On the emergence of homochirality and life itself

Jeffrey Skolnick and **Mu Gao** (Georgia Institute of Technology, USA)

Many of life's molecules including proteins are built from chiral building blocks. What drove homochiral building block selection? Simulations on demi-chiral proteins containing equal numbers of D- and L-amino acids show that they possess many modern homochiral protein properties. They have the same global folds and could do the same biochemistry, with ancient, essential functions being most prevalent. They could synthesize chiral RNA and lipids which formed vesicles. RNA eventually combined with proteins creating ribosomes for more efficient protein synthesis, and thus, life began. Increased native state stability from homochiral secondary structure hydrogen bonding helped drive proteins towards homochirality.

Essential macromolecules for life and which ones came first?

The cells of living systems are composed of different types of molecules. The molecules of life have a handedness, termed chirality. How chiral systems emerged from a background of equal populations of molecules of opposite chirality is one key question of the origin of life. Some, such as DNA, RNA and proteins that are essential for life as we know it, are composed almost entirely of building blocks with the same chirality, i.e., they are homochiral. As shown in Figure 1, the building blocks of RNA and DNA are right-handed, D-ribose and D-deoxyribose sugars, while proteins are composed of left-handed, L-amino acids or residues. However, by symmetry, in a random mixture, there should be roughly equal numbers of building blocks with D- and L-chirality; these molecules are referred to as demi-chiral. As life started, no matter which set of macromolecules came first, one key question is how did homochiral macromolecules emerge from a soup containing the same amounts of D and L building blocks?

To get a deeper understanding of how demi-chiral building blocks were preferentially selected for one chirality (D or L) over the other, it is important to review current ideas as to which macromolecules came first. The popular RNA world hypothesis conjectures that RNA was first. This model does not answer how proteins were synthesized without ribosomes which also contain proteins or how homochirality emerged. Another model invokes chiral crystal surfaces, but by symmetry, crystal surfaces of opposite chirality should occur with equal frequency. A third view asserts that metabolism, likely performed by proteins, came first, followed by replication. This model does not explain how proteins were synthesized from a soup that possibly contained an equimolar mixture of D- and L-amino acids. Reasonable-length proteins might have been synthesized via the foldamer hypothesis which conjectures that oligomers acted as templates catalysing the synthesis of likely demi-chiral proteins. Other mechanisms such as molecular mutualism or spontaneous peptide formation from aminonitriles might have occurred. Interestingly, if meteorites were the initial source of amino acids, due to an excess in the L-amino acid composition in some meteorites, this could cause L-amino acids to win. Similarly, the excess of d-sugar acids in some meteorites could be the source of chiral RNA. By whatever mechanism proteins formed, in this contribution we examine the consequences if demi-chiral proteins were present prior to the time when life began.

We undertook a computer experiment that examined the structure and functions of the low-energy, compact structures of protein sequences comprised of randomly arranged, equal amounts of D- and L-amino acids. We then explored how the structural and biochemical features of such demi-chiral proteins compare to modern homochiral proteins. Were they as stable? Is there any similarity between their global, compact conformations and modern proteins? Are there thermodynamic driving forces towards homochirality? Could demi-chiral proteins catalyse chiral chemical reactions? If so, do they have any inherent bias towards particular types of enzymes and biochemical pathways?

Increased thermodynamic stability drives proteins towards a more homochiral amino acid composition

One dominant interaction in protein folding is the hydrogen bond between the amino acid's backbone carbonyl oxygen and amide hydrogen (as shown in

