

Synthesis at the Interface of Chemistry and Biology*

by Peter G. Schultz

The feature that perhaps most distinguishes chemistry from the rest of the sciences is the ability of chemists to control the structure of matter at the molecular level. Unfortunately, we are not nearly as adept at the synthesis of molecules with defined functions as we are at the synthesis of molecules with defined structures. As the focus of chemistry increasingly shifts from structure to function, chemists will need to develop better strategies to efficiently generate molecules, and systems of molecules, with desired physical, chemical, or biological properties in order to meet the biomedical, energy, and environmental needs of the future. Indeed this challenge represents one of the great opportunities for synthesis in the coming years. One direction we can turn for help is Mother Nature: after all, living organisms carry out a remarkable array of complex functions using natural molecules and molecular assemblies. With this theme in mind, the focus of our work has been to exploit nature itself, i.e., use the synthetic strategies, molecules, and biosynthetic machinery of living organisms, together with more traditional chemical approaches, to generate molecules with properties that might be difficult to realize by chemical strategies alone.

An Expanded Genetic Code

As an illustration of this notion we asked the question whether our molecular level understanding and chemical/biological tools are sophisticated enough to begin to manipulate the genetic code itself, i.e., generate organisms that genetically encode 21 or more amino acids. Although the functional groups contained in the 20 amino acid code might be sufficient for life, they might not be optimal. Consequently, the development of a general method that allows us to genetically encode additional amino acids beyond the canonical 20 might facilitate the evolution of proteins, or even entire organisms, with new or enhanced properties. Moreover, the ability to incorporate amino acids with defined steric/electronic properties and chemical reactivity at unique sites in proteins should provide powerful new tools for exploring protein structure and function; much the same way physical organic chemists use synthesis to understand the chemical reactivity of organic molecules.

The incorporation of additional amino acids into proteins directly in a living organism requires the following new components of the protein translational machinery: a unique tRNA-codon pair, a corresponding aminoacyl-tRNA synthetase, and significant intracellular levels of the unnatural amino acid. To ensure that the unnatural amino acid is incorporated with high fidelity the tRNA must not be recognized by the endogenous aminoacyl-tRNA synthetases (aaRS) of the host but still function efficiently in translation (an orthogonal tRNA). Moreover, this tRNA must deliver the novel amino acid in response to a unique codon that does not encode any of the common 20 amino acids. This codon can be either one of the degenerate stop codons (e.g., an amber nonsense codon) or an efficient four-base frameshift codon. Another requirement for high fidelity is that the cognate aminoacyl-tRNA synthetase (an orthogonal synthetase) aminoacylates the orthogonal tRNA but does not aminoacylate any of the endogenous host tRNAs. Furthermore, this synthetase must aminoacylate the tRNA with only the desired unnatural amino acid, and not with any of the large number of endogenous amino acids of the host organism. Similarly, the unnatural amino acid cannot be a substrate for the endogenous synthetases if it is to be incorporated uniquely in response to its cognate codon. To this end we used a combination of structure-based design and large libraries of tRNAs and aminoacyl-tRNA synthetases, together with a series of positive and negative selections, to generate unique tRNA/aaRS pairs specific for the amino acid of interest. The positive selection is based on chloramphenicol resistance, which is conferred by the suppression of an amber mutation at a permissive site in the chloramphenicol acetyltransferase gene only in the presence of the unnatural amino acid. The negative selection uses the toxic barnase gene with amber mutations at permissive sites and is carried out in the absence of the unnatural amino acid to eliminate aaRS mutants that aminoacylate endogenous amino acids. This selection scheme and more facile variants have been used to develop orthogonal tRNA/aaRS pairs that are capable of selectively inserting one or more unnatural amino acids into proteins in *E. coli* in response to nonsense and/or four-base frameshift codons (with a cognate tRNA containing an expanded anticodon loop) in good yields (>1 g/L) and with high translational fidelities. This system has been

*Essay based on the presentation of the first "Chemistry for the Future Solvay Prize" awarded to Professor Peter G. Schultz on 4 December 2013. See more on page 5.



*Solvay Prize presentation at the Academy in Brussels
on 4 December 2013*

expanded to both yeast and mammalian cells; in addition transgenic flies and worms with a 21 amino acid code have been created. Most recently a “synthetic” *E. coli* strain has been generated in which the TAG codon has been deleted and used to uniquely specify unnatural amino acids.

