, M. Vignuzzi, A. A. Peden, P. Beltrao, N. J. Krogan, *Science* **2020**, *370*, eabe9403.

[15] B.-P. Li, Y.-T. Mao, Z. Wang, Y.-Y. Chen, Y. Wang, C.-Y. Zhai, B. Shi, S. Y. Liu, J. L. Liu, J. Q. Chen, *Cell. Physiol. Biochem.* **2018**, *46*, 907.

[16] Y. Lei, Y. Takahama, *Microb. Infect.* **2012**, *14*, 262.

[17] 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. Engl.* **2014**, *53*, 10624.

[18] L. Porter, A. K. Kim, L. L. Looger, A. Majumdar, M. R. Starich. Pervasive fold switching in a ubiquitous protein superfamily. *BioRxiv* **2021**.

[19] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, D. Hassabis, *Nature* **2020**, *577*, 706.

[20] A. F. Dishman, R. C. Tyler, J. C. Fox, A. B. Kleist, K. E. Prehoda, M. M. Babu, F. C. Peterson, B. F. Volkman, *Science* **2021**, *371*, 86.

[21] F. Dingfelder, S. Benke, D. Nettels, B. Schuler, *J. Phys. Chem. B.* **2018**, *122*, 11251.