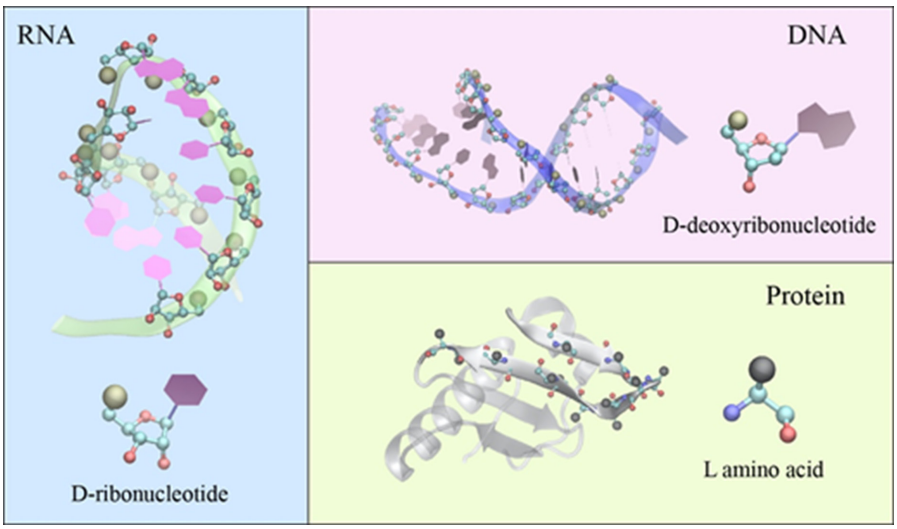

Figure 1. Building blocks of RNA, DNA and proteins. Carbon, oxygen and nitrogen atoms are shown in cyan, red and blue spheres, respectively. The phosphate group and protein side chain groups are represented by tan and black balls, respectively. Bases are shown in purple polygons. Hydrogen atoms are not shown.

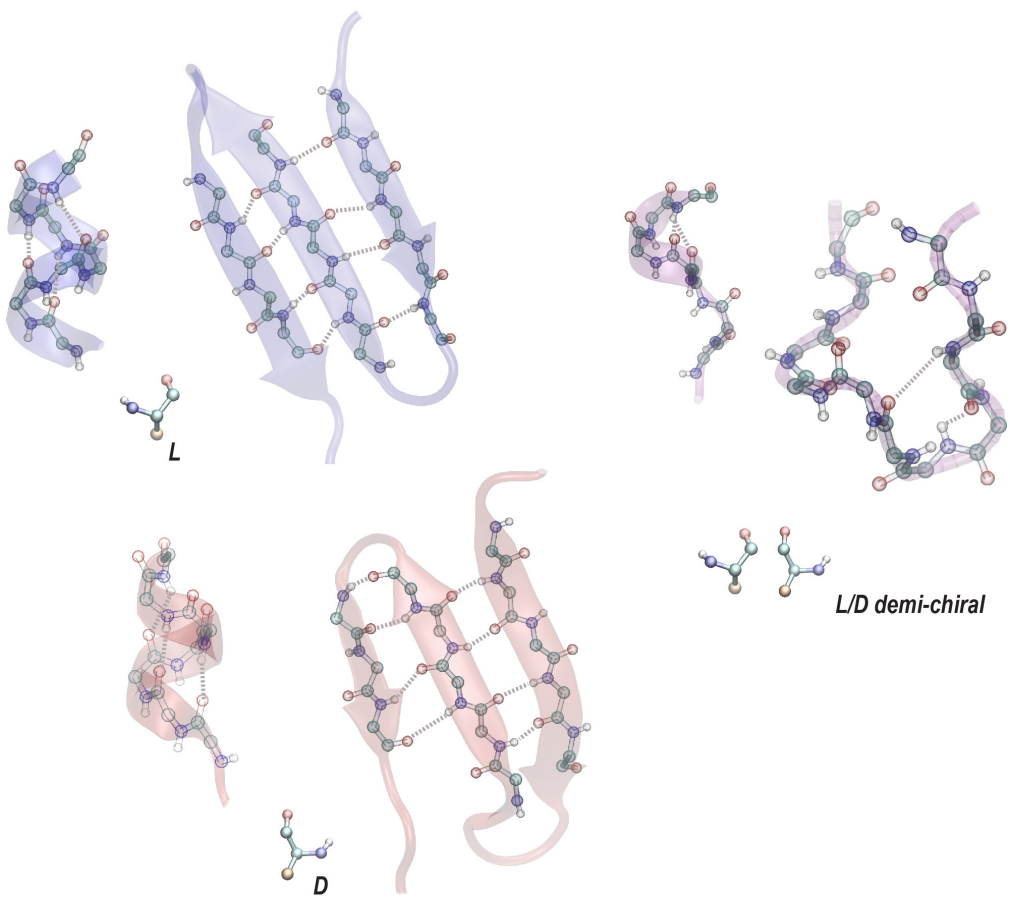

Figure 2. Hydrogen bond networks observed in L (purple), D (pink) homochiral and L/D demi-chiral artificial protein structures. Hydrogen bonds are shown as dashed lines between red oxygen and white hydrogen atoms. For clarity, only backbone atoms of the peptide structure are shown. The demi-chiral structure is a racemic mixture of L/D residues.

