Pervasive fold switching in a ubiquitous protein superfamily

Some proteins adopt more than one stable fold, a predictable and taxonomically diverse attribute.

Lauren L. Porter[1,2,*], Allen K. Kim[1], Loren L. Looger[3], Anaya Majumdar[4], and Mary Starich[2]

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

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

[3]Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA 20147, USA

[4]The Johns Hopkins University Biomolecular NMR Center, The Johns Hopkins University, 3400 N. Charles St. Baltimore, MD 21218, USA

[*]Correspondence: lauren.porter@nih.gov

Abstract. Fold-switching proteins challenge the one-sequence-one-structure paradigm by adopting multiple stable folds. Nevertheless, it is uncertain whether fold switchers are naturally pervasive or rare exceptions to the well-established rule. To address this question, we developed a predictive method and applied it to the NusG superfamily of >15,000 transcription factors. We predicted that a substantial population (25%) of the proteins in this family switch folds. Circular dichroism and nuclear magnetic resonance spectroscopies of 10 sequence-diverse variants confirmed our predictions. Thus, we leveraged family-wide predictions to determine both conserved contacts and taxonomic distributions of fold-switching proteins. Our results indicate that fold switching is pervasive in the NusG superfamily and that the single-fold paradigm significantly biases structure-prediction strategies.

Fold-switching proteins remodel their secondary and tertiary structures and change their functions in response to cellular stimuli (*1*). Their large, sometimes reversible, conformational changes challenge the long-held paradigm that globular proteins have a single fold specified by their amino acid sequences (*2, 3*). Fold switchers regulate diverse biological processes (*4*) and are associated with human diseases such as cancer (*5*), malaria (*6*), and COVID-19 (*7*).

Whereas approaches such as coevolutionary analysis (*8*) and deep learning (*9*), have greatly improved rapid prediction of secondary and tertiary structure for single-fold proteins (*10*), similar methods to classify fold switchers have lagged. (Design of fold-switching sequences has seen more, but limited, success (*11-13*)). This comparative lack of progress arises primarily from the small number of experimentally observed fold switchers (<100), hampering the discovery of generalizable characteristics that distinguish fold switchers from single folders. As a result, essentially all naturally occurring fold switchers have been discovered by chance (*14*), leaving their natural abundance an open question.

Nevertheless, prediction of fold switchers seems achievable. Previous work has shown that discrepancies between homology-based secondary structure predictions can be an effective predictor of fold switching (*15-17*). We sought to leverage these discrepancies to discover dissimilar fold-switching sequences in the NusG protein superfamily, the only family of transcriptional regulators known to be conserved from bacteria to humans (*18*). Housekeeping NusGs (hereafter called NusGs), which exist in nearly every known bacterial genome, associate with transcribing RNA polymerase (RNAP) at many operons, where they promote transcription elongation. Specialized NusG homologs ($NusG^{SP}$s), such as UpxY, LoaP, and RfaH, have also evolted to promote transcription elongation, but at specific operons (*19*). Additionally, NusG and $NusG^{SP}$s regulate translation to varying degrees (*20*). Atomic-level structures of NusG and RfaH have been determined (**Fig. 1A**). They share a two-domain architecture with an N-terminal NGN domain that binds RNAP,

and a C-terminal β-roll domain (also known as a Kyprides-Ouzounis-Woese, or KOW motif) that both recruits the Rho transcription terminator and binds the S10 ribosomal subunit to promote translational efficiency. By contrast, the C-terminal domain (CTD) of *Escherichia coli* RfaH switches between two disparate folds: an α-helical hairpin that inhibits RNAP binding except at operon polarity suppressor (*ops*) DNA sites and a β-roll that binds the S10 ribosomal subunit, fostering efficient translation (*21*). This reversible change in structure and function is triggered by binding to both *ops* DNA and RNAP (*22*).