On the order of 100 unnatural amino acids with novel chemical, biological, and physical properties have been genetically encoded in living organisms. These include amino acids with novel steric/packing and electronic properties for mechanistic studies; photo-cross-linking amino acids which have been used to probe protein-protein and protein-nucleic acid interactions in vitro or in vivo; keto, diketo, acetylene, azide, thioester, and boronate containing amino acids that contain functional groups with unique chemical reactivity which have been used to site-specifically introduce a large number of biophysical probes, tags, and drugs into proteins in vitro or in vivo; redox-active amino acids to modulate electron transfer in proteins; photocaged and photoisomerizable amino acids to photoregulate cellular processes; metal-binding amino acids for catalysis, protein folding and regulation; amino acids that contain NMR probes or fluorescent or IR-active side chains as local probes of protein structure and dynamics in vitro and in vivo; α -hydroxy acids and D-amino acids as probes of backbone conformation and hydrogen-bonding interactions; and sulfated amino acids and mimetics of phosphorylated amino acids as probes of protein post-translational modifications. Clearly this list will be further expanded to include many additional amino acids with novel chemical, physical, and biological properties.

In addition, we are beginning to examine the influence of an expanded genetic code on the evolution of peptides and proteins with new or enhanced properties. For example, a modified phage display system was used to evolve germline antibodies in strains that genetically encode sulfotyrosine. We found that antibodies containing the unnatural amino acid out-competed the other variants in binding HIV gp120. In a second experiment we generated a library of cyclic peptides containing unnatural amino acids using an intein-based method for cyclization. In a selection system based on inhibiting protease activity for cell survival, cyclic peptides containing an aryl ketone side chain were evolved that inhibited HIV protease by a novel mechanism involving formation of a Schiff base with a surface lysine residue and thereby destabilizing the protein. Most recently, we generated a library of β -lactamase mutants in which residues throughout the protein were randomly mutated to UAAs. In a selection scheme based on resistance to ceftazidime we isolated mutants containing UAAs with enhanced catalytic activity relative to the wild type protein, or canonical amino acid variants. Finally, we have also successfully “synthesized” an autonomous 21 amino acid bacterium that both biosynthesizes and genetically encodes the unnatural amino acid, p-aminophenylalanine. It will be of interest to compare its evolutionary fitness to that of wild-type *E. coli*. Thus, by seamlessly integrating the complex translational machinery of living cells with new chemistries and *in vitro* evolution methods, we have overcome an evolutionary constraint imposed by the universality of the genetic code. This advance may allow the generation of proteins and perhaps even living organisms with novel or enhanced properties, and underscores the power of co-opting (rather than mimicking) Nature to create novel new functions.

Harnessing the Immune System

Another example of synergy between chemistry and biology in the generation of molecules with novel functions is the development and application of diversity-based synthetic strategies, an approach inspired by the sophisticated combinatorial and mutational mechanisms by which antibodies are evolved to recognize foreign antigens with high affinity and selectivity. The notion that this natural diversity can be used to create novel chemical function was first illustrated with the generation of catalytic antibodies. Rather than attempting to design a synthetic host that selectively binds a substrate of interest and then modify it with catalytic auxiliaries, it was realized that one could

Synthesis at the Interface of Chemistry and Biology

simply co-opt the immune system to generate a highly selective natural host in the form of an antibody combining site. To generate a selective catalyst rather than a selective receptor, stable transition-state analogues (rather than substrates) were used as antigens on the basis of the Pauling notion that enzymes evolve maximum binding affinity to the transition state of a reaction. The early experiments by Lerner and co-workers and in our own laboratory involved the generation of esterolytic antibodies using phosphonate/phosphate transition-state analogues. Other approaches have since been developed to generate catalytic antibodies, including covalent catalysis, proximity effects, and general acid-base catalysis (thereby allowing us to dissect the contribution of each of these factors to biological catalysis). Using these approaches, antibod-

ies have been generated that catalyze a wide array of chemical reactions, from acyl transfer and redox reactions to pericyclic and photochemical reactions with specificities and, in some cases, rates rivaling those of enzymes.