Figure 2). Hydrogen bonding induces the formation of regular secondary structures such as α-helices and β-strands. To do so requires sequential stretches of amino acids having the same chirality. Thus, lacking longer homochiral stretches of amino acids, demi-chiral proteins have shorter and less regular secondary structures; consequently, their hydrogen bond energy is dramatically less than in homochiral proteins. Backbone hydrogen bond energy increases as the D:L ratio deviates from 0.5. For example, in demi-chiral proteins, 30% of the residues have hydrogen bonds between backbone oxygens and hydrogens. In contrast, in homochiral proteins, roughly 60% of residues have backbone hydrogen bonds. As such, the compact conformations of demi-chiral proteins are less stable. Relative to modern proteins, the free energy of demi-chiral compact protein structures has a proportionally larger contribution of hydrophobic interactions between residues that are local in structure but not necessarily local in sequence. This prediction is compatible with the observation that, on average, ancient protein superfamilies contain more hydrophobic residues. Importantly, increased thermodynamic stability of the compact folded structure could be the driving force towards homochiral systems and does not require any selection for function. However, in more stable compact structures, the biochemically active state of the protein is more populated, and thus, they would have a functional advantage.

Global folds of demi-chiral proteins are the same as native proteins

Having generated a library of compact, demi-chiral protein structures, how similar are their global folds to modern proteins? To answer this, structural alignments (that identify the most significant structural match between two protein structures) were performed using the widely used TM-align algorithm. The TM-score ranges from [0,1], with a TM-score above 0.4 (whose *p*-value is 3.4×10^{-5}) indicating that the two folds are very similar if not identical. The results shown in Figure 3 demonstrate that demi-chiral proteins have the same global folds as modern ones. Of particular interest are ancient ribosomal proteins. Are the structures of the ancient ribosomal proteins (L1-6, L10-16, L18, L22-24,L29, L30, S2-5, S7-15, and S19) present? Ignoring long structureless tails, the structures of the individual domains of the ancient ribosomal proteins are also in the demi-chiral protein library.

Ligand binding pockets in demi-chiral and homochiral proteins are the same

Could ancient demi-chiral proteins bind the same types of metabolites as modern proteins? As shown in Figure 4, the surface of a globular protein has craters or pockets that can bind small molecule ligands including

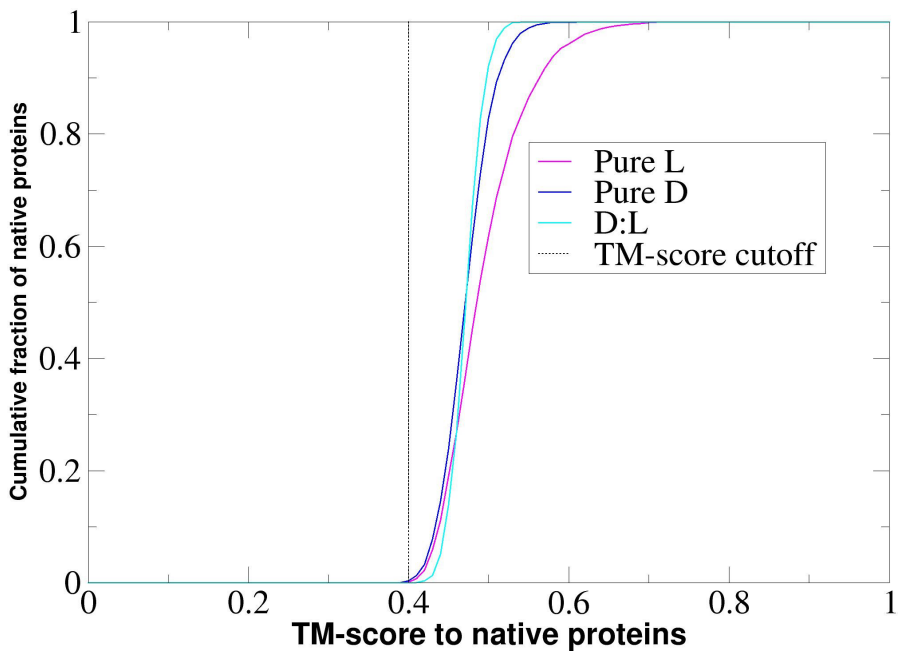

Figure 3. Cumulative fraction of native proteins whose best TM-score is less than or equal to the value on the abscissa obtained from aligning a representative PDB library to the 4516 protein structures in the pure L, pure D and D:L structural libraries. The TM-score cutoff for significant fold similarity of 0.4 is shown in the dashed line.

