Creating of biomolecular structures with proteins or nucleic acids

Genevieve C. Pugh, Jonathan R. Burns, Stefan Howorka*

Department of Chemistry, Institute of Structural and Molecular Biology,
University College London, 20 Gordon Street, London WC1H OAJ (UK)

*Corresponding author: s.howorka@ucl.ac.uk

The molecular building of protein, DNA, and RNA nanostructures is relevant in many areas of biological and material sciences. Nanoscale engineering was pioneered with proteins, yet DNA is rapidly gaining traction. But, what are the advantages of the different biopolymers and which is best suited for a given molecular structure, function, or application? In this Review, we evaluate proteins' and DNA/RNA's different structural properties, possible designs and synthetic routes for functional nanostructures. By comparing protein engineering and DNA nanotechnology, we highlight molecular architectures that are relevant for applied and fundamental science.

Introduction

Designing and engineering proteins is both scientifically exciting and of practical relevance in many areas including biotechnology and biomedicine. Protein engineering has provided, for example, new enzymes for industrial biocatalysis[1-4], highly tailored antibodies for cancer treatment[5] easily trackable fluorescent proteins for biological research[6,7], and efficient polymerases for forensic DNA detetection[8].

Molecular construction using DNA is rapidly catching up, as evidenced by DNA aptamers and rationally designed DNA nanostructures[9-18]. This use of DNA is surprising as its biological role lies in encoding genetic information and influencing gene expression. Nevertheless, nucleic acids have been re-tooled into artificial recognition agents and structural scaffolds for applications in materials science, biophysical research and increasingly in synthetic biology. DNA nanotechnology, in particular, has drastically expanded the repertoire of rationally designed nanostructures beyond RNA and DNA aptamers that can bind proteins or other ligands[19-24].

With protein engineering and DNA nanotechnology growing rapidly, several fundamental questions emerge. What are the relative strengths of polypeptides, DNA and RNA in terms of their structural and functional properties, as well as the ease of engineering? Which molecular building material is best suited for a given application? Moreover, considering that synthetic biology and nanobiotechnology are exciting areas, what are the most promising applications of new bio-nanostructures? Answering these questions is important as protein engineering and DNA nanotechnology are often seen as separate research areas with limited interaction.

In this Review, we synoptically compare protein, DNA and RNA as building materials for nanostructures with defined architecture and function. By answering the above questions, the Review aims for a unifying understanding of strengths and weaknesses of these biopolymers. The survey starts with a side-by-side evaluation of proteins and DNA/RNA in

terms of their physico-chemical properties, the principles for their rational design and engineering, and methods for their chemical and biochemical synthesis. The ensuing three sections compare how these differences influence the functional performance of protein or DNA nanostructures as biomolecular recognition agents, enzymatic catalysts, and structural scaffolds. This includes a discussion of applications in biotechnology, biomedicine, synthetic biology, and materials research. A concluding section outlines possible avenues of future research. Given the broad scope of the Review, only a selection but not all references of the field can be cited. The Review is written for a broad and interdisciplinary audience interested in biomolecular design with chemical precision.

Properties and structures of biopolymers

Chemical diversity, folding, structural complexity, compactness, and dimensions of proteins and DNA affect their engineering properties. The 20 different amino acids of proteins lead to a broad chemical diversity (Figure 1, Box 1) and a larger scope for non-covalent interactions during folding compared to DNA. H-bonding and non-covalent interactions in the polypeptide backbone enable many secondary structures and over 1300 known protein folds (Box 1)[25]. An undesirable consequence of structural complexity is misfolding and aggregation[26]. However, in successful outcomes, a charge-neutral and conformationally flexible backbone usually yields compact protein architectures of moderate dimensions (< 10 nm, average size of 34 kD in bacteria and 49 kD in eukaryotes[27]), which also reflects the small size of amino acids (Box 1, Figure 1).

DNA and RNA are each built from four standard nucleotides. Hence, the chemical diversity and scope for non-covalent interactions in folding of DNA and RNA are more contained compared to proteins (Figure 1, Box 1). This is exemplarily illustrated by the lack of positively charged residues and direct ion bridges even though metal cations can lead to indirect bridges. DNA strands interact primarily through hydrogen bonding between base pairs and hydrophobic base stacking. The contained scope of interactions leads to fewer folding arrangements that are possible for proteins (Figure 1, Box 1). This makes DNA folding easy to predict when Watson-Crick base-paired duplexes (Box 1) are the core element of the architectures. Additional Hoogsteen base pairs expand the range of structures to G-quadruplexes or triplexes (Box 1), and cytosine-cytosine pairs to the i-motif. Folded DNA structures are usually less compact than proteins partly due to the electrostatic repulsion of the negatively charged backbone. Furthermore, duplexes with a high persistence length of around 50 nm[28] easily yield up to 100 nm-large DNA nanostructures which also reflects the bigger size of nucleotides compared to amino acids (Figure 1, Box 1).

The different characteristics of proteins and DNA influence the engineering of new architectures, as different engineering approaches need to balance between the strengths and weaknesses of the biopolymers Due to the chemical diversity and complex folding of proteins, most engineering approaches rely on the modification of naturally occurring proteins. Chemical diversity can be controlled by means of rational engineering, which is used to alter few amino acids or peptide stretches and is often guided by structural insight or computational tools (Box 2). Compared to rational design, directed evolution explores greater chemical diversity by creating large gene libraries of randomized sequences, expressing the corresponding proteins, and screening their properties of interest[29-31]; misfolded proteins are discarded in the screen (Box 2). Directed evolution methods mRNA display[32,33] and yeast display cells[34-36] are of particular interest. The former achieves a large chemical diversity by translating proteins and chemically associating them with their encoding mRNA for screening and identification. By comparison, the former method obtains proteins with specific biomolecular functions by expressing large amounts of proteins, post-translational modifying them and displaying them on the cell wall of yeast for easy screening. Computer-aided de novo design explores the largest diversity without requiring biological

scaffolds as starting point,[37-41] even though the design can be inspired by natural templates, and use existing secondary structure elements[42] (Box 2, Figure 1). Given the difficult prediction of complex folding, mostly small proteins are designed even though higher computing power will help increase size and complexity[38]. A powerful strategy is to combine two or more of the above methods to compensate their isolated drawbacks[42-44].

Engineering with nucleic acids is different as the small chemical diversity can translate into simple folding. Novel structures of aptamers[21,45] are generated via directed evolution method SELEX (Box 2) that is simpler than analogous protein methods because in nucleic acids both function and encoding sequence are combined (Box 2). Complementary, completely rational de novo design[10,14,17] can build a nano-sized object of almost any geometry from scratch because folding relies on simple base-pair rules and a small set of standard structural building units including DNA duplex units, hairpins, and Holliday junctions (Box 2). The rationally designed DNA nanostructures are, unlike aptamers, usually composed of multiple DNA strands and reach sizes up to 100 nm even. Recently, single-chain DNA origami[46] and co-transcriptionally folded single-chain RNA nanostructures have been prepared[47]. In addition, microscale symmetric DNA origami assemblies[48] and similarly sized arrays with arbitrary patterns are accessible[49].

[H1] Designing and building function

Protein and DNA/RNA differ considerably and this contrast has consequences on the building of functional nanostructures, which are relevant in biomedicine, biotechnology, synthetic biology, and material sciences. The following three sections discuss these differences in light of three representative functions: biomolecular recognition, catalysis, and structural support. Other functions such as motor activity[50-54] or cellular signal processing and biocomputation[11,55,56] are not covered due to space limitations, whereas transmembrane transport by protein vs. DNA pores[57-60] has been compared previously[61,62].

[H2] Biomolecular recognition

Biomolecular recognition enables highly specific and tight binding of ligands. Molecular binding agents are hence of relevance in biosensing and biomedicine, and many research applications including analysis of biological cells. We evaluate binding agents by comparing antibodies, antibody derivatives, other protein-based recognition scaffolds to DNA/RNA-based aptamers. The chemical and structural properties of proteins and DNA are key because specific and tight recognition requires the side chains of the biopolymers to line a binding pocket that is structurally and chemically complementary to the ligand. Furthermore, the engineering strategies used for the different biopolymers strongly influence how easily binding sites are obtained.

Antibodies and antibody derivatives illustrate that proteins excel on specific and tight recognition, and suitable strategies to achieve both properties. Antibodies are Y-shaped structures with two identical arms termed Fab for fragment antigen-binding[63,64](Figure 2A); the bottom stem of the Y-shaped structure — the fragment crystallisable (Fc)-region — does not interact with antigens, but with other receptors (Figure 2A). Each Fab region is composed by a pair of two polypeptides— heavy and light chain — whereby the site for antigen binding is formed by two corresponding antigen-binding loops. Two rather than one loop better exploit the chemical and structural diversity of the amino acids that is required for highly specific recognition of ligands.

The chemical and structural diversity can be harnessed in biotechnology with the mutational mechanism of antibody-producing immune cells using hybridoma technology. In this method,

continuously growing hybrid cells each one producing a monoclonal antibody type specific for one recognition motif[65]. The hybrid cells are made from antibody-producing, yet short-lived B immune cells and immortal cancer cells. This approach uses mouse immune cells, therefore the surface of the mice-derived antibodies causes immunogenicity in humans[66-68]. However, antibodies can be humanized by genetically fusing the mice-derived antigen-binding loops into a human antibody framework[66-68] thereby expanding their therapeutic applications. Alternatively, antigen-binding units are sourced from bacterial phage-display libraries[69] or also yeast libraries[34,35]. Both routes generate gene libraries of randomized sequences outside cells and then expressing and screening the proteins. The advantage of using yeast libraries is that antibodies and other mammalian cell-surface and secreted proteins such as receptors and cytokines are post-translational modified with carbohydrates in the cell endoplasmic reticulum to achieve efficient folding and activity[34,35].

Antibody engineering also yielded variants with two different Fabs to recognize two distinct antigens. These bispecific antibodies[5,7063,71]. are used in immune cancer therapy to recruit via specific recognition cytotoxic T cells to the targeted tumour cells; the Fc part can also help bind another immune cell such as macrophage. Bispecific antibodies are used for other biomedical applications, such as to co-localize two proteins in cells and test the biological effect.[63]. Bispecific antibodies can be produced within a cell line. But this can result in a uncontrolled combination of the two light and heavy chains present within each cell and inactive isoforms[72]. Rational engineering can overcome the problem by heterodimerising the heavy chains[70] even though the pairing with light chains is still uncontrolled. One solution involves using of a single common light chain[5] however a better alternative is to produce in two separate cells two half antibodies, each composed of the matching heavy and light chain. After their purification, the matching antibody halves are then specifically coupled by exploiting the rationally engineered complementary 'knob-in-hole' surface structures on the two heavy chains[63].

The cumbersome fabrication of complete antibodies limits their use for many applications. Engineering approaches can truncate the Fc region and stabilize the remaining single Fabs by looping the light and shortened heavy chain into a single polypeptide. Due to the smaller size, these single-chain variable fragment antibodies can display higher tissue penetration without compromising affinity[73]. Their antigen-binding region can also be optimised with directed evolution using yeast display[34,35].The most important biomedical application of single-chain variable fragments is in cancer therapy (Rosenbaum L. "Tragedy, Perseverance, and Chance - The Story of CAR-T Therapy". The New England Journal of Medicine. 2017, 377, 1313–1315). The fragments are parts of chimeric antigen receptors (CAR) that are engineered onto immunological T-cells to recognise tumour cells. The chimeric receptors comprise the extracellular recognition domain, a genetically fused membrane-spanning domain, and an intracellular signalling endodomain (Gross G, Waks T, Eshhar Z. "Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity". Proceedings of the National Academy of Sciences of the United States of America. 1989, 86: 10024–8). When CARs on T-cells bind to cognate surface proteins on tumours, multiple CARs form lateral cluster in the membrane which in turn leads to signalling and activation and reprogramming of T-cells into cytotoxic cells that destroy the cancer cells (Rosenbaum L. "Tragedy, Perseverance, and Chance - The Story of CAR-T Therapy". The New England Journal of Medicine. 2017, 377, 1313–1315).

An alternative solution to large antibodies are camelid antibodies, as they naturally lack the light chain, which is compensated by a slightly longer heavy chain[74,75]. These compact binding agents, aptly termed nanobodies[76], have high structural stability[75,77] and are less prone to aggregation[76]. The small units can reach epitopes inaccessible to larger antibodies[78] and can also be expressed in a form that is fused to green fluorescent protein[79].