The detailed characterization of the immunological evolution, three-dimensional structures, and mechanisms of catalytic antibodies has also helped to dissect and quantify the relationship between binding energy and catalysis in the evolution of catalytic function. Indeed the use of transition-state analogues to elicit catalytic antibodies provided “proof by synthesis” of the Pauling notion of enzymatic catalysis. In another example, a “ferrochelataase” antibody, which catalyzes the efficient insertion of metal ions into porphyrin (the last step in heme biosynthesis), was generated

Chemistry for the Future Solvay Prize

On 4 December 2013, the presentation of the first “Chemistry for the Future Solvay Prize” took place in Brussels at the prestigious Palace of the Academy of Sciences and Arts in Belgium. This prize created by Solvay in 2013 is to reward every two years a major scientific discovery that could shape tomorrow’s chemistry and help human progress.

After the welcome address by Nicolas Boël (Chair of the Board of Solvay), the laureate Peter G. Schultz was invited on stage to receive the superb crystal trophy and the 300000 € prize from the hands of Jean-Pierre Clamadieu (Solvay CEO).

The large audience from scientific, academic, and political circles in Belgium had responded to the invitation of Solvay to congratulate Peter Schultz in the presence of US ambassador Mrs. Denise Campbell Bauer and to enjoy a Science and Innovation event.

Prof. Hakan Wennerstrom (University of Lund, Sweden), Chair of the Prize Jury, reminded us of the scope and goals of this new international award. After a short recap of Peter Schultz’s career, he highlighted the scientific work and his great originality particularly in the field of stem cell differentiation techniques and modification of biochemical processes of life to incorporate unnatural amino acids into proteins.

To complete this “Science and Innovation” afternoon, Peter Schultz discussed the scope of his research. As a wonderful scientist and teacher, he managed to attract the public’s attention and showed

many potential applications of his work in medicine and chemistry.

Then, as a chemist himself, the Belgian Prime Minister Elio di Rupo stressed the initiative of Solvay and emphasized the role of science and research in our modern society and industry to respond to the changes and needs of our world.

And finally, to illustrate the “Spirit of Innovation” Bertrand Piccard, a third generation explorer of our planet, invited us virtually into the cockpit of “Solar Impulse” (around the world solar airplane) stimulating the public by its dynamism and inspiring messages.

A high level meeting, a great laureate.... See you in 2015!

 www.solvay.com/en/innovation/solvay-prize/ceremony



Price-giving of the award to the Laureate Professor Peter G. Schultz, by Solvay CEO Jean-Pierre Clamadieu (left)

against an N-methyl porphyrin, which mimics the distorted porphyrin ring of the putative transition state for metalation. The crystal structure of the Michaelis complex indeed showed that the substrate is bound in a strained conformation, providing the first direct structural evidence for the theory of substrate strain proposed by Haldane over 70 years ago. The characterization of catalytic antibodies has also provided fundamental insight into the mechanisms by which the immune system itself evolves selective receptors. For example, the first detailed structural comparisons of germline and affinity-matured antibodies revealed the critical role of structural plasticity (in addition to genetic diversity) in determining the tremendous binding potential of the germline antibody repertoire. Germline antibodies appear to have a high degree of intrinsic combining site conformational flexibility (reminiscent of the chemical instruction theory of the immune response proposed by Haurowitz and Pauling) which allows them to bind multiple, distinct ligands in different conformational states. That conformational state which binds a specific antigen is then locked and further refined by somatic mutations which occur during affinity maturation (not protein folding as proposed by Pauling). Structural and biophysical analyses of the immunological evolution of catalytic antibodies also pointed to the critical role of mutations distal to the active site in controlling the binding and catalytic activity of proteins through complex networks of side chain and backbone interactions. Indeed these studies underscore a key aspect of diversity-based synthetic strategies—the fact that analyses of the relationship between molecular structure and properties in molecules obtained by combinatorial methods often lead to new chemical insights which further increase our ability to generate new molecular function from basic chemical principles.

Diversity-based Synthesis

The demonstration that the vast structural diversity of antibody molecules can be redirected with proper chemical instruction to generate selective catalysts illustrated the utility of molecular diversity (the antibody repertoire in this case) as a new, biologically inspired “synthetic strategy” to create novel chemical properties. Shortly thereafter, libraries of other biomolecules were designed and synthesized in order to identify molecules with new or enhanced functions. These included the use of phage display libraries to generate peptides, proteins, and antibody fragments with novel specificities, and libraries of random

RNA sequences (including those containing unnatural bases with novel functional groups) to identify RNAs that selectively bind ligands with high affinity, that catalyze chemical reactions such as acyl or phosphoryl transfers, or whose structure and transcription is regulated by the binding of small synthetic molecules. Today, combinatorial strategies are impacting many areas of chemistry. This method is particularly valuable when theory has insufficient predictive power to guide molecular design with precision, and quickly provides large amounts of experimental data to guide additional experiments and/or theoretical predictions.

