Synthetic biology is a rapidly growing field that has emerged in a global, multidisciplinary effort among biologists, chemists, engineers, physicists, and mathematicians. Broadly, the field has two complementary goals: To improve understanding of biological systems through mimicry and to produce bio-orthogonal systems with new functions. Here we review the area specifically with reference to the concept of synthetic biology space, that is, a hierarchy of components for, and approaches to generating new synthetic and functional systems to test, advance, and apply our understanding of biological systems. In keeping with this issue of *Current Opinion in Structural Biology*, we focus largely on the design and engineering of biomolecule-based components and systems.

**Addresses**

1 School of Chemistry, University of Bristol, Bristol BS8 1TS, UK  
2 Department of Biochemistry, University of Bristol, Bristol BS8 1TD, UK  
Corresponding author: Woolfson, Derek N (D.N.Woolfson@bristol.ac.uk)

**Current Opinion in Structural Biology 2008, 18:1–8**  
This review comes from a themed issue on Engineering and design  
Edited by Dek Woolfson and Lynne Regan

0959-440X/$ – see front matter  
© 2008 Elsevier Ltd. All rights reserved.  
DOI 10.1016/j.sbi.2008.06.006

**Introduction**

Complexity in Nature is astounding, and attempts to mimic it present considerable challenges and potential rewards. This complexity stems from a hierarchical organization of biomolecular components and layers of interactions between them. Encouragingly, many aspects of both the components and their interactions are becoming increasingly understood. As outlined below, this hierarchical view and our improved understanding of and ability to engineer biology are the cornerstones of synthetic biology.

We refer to the potentially vast arena in which synthetic biologists can operate as *synthetic biology space*. This is represented as a plot of complexity of components against some indicator of how divergent from nature these are (i.e. how ‘synthetic’ the components are) (Figure 1). We find this useful in two respects: First it provides a framework to chart routes toward the common goal of creating multi-component, encapsulated, functional systems; second, it allows a wide variety of studies to be grouped into a small number of general approaches. We believe that this will be useful in defining and, hopefully, helping to develop the exciting and broad area of synthetic biology.

For this review, because of the breadth of topics that contribute to this emerging field, we found it necessary to refer to classic studies from the past two decades, reviews in various areas from the past five years, as well as more recent work from the past three years.

**Synthetic biology space: hierarchies of components, interactions and approaches**

At the base of the hierarchy is a set of *basic units*—amino acids, nucleic acids, sugars and lipids1 (Figure 1). One level of complexity above these are what might be termed *tectons*. This term is borrowed from supramolecular chemistry [1], where it is used to describe programmed molecular components and nanoscale building blocks [2]. An example of a nucleic acid tecton would be a short oligonucleotide containing the information for further assembly through interactions with other tectons. Similarly, an amino acid based tecton would be a polypeptide designed to form stretches of self-assembling α-helix or β-strands. Importantly, a tecton is something more than a simple element of secondary structure: It implies that the element contains information about its further assembly into prescribed higher order structures. Combining tectons leads to the next level in the hierarchy, in which *self-assembled units* are formed through interactions programmed into tectons. For peptides and proteins, autonomous folding motifs would be self-assembling units. By prudent organization of such units one can arrive at *functional assemblies*. As with tectons, the definitions of *self-assembling unit* and *functional assembly* encompass functional protein and DNA tertiary and quaternary structures. With further organization, interacting networks of functional assemblies – that is, *systems* – can be constructed. In Nature, complex interacting components of a system are almost always contained, or encapsulated, within lipid membranes, which enable cells to maintain control over their environments, and the biochemical processes they conduct.

Here, we use the concept of *synthetic biology space* (Figure 1) to capture the assembly of these various components and

---

1. There are, of course, many other small molecules involved in biological processes. However, a large fraction of the structural complexity of organisms can be represented in this small subset of building blocks.
2 Engineering and design

Figure 1

Synthetic biology space [2]. An approach or study in synthetic biology is resolved according to where it (or its natural equivalent) appears in the natural hierarchy (y axis), and by some measure of how synthetic it is (x axis). Colored arrows indicate approximate routes through synthetic biology space taken by studies in any of the four approaches to synthetic biology described in the main text: genome engineering, biomolecular engineering, biomolecular design, and artificial protocell design. Blue arrows indicate approaches usually conducted in vivo and orange arrows indicate in vitro approaches. On the left, the various levels in the natural hierarchy and their ranges are described, along with illustrative natural examples from various points in the hierarchy.

the resulting hierarchy, as well as to highlight approaches and recent studies under the general umbrella of synthetic biology. At first, some of the studies appear to have entirely different origins and objectives. However, by placing them in a common framework, it is possible to recognize how each contributes toward the shared goal of generating complex synthetic systems.