Recognition can be engineered into a non-antibody scaffold as shown by designed ankyrin repeat proteins (DARPins) (Figure 2B) which are derived from the of integral membrane protein ankyrin[80]. A DARPIn recognition site is composed of up to five repetitive looped homologous units and, because of its modular structure, it mimics the modular recognition principle of ankyrins. The loops are selected by screening a library of around 1000 repeats of different sequences[81]. DARPins have been used as crystallization chaperone[82] and sensors of protein conformation, and to induce apoptosis in tumors[83]. The concept of repeat-based recognition scaffolds is common and has been implemented using sequences with leucine-rich-repeats[84], a 42-residue tetratricopeptide repeat variant[85], HEAT-like repeats[86], and others[87,88].

Non-antibody recognition scaffolds smaller than DARPins are peptide macrocycles including immunosuppressant cyclosporine[89]. Macrocycles feature around 8-12 amino acids including D-enantiomers and L-versions with non-standard residues. The recognition spectrum of macrocyles has been explored by sampling the available structure space by sequence design and energy landscape calculations[90].

As illustrated above, the chemical variety of polypeptides help create strong and specific molecular recognition agents. Both biological and in vitro generated libraries can be screened to identify highly functional binders, while rational engineering helps redesign the overall scaffold. How do RNA and DNA-based recognition agents compare to the impressive properties of proteins and peptides? RNA or DNA can form remarkably strong and specifically binding aptamers (Figure 2C) even though the chemical diversity of endogenous nucleic acids is lower than for proteins[21-24,45]. A compensatory factor is the relatively short length of DNA or DNA-aptamer strands, which makes it easy to explore via SELEX a greater physicochemical parameter space than would be possible for proteins[19,20]. Furthermore, aptamers can address some of the shortcomings of antibodies. For example, the folded strands are fabricated from commercial chemical synthesis, whereas antibodies require more complex manufacturing processes that make use of cell-based expression systems followed by purification to remove biological contaminants. As a drawback, DNA and RNA make aptamers unstable towards nucleases, which can limit their use in therapeutic applications.

Aptamers are used in diagnostics, therapeutics, biosensing and as research tools for cell biology[21-24,45]. One clinically approved RNA aptamer is discussed to highlight structural features as well as strategies to address nuclease instability[91]. Aptamer Pegaptanib is approved by the Food and Drug Administration and is used against age-related macular degeneration. It binds the molecular target vascular endothelial growth factor with an affinity of around 50 pM[91]. The RNA strand is 27 nt long and folds into a hairpin structure with an internal loop, which is bent into the biologically active 3D conformation . To increase nuclease stability, all of the 13 pyrimidines are 2′-fluorine-modified, while 12 of the 14 purines carry the 2′-O-methy modification. Furthermore, a 40 kDa polyethylene glycol moiety is linked to the 5′ terminus to sterically screen nucleases. More aptamers are developed for other diseases[23]. Aptamers can also be integrated into rationally designed nanostructures to achieve multiple recognition capabilities[92,93].

Compared to antibodies or related structures, aptamers are less widely used in biomedicine. This is not due to the insufficiency of nucleic acids to form aptamers with high selectivity and specificity or to the lack of suitable engineering methods that can harness the chemical and structural potential. Both proteins and nucleic acid rely on screening methods of randomized libraries because the engineering of binding sites is harder with rational design. The higher use of protein-based recognition agents in therapeutics is mostly due to the lower nuclease stability of DNA or RNA aptamers, and filtration by the renal system[21].

[H2] Biocatalysis

Biocatalysis is essential in industrial biotechnology. It helps to synthesize chemicals and pharmaceuticals[2,94-96], biofuels[97,98], foodstuff including proteins, but also in biomedicine and biosensing[3]. It furthermore provides an environmentally friendly and cost effective alternative to classical transition metal catalysts[99]. Biocatalysis is enabled by enzymes, but can additionally be mediated by RNAzymes or DNAzymes.

Efficient catalysts accelerate chemical reactions and control which part of the substrate undergoes a specific chemical transformation. The biopolymer residues are key in both cases because fast catalysis requires the active site to weakly bind the substrate but energetically stabilize the reaction transition state. Furthermore, residues achieve chemical selectivity by recognising the substrate. Industrial applications requires biocatalysts that exhibit affinity for a defined range of substrates [97,100], increased rate and yield for the chemical transformations, a high stereo-, chemo- and/or regioselectivity[99], or tolerance for extreme conditions including elevated or lower temperature, organic solvents, and/or inhibitory chemicals[4]. Enzymes are outstanding biocatalysts as they meet these requirements. With their diverse repertoire of amino acids (various sizes, polarity or charge), active sites are formed to specifically interact with substrates and transitions states[2,4,101]. Several engineering approaches can harness the advantages of proteins and turn them into valuable tools for industrial processes. For example, traditional rational design[102] on a nitrilase with a single amino acid replacement increased the enantioselectivity by 11% to 98.5% for the R-isomer of a precursor for a cholesterol-lowering drug[103].

Directed evolution with libraries exploits a wider chemical and structural repertoire than rational design[29,30]. Means to diversify the library are gene shuffling where a gene is randomly fragmented and the fragments are then homologously recombined back into a more diverse set of genes via polymerase chain reaction (PCR) reassembly[104] but also cassette mutagenesis to replace a small subset of a gene with synthetic DNA of randomized sequence[105,106], and error-prone PCR. Directed evolution helped to turn iron-haem containing cytochrome P450 into a biocatalyst for the highly enantioselective amination of benzylic C-H bonds[107], thereby addressing a demand for an efficient and renewable catalyst for this reaction. The cytochrome was also engineered to perform the industrially important anti-Markovnikov oxidation of styrenes with high efficiency[108].

The extent of structural re-arrangement afforded by directed evolution is illustrated by a retro-aldolase. In this case, the reactive lysine in the original active sites is replaced by a lysine in a new substrate-binding pocket[109] (Figure 3A). In a more drastic case, a non-catalytic scaffold was converted into an RNA ligase[32] through a complete remodelling of the active site[110], including enhanced conformational dynamics[110,111]. This drastic change is likely caused by the in vitro production of the protein library, which does not impose any of the constraints observed in cell-based expression. Indeed, the evolution of a metalloenzymes for olefin metathesis via bacterial protein production needed the screened enzymes to be located in the bacterial periplasm to avoid metalloenzyme-inhibiting glutathione in the cytoplasm[1]. Similarly, enzymes can also be generated with yeast to display the screened proteins on the cell surface[36,112].

To allow for more directed changes in enzyme structure and function, evolution and screening can be conducted with targeted libraries,[29,30] which involves combining mutations from different rational design variants with rounds of error-prone PCR[113], or sequencing library hits to identify gene segments to be genetically randomized[96], or guiding the creation of randomised libraries with computational design[114-116]. For example, focused mutagenesis resulted in polysialyltransferase enzymes to produce polysaccharides with narrow size distribution ideal for glycoengineering applications[117]. Computations also enabled the re-

design of the active site of an amine transaminase to attain marginal activity for the synthesis of the chiral amine precursor for antidiabetic drug sitagliptin; directed evolution followed to optimize the enzyme to the manufacturing settings[99]. In another case, computational approaches were used to develop a statistical model to establish the relationships between catalytic activity and the structures of interacting small-molecule and protein. This resulted in targeted libraries as demonstrated for a halohydrin dehalogenase to synthesize a precursor for a cholesterol-lowering drug[118]. Hydrolytic enzymes comprising a catalytic triade of cysteine-histidine-glutamic acid placed inside a bundle of seven peptide α-helices (Figure 3C) have been designed using solely computational approaches[119].

The production of Kemp eliminase illustrates the benefit of combining all previous engineering routes. Kemp eliminase was computationally designed, then subjected to in vitro evolution[114], which was followed by sequencing to identify gene hot spots of high enzymatic activity, and constructing small focused libraries resulting in a striking overall 6×10^8-fold rate increase (Figure 3B)[120].

Beyond isolated enzymes there is considerable interest to engineer multi-enzyme assemblies or biocatalytic pathways in which cascades of several enzymes catalyze interconnected reactions[121]. Sophisticated cascades can be obtained by exploiting the increase by several orders of the catalytic rates due to proximity and/or nanoconfinement of enzymes [122,123]. Other systems that are appealing for industrial use integrate engineered membrane transporters for whole-cell catalysis in order to efficiently transport both substrates and products across the cellular barrier[124].

Similar to proteins, DNAzymes and RNAzymes are obtained via the directed evolution approach. They can catalyse reactions on the cleavage and ligation of the nucleic acid backbone, photorepair of DNA, and modification of peptides and small molecules[125,126]. One exemplary DNAzyme is a RNA-ligase of known X-ray structure (Figure 3D)[126]. The active site and the substrate form a compact folding unit that is stabilized by numerous tertiary interactions. The regioselectivity of the reaction could additionally be elucidated by structure-guided mutagenesis, which also helped to manipulate substrate recognition and reaction rate. Similarly relevant are RNA polymerase[127] and reverse transcriptase ribozymes[128].

The absence of DNA or RNAzymes in industrial applications is surprising as the biotechnologically relevant production of antibodies and enzyme uses ribosomes, which are RNAzymes[129,130]. However, ribosomes are composed by multiple RNA strands and additional polypeptides (Figure 6A). Such a complex multimeric structure does not easily lend itself to drastic re-engineering or partial de novo design.

Catalysts made just of RNA or DNA usually carry-out nucleic acids-related reactions, as pointed out above, even though directed evolution achieves also non-RNA/RNA transformations[131]. Furthermore, compared to proteins, most DNAzymes and RNAzymes only modestly speed up reactions. The narrower catalytic scope and slow speed is a consequence of the smaller chemical parameter space of nucleotides than that of amino acids. Furthermore, the larger size of the nucleotides and the negatively charged DNA backbone can reduce the structural flexibility and electrostatic tuneability, which is required for versatile catalysis. Several of these points are being addressed with non-biological xeno nucleic acids, which allow for a greater chemical diversity than it would be case with native nuclei acids (Box 2)[132-134]. An alternative route uses DNA-based hybrid catalysis, where a transition metal complex is covalently or supramolecularly bound to a duplex. The chiral microenvironment of DNA in addition to further interactions enable high enantioselectivities and, often, additional rate accelerations[135,136]. This idea may in future be extended to organo-catalysis. In a separate approach, DNA can be used to spatially pre-assemble organic reagents to facilitate their proximity-induced covalent linkage. In this DNA-templated

synthesis, short DNA strands -"codons"- are linked to the organic building blocks and sequence-specifically hybridize to the DNA template which then links the blocks into oligomers, as demonstrated by the synthesis of a macrocycle[137]. By bringing together reagents, DNA only fulfils one function of a catalyst but it does not stabilize the transition state. DNA-templated synthesis has been developed into a tool in materials discovery[138] and can also be conducted with a DNA-based molecular machinery that controls and records the formation of covalent bonds[139].

[H2] Structural support and scaffold

Structural support is the third function of protein or DNA/RNA nanostructures. In biology, protein-based cytoskeletal supports define the shape of membranes, organelles, and cells, but also guide the arrange non-membrane components, such as chromosomes, at defined position at length scales of up to a few micrometres. Biological structural supports comprise up to thousands of subunits, in contrast to the monomeric or oligomeric nature of enzymes or antibodies. As other difference, the function of scaffold proteins relies largely on their nanoscale morphology and dimensions; catalysis is usually not involved. As similarity, interactions with other subunits require a fine-grained protein surface, which is the same for molecular recognition agents.

In this section, we first describe representative biological protein-based structural scaffolds. This is necessary as their functions are quite diverse. Then, we discuss new designed protein assemblies, and their applications. To further appreciate the influence of the building material, we finally describe DNA-based synthetic scaffolds along with their applications.

[H3] Natural protein assemblies

Biological protein-based scaffolds can be categorized according to their ultrastructural shape, which is related to their biomolecular functions and molecular architecture of the subunits. The shapes are fibres, sheets, or hollow shells.

Microscale rod-like and cylinder-like assemblies physically connect other cellular components and form the intracellular cytoskeleton (microtubule, intermediate filaments), the extracellular matrix, and bacterial pili that form hair-like appendages on the cell surface to form contact to other bacteria and can be involved in biofilm formation[140]. Rods and cylinders are also used for intra and intercellular transport such as actin tracks to move cargo with myosin and other molecular motors or as rotary hooks for propulsion[140].

