Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Are natural proteins special? Can we do that?

Michael H Hecht, Shlomo Zarzhitsky, Christina Karas and Sarangan Chari

Abstract

Natural proteins represent a minuscule fraction of possible sequence space. These very rare sequences display remarkable properties: They fold into many different stable structures, and perform a wide range of complex biological functions. These two considerations — rarity and functionality — may suggest that natural proteins are somehow special. Is this true? We address this question by exploring attempts to recapitulate the *special* structures and functions of natural proteins into sequences designed *de novo*.

Address
Departments of Chemistry and Molecular Biology, Princeton University, Princeton, NJ 08540, USA

Corresponding author: Hecht, Michael H (hecht@princeton.edu)

Current Opinion in Structural Biology 2018, **48**:124–132

This review comes from a themed issue on **Proteins**

Edited by **Birte Hocker** and **Jakob Winther**

https://doi.org/10.1016/j.sbi.2017.11.009

Introduction

When we ask whether something is 'special' we are implicitly asking two questions: Is it rare? ('she is one in a million'); and Is it easy/hard to replicate ('I can do *that!*'). Are natural proteins rare? Can we replicate their structures and properties?

For a relatively short protein of 100 amino acids, there are 20^{100} possible sequences. It has been estimated that a collection containing one molecule of all these sequences would fill a volume larger than a mole of universes [1] (Figure 1). While the exact number of existent natural sequences is unknown, it is dwarfed by this number of possible sequences. By this criterion, natural proteins — a miniscule fraction of possible proteins — are rare; far more unusual than your friend who is one in a million.

This unusual collection of natural proteins arose in response to selective pressures. The surviving sequences enhanced the fitness of their hosts, while an almost unimaginable number of alternative sequences were lost to extinction. This may lead one to speculate that the survivors of life-or-death selections that operated over billions of years in myriad cells and organisms must surely be *special*. But are they? *Can we do that?*

Recent advances in genome sequencing, proteomics, protein design, and synthetic biology enable us to address these questions with more data and (hopefully!) more insight than ever before. It is now possible to assess whether we can create entirely novel proteins that recapitulate the key features of naturally evolved proteins. Can we produce non-natural sequences? Will they fold? Will they bind, assemble, and catalyze? Can we create novel proteins that sustain life? Can we do *that?*

Early steps toward protein design

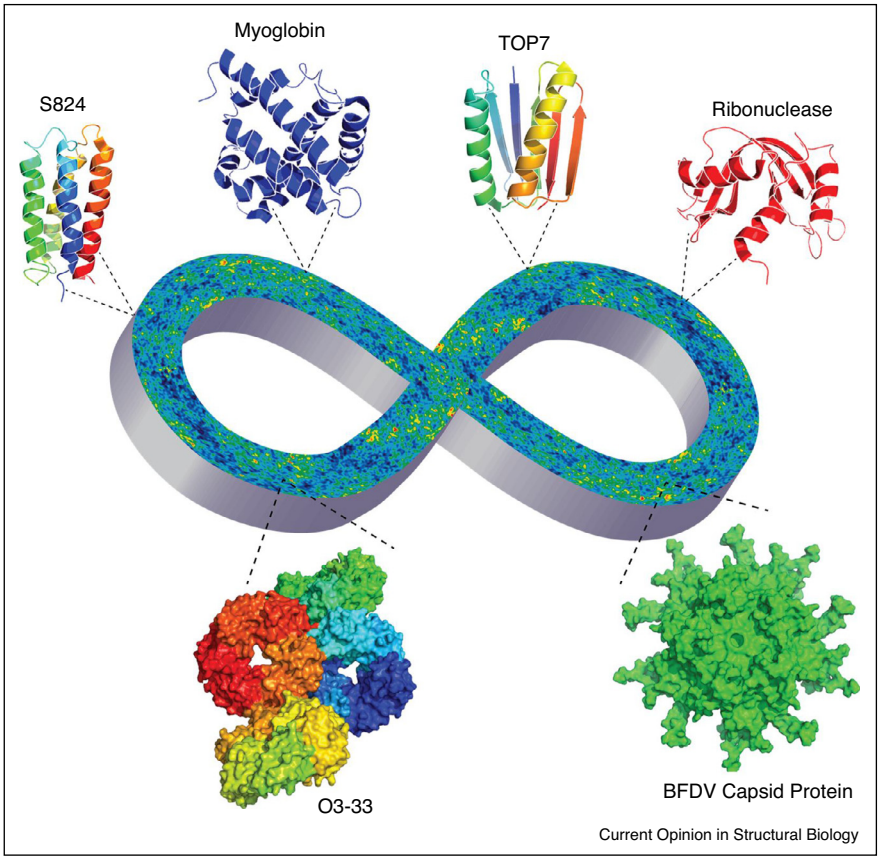
Since the seminal experiments in the 1960s by Anfinsen [2] and Merrifield [3], it has been clear that natural proteins can fold without assistance from any 'life force' provided by living organisms. A quarter of a century later, scientists began to ask whether *non*-natural sequences could also fold spontaneously. The initial goals were rather modest: Can one devise sequences that are unrelated to natural sequences, but nonetheless fold into simple 4-helix bundles? Early successes in the late 1980s and early 1990s included α-4, designed by Regan and DeGrado [4], and Felix, designed by Hecht, Ogden and the Richardsons [5]. These early studies showed that folding *per se* is not a *special* property of natural sequences. Artificial sequences fold too. So at least we can do *that.*

