Synthetic Proteins

E.-K. Ehmoser-Sinner, C.-W. D. Tan, *Lessons on Synthetic Bioarchitectures*, Learning Materials in Biosciences, https://doi.org/10.1007/978-3-319-73123-0_3

What You Will Learn in This Chapter

"Toward a Synthetic Genome" Section

We will now look at an example of human-designed proteins whose natural counterparts have never been observed. We will show how even nearly random protein sequences can have surprising biological functions. We will explain how the novelty of such proteins reflects the novelty of the genes encoding them, as well as how a minimal set of such genes capable of generating autopoiesis would constitute a novel and truly synthetic genome.

3.1 Synthetic Proteins: What Are They?

"Synthetic protein" is still not a precise terminology; it can describe a protein synthesized from living cells (e.g., harvested from a biofermenter), a hybrid macromolecule (e.g., a fluorescently labeled antibody), or even a synthetic alternative to a protein with a desired functionality (e.g., an enzyme), which can be based on completely synthetic compounds, synthesized in the laboratory. Let us start with the aim of such material development first: why bother with synthetic proteins when the proteins built by nature are obviously working quite well on our planet?

Such attempts are often related to pharmaceutical approaches, e.g., synthetic proteins as novel drugs, such as novel antibiotics, novel implant materials, and oncotargets. Antimicrobial peptides, for example, are on the way to revolutionizing the view on drug development, as inherent peptidic fragments of human albumin, for example, play a role in the context of virus infections and, moreover, in diabetes and Alzheimer disease, as recent findings have indicated.

Such molecules are targets for stabilizing and transformation endeavors in order to perform "communication" tasks in misregulatory contexts. This would be a major breakthrough and brings us back to the introduction of this book (see ▶ Chap. 1) regarding synthetic bioarchitectures as a converging field of research: communication with nature on the level of molecules.

Why is the molecular level of such interest and potential? It is not only to "understand" and "define" compounds; the atomistic/molecular level is simply the common denominator of life—whether a protein derives from humans or plants is irrelevant in the moment of its presence in a suitable environment. This is the prerequisite for biotechnology: when bacterial species produce active compounds based on the genetic code of humans—for example, human insulin, synthesized in bacterial species. Another example to demonstrate the potential of biotechnology in general is the possibility (and the ethical issue) of crossing the borders of theoretically all living species. The gene coding for an autofluorescent protein from *Photobacterium leiognathi* has been transferred into the higher plant specimen *Nicotiana tabacum*, in which it performs its function even it is translated in a very different biochemical and genetic context (Krichevsky et al., *PLoS One*, 2010, 5: Issue 11, e15461) (◘ Fig. 3.1).

The context of molecular biology offers a broad spectrum in employment of proteins for therapeutic use. Genetic manipulation strategies enable the synthesis of precise biological compounds, which are partly modified according to their desired function (optimized in specificity) or even combined with synthetic (e.g., carrier) materials.

And, of course, the race to obtain completely novel functionalities in designing proteins de novo has already started!

In the attempt to "build" synthetic proteins de novo, an interesting example demonstrates the capability of such autologous synthetic proteins.

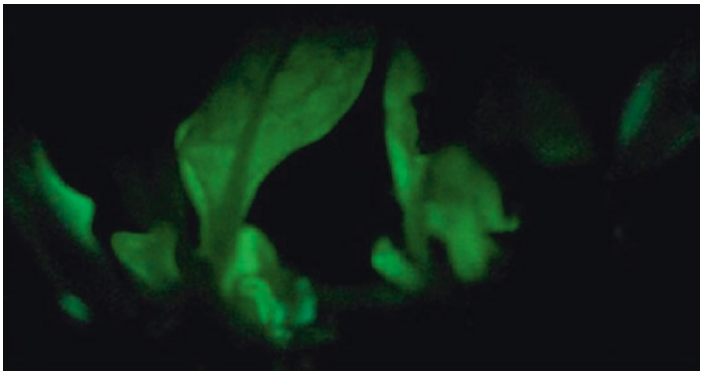

Fig. 3.1 Visual detection of autoluminescence in LUX-TrnI/TrnA plants. **a** Photograph taken in the dark with a handheld consumer camera (Nikon D200; AF-S Micro Nikkor 105.0 mm 1:2.8 G ED lens; exposures 5 min at f/4.5, 105 mm focal length, ISO 3200). **b** Photographs of transplastomic and wild-type plants taken with lights on or off (Krichevsky et al., *PLoS One*, ► https://doi.org/10.1371/journal.pone.0015461.g005)