Sheet-like assemblies are found in bacterial S-layers (Figure 4A), that function as sole structural exoskeleton for almost all archaea and in eubacteria as attachment matrix for exo-enzymes or to invade host cells, such as for *Bacillus anthracis* layers[141-143]. The SbsB protein lattice from a Gram-positive bacterium illustrates this molecular architecture (Figure 4A)[144]. The protein consists of an amino-terminal cell-wall attachment domain and six consecutive immunoglobulin-like domains. The latter domains arrange in a crystallization unit that is stabilized by interdomain Ca^{2+} coordination with histidine and other metal-chelating amino acids. This compact unit pre-positions the areas of intermolecular contact to facilitate lateral assembly into the protein lattice.

Hollow protein shells such as viral capsids define the shape of viral particles, protect the enclosed viral genetic material, and facilitate cell surface recognition by host membranes through the display of recognition motifs[145]. Typical viral shells have an icosahedral or rod ultrastructure such as the M13 bacteriophage (Figure 4B). The approximately 50 nm-wide and up to a micrometre-long shell is mostly assembled from the major p8 coat protein. p8

interacts via its inner-facing, positively charged side with the single-stranded circular viral DNA. Each two other proteins form the caps of the filament. A related class of protein shells are bacterial micro-compartments. These organelles contain different enzymes to initiate catalytic cascades to, for example, fix CO_2 or oxidise carbon sources for energy release[146,147]. The icosahedric compartment shell usually comprises at least two different protein types. The first forms pentamers positioned at the vertices of the icosahedral shell. The second assembles into cyclic hexagons that constitute the shell's facets[147]. The hexagonal circular units enclose a central pore, which is assumed to facilitate cross-boundary exchange of enzymatic substrate and products. The larger enzymes are retained inside the bacterial micro-compartment[147-149].

Another protein cage, which can adopt icosahedric symmetry is formed by clathrin. This protein covers cell membranes via a lattice can shape the membranes into drop-like intracellular vesicles to transport enclosed molecular cargo into and out of the cells[150]. The curved lattice is formed by triskelions in which three protrusions are linked with some structural flexibility to the central trimerisation domain. Due to the flexibility, interaction between the three protrusions can form hexamer as well as pentamer units, yielding lattices with varying numbers of triskelions and symmetries including icosahedrons. A different and smaller protein shell is ferritin. It is composed of 24 α-helical polypeptides and binds iron inside cells, yet it does not adhere to membranes. BAR domains that are found in all eukaryotes, by comparison, form crescent-shaped dimers that binds preferentially to highly curved negatively charged membranes[151] and thereby shape the morphology of the bilayer (Figure 4C) and facilitate the clathrin-independent formation of vesicle pits[152].

[H3] Engineered protein assemblies

Which type of protein scaffold can be synthetically produced? Rational design helped build fibres using β-strand peptides that assemble into β-sheets. Alternatively, amphipathic α-helices were designed to interact via coiled-coils to bundle into long fibres[153]. If the peptides are straight and short, it is possible to elongate the fibres without branching for up to hundreds of μm[153]. Supplementing additional peptides containing flexibly linked α-helices produced kinked, waved and branched fibres[154]. Furthermore, pre-assembled blunt-ended coiled-coil barrels with central channels were used as building blocks to form peptide nanotubes[155,156].

2D nanosheets could be assembled from an α-helical amphiphilic peptide. Its 3-fold screw symmetry led to hexagonal packing of helices[157]. The design of larger 2D protein lattices had to cope with subunit interfaces that are more corrugated than that of small peptide subunits. To overcome the limitation, longer ranging metal coordination bonds were installed at the protein interface via suitable electron-donors histidine and glutamic acid[158]. The resulting monomeric proteins assembled into planar arrays up to the micrometre scale upon addition of Zn^{2+} that formed metal coordination bonds with the amino acids to bridge the subunits. 2D protein arrays were also formed via covalent disulphide bonds between protein subunits that had an inherent C4 symmetry and assembled into a square lattice[159].

By contrast, a 3D structure was attained without cofactors by using a single-chain α-helical polypeptide that folded via coiled-coils into the final tetrahedron-shaped product[160]. In the alternative and more frequently used multimeric mode, protein cages are assembled from subunit -proteins that are genetically coupled from distinct oligomeric protein domains in defined geometric orientation[161]. For example, a cube-shaped cage of 22 nm side length was formed by having a trimer at the corners of the cube, a dimer at its edges, and linking the trimer and the dimer with a semi-rigid α-helical linker (Figure 4D)[162]. Alternatively, the linkage could be promoted by non-covalent protein interactions[163]. In this case,

complementary packing arrangements are first identified by means of computational simulations, that are followed by the design of low-energy interfaces between the building blocks,[164] often involving only a few amino acid alterations[164,165]. The approach facilitaed the fabrication of produce a 24-subunit, 13-nm diameter octahedron from a trimeric protein building block[163]. Even larger compartments were obtained with heterodimeric and homotrimeric coiled-coil peptide bundles that formed hexagonal networks, which closed into spherical cages of approximately 100 nm diameter[166].

[H3] Applications of protein assemblies

The spherical and multimeric nature of spherical protein assemblies can be exploited in biomedicine to display peptide epitopes for vaccines. The display of the antigen parts helps recognition by the immune system and was achieved by genetically linking the haemagglutinin protein to the ferritin subunits. Immunization with the resulting particles resulted in antibody titres tenfold higher than those from a reference vaccines[167], also because multiple copies of the epitope were on the vaccine. A related genetic fusion approach placed poliovirus neutralization epitopes onto nanoparticles formed by the hepatitis B surface antigen protein[168].

The high lateral density of planar S-layer protein lattices finds applications in biotechnology as immobilization matrix for metabolic enzymes. The resulting catalytic lattice displayed a longer shelf-life than the free enzyme[169], most likely due to the conformational stabilization by the lattice[170]. In a similar fashion, the protein shells of bacterial microcompartments have been exploited as high-density anchoring site for new enzymes[171].

Examples of applications of synthetic protein scaffolds in synthetic biology includes the fabrication of a virus-like protein assembly to encapsulate RNA genomes[43]. The capsid proteins were first computationally designed into icosahedral assemblies with positively charged inner surfaces to package full-length mRNA genomes. Then, directed evolution of capsid proteins led to a 100-fold increased genome packaging, close to reference standard of a recombinant adeno-associated virus. Existing virion particles that protect DNA were previously engineered as vectors for clinical gene therapy[172].

Natural and engineered protein assemblies can be also exploited for researching new materials. The regularity of the protein lattices can be used to spatially organize inorganic materials, such as nanocrystals with optical, electronic and magnetic properties for information technology. S-layers, for example, were used to nucleate in situ arrays of 5 nm-sized CdS nanocrystals within the pores of the lattice[173]. Different approaches have been explored to direct immobilize arrays of preformed nanoparticles[174,175].

Protein assemblies have further been employed to produce arrays of two different types of non-biological units, as shown by the engineered M13 phage into hybrid materials for bioenergy applications[176,177]. In one study, elongated virus capsids were used to mimic the photosynthetic light-driven oxidation of water. Porphyrin photosensitizers and catalytic reaction centres were chemically linked to the p8 coat proteins of the viruses. Immobilization at tuneable nanoscale spacing on the protein lattice was essential to improve the overall catalytic activity[176]. In a second study, the phage formed nanostructured electrodes for Li ion batteries. To interface with Li, amorphous $FePO_4$ nanoparticles were grown on the p8 coat protein that had been equipped with a particle-seeding glutamic acid tag. Further electrical conduction was introduced by placing at the virus cap peptides that exhibited affinity for single-walled carbon nanotubes. Combining the capsid with nanotubes resulted in a high-density network where electrochemically active nanoparticles were conductively connected through the carbon nanotubes to the metallic electrode base[177].

The above studies illustrate the strengths and limitations of large protein assemblies. Protein self-assembly only requires one to a few types of easily produced protein subunits, which give rise to larger-order repetitive structures. The regularity is a logical consequence of the small size of proteins and the short persistence length of polypeptides. Nevertheless, repetition can be a limitation when more complex ultrastructures are required. Biology features less symmetric multicomponent complexes such as in transport[178] and energy conversion, but their engineering is difficult. Hence, there is currently no protocol to obtain protein assemblies of any geometrical shape or any dimension from 20 to 100 nm.

[H3] DNA-based scaffolds

DNA is an alternative building material that can address some limitations of proteins. One noted advantage of DNA duplexes is their high persistence length, which makes production of large non-repetitive nanostructures easier. In addition, the rational design of DNA structures is greatly facilitated by the simple base-pairing rules and the predictable folding of DNA strands. Consequently, DNA nanotechnology relies on building de novo architectures with simple structural motifs such as DNA duplexes, DNA hairpins and Holliday junction-based cross-overs (Figure 5A) as well as motifs with non-Watson Crick base pairs. DNA nanotechnology also exploits the relative easy predictions of which interconnected units fold from a strand with controllable sequence.

A range of rational design and assembly strategies have been developed for de novo DNA nanostructures (Box 1). Several representative structures are described here to illustrate the scope of DNA nanotechnology. All structures are preceded by the tile-based design, in which short DNA strands assemble into repetitive arrays[179-182].

DNA assemblies into highly tuneable shapes and dimensions are obtained with the origami approach. A set of single stranded DNA oligonucleotides act as staples to direct the folding of a long ssDNA scaffold into a nanostructure of non-repetitive architecture[10]. The scaffolds are around 10,000 nucleotides long and can lead to nanostructures up to 100 nm. The structure composed of a layer of parallel aligned duplexes are held together by multiple cross-overs (Figure 5). Shapes accessible via the design route are rectangles, stars, and a smiley face (Figure 5A). A box composed of multiple side-ways interconnected origami walls was also made (Figure 5B)[183]. Its lid could be opened after adding a key oligonucleotide to unlocked the duplex-based closure mechanism.

The DNA origami approach can be extended to yield structures multiple duplex layers thick[184,185]. The parallel aligned duplexes can be arranged in square or honeycomb lattices. caDNAno software greatly facilitates the design of DNA origami structures[186]. The user inputs the desired geometry and dimensions of the nanostructure and caDNAno calculates the sequences of staple strands. However, user experience is required to reposition Holliday junctions to generate robust and stable constructs. Advances in software development have further automated the design process to minimize user adjustments[187,188]. In addition, the need for staple strands has recently been relieved with the new concept of single-strand DNA origami, in which a single long nucleic acid strand threads through the nanostructure[46] similar to co-transcriptional folding of a single stranded RNA into the final product[47].

A greater range of morphologies is accessible with curved DNA. In contrast to classical DNA origami with straight helixes, twisted or bent building units can be obtained by inserting or deleting bases within duplexes[189,190]. Bundles of neighbouring DNA duplexes can have the same curvature, and nanostructures resembling twisted ribbons or triangles have been attained using this kind of DNA duplexes[189]. Alternatively, DNA origamis can exhibit different curvatures in neighbouring duplexes that lead to variable distances between cross-overs[190]. This curved design approach can be used to form shapes that resemble Genie bottles[190] but

also arrays of clathrin-triskelion-like stars that assembled into a planar higher-order lattice on supported semifluid bilayer membranes (Figure 5C)[191].

Another route to obtain larger non-repetitive assemblies relies on 3D shape complementarity[192]. Here, DNA origamis are designed to assemble into even larger units than those obtained in standard DNA origami approaches by virtue of duplex protrusions that exhibit complementary shapes. The complementary units are stabilized by hydrophobic base-stacking between the corresponding duplexes, rather than direct base-pairing. Possible resulting structures include synthetic filaments composed of Lego-like bricks and nanostructures resembling the human body, which comprise three parts: the legs and two body halves with hands (Figure 5D). The idea of using blunt-end stacking to arrange large units has also been demonstrated with orthogonal DNA origami tiles. To achieve programmable assembly, the planar tiles are designed to feature complementary edges composed of alternating protruding blunt-ends forming duplex loops and inactive patches[193].

In DNA wireframes, straight duplex modules are connected at tuneable angles into a network allowing to form regular geometries such as tetrahedrons[194] or octahedrons[195] or more exotic architectures such as nanoscale rabbits, bottles, or spirals wireframes[187,196]. The creation of regular or irregular shapes can be guided depending on the use of solely short DNA oligonucleotides[194] or also a DNA scaffold that threads through the structure[187,195,196].

In brick-design[197,198] short oligonucleotides 'DNA bricks' assemble into 3D structures whereby design units are cubes, which include an 8 bp interaction between neighbouring strands. This design helped create a complete alphabet at the nanoscale.