Novel proteins fold into stable structures: both natural and unnatural

While early work on protein design focused on simple structures, the field progressed rapidly, and protein designers soon tackled more challenging problems. Less than a decade after publication of the first 4-helix bundles, Dahiyat and Mayo demonstrated that it is possible to design novel sequences that fold into zinc finger domains, containing alpha structure, beta structure, and a bound metal [6]. Their novel sequence was chosen by a fully automated computational algorithm, and the resulting protein (FSD-1) folded into a structure that closely matched the design target. Importantly, in contrast to the previous designs (α-4 and Felix), FSD-1 formed a well-ordered — rather than dynamic — structure.

While FSD-1 showed that natural *sequences* are not special in their ability to fold into native-like structures, a

Figure 1



A universe of natural and novel proteins. The central image shows the cosmic background radiation of the early universe [13] superimposed on the symbol of infinity. Natural sequences are a miniscule fraction of the astronomical size of possible sequence space. Ribbon diagrams show a natural protein that binds a cofactor (myoglobin, 1MBN); a natural enzyme (ribonuclease, 1FS3): a *de novo* protein from a combinatorial library that folds into a native-like structure (S-824, 1P68); and a computationally designed sequence that folds into a novel structure (TOP7, 1QYS). Space filling models show a natural protein assembly (BFDV capsid protein, 5J37), and a fully designed assembly (03-33, 4DDF).

question remained: Are the natural *structures* that were selected by evolution somehow special? Are other structures and topologies possible? This question was answered by Kuhlman, Baker, and colleagues, when they designed TOP7, a protein with a non-natural sequence that folds into a structure not previously seen in nature [7] (Figure 1).

In the intervening years, a wide range of novel sequences and novel structures have been designed *de novo*. These include idealized α-helical structures, and a range of α/β topologies [8,9•]. Fully β-sheet proteins have also been designed [10]; however, because β-strands are prone to aggregate [11], progress in this area has been slower. Nonetheless, it is clear that both natural and unnatural sequences can fold into a wide range of natural and unnatural structures [12••].

A hallmark of natural proteins is their tendency to fold cooperatively. Although this is not universally true, most natural proteins fold and unfold by a two state mechanism without stable intermediates. In the early days of protein design, this feature seemed special. Designing novel sequences that folded cooperatively was challenging [4,5] and became a gold standard for early workers in the field. However, as the field of protein design matured, many (although not all) novel sequences were shown to fold cooperatively [6–8,14–16]. Thus, it appears that cooperativity *per se* is not a special property of natural proteins. We too can produce cooperative systems.

Novel proteins by the millions and trillions

The preceding sections highlight achievements in the design of individual proteins. Natural proteins, however, were selected from feedstocks containing myriad sequences. Are vast collections of protein sequences a special property of natural ecosystems? Can such collections be generated *de novo?*