endogenous metabolites. Are the ligand binding pockets in modern proteins the same as in demi-chiral proteins? To address this issue, the pockets in the demi-chiral protein structural library were compared to a representative library of modern protein structures containing 213,100 pockets in approximately 36,800 proteins. Of native ligand binding pockets, 99.1% have a significant match to demi-chiral pockets. The reason this occurs is that the number of distinct ligand binding pockets in proteins is small (about 500) and results from defects in packing of secondary structural elements. Thus, demi-chiral proteins could bind (perhaps weakly) the same types of metabolites as contemporary proteins.

Native active sites are found in demi-chiral proteins

While their pockets are similar, one way for them to perform the same chemistry as modern proteins is to have L-amino acids located in a pocket that closely resembles a native protein's active site. In fact, the relevant protein backbone atoms are on average within 1 Å of the corresponding native protein's active site atoms. Thus, we examined whether there are appropriate constellations of L-amino acids in the library of demi-chiral protein structures that match the native active sites corresponding to 593 distinct Enzyme Commission

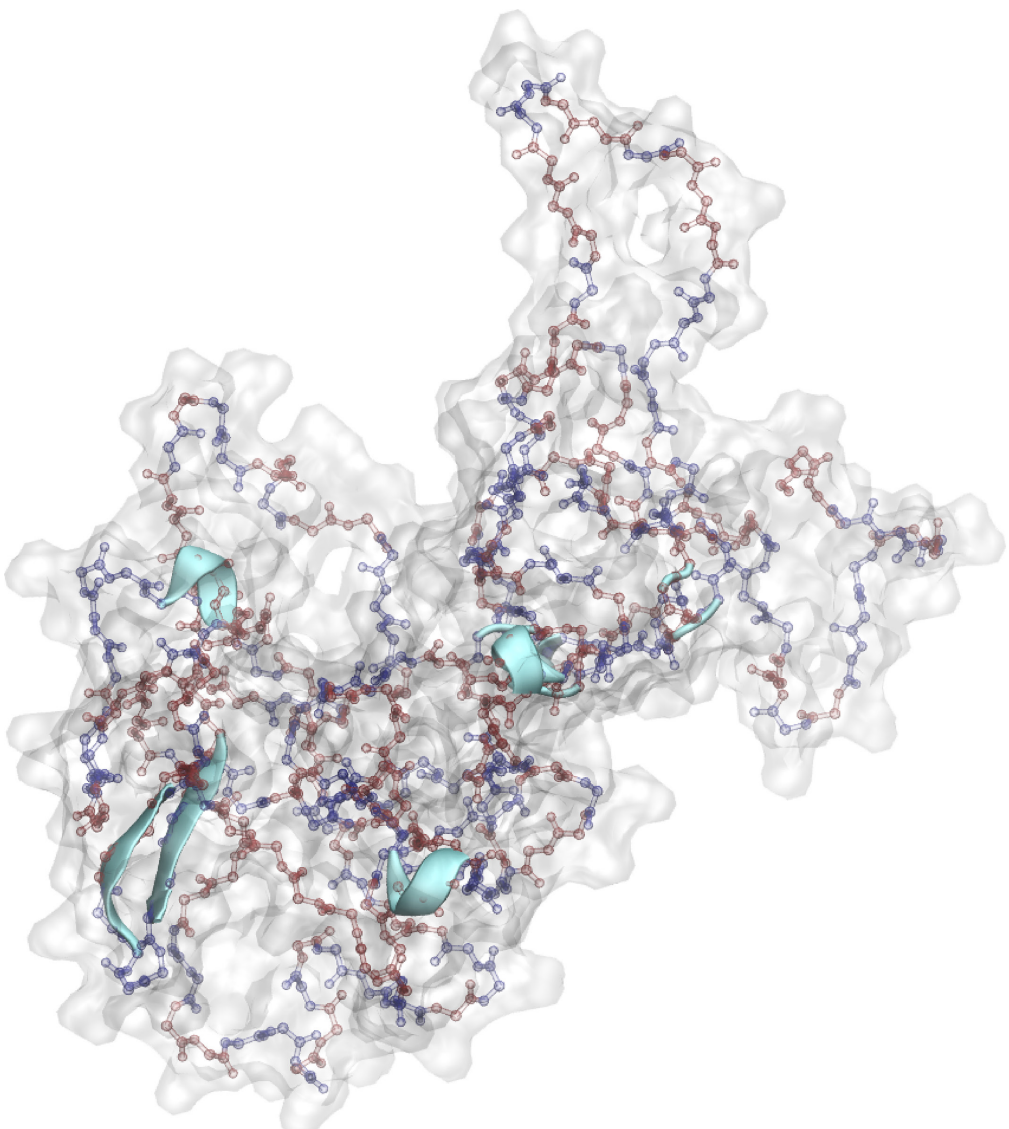

Figure 4. Pockets observed in a demi-chiral structure. Backbones of the L- and D-amino acids are coloured in red and blue, respectively. Regular α-helices and β-sheets are shown in a cyan cartoon representation. The white contour is the surface of the protein.

Table 1. List of enzymes in the demi-chiral protein library that are members of the minimal bacterial gene set

Functional category	EC number	Biochemical function	Number of sequences generated
DNA metabolism associated with the replication machinery	2.7.7.7	DNA-directed DNA polymerase	3
	6.5.1.2	DNA ligase	80