One particularly illustrative example is the application of diversity-based approaches to the generation of solid-state materials with novel properties. The properties of many functional materials, such as high-temperature superconductors, heterogeneous catalysts, ferroelectric materials, magnets, and even structural materials, arise from complex interactions involving the host structure, dopants, defects, and morphology, all of which are highly dependent on composition and processing. Unfortunately, our current level of theoretical understanding does not generally allow one to predict the structures and resulting properties of these materials. Given the large number of elements in the periodic table that can be used to



make compositions consisting of up to six elements, the universe of possible new compounds with interesting physical and chemical properties remains largely unexplored; combinatorial synthetic methods represent a powerful way for experimentalists and theorists alike to more effectively mine this huge chemical space for interesting new materials properties. The first application of combinatorial methods to materials science involved the synthesis and screening of libraries of thin-film copper oxides to identify high-temperature superconductors. More recently, a variety of thin

film, solution-based and physical methods (e.g., ball milling) have been used to make libraries of diverse solid state materials. In addition, a large number of scanning or parallel detection systems have been developed for rapidly screening materials libraries for optical, electronic, magnetic, adsorptive, or catalytic properties of interest. This combinatorial approach to materials discovery, is now practiced in many industries and has led to new olefin polymerization and oxidative catalysts, hydrogen storage materials, separations materials, dielectrics, phosphors, etc. and is now being applied to the optimization of complex integrated devices such as lithium ion batteries, solar cells, and computer chips. Indeed with the challenges we now face finding new environmentally friendly energy sources, combinatorial methods are likely to play a critical role in the development of enabling new materials. These include new hydrogen and methane storage materials, fuel cell catalysts, photovoltaic devices, CO₂ sequestrants, and high-energy-density batteries. This will likely be best achieved by a synergistic use of combinatorial approaches, more conventional solid-state chemistry, and theory.

Regenerative Medicine and Neglected Disease

Another particularly powerful application of combinatorial strategies, involves the synthesis of diverse libraries of nonoligomeric synthetic molecules. Just as large libraries of antibodies are genetically assembled from families of V, D, and J gene segments, it was realized that libraries of small organic molecules could be efficiently assembled from chemical building blocks. Although there are many examples of the rational design of biologically active small molecules, it remains a challenge to design *a priori* molecules that selectively activate or inhibit a desired enzyme or receptor, or modulate a specific cellular signaling pathway, regulatory circuit, or transcriptional program. As a consequence, the screening of synthetic chemical libraries offers a highly effective approach to identify biologically active molecules, especially molecules with novel cellular activities, which may not be predicted or even conceived of in hypothesis-driven experiments. However, with the increased availability and decreased cost of chemical libraries and the power of modern screening technologies, the question arises as to which opportunities should the academic chemistry community pursue with these new tools? One answer is to focus on those areas of biology which are still poorly understood and, as a conse-

quence, there exists a real need for small molecules as *in vitro* and *in vivo* probes; another is to focus on major unmet medical needs that have been largely ignored by industrial research efforts due to perceived risk or financial considerations.

A timely example (of both) is regenerative medicine, in which new cells (e.g., neurons, muscle, chondrocytes, etc.) are generated to replace tissues lost to degenerative diseases or aging. To this end, we and others are carrying out cell-based screens to identify molecules that control cell fate. For example, we have carried out image-based screens with one class of adult stem cells, hematopoietic stem cells (HSCs), to identify molecules that control self-renewal and differentiation (HSCs are adult stem cells that give rise to all the blood lineages such as macrophages, B and T cells, platelets, red blood cells, etc.). Molecules have been identified that are able to significantly expand cord blood HSCs in an undifferentiated state by antagonizing the aryl hydrocarbon receptor. These compounds will likely produce a robust source of HSCs for the large number of cancer, blood, and autoimmune disease patients for which no matched donors exist. In another experiment we have used image-based screens to identify a molecule that induces the selective neurogenesis of neural progenitor cells *in vitro* and *in vivo* in the rat dentate gyrus. This compound acts by selectively binding the centrosomal protein TACC3, which has been previously implicated in regulating the balance between progenitor cell renewal and differentiation. This and other such molecules may ultimately lead to new treatments for neurodegenerative disease. We have also identified molecules that selectively induce mesenchymal stem cells (adult stem cells which normally give rise to osteoblasts, adipocytes, and chondrocytes) to undergo osteogenesis to form bone, or chondrogenesis to form cartilage. The chondrogenic molecule, kartogenin, functions by blocking the interaction of the cystolic protein filamin A with the transcriptional coactivator CBF- β . CBF- β then traffics to the nucleus and selectively upregulates the expression of the master transcriptional regulator, Runx1. These molecules have shown excellent efficiency upon intra-articular injection in rodent osteoarthritis models. More recently, a molecule has been identified from image-based screens of oligodendrocyte precursor cells that induce their selective differentiation to oligodendrocytes. These molecules show excellent efficacy and function in both *in vitro* and *in vivo* by inducing remyelination of axons, rather than by an immunosuppressive mechanism.

