EVOLVABILITY OF STRUCTURE AND FUNCTION IN ANCIENT PROTEINS. A.D. Goldman[1,*], S.C. Page[1] , J.T. Beatty[1]. [1]Oberlin College, Oberlin, OH, USA. [*]agoldman@oberlin.edu.

Introduction: Many proteins are composed of multiple domains, which each perform separate subfunctions that together contribute to the overall function of the protein. Domains are also usually structurally independent and have a semi-independent evolutionary history from the rest of the protein. The early evolution of the proteome likely involved an exploration of domain structure and function prior to the combination of these domains into larger multidomain proteins. The early stages of protein evolution may have also been influenced by the use of enzyme cofactors that were important catalysts prior to the emergence of protein-mediated metabolism[1]. Here we present results from two parallel studies revealing trends that potentially shaped the early evolution of the proteome.

Structural complexity and the evolution of new functions: To study the evolutionary relationship between protein structure and the evolution of new functions, we created a database of single domain proteins for which the structure and function are both known. Single domain proteins contain only one tertiary structure and therefore allow us to attribute a protein's function directly to its structure. We analyzed this database with two measurements of structural complexity, one that ranks protein structural classes from simple to complex and another that measures the metric entropy of secondary structure elements in mixed α/β architectures. In both cases, we find a positive correlation between functional breadth and structural complexity in ancient[3,4] protein superfamilies. In the same superfamilies, we also find a positive correlation between functional breadth and cofactor usage.

Comparison	Proteins	Correlation	p-value
number of functions vs. structural complexity	Ancient	0.61	0.03
	All	-0.04	0.38
number of functions vs. number of cofactors used	Ancient	0.53	0.06
	All	0.18	0.06

Table 1: The relationships of the number of functions that can be performed by mixed α/β single domain proteins in a given superfamily and either the average structural complexity (defined as metric entropy of secondary structure) or the number of cofactors used by proteins in that superfamily. For predicted ancient protein superfamilies, there is a strong correlation between functional breadth and structural complexity as well as functional breadth and cofactor usage. These relationships are either weaker or non-existent when non-ancient protein superfamilies are included.

Designed proteins provide a new view of sequence space: Protein domains are grouped into superfamilies by sequence similarity. This view of protein evolution is restricted by a limited sampling from extant organisms. We used the Rosetta Design[5] platform to explore potential sequence space for domain structures as defined by both the SCOP[6] and CATH[7] databases. In many cases, seemingly unrelated protein superfamilies are connected by mutual sequence similarity to synthetic sequences. This effect is more frequent among simpler classes of protein structure (*e.g.*, 100% for those containing only α-secondary structure). This expanded sequence space thus appears to be more connected for simpler structures.

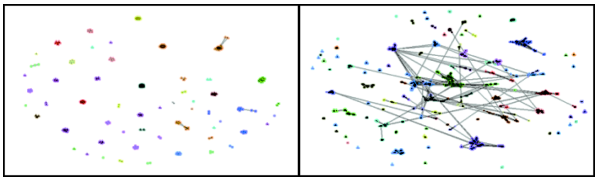

Figure 1: An example of sequence similarity between seemingly unrelated protein superfamilies from the four-helical up-and-down bundle fold. Naturally occurring sequences do not show similarity >50% between members of different superfamilies (left). Synthetic sequences designed to have a similar structure to one of the naturally occurring proteins are added and the new dataset is reanalyzed. In this new dataset, naturally occurring sequences from different superfamilies are bridged by mutual similarity to synthetic sequences (right). This effect is seen in both the SCOP and CATH definitions of this structure.

Conclusion: We have described several trends that are likely to have shaped the early proteome. Simple protein structures appear to have a more fluid sequence space, which could have facilitated evolutionary divergence. However, ancient protein families with complex structures are more capable of evolving a broad range of functions and using cofactors to achieve this functional breadth. Organisms able to synthesize more complex proteins could thus begin to explore a greater range of potential functions, ultimately leading to a protein repertoire as sophisticated as that observed in life, today.

References: [1] White H.B. (1976) *J Mol Evol, 7*, 101–104. [2] Wang M., et al. (2007) *Genome Res, 17*, 1572–1585. [3] Yang S., et al. (2005) *Proc Natl Acad Sci USA, 102*, 373–378. [4] Kuhlman B., et al. (2003) Science, 302, 1364-1368. [5] Murzin A.G., et al. (1995), J Mol Biol, 247, 536-540 [6] Sillitoe I., et al. (2013) Nucleic Acids Res, 41:D490-8.