DNA repair	4.2.99.18	Class I DNA (apurinic or apyrimidinic site) endonuclease	18
	3.2.2.23	DNA -*N*-glycosylase	1007
	3.2.2.21	DNA-3-methylguanine glycosylase	3
Translation: aminoacyl-t-RNA synthesis	6.1.1.1	Tyrosine-t-RNA ligase	56
	6.1.1.6	Lysine-t-RNA ligase	686
	6.1.1.10	Methionine-t-RNA ligase	7
	6.1.1.11	Serine-t-RNA ligase	19
	6.1.1.12	Aspartate-t-RNA ligase	544
	6.1.1.18	Glutamine-t-RNA ligase	966
	6.1.1.19	Arginine-t-RNA ligase	1901
	6.1.1.22	Asparagine-t-RNA ligase	5423
Ribosomal function	2.1.1.48	t-RNA (uracil-2'-*O*-)methyltransferase	1
Protein processing	3.4.11.18	Aminopeptidase	82
	3.4.11.1	Aminopeptidase	1363
Transport	2.7.1.69	D-Glucosamine PTS permease	48
Glycolysis	4.2.1.11	Enolase	2
	4.1.2.13	Fructose 1,6-biphosphate aldolase	3412
	1.2.1.12	Glyceraldehyde 3-phosphase dehydrogenase	1436
	5.4.2.1	Phosphoglycerate mutase	1010
	1.1.1.27	L-Lactate dehydrogenases	1540
	2.7.1.11	6-Phosphofructokinase	10
	2.7.2.3	Phosphoglycerate kinase	2054
	5.3.1.1	Triosephosphate isomerase	12
Pentose phosphate pathway	5.1.3.1	Ribulose-phosphate 3-epimerase	2
Lipid metabolism	1.1.1.94	Glycerol-3-phosphate dehydrogenase	989

Continued

Table 1. Continued

Functional category	EC number	Biochemical function	Number of sequences generated
Biosynthesis of nucleotides	2.7.4.6	Nucleoside diphosphate kinase	1709
	1.17.4.1	Ribonucleoside diphosphate reductase	38
	1.8.1.9	Thioredoxin-disulphide reductase	5320
Biosynthesis of cofactors	2.7.7.3	Pantetheine-phosphate adenylyltransferase	341
	1.5.1.3	Dihydrofolate reductase	723

(EC) numbers (a classification scheme where if all four digits match, the protein catalyses the production of the identical product) from the Catalytic Site Atlas (CSA) library. Even in this small structural library, 76% of native active sites have structural matches. Relaxing the Root-Mean-Square Deviation (RMSD) criterion to 3 Å, then 92% of all active site geometries are recovered.