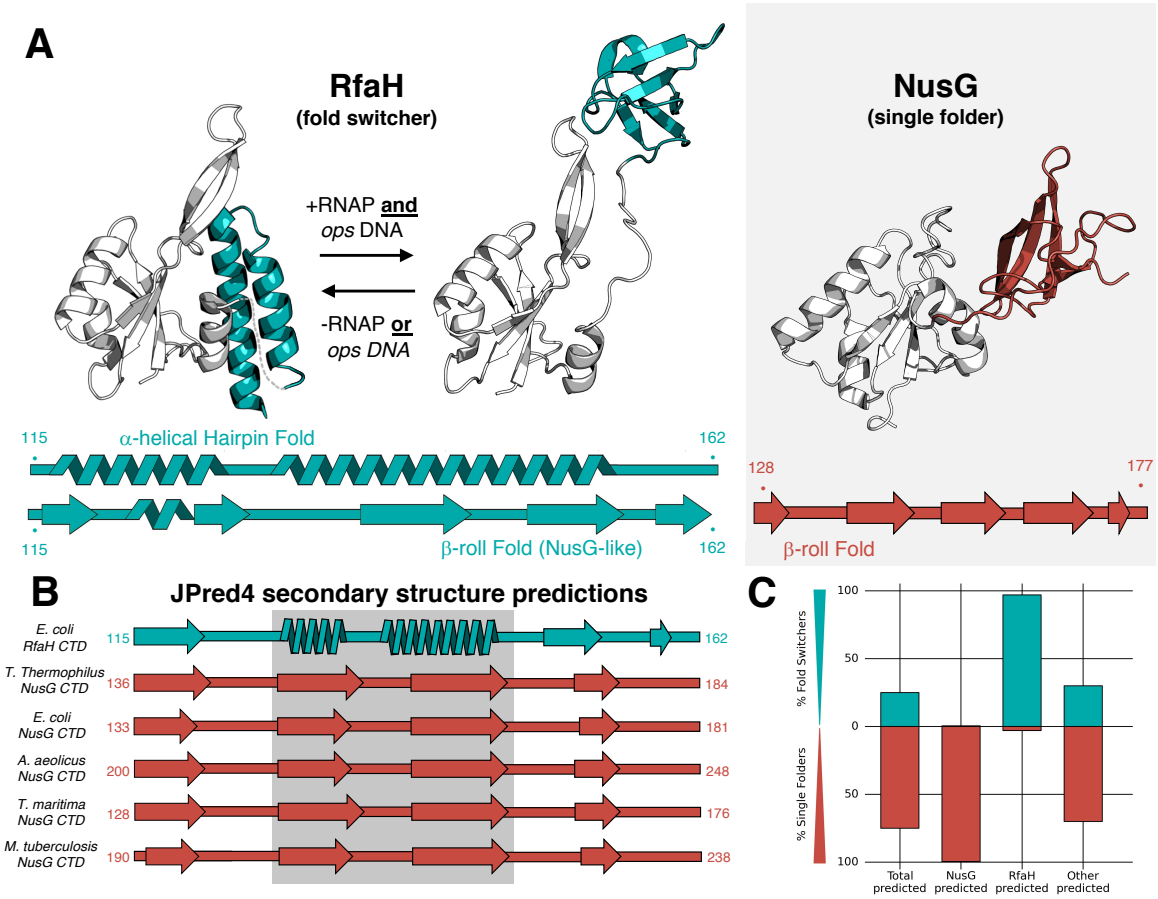

Figure 1. Sequence-based secondary structure predictions discriminate between fold switchers and single folders. (A). Experimentally determined folds and secondary structures of autoinhibited/active RfaH (α-helical hairpin/β-roll folds, respectively) and of single folder NusG. Dotted line represents the NTD-CTD linker missing in the RfaH crystal structure. RfaH/NusG CTDs are colored teal/red; NTDs are gray. **(B).** JPred4 secondary structure predictions discriminate between RfaH and NusGs with experimentally determined structures (teal and red, respectively). Gray box highlights secondary structure prediction discrepancies. Residue numbers are shown at ends of secondary structure diagrams. **(C).** Of the sequences in the NusG family, 25% were predicted to switch folds. Predictions of the houskeeping NusG subset of these sequences were highly consistent with the expectation that housekeeping NusGs do not switch folds. Furthermore, 97% of RfaH-like sequences and 30% of non-NusG, non-RfaH sequences were predicted to switch folds.

We observed that JPred4 (*23*) secondary structure predictions discriminate between the sequences of RfaHs and NusGs with experimentally determined atomic-level structures (**Figs. 1A&B**). These sequence-based calculations consistently indicate that NusG CTD sequences fold into β-strands connected by coils, whereas the *E. coli* RfaH CTD assumes a mixture of α-helix, β-strand, and coil. Thus, our results suggest that JPred4 can distinguish between the sequences of single-folding NusGs and fold-switching RfaHs.

To test the generality of our secondary-structure-based approach, we collected 15,516 non-redundant RfaH/NusG sequences from an iterative BLAST (*24*) search (**Methods**) and tested our predictive method separately on annotated RfaHs, NusGs, and other hits from the search. In total, 25% of proteins in the NusG superfamily were predicted to switch folds (**Figures 1C and S1, Data S1**), a considerable subpopulation with over 3500 sequences. To determine the accuracy of this estimate, we sought to determine the false-negative and false-positive rates of our predictions. For this, we used an orthogonal method to annotate sequences as NusGs or $NusG^{SP}$s – namely, exploiting known operon structures of NusG and several specialized orthologs (*19*). We mapped the sequences used for fold-switch prediction to sequenced bacterial genomes (Ensembl; **Methods**) and analyzed each sequence's local genomic environment for signatures of co-regulated genes. Of our 15,516 total sequences, 5,191 mapped to Ensembl contexts consistent with housekeeping NusG function. Only 28 of these were predicted to switch folds (**Figure 1C**), suggesting a false-positive rate of 0.5% for fold-switch predictions. We then performed a similar calculation in previously identified RfaHs (*19*) and found that of the 1,078 in our Uniprot database, 31 were predicted not to switch folds (**Figure 1C**). Furthermore, of the 474 of these sequences that mapped to an annotated Ensembl locus, only one fell within a NusG-like operon. These results suggest that fold switching is widely conserved among RfaHs, which, if correct, indicates a false positive rate of 3%. Of the remaining sequences with high-confidence predictions (**Methods**), 30% were predicted to switch folds (**Figure 1C**).