3.2 Toward a Synthetic Genome

In ► Chap. 2, we examined how the J. Craig Venter Institute (JCVI) attempted to determine the minimal set of genes encoding bacterial life. This work went on to encourage attempts at the institute to construct a synthetic bacterial genome, using raw materials. Although the JCVI's work resulted in a synthetically assembled genome, the product was not bona fide synthetic, since the proteins encoded were of existing types. A truly synthetic genome is novel not because of the way it is constructed but in the proteins it encodes. As such, a truly synthetic genome should encode proteins that have never been observed in our biosphere. This has not been achieved yet.

We present the work of Fisher et al. as an example of how this might be possible in the future. In their report, they attempted to show that novel synthetic proteins could be biologically functional. To do this, they first generated 1.5×10^6 partially random 102-residue protein sequences. The only condition they set was that the sequences had to be able to form stable globular structures (Fig. 3.2). This was based on the assumption that a tertiary structure was necessary for most biological functions.

Plasmids encoding each synthetic protein were then transformed into 27 auxotrophs, which can only grow in rich media. Some of the synthetic proteins were able to rescue four of the 27 mutants (Fig. 3.3). Where the auxotrophs were transformed with plasmids encoding *lacZ*, no colonies formed in minimal media. In contrast, rescued mutants were able to form colonies.

Each mutation was rescued by a different set of synthetic proteins. Although each rescue allowed the host to grow in minimal media, the rescues tended to grow more slowly than wild-type hosts (Fig. 3.4). This could be because, unlike wild-type proteins, the synthetic proteins had not been optimized for function through evolution. As such, they might have been selected against in nature.

In trying to understand what functions these synthetic proteins might have, the researchers looked at the mutations that were rescued. Each mutation was in a gene encoding an enzyme critical for the survival of the host in a minimal medium (Fig. 3.5).

The synthetic proteins, despite being completely artificial, could somehow complement the function of these missing enzymes.

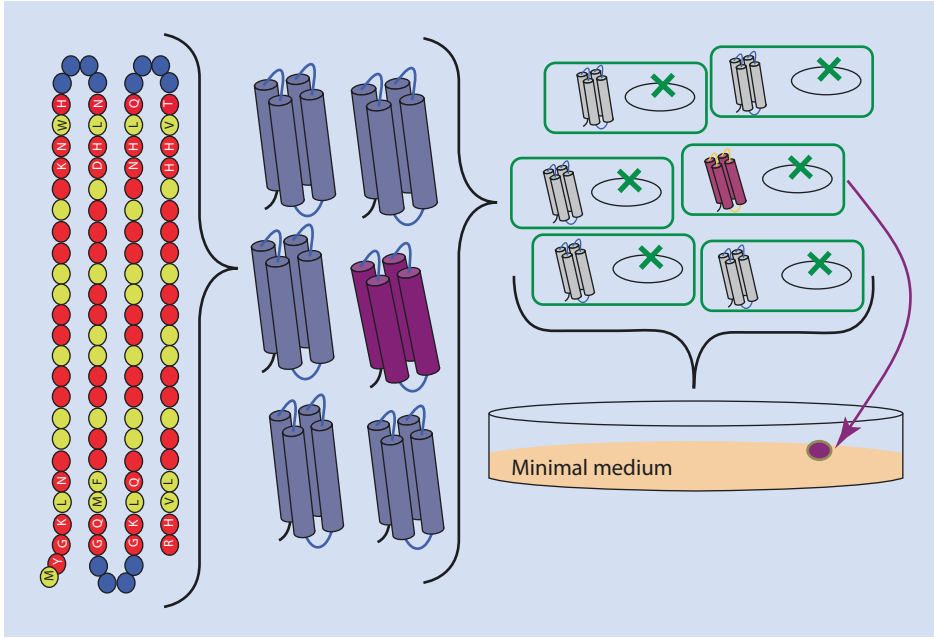

Fig. 3.2 Design of a collection of novel proteins and rescue of Escherichia coli auxotrophs (Fisher et al. 2011)

Complementation frequency and time required for colony formation on selective media (M9-glucose).