The above analysis did not require that the specific residue types (e.g. serine) match the corresponding active site residues. At a minimum, both the active site geometry and the appropriate residue types of an enzyme are required to perform the specified enzymatic catalysis. To explore if the appropriate L active site residues can be found in a random demi-chiral sequence library, 34,710,000 random sequences were generated for each demi-chiral protein structure containing the geometry of a CSA library active site. Of the CSA active sites, 88% have at least one randomly generated sequence where all catalytic residues with L-chirality exactly match. The smaller the number of active site residues, the more frequently that active site was independently discovered.

Table 1 shows a subset of enzymes that are members of the minimal bacterial gene set found in the demi-chiral protein library. Many functions needed for life are recovered including enzymes associated with DNA repair, protein processing, translation, glycolysis and cofactor and nucleotide biosynthesis; 8/10 enzymes involved in glycolysis, a pathway essential for energy production, are present. Even in this rather small demi-chiral protein library, many enzymatic functions necessary for life as we know it are generated, *without any selection for function*. We would expect that if the demi-chiral library were significantly expanded in size, the minimal bacterial gene set would become even more complete.

Table 2 presents a representative subset of contemporary metabolic pathways found in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database that are at least partially found in demi-chiral proteins. Key metabolic pathways that are at least partially recovered include purine, pyruvate, sugar and amino acid metabolism, the citrate cycle, fatty acid biosynthesis and lipid metabolism. On average, 17.7% of the enzymes in a given KEGG pathway have matches in the demi-chiral enzyme library. Of course, not all of these enzymes or for that matter structural types might have been present in the demi-chiral protein soup. Nevertheless, this study shows that demi-chiral proteins have the inherent capability of yielding a significant fraction of the biochemistry of life. Their biochemical function emerges from the bald requirements that they must be present and have minimal stability and activity.

Implications

This review focused on the results of computer experiments in which a compact fold library of demi-chiral proteins were generated without selection for function. The compact conformations of a protein are important as these have the capacity to form the pockets capable of binding small molecule ligands. In addition, such structures must be at least marginally stable so that they would be sufficiently populated to perform enzymatic function. Remarkably, compact demi-chiral protein structures recapitulate many key aspects of the folded structure and function of modern proteins but were likely less stable and catalytically efficient. Lacking long regular homochiral stretches of amino acids, their α-helical regions are shorter, and while extended states exist, they mostly lack the ability to form backbone hydrogen bonds. On average, the predicted stability of demi-chiral proteins is 53% of that of native proteins. They are relatively more stabilized by burial and pair interactions. This important qualitative result is independent of the particular force field used, suggesting it is true. Increased stability due to additional backbone hydrogen bonding would drive selection towards more chiral systems. As the ratio of L:D amino acids deviates from 0.5, proteins rapidly become more stable. The excess

Table 2. Summary of representative pathways found in the demi-chiral protein library which contain at least 10 matching distinct enzymes ranked by the number of enzymes generated at random

Number of enzymes	Type of pathway
246	Metabolic pathways
118	Biosynthesis of secondary metabolites
91	Microbial metabolism in diverse environments
27	Glycolysis/gluconeogenesis
21	Purine metabolism
20	Fructose and mannose metabolism
20	Carbon fixation in photosynthetic organisms
18	Pyruvate metabolism
18	Amino sugar and nucleotide sugar metabolism
17	Cysteine and methionine metabolism
16	Arginine and proline metabolism
16	Alanine, aspartate and glutamate metabolism
15	Pyrimidine metabolism
14	Glyoxylate and dicarboxylate metabolism
14	Glycine, serine and threonine metabolism
13	Propanoate metabolism
12	α-Linolenic acid metabolism
12	Tryptophan metabolism
12	Pentose phosphate pathway
12	Methane metabolism
11	Starch and sucrose metabolism
11	Glutathione metabolism
11	Galactose metabolism
11	Citrate cycle (TCA cycle)
10	Valine, leucine and isoleucine degradation
10	Phenylalanine metabolism
10	PI3K-Akt signalling pathway
10	Glycerophospholipid metabolism
10	Aminoacyl-t-RNA biosynthesis

L-amino acid composition in some meteorites could cause L-amino acid containing homochiral proteins to predominate. Similarly, the excess of d-sugar acids in some meteorites could be used to build chiral RNA using the primitive demi-chiral RNA polymerase found in the demi-chiral protein library. Alternatively, due a random fluctuation in D:L composition, some proteins might possess an excess of D- or L-amino acids. These would have more stable compact conformations and therefore were functionally superior. We would expect a similar conclusion to hold for demi-chiral DNA and RNA.