We then validated our predictions experimentally. To do this, we selected a representative group of variants with dissimilar sequences by clustering all NusG-superfamily sequences and plotting them on a force-directed graph, hereafter called NusG sequence space (**Figure 2A, Data S1, Figure S2**). The map of this space revealed that some putative fold-switching/single-folding nodes cluster together within sequence space (upper/lower groups of interconnected nodes), while other regions had mixed predictions (left and right groups of interconnected nodes). Candidates selected for experimental validation came from distinct nodes, had diverse genomic annotations, and originated from different bacterial phyla (**Table S1, Figure S3**).

The circular dichroism (CD) spectra of 10 full-length variants were then collected. We expected the spectra of fold switchers to have more helical content than single folders because their CTDs have completely different structures (RfaH: all α-helix, NusG: all β-sheet), while the secondary structure compositions of their single-fold NTDs are expected to be essentially identical. We initially compared *E. coli* RfaH (variant 3) and *E. coli* NusG (variant 9), whose atomic-level structures have been determined previously (*25, 26*). Indeed, their CD spectra were quite different (**Figure S4**). As expected, *E. coli* RfaH had a substantially higher α-helix:β-strand ratio (1.1, **Figure 2B, variant 3**) than *E. coli* NusG (0.54, **Figure 2B, variant 9**) – consistent with solved structures. In addition to *E. coli* RfaH and NusG, all 8 other predictions were confirmed by experiment (**Figure 2B**). Specifically, CD spectra with RfaH-like α-helix:β-strand ratios were found for five out of the eight variants: two RfaHs (variants 2, 6), a LoaP (variant 1), an annotated NusG (variant 4), and an annotated "NGN domain-containing protein" (variant 5) (**Table S1**). CD spectra with NusG-like α-helix:β-strand ratios were found for the remaining three variants: two annotated NusGs (variants 8, 10, both genomically consistent with being NusGs) and one UpbY (variant 7).

Although the CD spectra of all full-length variants were consistent with our predictions, they do not conclusively demonstrate that any of the 6 RfaH-like variants switch folds, only that their

CTDs likely fold into α-helical hairpins. Previous work (*21*) has shown that the full-length RfaH CTD folds into an α-helical hairpin while its isolated CTD folds into a stable β-roll. Thus, we determined the CD spectra of five isolated CTDs: three from putative fold switchers, and two from putative single folders. All of them had low helical content and high β-sheet content (**Figure S5**), strongly suggesting that all three of the RfaH-like proteins tested have CTDs that switch from the α-helical hairpin in the full-length form to the β-roll as isolated CTDs.

We then used solution nuclear magnetic resonance (NMR) spectroscopy to characterize two variants at higher resolution. Previous work (*21*) has shown that the isolated CTD of RfaH has a significantly different ^{1}H-^{15}N HSQC than full-length RfaH, whose CTD folds into an α-helical hairpin. Thus, we conducted similar experiments on one single-fold variant (#8) and one putative fold switcher (#5). We found that the backbone amide resonances of the full-length and CTD forms of variant #8 were nearly superimposable, whereas the full-length and CTD forms of variant #5 shared only 7/58 common backbone amide peaks (**Figure 2C**). This result demonstrates that, as predicted, variant 8 does not switch folds. It is also consistent with the prediction that variant 5 switches folds, because large backbone amide shifts indicate either refolding or a large change in protein interface. Both occur in fold-switching RfaH but not in single-folding NusG. Subsequently, we used the assigned backbone amide resonances to characterize the secondary structures of both CTD variants at higher resolution (**Figure S6, Table S2**). Both were consistent with the β-roll fold, which may indicate that the CTD of full-length #5 adopts an alternative fold.