[H3] Applications of DNA-based scaffolds

One application of rationally designed DNA lies in synthetic biology, in which it is used to mimic the function of proteinaceous cytoskeletons that shape lipid bilayers[199-201]. DNA scaffolds offer more extensive control over morphology and size of the membrane supports compared to protein structural supports. Membrane anchoring is achieved by attaching tags such as cholesterol[202,203] as observed in the case of flat nanobarges[191,204-206] and origami bricks that can also deform vesicle bilayers[207]. A related study has showed, however, that hydrophobic tags attached to cuboid DNA cages interact themselves without membranes. The cages form defined, quantized oligomers such as dimer, trimer or tetramer that are stabilized by the hydrophobic interactions[208].

DNA nanostructures can be used to modulate the shape of the lipid bilayers. For example, simple DNA rings with a size of up to 200 nm were built to control the diameter of spherical bilayer vesicles formed within the ring[209,210]. The DNA templates helped to homogenize the size of conventional vesicles. Unconstrained vesicles are heterogeneous in size and this can be an issue for biophysical research, but also for the use of cargo-loaded vesicles in bioimaging and biomedicine. In a different route, a spherical nanoscaffold was coated with a bilayer membrane to generate a virus-like particle[211]. With further development, the approach could prolong the half-life of vesicle-coated drugs.

Membrane morphologies including tubes, rings, spirals, and organelle-like vesicle arrays were made with a modular DNA design[199]. An example of possible DNA scaffold was assembled from multiple interconnected cylinder-like nanocages, each composed of two open DNA squares and four connecting pillars (Figure 5E). The presence of lipid anchors inside the cages enable liposomes to assemble in this confined space. By shortening the pillars, the vesicles were brought together and fused into elongated membrane tubes. Asymmetric shortening formed wedge-like cages that produced nanoscale doughnuts and

similarly shaped membrane rings (Figure 5E). In a related approach, BAR-like DNA origami rods with different degrees of curvature were shown to induce membrane bending and deformation of vesicle on the micrometre-scale[201]. When added in excess, the curved DNA rods wrapped around closely, stabilizing the lipid nanotubes[201]. The membrane vesicles with controlled dimension and geometry could help understanding the relation between the morphology of membranes, its lipid composition and dynamics. A related topic of interest is how the curvature of the membrane alters the activity of membrane proteins within the photosynthetic or oxidative respiration systems[212].

DNA origami also helped pioneering a new fluorescence microscopic imaging method that reaches high spatial resolution[213]. Classical high-resolution methods can suffer from the short life-time of the fluorophores and insufficient localization. By attaching a single-stranded DNA tag to the molecule of interest, multiple transient binding of complementary oligonucleotides with fluorescent labels could enhance the overall signal strength and improve resolution. The method was pioneered with a DNA origami plate featuring multiple DNA extensions for hybridization[213]. It can also help to structurally characterize 3D DNA polyhedra in solution[214].

In another cell biological application, DNA nanostructures have been designed to probe changes in intracellular pH. The sensor featured an i-quadruplex section, which switches between an open and closed state, depending on the pH. The fluorescent emission of a a fluorophore pair placed on the i-quadruplex provides information of pH changes between 5.8 and 7. The DNA nanostructure helped mapping the pH of early endosomes and the trans-Golgi network in real time[215,216].

DNA nanostructures can also be used in the field of material sciences as templates to guide the assembly of individual polymers in designed patterns. This approach has facilitated the study of the properties of single polymer strands useful to generate molecular-scale electronic devices or optical wires[217]. DNA nanostructure can also act as scaffold to spatially control assembly of nanoparticles[14]. Nanoparticles with plasmonic properties are of particular interest due to the possible applications in optical[218], electronic[219] or photonic devices[219,220]. As the collective properties are affected by the particle distance and arrangement[14,220], DNA strands has been used as spacer to control the lattice periodicity and symmetry of nanoparticle assemblies[221,222].

DNA origami can help attain more complex arrangements of nanoparticles than accessible with DNA strands. For example, a 2D DNA array was used to assemble a chess-board pattern of alternating 5 nm and 10 nm gold nanoparticles[223]. Strand displacement allowed switching of particles and changes in the geometry of the structures[224]. For more complex 3D patterning, nanoparticles were organized in a regular pattern around DNA nanotubes[225] and in a helical arrangement around nanorods[226]. The precise position of the particles is reflected in the circular dichroism spectrum. DNA origami nanoantennas were also built to enhance the fluorescence intensity dyes. A dye molecule was placed in a plasmonic hotspot of zeptolitre volume between two 100 nm gold nanoparticles and showed a 117-fold increase of the fluorescence[227].

DNA origami has also been used as a mould to create nanoparticles with predesigned 3D shapes and surface modifications[228]. Stiff DNA origami structures were designed and the scaffold was used to enclose a small gold cluster, which was then grown into a nanoparticle that assumed the shape of the container. Cuboid, triangular and nanoparticles of more complex geometries were generated with this approach. The moulding strategy can be extended to grow polymers inside of the DNA cage[229]. The remaining DNA strands on the exterior lead to programmable polymer assembly. DNA cages can also be used to print DNA sequences onto gold particles[230].

DNA origami has furthermore guided the production of 3D plasmonic crystals by enclosing gold nanoparticles inside DNA origami tetrahedra, which were then linked to a diamond superlattice[231]. The placement strategy has been adapted to couple molecular emitters to optically resonating photonic crystal cavities. This approach was based on the predictable binding of DNA origami onto nanopatterned surfaces with cavities [232,233]. DNA scaffolds modified with different numbers of dyes were positioned onto the photonic crystal cavities. The tuning of the light intensity by each resonating cavity resulted in the nanoscale reproduction "The starry night" painting by Van Gogh[234].

Conclusions and outlook

This Review has raised and addressed a simple question: how do the different chemical properties of proteins and DNA/RNA influence the engineering of nanostructures and their functions? To answer the question, the two building materials were compared in terms of biochemical properties and ease of engineering. These two aspects have been compared by looking at two main features of the biopolymers: chemical and structural diversity, and predictable easy folding. Although both features are desirable, they are not always exhibited by a single biopolymer type. Proteins are chemically and structurally diverse but this inherently makes rational design challenging. By comparison, DNA is less diverse but stands out because of its programmable folding into the predictable nanostructures. This is a general comparison, as highly functional structures can also be obtained by means of advanced directed evolution methods.

The biopolymers have been evaluated for three exemplary functions. In the case of biomolecular recognition, both proteins and nucleic acids achieve strong and specific binding. This is easy to understand for proteins due their great chemical parameter space. For nucleic acids, the limited chemical diversity is compensated by rich directed evolution libraries. Biocatalysis in industry is dominated by enzymes due to their larger chemical and structural diversity, and fast catalytic speed; directed evolution is the central approach even though computational design gains traction. Finally, the building of structural supports can be achieved with proteins but more easily with DNA nanostructures that often outperform proteins in several applications. For both protein and DNA-based structural supports, rational design with computational tools is the principal approach in use thus far.

What are areas of growth and possible future trends of research? Within protein engineering, the ever-increasing computational power will facilitate de novo design of proteins by better predicting polypeptide folding[38]. In addition, combining rational design with directed evolution will gain popularity given the considerable benefits. The advantages of this combined approach have been highlighted by the fabrication of artificial proto-viruses and efficient enzymes whose essential protein scaffolds were computationally designed but then functionally optimized via directed evolution[43,90].

Another area of interest are protein hybrids that integrate two functions such as structural support and catalysis. Inspirational templates are bacterial microcompartments, where protein scaffolds spatially confine enzymes to accelerate chemical transformations due to the reduced diffusion distances[146-149]. Synthetic analogues are intracellular filaments made from assembled bacterial microcompartment shell protein engineered to capture enzymes[235]. Furthermore, the biocatalytically important stoichiometry of enzyme components within a pathway could be controlled by using an scaffold that was engineered with interaction domains to specifically recruit the enzymes fused to cognate peptide ligands[236]. In both cases, enzyme confinement led to an overall increased catalytic activity compared to isolated enzymes. Future research on metabolite membrane transporters could

help organize more enzyme pathways to improve metabolic flux for industrially relevant chemical synthesis.

The functionality of proteins can also be enhanced with non-natural amino acids that are introduced via peptide synthesis[237,238], a reprogrammed genetic code[239] or by post-synthesis chemical modifications[240]. Such chemically precision can turn proteins into tools to probe, image and control protein function, or provide valuable protein therapeutics[239,240]. The benefit of chemical coupling small molecules is illustrated by therapeutically potent antibody-drug conjugates[241]. The conjugates specifically recognize cancer cells and simultaneously deliver their cytotoxic drug payloads. A considerable research interest is to improve the conjugate quality by avoiding the semi-random coupling of the drugs with the lysine residues at the sensitive antigen binding sites[242]. Current research hence aims to develop and apply site-selective reactions that link drugs at defined positions including the antibody disulphides bonds[243].

Advances in aptamer and DNA nanotechnology have achieved remarkable results. To further exploit their potential, the limited chemical variety could be addressed by using chemically modified nucleic acids[244] that can improve, for example, molecular recognition of aptamers[133]. Furthermore, the scalable and inexpensive production of DNA origami is of relevance to open up applications in biomedicine and therapy. For classical DNA origami, staple strands are chemically synthesized at low-scale, whereas scaffolds strands are sourced from cells. Using a new approach, staples can be made within bacterial hosts via bacteriophages, avoiding synthetic reactions. The encoded staple sequences were enzymatically synthetized as they were cut by two lateral Zn^{2+}-dependent DNAzymes, thereby allowing biological production of both staples and scaffold[245]**.** In another route, staple strands were made by a new design and folding programme in order to exploit the self-folding of single DNA or RNA strands into complex yet unknotted structures[46]. Both approaches will help deliver applications of nucleic acid nanostructures in, for example, therapy or intracellular reorganization of biological components[11,1418]. Biomedical and therapeutic applications can further exploit strategies to make DNA nanostructures compatible with the immune system of higher organisms[246] and a fundamental understanding of the cellular uptake and processing of DNA nanostructures[247] as well as their designed function inside cells[56]. As further advantage, nucleic acid structures can help advance biomedical applications with programmable intracellular biocomputation for triggered delivery of therapeutic cargo to target cells[55]. DNA nanotechnology can also be used for biosensing[24861]. Other areas of growth include the development of computational tools to guide and enhance the folding accuracy of DNA origami [249], or to increase the stability of DNA strands towards nucleases for biological or therapeutic applications[18].

Hybrid nanostructures composed of proteins and nucleic acids are also of considerable interest[250], as the combined strength of the two components can compensate their isolated weaknesses. This is well illustrated by the ribosome, which is composed of RNA and polypeptides strands (Figure 6A). However, simpler hybrid systems have been engineered. For example, DNA scaffold have been used to arrange multiple enzymes or cofactors in a nanostructures enhancing the catalytic turn-over[251], like biological catalytic cascades in confined environments[122,123]. Similar benefits were obtained for a DNA nanocage that encloses enzymes (Figure 6B)[252,253], or a DNA nanorod that link two enzymes and an artificial swinging arm to channel substrates between the enzymes[254]. Furthermore, the activity of an enzyme was controlled by tethering it into a closable DNA vault. The unit could be opened by a DNA key and thereby allow the enzyme to regain steric access to the chemical substrates and restore catalytic activity[255]. However, structurally more less ordered assemblies can also result in catalytic enhancement as shown by linking enzymes to intracellular flexible RNA assemblies[256]. Hybrid nanostructures are also excellent tools in cell biological research. For example, DNA scaffolds can be used to expose activating protein-

ligands to cells in well-defined arrangements, with full control over protein number, stoichiometry, and nanoscale distance[257]. DNA nanostructures are also templates to study movement of molecular motors[53]. Hybrids protein-DNA scaffolds of new defined geometry[258] can be created by stabilizing selected vertices of a DNA tetrahedron with duplex-binding RecA protein[259]. Furthermore, new function can also be enabled by puncturing lipid bilayer with hybrid pores to facilitate transport of large molecular cargo between the membrane barrier (Figure 6C). In these hybrids, a ring-shaped DNA scaffold arranges multiple tethered membrane-spanning peptide subunits into a contiguous pore of unusually large nanoscale diameter[260].

In conclusion, this Review has described how the different properties of proteins and nucleic acids define their engineering, function and suitability for specific applications. Although the building materials are different, there is considerable scope for working together and synergistically combining their strengths.