We found that demi-chiral proteins have many native-like protein properties. The approximate global folds of all native single-domain, protein structures are found in the demi-chiral protein library. This is consistent with previous work suggesting that protein compactness dictates a number of distinct folds of protein domains, which is remarkably small, about 1000. Consistent with this conclusion, a subset of demi-chiral proteins has the global folds of early ribosomal proteins. Which particular protein folds were adopted and when would depend on which demi-chiral proteins were present in the primordial soup.

Could demi-chiral proteins perform the chiral chemical reactions responsible for contemporary metabolism? While a demi-chiral protein contains an equimolar mixture of D- and L-amino acids, sometimes by chance all L- (or D-) amino acids are found in backbone geometries that are very close to the active sites of native proteins. Remarkably, without any selection for function, in this library of demi-chiral proteins, 86% of the 550 active sites associated with 456 distinct EC numbers have exact sequence matches, and all but two have matches if similar amino acids are also allowed. The most frequently generated enzymatic functions are those found in ancient proteins and involve the key biochemical processes essential for life including glycolysis, ribosomal function, translation and DNA synthesis. Thus, demi-chiral proteins possess the inherent ability to discover such functions at random. What was previously assumed to be emergent properties driven by protein evolution are actually just the intrinsic properties of compact proteins.

Our study on the origins of homochirality suggests that the RNA and metabolism first world ideas might actually be synergistic. The demi-chiral proteins composed of equal numbers of D- and L-amino acids that might have been present in the primordial soup possess many properties of contemporary homochiral proteins. Their global folds are the same and include those of ancient ribosomal proteins necessary for protein transcription. They form cavities like ordinary proteins and could perform the same biochemical functions, with the most ancient, essential ones being most prevalent. The biochemistry of life as we know it and

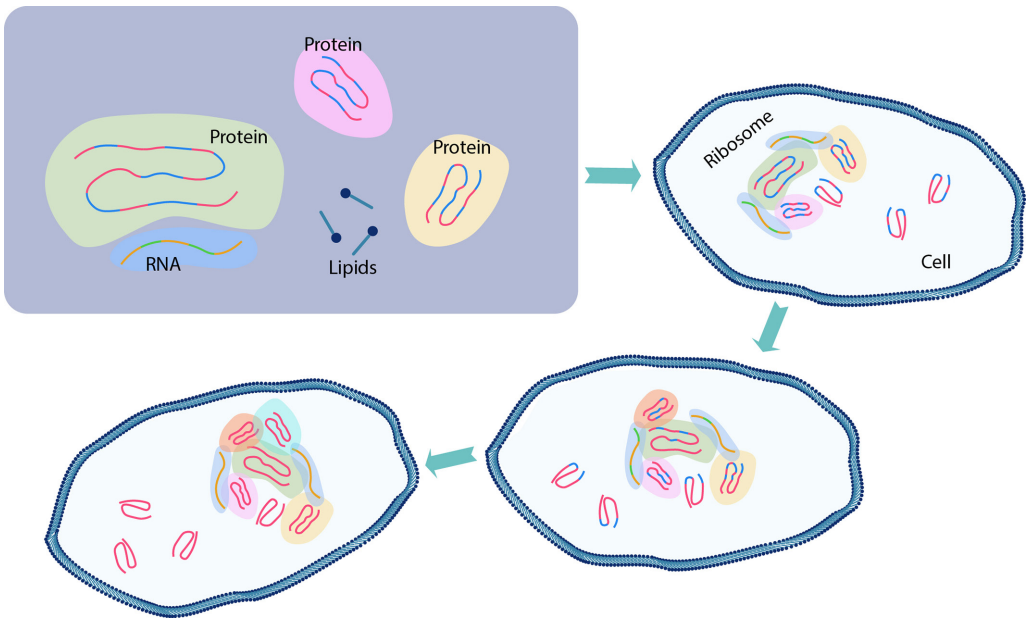

Figure 5. A possible path from ancient demi-chiral proteins and RNAs to modern homochiral biomolecules. The different colours of the backbones indicate the different chiralities of the basic building blocks of these molecules.

protein homochirality possibly result from the increased stability of the native state driven by backbone hydrogen bonds. The selection of D- vs L-chirality in proteins might have emerged from a fluctuation in composition or from the excess of L-amino acids in some meteorites. As schematically depicted in Figure 5, early demi-chiral proteins, while not as stable or functionally efficient as modern ones, could have synthesized chiral RNA as well as lipids which could form vesicles.

