Origin of a folded protein from an intrinsically disordered ancestral peptide

Hongbo Zhu, Edgardo Sepulveda, Marcus D. Hartmann,
Manjunatha Kogenaru, Jörg Martin, Andrei N. Lupas
Max Planck Institute for Developmental Biology, Department of Protein Evolution
Spemannstr. 35, D-72076 Tübingen, Germany
hongbo.zhu@tuebingen.mpg.de

For the most part, contemporary proteins can be traced back to a basic set of a few thousand domain prototypes, many of which were already established in the Last Universal Common Ancestor of life on earth, around 3.5 billion years ago. The origin of these domain prototypes, however, remains poorly understood. We have proposed that they arose from an ancestral set of peptides, which acted as cofactors of RNA-mediated catalysis and replication [ASL15]. Initially, these peptides were entirely dependent on the RNA scaffold for their structure, but as their complexity increased, they became able to form structures by excluding water through hydrophobic contacts, making them independent of the RNA scaffold. Their ability to fold was thus an emergent property of peptide-RNA coevolution.

The ribosome is the main survivor of this primordial RNA world and offers an excellent model system for retracing the steps that led to the folded proteins of today, due to its very slow rate of change [LA17]. Collectively, ribosomal proteins chart a path of progressive emancipation from the RNA scaffold, offering a window onto the time when proteins were acquiring the ability to fold.

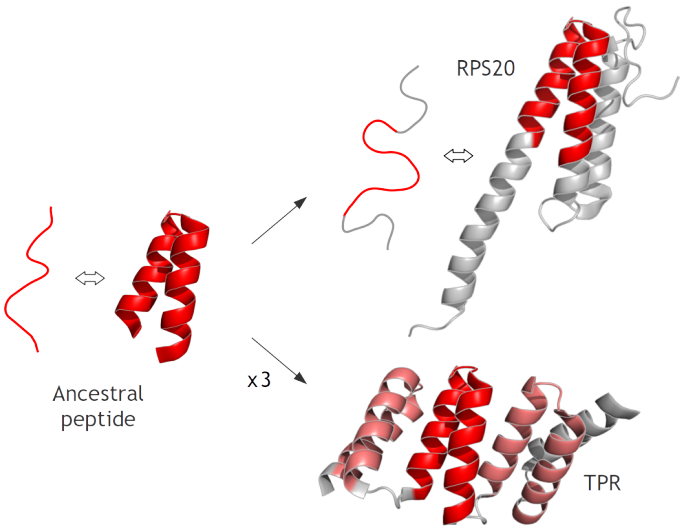

Figure 1: Scenario for the divergent evolution of ribosomal protein RPS20 and the cytosolic TPR fold from an ancestral, ribosome-associated $\alpha\alpha$-hairpin (red). The $\alpha\alpha$-hairpin and RPS20 are unstructured in the absence of their cognate rRNAs.

We have retraced this emancipation from the RNA scaffold computationally and experimentally, by investigating the plausible ancestor in ribosomal proteins for a cytosolic protein fold, the tetratricopeptide repeat (TPR). The crystal structure of TPR domain reveals that the repeat units of the fold are

helical hairpins, stacked into a continuous, right-handed superhelical architecture (Figure 1). These $\alpha\alpha$-hairpins in all known TPR-containing proteins can be detected using a single sequence profile, underscoring their homologous origin. More importantly, the $\alpha\alpha$-hairpin has been discovered to occur in multiple seemingly unrelated protein folds. All these occurrences are very similar in their sequence and structure, suggesting that the $\alpha\alpha$-hairpin is likely to be one of the remnants of the ancestral peptides [ASL15].

In searching for the origin of TPRs, we hypothesized that the hairpin at the root of the fold might either have been part of a different, non-repetitive fold or have given rise to both repetitive and non-repetitive folds at the origin of folded domains (Figure 1). Either way, we hoped that we might find $\alpha\alpha$-hairpins in non-repetitive proteins that are similar in both sequence and structure to the TPR unit, suggesting a common origin. Through the sequence and structure comparison of the TPR $\alpha\alpha$-hairpins to all proteins of known structure, we could indeed detect such hairpins, including one from a ribosomal protein S20 (RPS20).

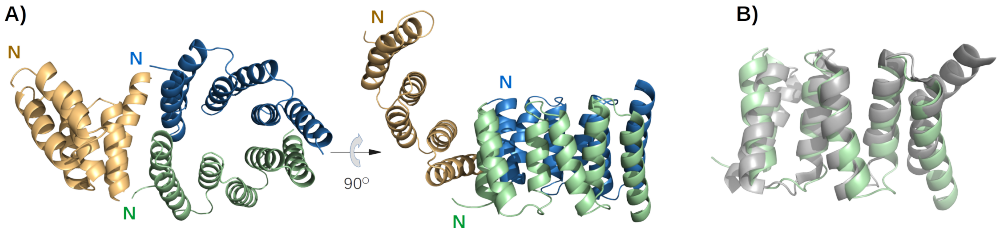

Figure 2: (A) The crystal structure of the RPS20-derived repeat protein, M4N. The three chains in the asymmetric unit are colored green, blue and yellow, respectively. Chains A and B form a dimer. (B) Superposition of the RPS20-derived repeat protein (green) and the TPR protein CTPR3 (PDB: 1na0, chain A, gray).

Subsequently, by amplifying an $\alpha\alpha$-hairpin from a RPS20, which is unstructured in the absence of the cognate ribosomal RNA, we explored whether an intrinsically disordered peptide could form a folded protein through an increase in complexity afforded by repetition. Simple repetition was not sufficient in our case, but the repeat protein was so close to a folded structure that only two point mutations per repeat were necessary to allow it to fold reliably (Figure 2) [ZSH$^+$16]. The mutations needed for this transition did not appear to affect negatively the interaction with the RNA scaffold and were neutral for survival and growth in the parent organism, raising the possibility that they could have been among the variants sampled multiply in the course of evolution. TPRs could thus have plausibly arisen by amplification from an ancestral, RNA-dependent helical hairpin, as proposed by our theory.

References

[ASL15] Vikram Alva, Johannes Sding, and Andrei N Lupas. A vocabulary of ancient peptides at the origin of folded proteins. *eLife*, 4:e09410, Dec 2015.

[LA17] Andrei N. Lupas and Vikram Alva. Ribosomal proteins as documents of the transition from unstructured (poly)peptides to folded proteins. *Journal of Structural Biology*, 198(2):74 – 81, 2017.

[ZSH$^+$16] Hongbo Zhu, Edgardo Sepulveda, Marcus D Hartmann, Manjunatha Kogenaru, Astrid Ursinus, Eva Sulz, Reinhard Albrecht, Murray Coles, Jrg Martin, and Andrei N Lupas. Origin of a folded repeat protein from an intrinsically disordered ancestor. *eLife*, 5:e16761, Sep 2016.