Box 1:
Comparison of polypeptides' and nucleic acids' chemistry and structure

Protein	DNA/RNA
Monomers and chemical properties	
The side chains of the 21 proteinogenic amino acids are structurally and chemically diverse. They cover a 2.8-fold size range (van der Waals volumes: 28 $Å^3$, alanine; 79 $Å^3$, tryptophan)[261] and a large chemical spectrum from aliphatic, aromatic, and polar to acidic and basic as shown in Figure 1.	The four Watson–Crick bases are very similar in their flat aromatic structure. They cover a 1.4-fold size range (volumes: 66 $Å^3$, thymine; 93 $Å^3$, guanine. Only the number and position of exocyclic substituents changes but there is no switch from apolar to polar, or negative to positive charges.
Structural motives of polypeptides and nucleic acids	
The α-helix is stabilized by intra-backbone hydrogen bonding between N-H and C=O groups spaced by three residues. An α-helical turn is completed by 3.6 residues with a rise of 0.15 nm per residue along the helical axis, and an advance per turn of 0.54 nm. The β-sheet is formed by extended strands that interact via inter-backbone H-bonding. The strands are either parallel aligned with the N-termini at the same side, or antiparallel strands. Two residues complete a turn, the rise per residue is approx. 0.6 nm, and the distance between two strands is 0.5 nm. Coiled-coils are formed by two to up to seven α-helices stabilized by hydrophobic interactions. The helices feature heptead repeats with aliphatic amino acids in the position 1 and 4, and charged residues in positions 5 and 7. Polypeptide chains alter direction with turns that are named with a Greek letter to indicate the number of bonds: 5 to 1; π, α, β, γ, δ-turns.	The most prominent secondary structure for DNA is the B-type duplex and is stabilized by pairing between complementary bases. It has a diameter of 2 nm, 10.4 bp/turn, and a pitch of ~1.04 nm. The persistence length for the DNA duplex is approx. 50 nm[28] which is around 50 times higher than for ssDNA. Hairpin, triplex, and quadruplex[9,262,263] are other important secondary structures. Holliday junctions resemble a cross with four double stranded DNA arms and are relevant in DNA nanotechnology. The related double cross-over (DX) motif[14] is widely used in DNA origami structures. It consists of two helical domains linked at two cross-over points that are similar to Holliday junctions. Other motives used in DNA nanotechnology are parameic crossover (PX) and JX^2 motives[14].

837 Box 2:
838 Comparison of polypeptides' and nucleic acids' engineering and synthesis approaches

	Protein	**DNA/RNA**
Engineering and design approaches		
	Directed evolution: The directed evolution involves the generation of a large library of randomized mutants of a gene with known corresponding protein structure. The library members are expressed in cells[1] or small cell-like vesicles[264,265], via ribosome-display[266], mRNA display[32,33], or with phage-particles[267] or yeast cells[34-36] displaying the selectable proteins. Small libraries of peptide-like oligomers can be obtained via DNA-templated synthesis[137]. Screening is performed to identify variants with the desired functional property[29,31,268,269], and are usually followed by the amplification of the hit mutant genes. The process is often iterated.	Directed evolution: SELEX technology (systematic evolution of ligands by exponential enrichment)[19,20,24] screens a RNA or DNA oligonucleotide library of randomized sequences to isolate binders to a target ligand, followed by polymerase chain reaction (PCR) amplification, and more cycles of screening and amplification.
	Rational design approaches: Rational engineering and computer-aided design are best conducted with a high-resolution structure of the protein scaffold that helps to select those amino acids residues, peptide stretches, or even larger units that should be altered via site-directed mutagenesis. De novo protein design covers a wide range of approaches[37-40,270]. Its purest and most challenging form involves ab initio calculations to explore the full protein sequence space for a chosen architecture or function. Suitable structures and sequences are computationally screened by iteratively sampling the side chains (via the rotamers of all amino acids) and checking the structural validity of the backbone[38]. Some methods utilize as constraints secondary structure and contact predictions. Results can be cross-checked by comparison with approaches that predict structures from known sequences[38]. Template-based methods begin with a sequence, predict the secondary structure, and attempt to find a template structure and/or fragments from existing scaffolds in	Rational design approaches: Tile assembly [179-182]: Tiles composed of a few hybridized DNA oligonucleotides assemble into a regular array. Classical DNA origami[10,184]: A set of single stranded DNA oligonucleotides act as staples to direct the folding of a long ssDNA scaffold strand into a nanostructure with a single[10] or multiple duplex layers[184,185] arranged in square or honeycomb lattice. Design software caDNAno[186] predict the sequences of staple strands from a user-defined nanoshape of given geometry and dimensions. Curved DNA[189,190]: Within DNA origamis bases are inserted or deleted to uniformly twist or bend packs of duplexes[189]. Alternatively, DNA origamis with different curvatures in neighbouring duplexes are designed due to variable distances between cross-overs[190]. Brick-design[197,198]: Short oligonucleotides 'DNA bricks', assemble into 3D structures without a long scaffold strand. The design units are cubes, which represents an 8 bp interaction between neighbouring strands. Wireframes[187,194-196]: Straight duplex modules are connected at tuneable angles into a network whereby the sides are not filled with DNA, unlike classical DNA

the protein data bank that will fold similar to the target sequence[41]. Folding of the polypeptide into the defined architecture can be an issue due to the complexity of the folding pathways but can be engineered[271] with protein folding in vivo as **[Au:OK?OK]** a correction mechanism[272].	origami. Shape-complementarity[192]: DNA origamis assemble into larger units by shape-complementarity with duplex protrusions fitting into the designed recession, independent of pairing.

Synthetic approaches and chemical modifications	
Synthetic peptides of up to 50 residue can be generated via solid-phase peptide synthesis[273]. One or more synthetic fragments can be coupled via native chemical ligation[238] or linked to biologically generated proteins via expressed protein ligation[237] but also via Staudinger ligation[273]. Biologically generated proteins can be selectively chemically modified[240] also by carrying out directed evolution with unnatural amino acids[29].	Solid-phase DNA oligonucleotide synthesis[274] can yield up to 200 nt-long strands. Longer genes are assembled from partly overlapping oligonucleotides[275,276] via PCR. Base and backbone modifications, generally summarized as xeno nucleic acid (XNA) can include backbone analogues peptide nucleic acid, 1,5-anhydrohexitol nucleic acid, cyclohexene nucleic acid, glycol nucleic acid, locked nucleic acid, and threose nucleic acid. These and nucleobase analogues can be generated using chemical synthesis[132,134,244,277] and optionally incorporated via enzyme-mediated polymerisation[132,244] including an expanded genetic code[239]. Chemically modified DNA strands can also be obtained via DNA-templated synthesis to pre-arrange and ligate sequence-defined “codons” carrying chemically diverse side chains; and in vitro selection system screen for functional hybrid polymers that can be identified via DNA sequencing[278]. .

839
840

Figure 1

Proteins and DNA/RNA. Structures of amino acids and polypeptide secondary structures (left column), and nucleotides and nucleic acid secondary structures (right column). Amino acid with 20 naturally occurring side chains (middle left) grouped according to similar chemical properties. Polypeptides' secondary structures α-helix, β-sheet and coiled coil (bottom left, side and helix view, polypeptides shown in red, and backbone represented as a grey ribbon). DNA nucleotides with the 4 naturally occurring DNA bases (top right). The 4 bases interact via hydrogen bonding to yield Watson-Crick base pairs G-C and A-T (middle right). Anti-parallel DNA duplexes, A-, B and Z-forms, and other structures including hairpin junctions, triplexes and G-quadruplex (bottom right, DNA bases shown in red, and phosphodiester backbone represented as a grey ribbon). Scale bars, 1 nm.

Figure 2

Natural and engineered proteins, and DNA nanostructures used for biomolecular recognition. (A) Antibody immunoglobin G structure as illustrated with a schematic drawing (bottom) and molecular model shown with a red ribbon (top). One light and half of one of the two heavy chains form one fragment antibody binding (Fab) region. The Fab also features the antigen-binding loops (ABL) which bind the antigen (grey). Two halves of the heavy chain form the fragment crystallisable (Fc) region[279]. The inset shows the recognition of the HIV envelope glycoprotein (left) with an antigen binding loop (right)[280]. (B) A designed ankyrin repeat protein (DARPin, red/blue ribbon) interacting with signal transduction Ras GDP protein variant G12V (grey) that is involved in cancer development[281]. DARPin repeats have a structural motif of helix-turn-helix-β-hairpin. (C) A DNA aptamer inhibitor(red/blue) for lysophospholipase D Autotaxin (grey) which is a protein factor involved in cancer cell motility but also the adhesion of lymphocytes to immune organs[282].

Figure 3

Engineered bio-catalysts. (A) Enzyme retro-aldolase variant RA-61 (top) whose active site (green position in top, details see bottom) has been optimized with computer-based design including quantum mechanics and in vitro evolution and features lysine residues that interact with the substrate[109]. (B) Structure of the Kemp eliminase enzyme variant HG3.17 (top) with substrate 6-nitrobenzotriazole (green) bound in the active site shown in greater details at the bottom[120]. (C) Structure of a de novo designed enzyme (top) with a hydrolytically active catalytic triade Glu–His–Cys (position in green, detail at the bottom) embedded within the lumen of the α–helical bundle[126]. (D) Structure of the in vitro evolved ribozyme 9DB1 with ligase activity (top, active site green) with tertiary contacts within the catalytic domain (bottom)[126].

Figure 4

Natural and engineered proteins that fulfil structural and cytoskeletal roles. (A) SbsB S-layer structure and lattice[144]. The image of the bacterium was rendered following ref[283]. (B) M13 bacteriophage filament (grey) composed of p8 subunit (red/blue ribbon)[284]. (C) BAR domain (red ribbon subunit), and rendering illustrating their binding to a curved membrane (grey)[151]. (D) A synthetic protein cage with subunit shown as red/blue ribbon (top) and as space-filling model (grey, with subunit shown in red/blue)[162].

Figure 5

Rationally designed DNA nanostructures. (A) A smiley face made by DNA origami[10]. The inset shows a Holliday junction where the scaffold strand is grey and the staple strands are colored. (B) Open and closed DNA origami box[183]. (C) Clathrin-triskelion-like DNA origami nanostructures (subunit red/blue) assemble via adhesion on a supported lipid bilayer membrane into a regular array (grey)[191]. (D) DNA robots with closed and open arms (blue shades show the shape-complementary-based interactions)[192]. The open and closing is induced by modulating the electrostatic repulsion via the salt concentration. (E) Reconfigurable DNA nanocages (top, blue) assemble to complete a ring-like exoskeleton (grey) that templates a circular lipid tube (green)[199].

Figure 6

Natural and engineered hybrid systems that combine proteins and DNA/RNA. (A) Ribosome composed of RNA and polypeptide strands[285]. (B) Nanocaged enzymes with enhanced catalytic activity and increased stability against protease digestion[253]. (C) Hybrid nanopores constructs with the ring-like DNA domain (red) and membrane-spanning peptide domain (blue)[260].

TOC graphic

REFERENCES:

1. Jeschek, M. *et al.* Directed evolution of artificial metalloenzymes for in vivo metathesis. *Nature* **537**, 661-665, (2016).
2. Renata, H., Wang, Z. J. & Arnold, F. H. Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. *Angew. Chem. Int. Ed.* **54**, 3351-3367, (2015).
3. Denard, C. A., Ren, H. & Zhao, H. Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Chem. Biol.* **25**, 55-64, (2015).
4. Liszka, M. J., Clark, M. E., Schneider, E. & Clark, D. S. Nature versus nurture: developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.* **3**, 77-102, (2012).
5. Tiller, K. E. & Tessier, P. M. Advances in antibody design. *Annu. Rev. Biomed. Eng.* **17**, 191-216, (2015).
6. Rodriguez, E. A. *et al.* The growing and glowing toolbox of fluorescent and photoactive proteins. *Trends Biochem. Sci.* **42**, 111-129, (2017).
7. Martell, J. D. *et al.* A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat. Biotechnol.* **34**, 774-780, (2016).
8. Ellefson, J. W. *et al.* Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science* **352**, 1590-1593, (2016).
9. Seeman, N. C. Nanomaterials Based on DNA. *Annu. Rev. Biochem.* **79**, 65-87, (2010).
10. Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302, (2006).
11. Chen, Y. J., Groves, B., Muscat, R. A. & Seelig, G. DNA nanotechnology from the test tube to the cell. *Nat. Nanotechnol.* **10**, 748-760, (2015).
12. Pinheiro, A. V., Han, D., Shih, W. M. & Yan, H. Challenges and opportunities for structural DNA nanotechnology. *Nat. Nanotechnol.* **6**, 763-772, (2011).
13. Jones, M. R., Seeman, N. C. & Mirkin, C. A. Nanomaterials. Programmable materials and the nature of the DNA bond. *Science* **347**, 1260901, (2015).
14. Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nat. Rev. Mater.* **3**, (2017).
15. Torring, T., Voigt, N. V., Nangreave, J., Yan, H. & Gothelf, K. V. DNA origami: a quantum leap for self-assembly of complex structures. *Chem. Soc. Rev.* **40**, 5636-5646, (2011).
16. Sacca, B. & Niemeyer, C. M. DNA origami: the art of folding DNA. *Angew. Chem. Int. Ed.* **51**, 58-66, (2012).
17. Hong, F., Zhang, F., Liu, Y. & Yan, H. DNA origami: Scaffolds for creating higher order structures. *Chem. Rev.* **117**, 12584-12640, (2017).
18. Chen, Y. J., Groves, B., Muscat, R. A. & Seelig, G. DNA nanotechnology from the test tube to the cell. *Nat. Nanotechnol.* **10**, 748-760, (2015).
19. Ellington, A. D. & Szostak, J. W. Invitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822, (1990).
20. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510, (1990).
21. Zhou, J. & Rossi, J. Aptamers as targeted therapeutics: current potential and challenges. *Nat. Rev. Drug. Discov.* **16**, 181-202, (2017).