The resulting chiral RNA eventually combined with the early universal, ribosomal proteins (also present in demi-chiral structures) to make primitive ribosomes and enable more efficient, more chiral protein synthesis. This conjecture is supported by this work which suggests that the enzymes needed for both m-RNA and t-RNA synthesis and ribosomal function are found in the random demi-chiral protein library. This might yield a positive feedback loop where the breaking of chirality and emergence of metabolism and replication could have occurred quite close together in the primordial soup. These results suggest that the RNA and metabolism first world ideas might be synergistic. Early demi-chiral proteins could have synthesized chiral RNA as well as lipids which formed vesicles. RNA eventually combined with proteins to form ribosomes to enable more efficient protein synthesis, and thus, life started.

Finally, we note that other studies on related artificial protein systems have yielded many insights into the design principles of modern proteins. Thus, it is worthwhile to experimentally explore these predictions about the breaking of demi-chirality and the possible origins of the biochemistry of life.

Acknowledgements

We wish to thank Hongyi Zhou for performing many of the folding simulations done in this contribution. ■

Further reading

- Robertson, M.P. and Joyce, G.F. (2012) The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, DOI: 10.1101/cshperspect.a003608.
- Dyson, F.J. (1982) A model for the origin of life. *J. Mol. Evol.* **18**, 344–350.
- Lanier, K.A., Petrov, A.S. and Williams, L.D. (2017) The central symbiosis of molecular biology: molecules in mutualism. *J. Mol. Evol.* **85**, 8–13, DOI: 10.1007/s00239-017-9804-x.
- Skolnick, J., Zhou, H. and Gao, M. (2019) On the possible origin of protein homochirality, structure, and biochemical function. *Proc. Natl. Acad. Sci. U.S.A.*, DOI: 10.1073/pnas.1908241116.

Continued

- Canavelli, P., Islam, S. and Powner, M.W. (2019) Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **571**, 546–549, DOI:10.1038/s41586-019-1371-4.
- Glavin, D.P., Burton, A.S., Elsila, J.E., et al. (2020) The search for chiral asymmetry as a potential biosignature in our solar system. *Chem. Rev.* **120**, 4660–4689, DOI: 10.1021/acs.chemrev.9b00474.
- Skolnick, J. and Gao, M. (2013) Interplay of physics and evolution in the likely origin of protein biochemical function. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9344–9349, DOI: 10.1073/pnas.1300011110.
- Furnham, N., Holliday, G.L, de Beer, T.A.P et al. (2014) The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.* **42**, D485–489, DOI: 10.1093/nar/gkt1243.
- Kanehisa, M., Furumichi, M., Tanabe, M., et al. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361, DOI: 10.1093/nar/gkw1092.
- Skolnick, J. and Gao, M. (2020) The role of local versus nonlocal physicochemical restraints in determining protein native structure. *Curr. Opin. Struct. Biol.* **68**, 1–8, DOI: 10.1016/j.sbi.2020.10.008.

Jeffrey Skolnick is a Regent's Professor in the School of Biological Sciences in the Georgia Institute of Technology. He's had a long-standing interest in understanding the design principles of proteins with applications to the possible origins of the biochemistry of life. He has developed and applied approaches to proteomes for the prediction of protein structure and function, personalized medicine and predicting disease mode of action proteins, and small molecule drug efficacy and side effects with the goal of accelerating drug discovery. Email: skolnick@gatech.edu

Mu Gao is a Senior Research Scientist in the School of Biological Sciences in the Georgia Institute of Technology. Trained as a biophysicist, he has contributed to many computational algorithms for protein structure and function prediction, structural genomics analysis and biomolecular interaction studies. By applying these computational methods, he is interested in deciphering the basic principles of a living cell and in discovering effective therapeutics to take on health challenges.