Synthetic biologists may access the hierarchy of Nature at any level by making alterations to existing natural systems at one, or a number of levels. Studies in synthetic biology can generally be classed as either in vivo or in vitro [3] and may be further subdivided according to the approach they take to the problem at hand: genome engineering, biomolecular engineering, biomolecular design or protocell-design projects (Figure 1) [2]. These are not sharp-edged definitions, but broad classifications that we find useful. For example, genome engineering refers to approaches like that taken by Venter and colleagues to construct synthetic chromosomes for whole or minimal organisms; biomolecular engineering includes approaches such as the BioBricks initiative, which aims to create a toolkit of functional units (usually natural protein components) that can be introduced to present new orthogonal functions in living cells; the biomolecular-design approach refers to the general idea of the de novo design and combination of biomolecular components; and the protocell approach includes ambitious projects to make self-replicating, encapsulated systems from entirely synthetic components.

The task of each approach is similar: To create a more synthetic entry at a higher level of complexity by manipulating a part of the preceding level. Hence, advances in synthetic biology tend to take us up and to the right in the synthetic biology space of Figure 1. The most complex and least natural systems – which are likely to be the most difficult to achieve – tend to be found in the top-right portion of synthetic biology space. Indeed, this region currently remains unoccupied, and we speculate that this will be reached only by using a combination of basis sets² contained by a membrane (or membrane-like) layer.

Routes to complex systems

Basic units

One challenge is to increase the number of building blocks, and by implication the repertoire of chemistries, that are accessible to synthetic biology. For example, bio-inspired building blocks include the DNA analog PNA, in which purine and pyrimidine base-pairing is natural, but the sugar–phosphate backbone is replaced by N-(2-aminoethyl)-glycine units linked by peptide bonds [4]. Various non-proteinogenic α-amino acids can also be used to bring new chemistry to proteins within the more conventional polypeptide framework [5]. Further along the divergence axis (Figure 1) β-peptides have been shown to adopt distinct secondary structures [6] making them potential tectons (Figure 2) and recently tertiary structures bringing them to the self-assembled unit level [7,8]. This has been extended yet further up the hierarchy with the recent assembly of β-peptide-based fibers [9].

It is also possible to design building blocks that diverge completely from Nature. Examples include pyridine dicarboxamides [10]; helicogenic polyisocyantide systems [11]; anthranilamides [12]; quinoline oligamides [13]; benzoylurea oligomers [14] (Figure 2); and naphthyridine foldamers [15]. Onto these relatively rigid templates, a variety of functional side groups can be appended, analogous to natural side-chain functionality. This ‘blank-sheet’ approach is very tempting, as it presents the possibility that synthetic systems could effectively isolate structure from function, and hence makes nanostucture design much more routine. However, there are inevitably drawbacks: First, there is no guarantee that such systems could reproduce the complexity and dynamics inherent in, and so important to natural assemblies; second, for purely synthetic building blocks there is currently no analogous infrastructure for their replication, regulation, segregation, and turnover, as there is for the natural and, in some cases, non-natural [5,16,17] building blocks of cells.

² This could be an ‘all-natural’ basis – mixtures of proteins, nucleic acids, sugars, and lipids, among others – or an entirely synthetic basis, or indeed anywhere in between.
Progress on biomolecular-design approaches to synthetic biology. The interactions in DNA are well understood, and have led to new DNA-based nanostructures (self-assembled units) [22], and functional assemblies such as a DNA motor [24**]. Moving right to ‘less natural’ building blocks, an example of a self-assembled helical β-peptide foldamer is shown. Taking a further step, there are studies attempting to form complex structures from a large variety of building blocks that are entirely unrelated to natural ones, as illustrated by the oligobenzoylurea-based system [14].

**Synthetic tectons**

The impact in molecular, cell, and now synthetic biology of stringing together natural bases to form new DNA sequences cannot be overstated. Firstly, through recombinant DNA technology, it allows the facile production of mutated polypeptides at all levels of the hierarchy, which molecular, cell, and synthetic biologists can exploit to manipulate the exquisitely evolved apparatus of the cell. Indeed, this ability led directly to genome engineering and biomolecular engineering approaches in synthetic biology [18**,15**]. Secondly – because of the straightforward relationship between sequence and structure through Watson-Crick base-pairing – DNA fragments can be used as tectons for self-assembly in their own right (Figure 2). As a result, understanding how to design and produce structures from DNA based building blocks far exceeds that for protein-based equivalents [19**,20]. It is now possible to self-assemble a DNA nanostructure in virtually any two-dimensional structure [21**], a growing number of three-dimensional forms [22,23], and even molecular motors [20,24**].