22. Mor-Vaknin, N. *et al.* DEK-targeting DNA aptamers as therapeutics for inflammatory arthritis. *Nat. Commun.* **8**, 14252, (2017).
23. Nimjee, S. M., White, R. R., Becker, R. C. & Sullenger, B. A. Aptamers as therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **57**, 61-79, (2017).
24. Gotrik, M. R., Feagin, T. A., Csordas, A. T., Nakamoto, M. A. & Soh, H. T. Advancements in aptamer discovery technologies. *Acc. Chem. Res.* **49**, 1903-1910, (2016).
25. Attwood, T. K., Peffifer, S. R. & Thorne, D. *Bioinformatics challenges at the interface of biology and computer science: Mind the gap*. (John Wiley & Sons, 2016).
26. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884-890, (2003).
27. Kozlowski, L. P. Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res.* **45**, D1112-D1116, (2017).
28. Manning, G. S. The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophys. J.* **91**, 3607-3616, (2006).
29. Lane, M. D. & Seelig, B. Advances in the directed evolution of proteins. *Curr. Opin. Chem. Biol.* **22**, 129-136, (2014).
30. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567-573, (2009).
31. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379-394, (2015).
32. Seelig, B. & Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448**, 828-831, (2007).
33. Wilson, D. S., Keefe, A. D. & Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. U S A* **98**, 3750-3755, (2001).
34. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553-557, (1997).
35. Wang, X. X., Cho, Y. K. & Shusta, E. V. Mining a yeast library for brain endothelial cell-binding antibodies. *Nat. Methods* **4**, 143-145, (2007).
36. Yi, L. *et al.* Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proc. Natl. Acad. Sci. U S A* **110**, 7229-7234, (2013).
37. Woolfson, D. N. *et al.* De novo protein design: how do we expand into the universe of possible protein structures? *Curr. Opin. Struct. Biol.* **33**, 16-26, (2015).
38. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320-327, (2016).
39. Damborsky, J. & Brezovsky, J. Computational tools for designing and engineering enzymes. *Curr. Opin. Chem. Biol.* **19**, 8-16, (2014).
40. Hilvert, D. Design of protein catalysts. *Annu. Rev. Biochem.* **82**, 447-470, (2013).
41. Khoury, G. A., Smadbeck, J., Kieslich, C. A. & Floudas, C. A. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol.* **32**, 99-109, (2014).
42. Rocklin, G. J. *et al.* Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168-175, (2017).
43. Butterfield, G. L. *et al.* Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature*, (2017).
44. Chevalier, A. *et al.* Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74-79, (2017).

45. Keefe, A. D., Pai, S. & Ellington, A. Aptamers as therapeutics. *Nat. Rev. Drug. Discov.* **9**, 537-550, (2010).
46. Han, D. *et al.* Single-stranded DNA and RNA origami. *Science* **358**, (2017).
47. Geary, C., Rothemund, P. W. & Andersen, E. S. RNA nanostructures. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* **345**, 799-804, (2014).
48. Wagenbauer, K. F., Sigl, C. & Dietz, H. Gigadalton-scale shape-programmable DNA assemblies. *Nature* **552**, 78-83, (2017).
49. Tikhomirov, G., Petersen, P. & Qian, L. Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns. *Nature* **552**, 67-71, (2017).
50. Wickham, S. F. *et al.* A DNA-based molecular motor that can navigate a network of tracks. *Nat. Nanotechnol.* **7**, 169-173, (2012).
51. Omabegho, T., Sha, R. & Seeman, N. C. A bipedal DNA Brownian motor with coordinated legs. *Science* **324**, 67-71, (2009).
52. Derr, N. D. *et al.* Tug-of-war in motor protein ensembles revealed with a programmable DNA origami scaffold. *Science* **338**, 662-665, (2012).
53. Wollman, A. J., Sanchez-Cano, C., Carstairs, H. M., Cross, R. A. & Turberfield, A. J. Transport and self-organization across different length scales powered by motor proteins and programmed by DNA. *Nat. Nanotechnol.* **9**, 44-47, (2014).
54. Jung, C., Allen, P. B. & Ellington, A. D. A stochastic DNA walker that traverses a microparticle surface. *Nat. Nanotechnol.* **11**, 157-163, (2016).
55. Li, J., Green, A. A., Yan, H. & Fan, C. Engineering nucleic acid structures for programmable molecular circuitry and intracellular biocomputation. *Nat. Chem.* **9**, 1056-1067, (2017).
56. Amir, Y. *et al.* Universal computing by DNA origami robots in a living animal. *Nat. Nanotechnol.* **9**, 353-357, (2014).
57. Burns, J. R., Seifert, A., Fertig, N. & Howorka, S. A biomimetic DNA-based channel for the ligand-controlled transport of charged molecular cargo across a biological membrane. *Nat. Nanotechnol.* **11**, 152-156, (2016).
58. Langecker, M. *et al.* Synthetic lipid membrane channels formed by designed DNA nanostructures. *Science* **338**, 932-936, (2012).
59. Bell, N. A. *et al.* DNA origami nanopores. *Nano Lett.* **12**, 512-517, (2012).
60. Burns, J., Stulz, E. & Howorka, S. Self-assembled DNA nanopores that span lipid bilayers. *Nano Lett.* **13**, 2351-2356, (2013).
61. Howorka, S. Building membrane nanopores. *Nat. Nanotechnol.* **12**, 619-630, (2017).
62. Ayub, M. & Bayley, H. Engineered transmembrane pores. *Curr. Opin. Chem. Biol.* **34**, 117-126, (2016).
63. Spiess, C. *et al.* Bispecific antibodies with natural architecture produced by co-culture of bacteria expressing two distinct half-antibodies. *Nat. Biotechnol.* **31**, 753-758, (2013).
64. Brekke, O. H. & Sandlie, I. Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nat. Rev. Drug. Discov.* **2**, 52-62, (2003).
65. Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497, (1975).
66. Boulianne, G. L., Hozumi, N. & Shulman, M. J. Production of functional chimaeric mouse/human antibody. *Nature* **312**, 643-646, (1984).

67. Better, M., Chang, C. P., Robinson, R. R. & Horwitz, A. H. Escherichia coli secretion of an active chimeric antibody fragment. *Science* **240**, 1041-1043, (1988).
68. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321**, 522-525, (1986).
69. de Wildt, R. M., Mundy, C. R., Gorick, B. D. & Tomlinson, I. M. Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat. Biotechnol.* **18**, 989-994, (2000).
70. Kontermann, R. E. & Brinkmann, U. Bispecific antibodies. *Drug Discov. Today* **20**, 838-847, (2015).
71. Bargou, R. *et al.* Tumor regression in cancer patients by very low doses of a T cell-engaging antibody. *Science* **321**, 974-977, (2008).
72. Brinkmann, U. & Kontermann, R. E. The making of bispecific antibodies. *MAbs* **9**, 182-212, (2017).
73. Wu, A. M. & Senter, P. D. Arming antibodies: prospects and challenges for immunoconjugates. *Nat. Biotechnol.* **23**, 1137-1146, (2005).
74. Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446-448, (1993).
75. Stanfield, R. L., Dooley, H., Flajnik, M. F. & Wilson, I. A. Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science* **305**, 1770-1773, (2004).
76. Muyldermans, S. Nanobodies: natural single-domain antibodies. *Annu. Rev. Biochem.* **82**, 775-797, (2013).
77. Dolk, E. *et al.* Isolation of llama antibody fragments for prevention of dandruff by phage display in shampoo. *Appl. Environ. Microbiol.* **71**, 442-450, (2005).
78. Hussack, G. *et al.* Neutralization of Clostridium difficile toxin A with single-domain antibodies targeting the cell receptor binding domain. *J. Biol. Chem.* **286**, 8961-8976, (2011).
79. Rothbauer, U. *et al.* Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat. Methods* **3**, 887-889, (2006).
80. Jost, C. & Pluckthun, A. Engineered proteins with desired specificity: DARPins, other alternative scaffolds and bispecific IgGs. *Curr. Opin. Struct. Biol.* **27**, 102-112, (2014).
81. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575-582, (2004).
82. Huber, T., Steiner, D., Rothlisberger, D. & Pluckthun, A. In vitro selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: The Na(+)-citrate symporter CitS as an example. *J. Struct. Biol.* **159**, 206-221, (2007).
83. Pluckthun, A. Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol.* **55**, 489-511, (2015).
84. Park, K. *et al.* Control of repeat-protein curvature by computational protein design. *Nat. Struct. Mol. Biol.* **22**, 167-174, (2015).
85. Marold, J. D., Kavran, J. M., Bowman, G. D. & Barrick, D. A Naturally Occurring Repeat Protein with High Internal Sequence Identity Defines a New Class of TPR-like Proteins. *Structure* **23**, 2055-2065, (2015).

86. Urvoas, A. *et al.* Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (alphaRep) based on thermostable HEAT-like repeats. *J. Mol. Biol.* **404**, 307-327, (2010).
87. Boersma, Y. L. & Pluckthun, A. DARPins and other repeat protein scaffolds: advances in engineering and applications. *Curr. Opin. Biotechnol.* **22**, 849-857, (2011).
88. Parmeggiani, F. & Huang, P. S. Designing repeat proteins: a modular approach to protein design. *Curr. Opin. Struct. Biol.* **45**, 116-123, (2017).
89. Driggers, E. M., Hale, S. P., Lee, J. & Terrett, N. K. The exploration of macrocycles for drug discovery--an underexploited structural class. *Nat. Rev. Drug. Discov.* **7**, 608-624, (2008).
90. Hosseinzadeh, P. *et al.* Comprehensive computational design of ordered peptide macrocycles. *Science* **358**, 1461-1466, (2017).
91. Ng, E. W. *et al.* Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat. Rev. Drug. Discov.* **5**, 123-132, (2006).
92. Meng, H. M. *et al.* Aptamer-integrated DNA nanostructures for biosensing, bioimaging and cancer therapy. *Chem. Soc. Rev.* **45**, 2583-2602, (2016).
93. Rinker, S., Ke, Y. G., Liu, Y., Chhabra, R. & Yan, H. Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding. *Nat. Nanotechnol.* **3**, 418-422, (2008).
94. Fisher, M. A. & Tullman-Ercek, D. Change, exchange, and rearrange: protein engineering for the biotechnological production of fuels, pharmaceuticals, and other chemicals. *Curr. Opin. Biotechnol.* **24**, 1010-1016, (2013).
95. Porter, J. L., Rusli, R. A. & Ollis, D. L. Directed evolution of enzymes for industrial biocatalysis. *ChemBioChem* **17**, 197-203, (2016).
96. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185-194, (2012).
97. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86-89, (2008).
98. Steen, E. J. *et al.* Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559-562, (2010).
99. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305-309, (2010).
100. O'Maille, P. E. *et al.* Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol.* **4**, 617-623, (2008).
101. Lichman, B. R., Zhao, J., Hailes, H. C. & Ward, J. M. Enzyme catalysed Pictet-Spengler formation of chiral 1,1'-disubstituted- and spiro-tetrahydroisoquinolines. *Nat. Commun.* **8**, 14883, (2017).
102. Griengl, H., Schwab, H. & Fechter, M. The synthesis of chiral cyanohydrins by oxynitrilases. *Trends Biotechnol.* **18**, 252-256, (2000).
103. DeSantis, G. *et al.* Creation of a productive, highly enantioselective nitrilase through gene site saturation mutagenesis (GSSM). *J. Am. Chem. Soc.* **125**, 11476-11477, (2003).
104. Stemmer, W. P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389-391, (1994).
105. Wells, J. A., Vasser, M. & Powers, D. B. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* **34**, 315-323, (1985).