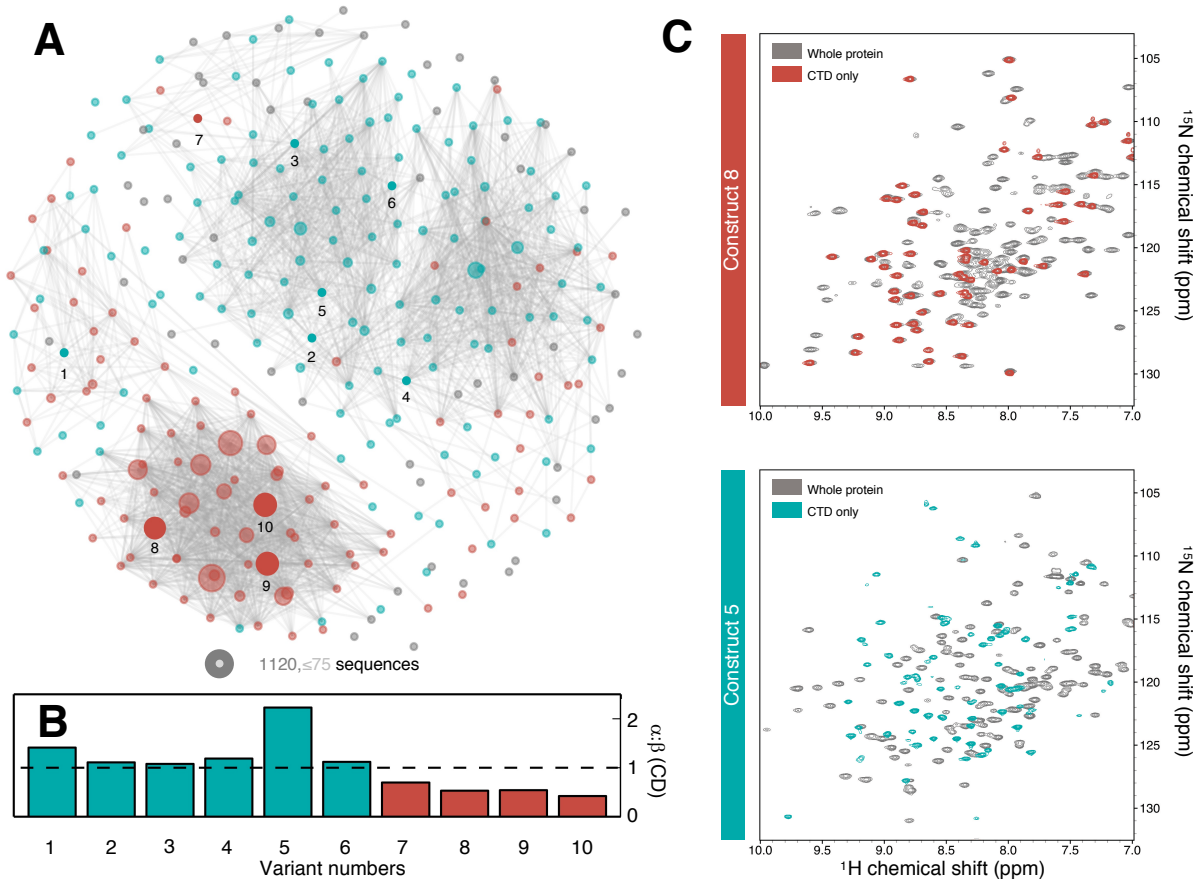

Figure 2. RfaH/NusG sequence space. (A). Force-directed graph of 15,516 RfaH/NusG sequences. The largest node contains 1120 sequences; all nodes with 75 sequences or fewer are the same (smallest) size. Edges connecting the graph represent an average aligned sequence identity between the sequences in two nodes that exceeds 22%. Nodes labeled in teal/red were predicted to be fold switchers/single folders, on average; gray nodes contained only sequences with low-confidence JPred4 predictions. **(B).** Fraction of α-helix:β-sheet measured from CD. Dotted line (1.0) represents threshold α-helix:β-sheet ratio for fold switcher definition. All ratios for predicted fold switchers are above the cutoff; all ratios for predicted single folders fall below the cutoff. Numerical labels shown in (A) correspond to variant numbers. **(C).** The ^{1}H-^{15}N HSQCs of full-length and CTD variants of a putative single-folder (Variant 8) are nearly superimposable, while the HSQCs of full-length and CTD variants of a putative fold switcher (Variant 5) differ significantly.

These results, though a very small proportion of the sequences in this superfamily, support the accuracy of our predictions and demonstrate that:

(1) Some but not all NusG^{SP}s besides RfaH, are likely to switch folds. Specifically, full-length LoaP (variant 1), which regulates the expression of antibiotic gene clusters (*27*), was found to have an RfaH-like CD spectrum, whereas full-length UpbY from *B. fragilis* (variant 7) appears to assume a NusG-like fold.

(2) Some annotated NusGs have RfaH-like CD spectra (variant 4), the result of incorrect annotation. Indeed, the genomic environemnt of variant 4 (**Methods**) suggests that is a UpxY, not a NusG.

(3) The fold-switching mechanism is conserved among annotated RfaHs with low sequence identity (≤32%, variants 2, 3, and 6), a possibility proposed previously (*20*), though without experimental validation. Also, "NGN domain-containing protein" variant 5 is genomically inconsistent with being a NusG and is likely another RfaH.

To benchmark the performance of our secondary-structure-based method, we assessed whether coevolutionary and template-based methods could also distinguish between fold switchers and single folders in the NusG superfamily. Specifically, we tested Robetta (*28*), EVCouplings (*29*), and Phyre2 (*30*) on variants 1-6 (**Figure 3A&B**). No method suggested a helical hairpin conformation in any variant (**Figure S7**), including *E. coli* RfaH (variant 3), which has an experimentally determined structure (*25*). To determine why these alternative methods failed, we examined the sequence alignments used to generate their predictions. In all cases, the alignments contained mixtures of RfaH and NusG sequences, and the resulting residue-residue couplings from Robetta and EVCouplings corresponded with the NusG-like β-roll fold (**Figure 3A**). These results suggest that β-roll couplings present in both single-folding and fold-switching sequences might overwhelm any α-helical couplings unique to fold-switching sequences.