Interactions between residues in polypeptides are much less well understood, which has consequences in designing peptide-based and protein-based tectons as self-assembling systems in synthetic biology. However, there has recently been progress in understanding sequence-to-structure relationships in proteins. It is relatively simple to design a peptide sequence that adopts one of α-helical, β-strand, or disordered peptide structure, but, as explained above, this is not tecton design since there is no necessary information for further assembly in such blocks. However, peptides such as coiled-coil-forming α-helices are good examples since they contain information for further assembly; in this case, binding to other coiled-coil-forming helices to form helical bundles. For these structures, rules for assembly are becoming very well established [25], and detailed information about the energetics of coiled coil interactions is now available [26,27**,28]. With such understanding, tectons can be created to spontaneously form self-assembled units including prescribed oligomers and a wide variety of novel coiled coils [25]. For this α-helical motif as well as for β-strand and collagen motifs, considerable progress has been made toward the assembly of tecton based fibrous materials, which resemble components of biological cytoskeletons and exoskeletons [29,30]. Recent progress in this area includes the definition of supramolecular order and ultrastructure in designed coiled coil and collagen-based systems [31*,32*], the introduction of switchable elements in amyloid-like and coiled coils systems [33,34], and the control of coiled coil fiber assembly and properties [35]. Similar progress has been made using peptide-containing hybrids to render a wide range of fibrous materials [36,37].
4 Engineering and design

Self-assembled units
A small, autonomously folding domain such as a zinc finger (Figure 3) contains a few small tectons (an α-helix and two β-strands) with the information necessary to fold the peptide into discrete, compact, functional self-assembled units. By collating sequence information for such domains, a consensus sequence can be created that captures this information in terms of the preference of each residue at each sequence position. In some cases, as with zinc fingers [38], this consensus sequence is all that is necessary to redesign existing domains. More recently, a variety of more sophisticated knowledge-based potentials have been derived from sequence databases [39]. These extract correlations between amino acid positions and extract can capture more information than a flat consensus sequence. This technique has been demonstrated on a variety of folds including PDZ [40], SH3 [41], and WW domains [42,43].

Structure-based energy potentials build further on this information from sequence alignments. In their landmark study, Dahiyat and Mayo use such an approach to create a de novo peptide with three-dimensional elements of a zinc finger, but with little homology to the natural sequence, and notably without the usual zinc coordination [44]. Kuhlman et al. extend this principle to the design of non-natural self-assembled units with Rosetta Design [45], a computer program that employs steric and residue-specific bonding parameters to design sequences to assume a target backbone trajectory. In principle, this allows one to draw almost any desired topology of final self-assembled unit and arrive at a primary sequence to produce it. Indeed, The Baker group use it to design an entirely new protein fold, TOP-7 [45].

This ability in design also opens up possibilities for engineering new ‘switching sequences’ (Figure 3), that is, a single polypeptide that adopts two folds, depending on the prevailing conditions. This property is crucial for the function of many natural proteins, as it allows sensing and signal transduction. As such, switching elements will undoubtedly become important components in synthetic biology [18**]. There are now a number of designed switches in the literature [46], including zinc finger-to-coiled-coil [47,48], antiparallel-to-parallel-coiled-coils [49], and coiled-coil-to-β-sheet switches [50].

Functional assemblies
In Nature many complex functions have evolved through combining self-assembled units either covalently or by oligomerization. Synthetic biology can expand on this by fusing functional units together in new combinations. In essence, molecular and cell biologists have already applied this approach to probe and report on biochemical interactions and events in the cell; for instance, the established use of GFP-fusion proteins [51], the yeast two-hybrid system [52] to assess the localization of and interactions made by proteins of interest.

In the case of engineering structural scaffolds that may be of use in synthetic biology, proteins with different oligomerization symmetries have been fused to create range of supramolecular assemblies, including cages [53], filaments [54], arrays [55,56], and rings [57].

With the view of making functional assemblies and systems, the BioBricks project (www.biobricks.org), and its registry of parts, epitomizes a plug-and-play approach to genetic engineering to generate orthogonal functions in cellular hosts (a so-called chassis) [58] (Figure 4). This is a global effort to standardize what we term here the biomolecular engineering approach (Figure 1) and provide a ‘toolkit’ of standard biological parts. The development of such a toolkit would help transform the field of synthetic biology into a true engineering discipline. Indeed, this type of approach is already delivering interesting systems, including an engineered E. coli that senses light and records images [59]. Developments in this aspect of synthetic biology have been summarized by Voigt [18**]. Arguably, combinations of genes of this type are systems in themselves, which leads to the top-most level in the hierarchy of Figure 1.