	Complementation Frequency	Time to Grow (Days)
Δ*serB*	~1/50,000	4
Δ*gltA*	~1/1,000,000	4
Δ*ilvA*	~1/3,000	3
Δ*fes*	~1/10,000	2

27 strains that grow in rich media (LB) on minimal media (M9 minimal glucose).

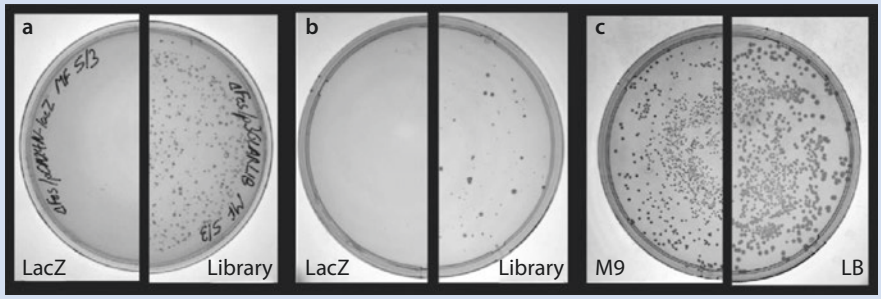

Fig. 3.3 Plasmids encoding each synthetic protein were transformed into 27 auxotrophs, which can only grow on rich media. Some of the synthetic proteins were able to rescue four of the 27 mutants (Δ*serB*, Δ*gltA*, Δ*ilvA*, Δ*fes*). Rescued mutants were able to form colonies. This demonstrates how even partially, randomly designed proteins can have effective biological functions

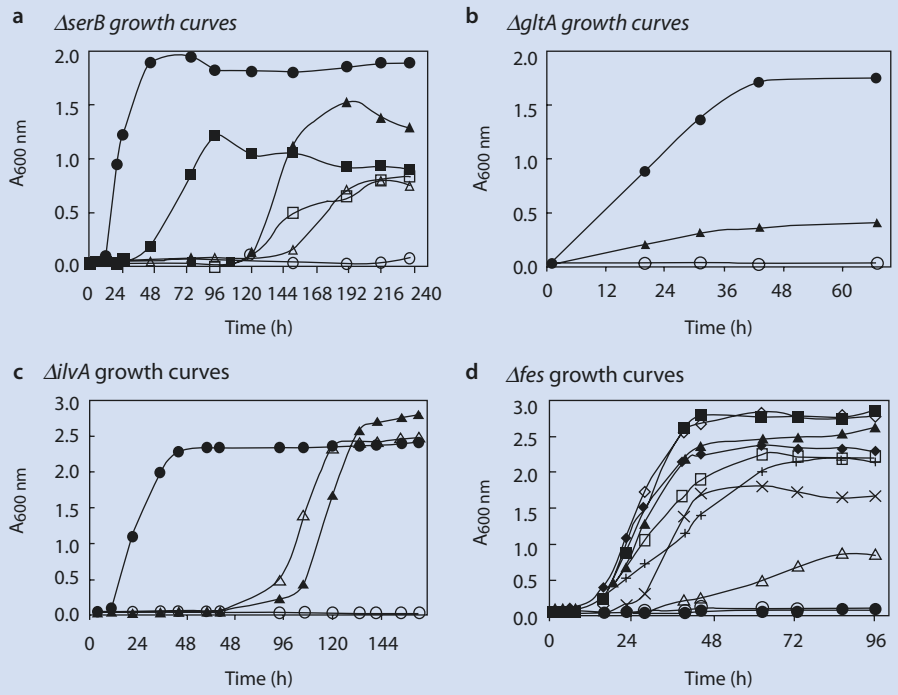

Fig. 3.4 Growth of auxotrophic strains of Escherichia coli in selective liquid media (Fisher et al. 2011)

- *serB*
  - phosphoserine phosphatase
  - responsible for the final step in serine biosynthesis
- *gltA*
  - citrate synthase
  - catalyzes an early step in glutamate biosynthesis:
- *ilvA*
  - threonine deaminase
  - catalyzes the first step in the production of isoleucine from threonine
- *fes*
  - enterobactin esterase
  - enables cells to aquire iron in iron-limited environments

(Over-expression is toxic.)

Survival under starvation!