106. Estell, D. A., Graycar, T. P. & Wells, J. A. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J. Biol. Chem.* **260**, 6518-6521, (1985).
107. Prier, C. K., Zhang, R. K., Buller, A. R., Brinkmann-Chen, S. & Arnold, F. H. Enantioselective, intermolecular benzylic C-H amination catalysed by an engineered iron-haem enzyme. *Nat. Chem.* **9**, 629-634, (2017).
108. Hammer, S. C. *et al.* Anti-Markovnikov alkene oxidation by metal-oxo-mediated enzyme catalysis. *Science* **358**, 215-218, (2017).
109. Giger, L. *et al.* Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494-498, (2013).
110. Chao, F. A. *et al.* Structure and dynamics of a primordial catalytic fold generated by in vitro evolution. *Nat. Chem. Biol.* **9**, 81-83, (2013).
111. Dodani, S. C. *et al.* Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models. *Nat. Chem.* **8**, 419-425, (2016).
112. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U S A* **108**, 11399-11404, (2011).
113. Preiswerk, N. *et al.* Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc. Natl. Acad. Sci. U S A* **111**, 8013-8018, (2014).
114. Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195, (2008).
115. Kries, H., Blomberg, R. & Hilvert, D. De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* **17**, 221-228, (2013).
116. Lu, Y., Yeung, N., Sieracki, N. & Marshall, N. M. Design of functional metalloproteins. *Nature* **460**, 855-862, (2009).
117. Keys, T. G. *et al.* Engineering the product profile of a polysialyltransferase. *Nat. Chem. Biol.* **10**, 437-442, (2014).
118. Fox, R. J. *et al.* Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **25**, 338-344, (2007).
119. Burton, A. J., Thomson, A. R., Dawson, W. M., Brady, R. L. & Woolfson, D. N. Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.* **8**, 837-844, (2016).
120. Blomberg, R. *et al.* Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature*, (2013).
121. Bhan, N., Xu, P. & Koffas, M. A. Pathway and protein engineering approaches to produce novel and commodity small molecules. *Curr. Opin. Biotechnol.* **24**, 1137-1143, (2013).
122. Kuchler, A., Yoshimoto, M., Luginbuhl, S., Mavelli, F. & Walde, P. Enzymatic reactions in confined environments. *Nat. Nanotechnol.* **11**, 409-420, (2016).
123. Agapakis, C. M., Boyle, P. M. & Silver, P. A. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.* **8**, 527-535, (2012).
124. Kell, D. B., Swainston, N., Pir, P. & Oliver, S. G. Membrane transporter engineering in industrial biotechnology and whole cell biocatalysis. *Trends Biotechnol.* **33**, 237-246, (2015).
125. Cuenoud, B. & Szostak, J. W. A DNA metalloenzyme with DNA ligase activity. *Nature* **375**, 611-614, (1995).
126. Ponce-Salvatierra, A., Wawrzyniak-Turek, K., Steuerwald, U., Hobartner, C. & Pena, V. Crystal structure of a DNA catalyst. *Nature* **529**, 231-234, (2016).

127. Horning, D. P. & Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. U S A* **113**, 9786-9791, (2016).
128. Samanta, B. & Joyce, G. F. A reverse transcriptase ribozyme. *eLife* **6**, (2017).
129. Cech, T. R. Structural biology. The ribosome is a ribozyme. *Science* **289**, 878-879, (2000).
130. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905-920, (2000).
131. Silverman, S. K. Catalytic DNA: Scope, applications, and biochemistry of deoxyribozymes. *Trends Biochem. Sci.* **41**, 595-609, (2016).
132. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341-344, (2012).
133. Pinheiro, V. B. & Holliger, P. Towards XNA nanotechnology: new materials from synthetic genetic polymers. *Trends Biotechnol.* **32**, 321-328, (2014).
134. Winnacker, M. & Kool, E. T. Artificial genetic sets composed of size-expanded base pairs. *Angew. Chem. Int. Ed.* **52**, 12498-12508, (2013).
135. Rioz-Martinez, A. & Roelfes, G. DNA-based hybrid catalysis. *Curr. Opin. Chem. Biol.* **25**, 80-87, (2015).
136. Boersma, A. J. *et al.* Catalytic enantioselective syn hydration of enones in water using a DNA-based catalyst. *Nat. Chem.* **2**, 991-995, (2010).
137. Gartner, Z. J. *et al.* DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**, 1601-1605, (2004).
138. O'Reilly, R. K., Turberfield, A. J. & Wilks, T. R. The Evolution of DNA-Templated Synthesis as a Tool for Materials Discovery. *Acc. Chem. Res.* **50**, 2496-2509, (2017).
139. Meng, W. *et al.* An autonomous molecular assembler for programmable chemical synthesis. *Nat. Chem.* **8**, 542-548, (2016).
140. Alberts, B. *et al. Molecular biology of the cell, Sixth edition*. (Garland Publishing, 2014).
141. Sleytr, U. B., Schuster, B., Egelseer, E. M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol. Rev.* **38**, 823-864, (2014).
142. Fagan, R. P. & Fairweather, N. F. Biogenesis and functions of bacterial S-layers. *Nat. Rev. Microbiol.* **12**, 211-222, (2014).
143. Missiakas, D. & Schneewind, O. Assembly and function of the Bacillus anthracis S-layer. *Annu. Rev. Microbiol.*, (2017).
144. Baranova, E. *et al.* SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly. *Nature* **487**, 119-122, (2012).
145. van Raaij, M. J., Mitraki, A., Lavigne, G. & Cusack, S. A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**, 935-938, (1999).
146. Tanaka, S., Sawaya, M. R. & Yeates, T. O. Structure and mechanisms of a protein-based organelle in Escherichia coli. *Science* **327**, 81-84, (2010).
147. Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**, 1293-1297, (2017).
148. Kerfeld, C. A., Heinhorst, S. & Cannon, G. C. Bacterial microcompartments. *Annu. Rev. Microbiol.* **64**, 391-408, (2010).

149. Yeates, T. O., Crowley, C. S. & Tanaka, S. Bacterial microcompartment organelles: Protein shell structure and evolution. *Annu. Rev. Biophys.* **39**, 185-205, (2010).
150. McMahon, H. T. & Boucrot, E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell. Biol.* **12**, 517-533, (2011).
151. Peter, B. J. *et al.* BAR domains as sensors of membrane curvature: the amphiphysin BAR structure. *Science* **303**, 495-499, (2004).
152. Johannes, L., Parton, R. G., Bassereau, P. & Mayor, S. Building endocytic pits without clathrin. *Nat. Rev. Mol. Cell. Biol.* **16**, 311-321, (2015).
153. Woolfson, D. N. & Mahmoud, Z. N. More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials. *Chem. Soc. Rev.* **39**, 3464-3479, (2010).
154. Ryadnov, M. G. & Woolfson, D. N. Engineering the morphology of a self-assembling protein fibre. *Nat. Mater.* **2**, 329-332, (2003).
155. Burgess, N. C. *et al.* Modular Design of Self-Assembling Peptide-Based Nanotubes. *J. Am. Chem. Soc.* **137**, 10554-10562, (2015).
156. Xu, C. *et al.* Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. *J. Am. Chem. Soc.* **135**, 15565-15578, (2013).
157. Magnotti, E. L. *et al.* Self-Assembly of an alpha-Helical Peptide into a Crystalline Two-Dimensional Nanoporous Framework. *J. Am. Chem. Soc.* **138**, 16274-16282, (2016).
158. Brodin, J. D. *et al.* Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375-382, (2012).
159. Suzuki, Y. *et al.* Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369-373, (2016).
160. Gradisar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* **9**, 362-366, (2013).
161. Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U S A* **98**, 2217-2221, (2001).
162. Lai, Y. T. *et al.* Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6**, 1065-1071, (2014).
163. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171-1174, (2012).
164. Mandell, D. J. & Kortemme, T. Computer-aided design of functional protein interactions. *Nat. Chem. Biol.* **5**, 797-807, (2009).
165. Grueninger, D. *et al.* Designed protein-protein association. *Science* **319**, 206-209, (2008).
166. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595-599, (2013).
167. Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102-106, (2013).
168. Delpeyroux, F. *et al.* A poliovirus neutralization epitope expressed on hybrid hepatitis B surface antigen particles. *Science* **233**, 472-475, (1986).
169. Schäffer, C. *et al.* Novel biocatalysts based on S-layer self-assembly of Geobacillus stearothermophilus NRS 2004/3a: a nanobiotechnological approach. *Small* **3**, 1549-1559, (2007).
170. Ilk, N., Egelseer, E. M. & Sleytr, U. B. S-layer fusion proteins--construction principles and applications. *Curr. Opin. Biotechnol.* **22**, 824-831, (2011).

171. Fan, C. *et al.* Short N-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci. U S A* **107**, 7509-7514, (2010).
172. Kotterman, M. A. & Schaffer, D. V. Engineering adeno-associated viruses for clinical gene therapy. *Nat. Rev. Genet.* **15**, 445-451, (2014).
173. Shenton, W., Pum, D., Sleytr, U. B. & Mann, S. Synthesis of cadmium sulphide superlattices using self-assembled bacterial S-layers. *Nature* **389**, 585-587, (1997).
174. Baneyx, F. & Matthaei, J. F. Self-assembled two-dimensional protein arrays in bionanotechnology: from S-layers to designed lattices. *Curr. Opin. Biotechnol.* **28**, 39-45, (2014).
175. Howorka, S. Creating regular arrays of nanoparticles with self-assembling protein building blocks. *J. Mater. Chem.* **17**, 2049-2053, (2007).
176. Nam, Y. S. *et al.* Biologically templated photocatalytic nanostructures for sustained light-driven water oxidation. *Nat. Nanotechnol.* **5**, 340-344, (2010).
177. Lee, Y. J. *et al.* Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes. *Science* **324**, 1051-1055, (2009).
178. Low, H. H. *et al.* Structure of a type IV secretion system. *Nature* **508**, 550-553, (2014).
179. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539-544, (1998).
180. Yan, H., Park, S. H., Finkelstein, G., Reif, J. H. & LaBean, T. H. DNA-templated self-assembly of protein arrays and highly conductive nanowires. *Science* **301**, 1882-1884, (2003).
181. He, Y. *et al.* Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **452**, 198-201, (2008).
182. Zheng, J. P. *et al.* From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* **461**, 74-77, (2009).
183. Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73-76, (2009).
184. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414-418, (2009).
185. Castro, C. E. *et al.* A primer to scaffolded DNA origami. *Nat. Methods* **8**, 221-229, (2011).
186. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **37**, 5001-5006, (2009).
187. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441-444, (2015).
188. Veneziano, R. *et al.* Designer nanoscale DNA assemblies programmed from the top down. *Science* **352**, 1534, (2016).
189. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725-730, (2009).
190. Han, D. *et al.* DNA origami with complex curvatures in three-dimensional space. *Science* **332**, 342-346, (2011).
191. Kocabey, S. *et al.* Membrane-assisted growth of DNA origami nanostructure arrays. *ACS Nano* **9**, 3530-3539, (2015).
192. Gerling, T., Wagenbauer, K. F., Neuner, A. M. & Dietz, H. Dynamic DNA devices and assemblies formed by shape-complementary, non-base pairing 3D components. *Science* **347**, 1446-1452, (2015).