Toward complex systems
The design of complex systems will require multiple structural and functional components to be encapsulated in some fashion. In Nature this is achieved largely using
lipid membranes. Although lipids should be counted among the basic building blocks from Nature, they are distinctly different in how they assemble; they associate rather than polymerize, which limits their information content and potential as tectons. However, there is enough information in the head and tail groups to generate a range of lipid-based 3-D assemblies, such as bilayers and micelles, and to control the physical behavior of the resultant structures. The physical chemistry of lipids is relatively well understood, which allows prediction and design [60,61]. There are, as always, intricacies to the resulting structures that are more difficult to predict, such as membrane permeability or fluidity of a mixed lipid system [62]. Advances in synthetic biology will probably involve learning how to bridge the water-membrane interface, for example, with the reliable design of integral membrane proteins [61,63].

Of course, synthetic biologists can make use of a wide range of additional methods for encapsulation [64], such as non-lipid-based systems, polymersomes [65], and water-in-oil emulsions [66]. Emulsions have already shown great promise in encapsulation, particularly for in vitro evolution of new enzymes [66,67–69]. Recently, Bayley and colleagues have shown that lipid-coated water droplets immersed in oil offer potential for studying integral membrane proteins and as routes to new systems – in this case, networks of droplets – with possible functions including biobatteries [70**].

Perhaps at the current extreme of such approaches, there are efforts that aim ambitiously to produce complex systems using completely synthetic basic units. For example, the ‘Protocell’ project [71], which is placed toward the far-right of synthetic biology space (Figure 1) and will entail the development of new genetic information carriers, metabolites, metabolic processes, functional assemblies, and encapsulation material [72] (Figure 4).

Returning briefly to more biological systems, genome engineering approaches are now emerging to deliver minimal genomes [73*], and the chemical synthesis of whole genomes [74**]. Together these technologies could lead the introduction of synthetic chromosomes into natural cellular hosts, or cell-like compartments.

**Conclusions**

By viewing synthetic biology in terms of synthetic biology space, it is possible to unify a diverse range of investigations, highlight their interrelationships, and see routes toward the production of complex functional systems. Also, it is possible to identify how some studies may
6 Engineering and design

contribute to the development of synthetic biology, and at which points in Nature’s hierarchy such design and engineering studies can provide useful information. Finally, it should be noted that the technological aspect of synthetic biology is clear and exciting to focus on. However, such advances will only be brought about through increased understanding of biological systems. This link between developing and testing our understanding of biology will be vital to the success of synthetic biology.

Acknowledgements

We thank members of the Woolfson group for many helpful discussions, and the BBSRC (grant number E022359) and the HFSP (RGP512007) for funding.

References and recommended reading

Papers of particular interest, published within period of review, have been highlighted as:

- of special interest
- of outstanding interest


21. Rothemund PWK: Folding DNA to create nanoscale shapes and patterns. Nature 2006, 440:297-302. This landmark paper demonstrates how virtually any two-dimensional shape can be designed and self-assembled from oligo-DNA strands. Additionally, surface features can be included in the design.


27. Acharya A, Rishi V, Vinson C: Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). Biochemistry 2006, 45:11324-11332. The authors have exhaustively characterized effects on coiled coil stability of the substitution of 10 different amino acids into a & a' positions of the coiled coiled heptad repeat.


31. Papastolou D, Smith AM, Atkins ED, Oliver SJ, Ryadnov MG, Serpell LC, Woolfson DN: Engineering nanoscale order into a
32. Rele S, Song Y, Apkarian RP, Qu Z, Conticello VP, Chaikof EL: o-Periodic collagen-mimetic microfibers. J Am Chem Soc 2007, 129:14780-14787. Like reference [31], this paper presents the design of a peptide-based fibrous system with exquisite nanoscale ultrastructure. In this case, the design is based on the tripeptide-motif for fibrillar collagens.


This review is an excellent introduction into using water-in-oil emulsions for femtolitre-scale chemistry, and high-throughput techniques. Emphasis is on using these techniques for the study of in vitro directed evolution.


A novel idea for encapsulation and study of both water-soluble and membrane-soluble components is presented. Water droplets coated in lipid monolayers are generated by adding water-to-oil-lipid mixtures. The droplets can then be fused to generate networks of droplets connected by lipid bilayers, which can house integral membrane proteins.


8 Engineering and design


In this article the authors probe the entire genome of *Mycoplasma genitalium* and identify 382 genes essential for the viability of the organism.


On the basis of the work presented in reference [73], the authors are able to demonstrate the complete chemical synthesis of the genome of an entire organism, *Mycoplasma genitalium*. To date, this is the largest and most complex example of the complete chemical synthesis of genomic DNA.


Here the authors present a detailed review of protocell design in synthetic biology, with a large number of references and discussions of the rational behind this approach as well as many practical and theoretical considerations.