To explicitly test this hypothesis, we performed coevolutionary analysis on a set of sequences that our method predicted to switch folds. Specifically, we clustered our putative fold switchers by their JPred4 predictions and did coevolutionary analysis on the largest cluster using GREMLIN (*31*). The residue-residue couplings generated from these sequences differed substantially from the NusG-like couplings generated previously (**Figures 3A&B**). Furthermore, GREMLIN couplings calculated from the alignments used by EVCouplings and Robetta corresponded to the β-roll fold only (**Figure**

S8), demonstrating that GREMLIN identified helical contacts from the JPred-filtered sequence alignment, not because its algorithm was inherently more sensitive to fold switching than EVCouplings or Robetta**.**

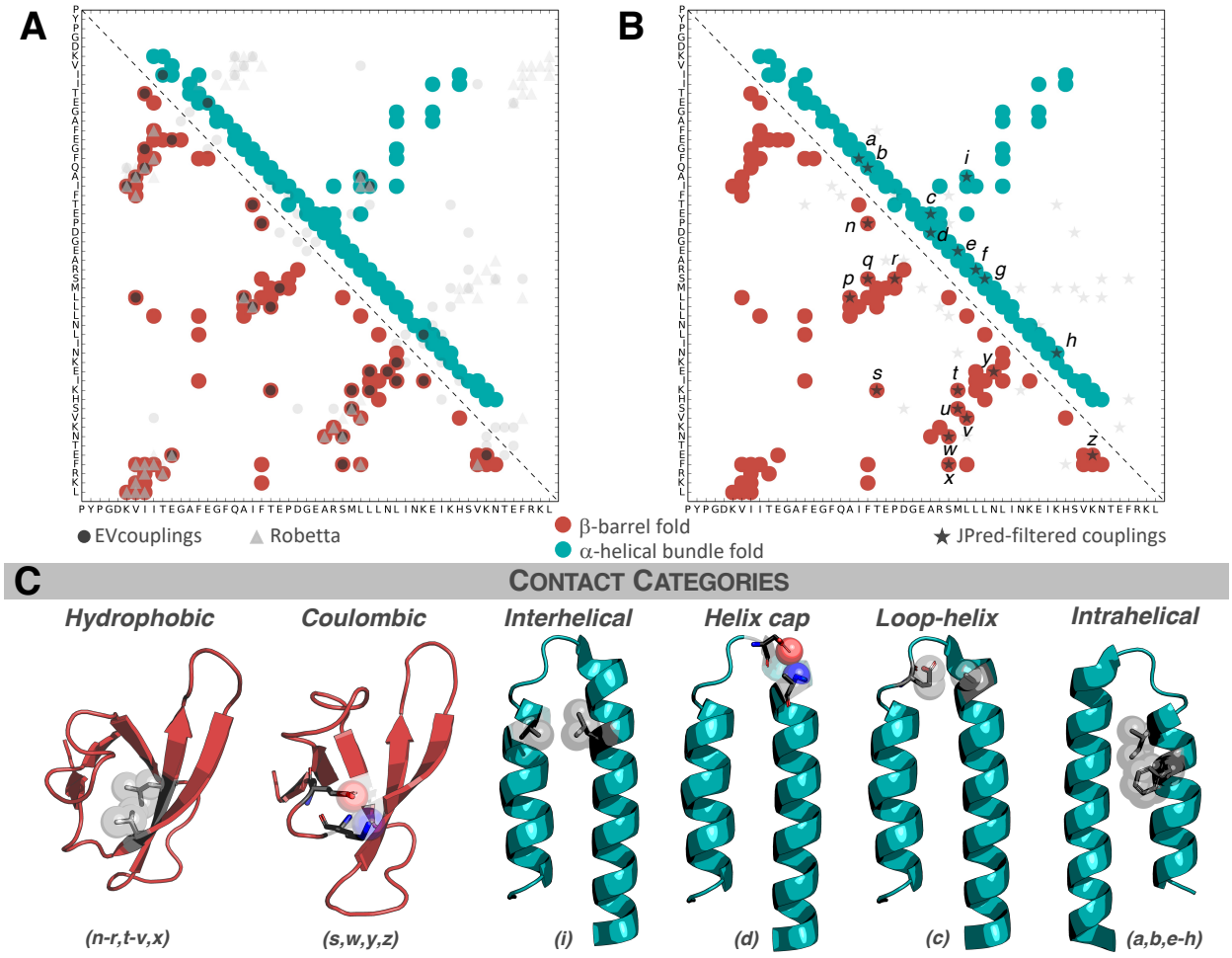

Figure 3. The single-fold paradigm biases protein structure predictions. (A). RfaH-like residue couplings from EVCouplings and Robetta are biased toward the β-roll fold (PDB ID: 2LCL, red); no couplings unique to the experimentally determined α-helical hairpin (PDB ID: 2OUG, chain A, teal) were identified. Both sets of predictions were derived from alignments with mixtures of single-folding and fold-switching sequences. Couplings that do not correspond to experimentally observed contacts are lighter shades of gray. (B) Fold-switching sequences predicted by JPred4 yielded residue-residue couplings from both the β-roll and the α-helical hairpin. Italicized letters correspond to individual couplings in the two folds. (C). Categories of residue-residue contacts from both folds using the alphabetically labeled contacts in (B), listed below each category.