193. Woo, S. & Rothemund, P. W. Programmable molecular recognition based on the geometry of DNA nanostructures. *Nat. Chem.* **3**, 620-627, (2011).
194. Goodman, R. P. *et al.* Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **310**, 1661-1665, (2005).
195. Shih, W. M., Quispe, J. D. & Joyce, G. F. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* **427**, 618-621, (2004).
196. Zhang, F. *et al.* Complex wireframe DNA origami nanostructures with multi-arm junction vertices. *Nat. Nanotechnol.* **10**, 779-784, (2015).
197. Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623-626, (2012).
198. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177-1183, (2012).
199. Zhang, Z., Yang, Y., Pincet, F., M, C. L. & Lin, C. Placing and shaping liposomes with reconfigurable DNA nanocages. *Nat. Chem.* **9**, 653-659, (2017).
200. Howorka, S. Nanotechnology. Changing of the guard. *Science* **352**, 890-891, (2016).
201. Franquelim, H. G., Khmelinskaia, A., Sobczak, J. P., Dietz, H. & Schwille, P. Membrane sculpting by curved DNA origami scaffolds. *Nat. Commun.* **9**, 811, (2018).
202. Yoshina-Ishii, C. & Boxer, S. G. Arrays of mobile tethered vesicles on supported lipid bilayers. *J. Am. Chem. Soc.* **125**, 3696-3697, (2003).
203. Baldelli Bombelli, F. *et al.* Closed nanoconstructs assembled by step-by-step ss-DNA coupling assisted by phospholipid membranes. *Soft Matt.* **5**, 1639-1645, (2009).
204. Johnson-Buck, A., Jiang, S., Yan, H. & Walter, N. G. DNA-cholesterol barges as programmable membrane-exploring agents. *ACS Nano* **8**, 5641-5649, (2014).
205. Suzuki, Y., Endo, M. & Sugiyama, H. Lipid-bilayer-assisted two-dimensional self-assembly of DNA origami nanostructures. *Nat. Commun.* **6**, 8052, (2015).
206. List, J., Weber, M. & Simmel, F. C. Hydrophobic actuation of a DNA origami bilayer structure. *Angew. Chem. Int. Ed.* **53**, 4236-4239, (2014).
207. Czogalla, A. *et al.* Amphipathic DNA origami nanoparticles to scaffold and deform lipid membrane vesicles. *Angew. Chem. Int. Ed.* **54**, 6501-6505, (2015).
208. Edwardson, T. G., Carneiro, K. M., McLaughlin, C. K., Serpell, C. J. & Sleiman, H. F. Site-specific positioning of dendritic alkyl chains on DNA cages enables their geometry-dependent self-assembly. *Nat. Chem.* **5**, 868-875, (2013).
209. Yang, Y. *et al.* Self-assembly of size-controlled liposomes on DNA nanotemplates. *Nat. Chem.* **8**, 476-483, (2016).
210. Xu, W. *et al.* A programmable DNA origami platform to organize SNAREs for membrane fusion. *J. Am. Chem. Soc.* **138**, 4439-4447, (2016).
211. Perrault, S. D. & Shih, W. M. Virus-inspired membrane encapsulation of DNA nanostructures to achieve in vivo stability. *ACS Nano* **8**, 5132-5140, (2014).
212. Howorka, S. DNA nanotechnology: Bringing lipid bilayers into shape. *Nat. Chem.* **9**, 611-613, (2017).
213. Jungmann, R. *et al.* Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano Lett.* **10**, 4756-4761, (2010).
214. Iinuma, R. *et al.* Polyhedra self-assembled from DNA tripods and characterized with 3D DNA-PAINT. *Science* **344**, 65-69, (2014).
215. Modi, S. *et al.* A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nat. Nanotechnol.* **4**, 325-330, (2009).

216. Modi, S., Nizak, C., Surana, S., Halder, S. & Krishnan, Y. Two DNA nanomachines map pH changes along intersecting endocytic pathways inside the same cell. *Nat. Nanotechnol.* **8**, 459-467, (2013).
217. Knudsen, J. B. *et al.* Routing of individual polymers in designed patterns. *Nat. Nanotechnol.* **10**, 892-898, (2015).
218. Tan, S. J., Campolongo, M. J., Luo, D. & Cheng, W. Building plasmonic nanostructures with DNA. *Nat. Nanotechnol.* **6**, 268-276, (2011).
219. Ozbay, E. Plasmonics: merging photonics and electronics at nanoscale dimensions. *Science* **311**, 189-193, (2006).
220. Klinkova, A., Choueiri, R. M. & Kumacheva, E. Self-assembled plasmonic nanostructures. *Chem. Soc. Rev.* **43**, 3976-3991, (2014).
221. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607-609, (1996).
222. Alivisatos, A. P. *et al.* Organization of 'nanocrystal molecules' using DNA. *Nature* **382**, 609-611, (1996).
223. Zheng, J. W. *et al.* Two-dimensional nanoparticle arrays show the organizational power of robust DNA motifs. *Nano Lett.* **6**, 1502-1504, (2006).
224. Aldaye, F. A., Palmer, A. L. & Sleiman, H. F. Assembling materials with DNA as the guide. *Science* **321**, 1795-1799, (2008).
225. Sharma, J. *et al.* Control of self-assembly of DNA tubules through integration of gold nanoparticles. *Science* **323**, 112-116, (2009).
226. Kuzyk, A. *et al.* DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response. *Nature* **483**, 311-314, (2012).
227. Acuna, G. P. *et al.* Fluorescence enhancement at docking sites of DNA-directed self-assembled nanoantennas. *Science* **338**, 506-510, (2012).
228. Sun, W. *et al.* Casting inorganic structures with DNA molds. *Science* **346**, 1258361, (2014).
229. Trinh, T. *et al.* DNA-imprinted polymer nanoparticles with monodispersity and prescribed DNA-strand patterns. *Nat. Chem.* **10**, 184-192, (2018).
230. Edwardson, T. G., Lau, K. L., Bousmail, D., Serpell, C. J. & Sleiman, H. F. Transfer of molecular recognition information from DNA nanostructures to gold nanoparticles. *Nat. Chem.* **8**, 162-170, (2016).
231. Liu, W. *et al.* Diamond family of nanoparticle superlattices. *Science* **351**, 582-586, (2016).
232. Kershner, R. J. *et al.* Placement and orientation of individual DNA shapes on lithographically patterned surfaces. *Nat. Nanotechnol.* **4**, 557-561, (2009).
233. Hung, A. M. *et al.* Large-area spatially ordered arrays of gold nanoparticles directed by lithographically confined DNA origami. *Nat. Nanotechnol.* **5**, 121 - 126, (2010).
234. Gopinath, A., Miyazono, E., Faraon, A. & Rothemund, P. W. Engineering and mapping nanocavity emission via precision placement of DNA origami. *Nature* **535**, 401-405, (2016).
235. Lee, M. J. *et al.* Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. *Nat. Chem. Biol.*, (2017).
236. Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753-759, (2009).

237. Flavell, R. R. & Muir, T. W. Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. *Acc. Chem. Res.* **42**, 107-116, (2009).
238. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779, (1994).
239. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53-60, (2017).
240. Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5**, 4740, (2014).
241. Zolot, R. S., Basu, S. & Million, R. P. Antibody-drug conjugates. *Nat. Rev. Drug. Discov.* **12**, 259-260, (2013).
242. Chudasama, V., Maruani, A. & Caddick, S. Recent advances in the construction of antibody-drug conjugates. *Nat. Chem.* **8**, 114-119, (2016).
243. Maruani, A. *et al.* A plug-and-play approach to antibody-based therapeutics via a chemoselective dual click strategy. *Nat. Commun.* **6**, 6645, (2015).
244. Chen, T., Hongdilokkul, N., Liu, Z., Thirunavukarasu, D. & Romesberg, F. E. The expanding world of DNA and RNA. *Curr. Opin. Chem. Biol.* **34**, 80-87, (2016).
245. Praetorius, F. *et al.* Biotechnological mass production of DNA origami. *Nature* **552**, 84-87, (2017).
246. Surana, S., Shenoy, A. R. & Krishnan, Y. Designing DNA nanodevices for compatibility with the immune system of higher organisms. *Nat. Nanotechnol.* **10**, 741-747, (2015).
247. Lee, D. S., Qian, H., Tay, C. Y. & Leong, D. T. Cellular processing and destinies of artificial DNA nanostructures. *Chem. Soc. Rev.* **45**, 4199-4225, (2016).
248. Bell, N. A. & Keyser, U. F. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores. *Nat. Nanotechnol.* **11**, 645-651, (2016).
249. Dunn, K. E. *et al.* Guiding the folding pathway of DNA origami. *Nature* **525**, 82-86, (2015).
250. Sacca, B. *et al.* Orthogonal Protein Decoration of DNA Origami. *Angew. Chem. Int. Ed.* **49**, 9378-9383, (2010).
251. Wilner, O. I. *et al.* Enzyme cascades activated on topologically programmed DNA scaffolds. *Nat. Nanotechnol.* **4**, 249-254, (2009).
252. Linko, V., Eerikainen, M. & Kostiainen, M. A. A modular DNA origami-based enzyme cascade nanoreactor. *Chem. Commun.* **51**, 5351-5354, (2015).
253. Zhao, Z. *et al.* Nanocaged enzymes with enhanced catalytic activity and increased stability against protease digestion. *Nat. Commun.* **7**, 10619, (2016).
254. Fu, J. *et al.* Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nat. Nanotechnol.* **9**, 531-536, (2014).
255. Grossi, G., Dalgaard Ebbesen Jepsen, M., Kjems, J. & Andersen, E. S. Control of enzyme reactions by a reconfigurable DNA nanovault. *Nat. Commun.* **8**, 992, (2017).
256. Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470-474, (2011).
257. Angelin, A. *et al.* Multiscale Origami Structures as Interface for Cells. *Angew. Chem. Int. Ed.* **54**, 15813-15817, (2015).

258. Praetorius, F. & Dietz, H. Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes. *Science* **355**, (2017).
259. Schiffels, D., Szalai, V. A. & Liddle, J. A. Molecular Precision at Micrometer Length Scales: Hierarchical Assembly of DNA-Protein Nanostructures. *ACS Nano* **11**, 6623-6629, (2017).
260. Henning-Knechtel, A., Knechtel, J. & Magzoub, M. DNA-assisted oligomerization of pore-forming toxin monomers into precisely-controlled protein channels. *Nucleic Acids Res.* **45**, 12057-12068, (2017).
261. Zhao, Y. H., Abraham, M. H. & Zissimos, A. M. Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds. *J. Org. Chem.* **68**, 7368-7373, (2003).
262. Choi, J. & Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **40**, 5893-5909, (2011).
263. Hansel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267-1272, (2016).
264. Colin, P. Y., Zinchenko, A. & Hollfelder, F. Enzyme engineering in biomimetic compartments. *Curr. Opin. Struct. Biol.* **33**, 42-51, (2015).
265. Fischlechner, M. *et al.* Evolution of enzyme catalysts caged in biomimetic gel-shell beads. *Nat. Chem.* **6**, 791-796, (2014).
266. He, M. & Taussig, M. J. Eukaryotic ribosome display with in situ DNA recovery. *Nat. Methods* **4**, 281-288, (2007).
267. Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* **10**, 216-222, (2014).
268. Currin, A., Swainston, N., Day, P. J. & Kell, D. B. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* **44**, 1172-1239, (2015).
269. Davids, T., Schmidt, M., Bottcher, D. & Bornscheuer, U. T. Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr. Opin. Chem. Biol.* **17**, 215-220, (2013).
270. Frushicheva, M. P. *et al.* Computer aided enzyme design and catalytic concepts. *Curr. Opin. Chem. Biol.* **21**, 56-62, (2014).
271. Durao, P. *et al.* Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat. Chem. Biol.* **11**, 148-155, (2015).
272. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354, (2016).
273. Ramakers, B. E., van Hest, J. C. & Lowik, D. W. Molecular tools for the construction of peptide-based materials. *Chem. Soc. Rev.* **43**, 2743-2756, (2014).
274. Matteucci, M. D. & Caruthers, M. H. Synthesis of deoxyoligonucleotides on a polymer support. *J. Am. Chem. Soc.* **103**, 3185-3191, (1981).
275. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215-1220, (2008).
276. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52-56, (2010).
277. Shivalingam, A. & Brown, T. Synthesis of chemically modified DNA. *Biochem. Soc. T.* **44**, 709-715, (2016).

278. Chen, Z., Lichtor, P. A., Berliner, A. P., Chen, J. C. & Liu, D. R. Evolution of sequence-defined highly functionalized nucleic acid polymers. *Nat. Chem.* **10**, 420-427, (2018).
279. Harris, L. J., Larson, S. B., Hasel, K. W. & McPherson, A. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry* **36**, 1581-1597, (1997).
280. Garces, F. *et al.* Structural evolution of glycan recognition by a family of potent HIV antibodies. *Cell* **159**, 69-79, (2014).
281. Guillard, S. *et al.* Structural and functional characterization of a DARPin which inhibits Ras nucleotide exchange. *Nat. Commun.* **8**, 16111, (2017).
282. Kato, K. *et al.* Structural basis for specific inhibition of Autotaxin by a DNA aptamer. *Nat. Struct. Mol. Biol.* **23**, 395-401, (2016).
283. Sleytr, U. B. *et al.* Nanobiotechnology with S-layer proteins as building blocks. *Prog. Mol. Biol. Transl. Sci.* **103**, 277-352, (2011).
284. Marvin, D. A., Hale, R. D., Nave, C. & Helmer-Citterich, M. Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages Ff (fd, f1, M13), If1 and IKe. *J. Mol. Biol.* **235**, 260-286, (1994).
285. Schluenzen, F. *et al.* Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* **102**, 615-623, (2000).