Fig. 3.5 The figure details the product and function of the four mutated genes which the synthetic peptides had rescued. Interestingly, the four mutations were in genes encoding metabolic enzymes

It is still not clear how the synthetic proteins actually rescued the mutants. The following possibilities were suggested: that the synthetic proteins (1) perform the function of the missing enzymes; (2) allow the host to bypass the compromised metabolic pathway; or (3) are not, themselves, complementary but trigger the expression of complementary genes.

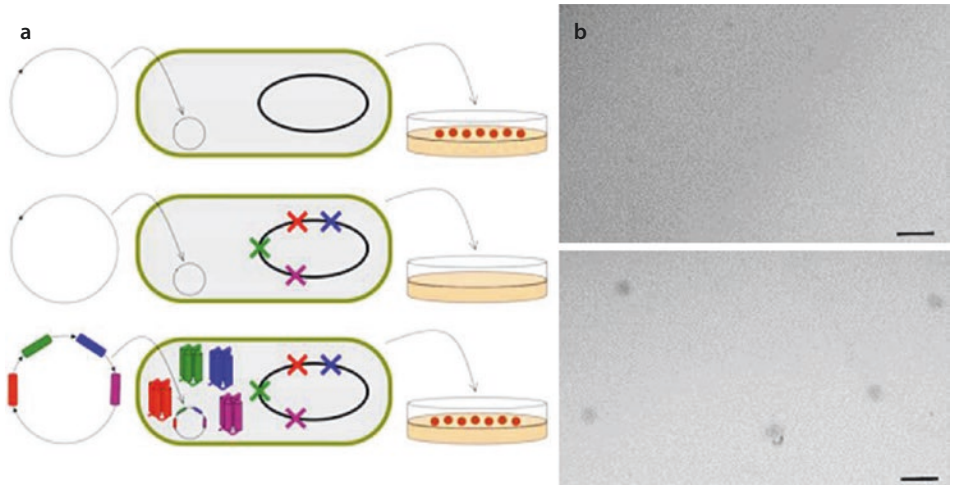

Fig. 3.6 The synthetic proteins were able to rescue auxotrophs carrying all four of the metabolic enzyme mutations. This demonstrates how multiple synthetic proteins can partially replace the function of portions of a genome, suggesting that an entire genome might be similarly replaced, or designed, to produce a viable system

The experiment was modified to include auxotrophs that carry all four of the rescuable mutations (Fig. 3.6). This mutant was then transformed with four of the complementary synthetic proteins. The synthetic proteins were able to rescue the auxotrophs, despite the fact that nearly 0.1% of the host genome had been compromised by the mutations. Might it be possible to similarly replace the remaining 0.9% of the host genome?

This work demonstrates that existing genes and proteins may not constitute all that is functional in existing living systems. Genes encoding novel proteins may be able to address existing needs and may even confer novel functions on their hosts without compromising their viability. As we approach a minimal library of synthetic proteins capable of sustaining life, we approach the emergence of a truly synthetic genome.

Conclusion

The term "synthetic proteins" describes amino acid–derived compounds that are (1) synthesized by a species under defined conditions (e.g., human insulin synthesized by bacteria); (2) manipulated with additional materials; or (3) derived from de novo design concepts based on the conventional amino acid material context or from noncanonical or even synthetic materials.

Take-Home Messages

"Toward a Synthetic Genome" Section

1. It is argued that a truly synthetic genome should be one that encodes synthetic proteins, not just one that is constructed by humans.
2. It is argued that for proteins to have a biological function, they require tertiary structures at least as complex as stable globular forms.
3. It is possible for specific, de novo globular proteins, comprising partially randomly generated sequences, to rescue conditional mutants.
4. Such synthetic proteins are not as effective as their natural counterparts, suggesting that they might have been selected against in nature, hence their scarcity or novelty.
5. A genome comprising genes that encode only synthetic proteins, and that would give rise to an autopoietic system, might be considered a truly synthetic one.

Further Reading

Fisher MA, McKinley KL, Bradley LH, Viola SR, Hecht MH (2011) De novo designed proteins from a library of artificial sequences function in Escherichia coli and enable cell growth. PLoS One 6:e15364

Ow DW, Wood KV, Deluca M, Dewet JR, Helinski DR, Howell SH (1986) Transient and stable expression of the firefly luciferase gene in plant-cells and transgenic plants. Science 234:856–859