The coevolutionary analysis of putative fold switchers suggests that their CTDs encode contacts unique to both the α-helical and β-roll folds – only one contact was shared between both folds (*i*/*q*, **Figures 3B&C**). Upon examining these contacts more closely, several categories emerged.

For the β-roll fold, eight hydrophobic and four Coulombic contacts were observed, and for the α-helical fold, six intrahelical hydrophobic contacts and one set each of interhelical contacts, strand-helix contacts, and helix-capping contacts were observed (**Figure 3C**). To assess statistical significance of these contacts, we analyzed the residue pairs in all sequences used for this coevolutionary analysis. For the β-roll fold, 97%/96% of residue pairs making strand contacts *q* and *v* were hydrophobic, and 76%/53% of contacts *y* and *s* could potentially form Coulombic interactions. For the helical hairpin fold, 96% of interhelical contacts were hydrophobic, 94% of helix-capping residues could potentially form an i-4→i or i→i backbone-to-sidechain hydrogen bond, 85% of residues in the helix-loop interaction had a charged residue in one position, but not both, and 80% of residues in intrahelical contact *a* were both hydrophobic. The remaining contacts gave more mixed results, perhaps due to hydrophobic residues making contact with the hydrophobic portion of their hydrophilic partners. Previous work has shown that interdomain interactions also contribute significantly to RfaH fold switching (*21*). Unfortunately, these interactions could not be identified by coevolutionary analysis (**Figure S9**), a likely result of the limited number of JPred-filtered sequences used in the analysis.

These results suggest that the sequences of fold-switching CTDs poise them to assume two disparate folds. Thus, it might be reasonable to expect these sequences to be relatively homogeneous, especially since variants of another fold switcher, human XCL1, lose their ability to switch folds below a relatively high identity threshold (60%, (*32*)). The opposite is true. The sequences of putative fold-switching CTDs are significantly more heterogeneous (20.4% mean/19.4% median sequence identity) than the sequences of predicted single folders (40.5% mean/42.5% median sequence identity, **Figure 4A**). Accordingly, among the sequences tested experimentally, similar mean/median sequence identities were observed: 21.0%/21.1% (fold switchers), 43.2%/41.2% (single folders, **Figure 4B**). Additionally, fold-switching CTDs were predicted in most bacterial phyla, and many were predicted in archaea and eukaryotes as well (**Figure 4C**). Together, these results suggest that many highly diverse

CTD sequences can switch folds between an α-helical hairpin and a β-roll in organisms from all kingdoms of life.