Another exciting opportunity for the academic

Synthesis at the Interface of Chemistry and Biology

community to exploit chemical libraries and screening technologies that is not generally competitive with pharmaceutical or biotechnology research interests is in the area of orphan and neglected diseases. For example, there exist both a large research opportunity and a major unmet medical need with respect to molecules that kill persistent *Mycobacterium tuberculosis* (the biology of persistors is largely unknown), or molecules that target nonessential host factors that are required for viral replication (HIV, HCV, Dengue, etc.), but which will not mutate rapidly. To this end, we recently identified a molecule from a cell-based biofilm screen using *Mycobacterium smegmatis* that kills both replicating and nonreplicating Mtb as well as XDR drug resistant strains. These molecules down-regulate key mycobacterial persistence genes, and function by inhibiting two activities—the cell wall biosynthetic enzyme DprE1 and the biosynthesis of an essential molybdenum cofactor. Analogues of this compound show excellent activity in chronic models of Mtb in rodents. In addition, there are a large number of orphan diseases (type I diabetes, muscular dystrophies, spinal muscular atrophy, childhood cancers, Rett syndrome, Fragile X, Huntington disease, etc.) for which no good treatments exist. The identification of

molecules that modulate these disease processes may ultimately lead to new therapies as well as provide new insight into the biology of many of these diseases.

Conclusion

Chemistry continues to evolve from its historical focus on molecular structure, reactivity, and synthesis to take on the challenge of making small and large molecules and even systems of molecules with tailored properties and functions. This requires improved theoretical and analytical tools, as well as innovative new synthetic strategies. Given the remarkable array of functions found in biological molecules, Mother Nature offers help in this regard through an approach to synthesis that seamlessly interfaces biology and chemistry. Hopefully, the examples illustrated above from our work, and the many other elegant examples in the literature, convey the exciting and highly relevant opportunities that exist for chemical synthesis at the interface of the chemical and biological sciences. 

Peter G. Schultz <schultz@scripps.edu> is professor at The Scripps Research Institute and Director at the California Institute for Biomedical Research, in La Jolla, California.

Treasurer's Column (continued from page 2)

Even if done grudgingly a Treasurer can sometimes admit that there are other things that are as important as and perhaps even more important than money to the lifeblood of an organization. Good communications among its members and with the world at large is one of these and, in terms of general renewal, the Bureau has set as a major objective for the coming year to have a new or updated website in place by the time of the General Assembly in August of 2015. A task force, which will consult widely with the membership to establish its needs, has been established under the leadership of Bonnie Lawlor, Chair of the Committee on Publications and Cheminformatics Data Standards. This will create the vision for the website and develop informed recommendations. Every organization also needs a constant supply of new volunteers to renew and keep its business abreast of the new developments. To this end, IUPAC is streamlining the election process whereby the membership of its Divisions and

Standing Committees are chosen so that we have a more transparent and effective recruitment process. Another vital necessity for success is a coherent and far-seeing strategic plan for the future and the need has been identified to draw up a contemporary strategic plan for IUPAC that will better reflect our objectives and define how these can be achieved in the modern environment. Our President, Mark Cesa, is leading a group that is formulating a new strategic plan for IUPAC for the coming years to ensure that it continues to serve the world community of chemists in both academic and commercial endeavours and continues to provide the language and standards that our subject needs to further enhance its unique and vital contribution to the progress and betterment of humankind. An IUPAC renewal is indeed very much in the air. 

John Corish <jcorish@tcd.ie> has been treasurer of IUPAC since January 2008. He has served IUPAC at many levels since 1979, including chair of the Subcommittee on Materials Chemistry, president of the Inorganic Chemistry Division, and member of the Finance Committee.