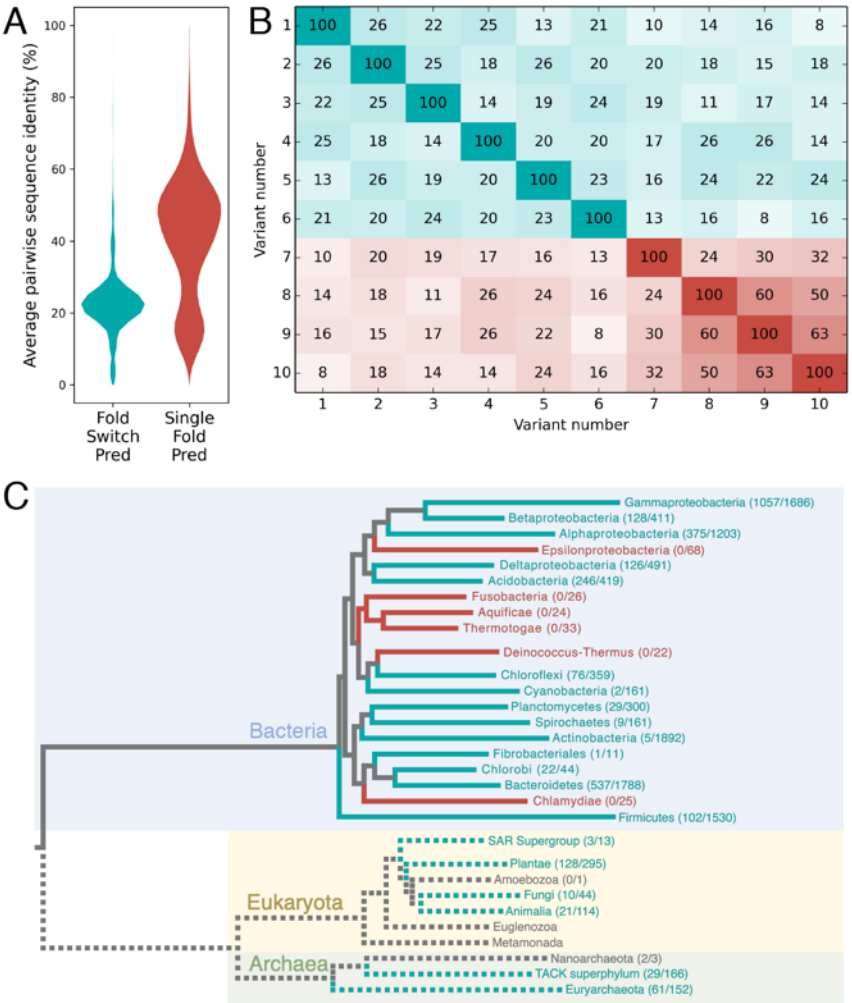

Figure 4. The sequences of fold-switching CTDs are highly diverse and found in a wide variety of bacterial phyla. (A). Violin plots of pairwise sequence identities differ significantly for putative fold switchers and putative single folders. On average, pairwise sequence identities are lower for putative fold switchers (20.4%) than single folders (40.5%). (B). Sequence identity matrix of the CTDs of variants 1-10 in **Figure 2**. Numbers in each box represent the % identity of the two variants compared. Darker boxes represent higher identity levels. (C). Fold-switching CTDs are found in many bacterial phyla and other kingdoms of life. Numbers next to taxa represent #predicted fold switchers/ #total sequences. Gray branches represent unidentified common ancestors, since the evolution of fold-switching NusGs is unknown. Dotted lines represent lower-confidence predictions since fold switching has not been confirmed experimentally in archaea and eukaryota. Fold-switching/single-folding predictions are represented by teal/red colorings; predictions in branches with fewer than 10 sequences are gray.

Why might the sequence diversity of fold-switching CTDs exceed those of single folders? Functional diversity is one likely explanation. Previous work has shown that NusG^{SP}s drive the expression of diverse molecules from antibiotics to toxins (*19*). Our method predicts that many of these switch folds. Furthermore, since helical contacts are conserved among at least some fold-switching CTDs, it may be possible that CTD sequence variation is less constrained in other function-specific positions. The fold-switching mechanism of RfaH allows it to function first as a transcriptional activator and subsequently as a translational activator, presumably speeding up the activation of downstream genes. Such NusG^{SP}s are likely under

strong selective pressure to evolve such fold-switching mechanisms when the regulated products control life-or-death events, such as the appearance of rival microbes or desiccation. Supporting this possibility, RfaH, LoaP, and UpxY usually drive operons controlling rapid response to changing environmental conditions, such as macrolide antibiotic production (*27*), antibiotic-resistance gene expression (*19*), virulence activation (*33*), and biofilm formation (*34*).

Our method was sensitive enough to predict fold switching, although the phenomenon was imperceptible to other state-of-the-art methods. These other methods assume that homologous sequences adopt the same fold, as evidenced by their use of sequence alignments that contained both fold-switching and single-folding sequences. These mixed sequence alignments biased their predictions. While those predictions are partially true since both fold-switching and single-folding CTDs can fold into β-rolls, they miss the alternative helical hairpin conformation and its regulatory function (*18*). Alternative approaches that account for such conformational variability and dynamics, a weakness in even the best predictors of protein structure (*10*), could lead to improved predictions. This need is especially acute in light of recent work showing how protein structure is influenced by the cellular environment (*35*).

Our results indicate that fold switching is a pervasive, evolutionarily conserved mechanism. Specifically, we predicted that 25% of the sequences within a ubiquitous protein family switch folds and showed that residue-residue contacts unique to each fold-switching conformation are conserved over the course of evolution. The success of our method in the NusG superfamily suggests that it may have enough predictive power to identify fold switching in protein families believed to contain only single folders. Such predictions would be particularly useful since many fold switchers are associated with human disease (*4-7*). Given the unexpected abundance of fold switching in the NusG superfamily, there may be many more unrelated fold switchers to discover.

Acknowledgements

L.L.P. thanks Drs. Marius Clore and Carolyn Ott for constructive input throughout this project and for many excellent suggestions concerning the text and figures. Thanks to Dr. Nico Tjandra for generously sharing his lab resources and office space and to Dr. David Landsman for administratively supporting this research, especially the experimental component. We also thank Drs. Liskin Swint-Kruse, Nico Tjandra, David Nyenhuis, and Daniel Morris for helpful comments concerning the text, Dr. Christos Kougentakis for helping to collect NUS NMR data, Dr. Nicholas Fitzkee for sharing his isotope labeling protocol, and Dr. Xiaofang Jiang for recommending TaxonKit and ITOL. This work utilized resources from the NHLBI Biophysics Core, the NHLBI Protein Expression Facility, and the NIH HPS Biowulf cluster (http://hpc.nih.gov), and it was supported in part by the Intramural Research Program of the National Library of Medicine, National Institutes of Health.

References

1. L. L. Porter, L. L. Looger, Extant fold-switching proteins are widespread. *Proc Natl Acad Sci U S A* **115**, 5968-5973 (2018).
2. C. B. Anfinsen, Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).
3. P. N. Bryan, J. Orban, Proteins that switch folds. *Curr Opin Struct Biol* **20**, 482-488 (2010).
4. A. K. Kim, L. L. Porter, Functional and Regulatory Roles of Fold-Switching Proteins. *Structure* **29**, 6-14 (2021).
5. B. P. Li *et al.*, CLIC1 Promotes the Progression of Gastric Cancer by Regulating the MAPK/AKT Pathways. *Cell Physiol Biochem* **46**, 907-924 (2018).
6. D. Giganti *et al.*, Secondary structure reshuffling modulates glycosyltransferase function at the membrane. *Nat Chem Biol* **11**, 16-18 (2015).
7. D. E. Gordon *et al.*, Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, (2020).
8. D. S. Marks, T. A. Hopf, C. Sander, Protein structure prediction from sequence variation. *Nat Biotechnol* **30**, 1072-1080 (2012).
9. J. Yang *et al.*, Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci U S A* **117**, 1496-1503 (2020).
10. A. W. Senior *et al.*, Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706-710 (2020).
11. P. A. Alexander, Y. He, Y. Chen, J. Orban, P. N. Bryan, A minimal sequence code for switching protein structure and function. *Proc Natl Acad Sci U S A* **106**, 21149-21154 (2009).
12. X. I. Ambroggio, B. Kuhlman, Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J Am Chem Soc* **128**, 1154-1161 (2006).

13. K. Y. Wei *et al.*, Computational design of closely related proteins that adopt two well-defined but structurally divergent folds. *Proc Natl Acad Sci U S A* **117**, 7208-7215 (2020).
14. M. Lopez-Pelegrin *et al.*, Multiple stable conformations account for reversible concentration-dependent oligomerization and autoinhibition of a metamorphic metallopeptidase. *Angew Chem Int Ed Engl* **53**, 10624-10630 (2014).
15. N. Chen, M. Das, A. LiWang, L. P. Wang, Sequence-Based Prediction of Metamorphic Behavior in Proteins. *Biophys J* **119**, 1380-1390 (2020).
16. A. K. Kim, L. L. Looger, L. L. Porter, A high-throughput predictive method for sequence-similar fold switchers. *Biopolymers*, e23416 (2021).
17. S. Mishra, L. L. Looger, L. L. Porter, Inaccurate secondary structure predictions often indicate protein fold switching. *Protein Sci* **28**, 1487-1493 (2019).
18. J. Y. Kang *et al.*, Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell* **173**, 1650-1662 e1614 (2018).
19. B. Wang, V. M. Gumerov, E. P. Andrianova, I. B. Zhulin, I. Artsimovitch, Origins and Molecular Evolution of the NusG Paralog RfaH. *mBio* **11**, (2020).
20. B. Wang, I. Artsimovitch, NusG, an Ancient Yet Rapidly Evolving Transcription Factor. *Front Microbiol* **11**, 619618 (2020).
21. B. M. Burmann *et al.*, An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* **150**, 291-303 (2012).
22. P. K. Zuber, K. Schweimer, P. Rosch, I. Artsimovitch, S. H. Knauer, Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. *Nat Commun* **10**, 702 (2019).
23. A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton, JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43**, W389-394 (2015).
24. S. F. Altschul *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
25. G. A. Belogurov *et al.*, Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol Cell* **26**, 117-129 (2007).
26. C. Wang *et al.*, Structural basis of transcription-translation coupling. *Science* **369**, 1359-1365 (2020).
27. J. R. Goodson, S. Klupt, C. Zhang, P. Straight, W. C. Winkler, LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in Bacillus amyloliquefaciens. *Nat Microbiol* **2**, 17003 (2017).
28. D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**, W526-531 (2004).
29. T. A. Hopf *et al.*, The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics* **35**, 1582-1584 (2019).
30. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858 (2015).
31. S. Ovchinnikov, H. Kamisetty, D. Baker, Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **3**, e02030 (2014).
32. A. F. Dishman *et al.*, Evolution of fold switching in a metamorphic protein. *Science* **371**, 86-90 (2021).
33. J. A. Leeds, R. A. Welch, RfaH enhances elongation of Escherichia coli hlyCABD mRNA. *J Bacteriol* **178**, 1850-1857 (1996).
34. C. Beloin *et al.*, The transcriptional antiterminator RfaH represses biofilm formation in Escherichia coli. *J Bacteriol* **188**, 1316-1331 (2006).
35. W. B. Monteith, R. D. Cohen, A. E. Smith, E. Guzman-Cisneros, G. J. Pielak, Quinary structure modulates protein stability in cells. *Proc Natl Acad Sci U S A* **112**, 1739-1742 (2015).