Biopolymers
Multifaceted biopolymers
Editorial overview
Fritz Eckstein, Daniel Kahne and William F DeGrado

Current Opinion in Chemical Biology 2002, **6**:805–808

1367-5931/02/$ - see front matter
© 2002 Elsevier Science Ltd. All rights reserved.

Fritz Eckstein

Max-Planck-Institut für experimentelle Medizin, Hermann-Rein-Str. 3, 37075 Goettingen, Germany; e-mail: eckstein@em.mpg.de

Fritz Eckstein's research interest is in the structure–function relationship and mechanism of action of catalytic RNAs and their potential for the sequence-specific inhibition of gene expression as well as in the antisense methodology for this purpose.

Daniel Kahne

Princeton University Department of Chemistry, Princeton, NJ 08540, USA; e-mail: dkahne@princeton.edu

Dan Kahne's research is focused in the area of carbohydrate chemistry and biochemistry, with an emphasis on using chemical synthesis to understand the role of functionally important carbohydrate moieties in biologically active natural products.

William F DeGrado

317 Anatomy Chemistry Building, Department of Biochemistry and Biophysics, 3700 Hamilton Walk, University of Pennsylvania, PA 19104-6059, USA; e-mail: wdegrado@mail.med.upenn.edu

William DeGrado's primary research interest is in the *de novo* design of proteins, in which one attempts to design proteins from first principals. This approach critically tests his understanding of protein folding and function, while also laying the groundwork for the design of proteins with properties not precedented in nature.

This section of *Current Opinion in Chemical Biology* is devoted to the recent highlights of research into biological polymers. To bring some order to this diverse subject, the section has been sub-divided into the three fields of nucleic acid, carbohydrate and protein biopolymer research. While it is impossible to comprehensively cover all the recent developments in this vast field, this year's *Biopolymers* section is the largest ever, reflecting the importance and growth of this area.

Nucleic-acid biopolymers

The nucleic acid coverage of this section comprises several areas of timely interest. Hohsaka and Sisido (pp 809–815) review the various approaches to incorporate non-canonical amino acids into proteins. The aim of such attempts is the production of proteins with new properties. Examples of this are the introduction of fluorescent groups, of derivatives for photocrosslinking or for photoactivation. Amber stop codons are often used in conjunction with a chemically amino-acylated tRNA to achieve this goal. However, extension of the codon–anticodon pairs to four or even five nucleotides offers the advantage of potentially incorporating more than one unusual amino acid into one protein. Both these methods rely on the chemical aminoacylation of the tRNA, a cumbersome step and not applicable *in vivo*. Thus, a further development is to mutate a particular synthetase to charge an amber suppressor tRNA with a given non-natural amino acid. This approach has been developed to the stage where a large number of new functional groups can be incorporated. A disadvantage of this approach is that for each new amino acid, the synthetase would have to be mutated for recognition as substrate. In any case, we can look forward to the production of proteins with hitherto unknown properties.

The rapid and reliable detection of sequence abnormalities in genes, such as single-nucleotide polymorphisms, is of particular importance not only for gene functional analysis but also in a medical setting for personalised medicine. This field is undergoing a remarkable development, particularly in the combination of biochemical with physical techniques. The review by Vercoutere and Akeson (pp 816–822) on biosensors provides some examples of such efforts. Although biosensors have certain features in common with arrays, they differ in important aspects such as being based on fibre optics, acoustic wave or electrochemical sensors for read-out. Many of the latter type rely on the electrochemical reduction of a molecule attached to the terminus of DNA to generate a current that differs for fully complementary DNA and DNA with a mismatch. Others use gold-particle-linked oligonucleotides as probes for hybridisation of complementary oligonucleotides for electrical detection. Quite different is the approach using a protein-formed nanopore. An ionic current passing through the pore depends on the double-stranded nature of the DNA. Some of these formats are also compatible with single-molecule detection. Activities in the area of biosensors are in a phase of rapid growth so that we can look forward to more exciting developments.

Single-molecule fluorescence of nucleic acids to study dynamics and conformational rearrangements is reviewed by Molova (pp 823–828). For such investigations, the single molecules of nucleic acids can be restricted by surface attachment or trapping in a matrix to a small volume. In most cases, FRET provides the read-out. The review centers around studies of the *Tetrahymena* ribozyme for which extensive and even X-ray structural and kinetic folding data exist. The advantage of single-molecule measurements, in conjunction with time resolution, obviously is that events suggested to be heterogeneous by looking at the whole population can be dissected into individual steps and assigned to a certain percentage of molecules. Having identified, for example, several docking steps and different pathways to reach the catalytically competent conformation, efforts can then be undertaken to understand the basis for such diversity.

Small non-messenger RNAs constitute a rapidly growing field of interest. They are discussed in two reviews by Agami (pp 829–834) and by Hütttenhofer *et al.* (pp 835–843). It was reported in 1998 that double-stranded RNA can regulate the expression of developmental genes in *C. elegans*. Research since then has shown that small RNAs of 21–23 nucleotides, now known as microRNAs (miRNAs) or small interfering RNAs (siRNAs), are responsible for this effect. They are processed from endogenously expressed RNA stem–loop structures of approximately 100 nucleotides by Dicer, a multi-domain RNase III. Strangely enough, only one of the strands of the miRNA can be observed in the cell as a stable entity. This short strand usually has near perfect antisense complementarity to the 3′ UTR of the target mRNA. However, how such a binding prevents translation is as yet unknown.

An additional class of RNAs of the same size have been detected more recently, the small interfering RNAs (siRNAs). Contrary to miRNAs, siRNAs induce cleavage of the target RNA. This occurs in the center of the siRNA–mRNA hybrid as part of a RISC complex. Other than the miRNAs, the siRNAs have to be fully complementary to the target for cleavage. The siRNAs are processed from dsRNA of foreign genomes by apparently the same or very similar process as the miRNAs; however, both strands are observed in the cell. In lower eukaryotes and plants, the siRNAs protect against the foreign intruders. A most exciting development in this area is the discovery that synthetic or transcribed 21-nucleotide long double-stranded oligoribonucleotides can perform this mRNA cleavage, opening a wide window for gene functional analysis. Because of its apparent ease of application, the siRNA approach rivals the antisense method with oligodeoxynucleotides or with ribozymes for the sequence-specific inhibition of gene expression.

Yet another class of small RNAs are the small nucleolar RNAs. As the name implies, they are in most cases localised in the nucleolus where they direct site-specific 2′-*O*-ribose methylation and pseudouridylation of ribosomal RNAs and spliceosomal snRNAs in conjunction with the respective enzymes on the basis of complementarity with the target site. Such modifications can have considerable consequences that are not fully understood yet. For example, a role in genomic imprinting is discussed at present. Other functions might go beyond modification and relate to pre-rRNA folding.

Thus, there is no doubt that short RNAs have many surprises still in store and will remain the subject of exciting research for quite a while to come.

Carbohydrate biopolymers

The carbohydrate coverage of this Biopolymers section contains two items of current interest. Imperiali and Dempski (pp 844–850) review what is known about oligosaccharyl transferase, which is responsible for attaching the tetradecasaccharide GlcNAc2Man9Glc3 to nascent polypeptides in the lumen of the endoplasmic reticulum (ER). Hart and co-workers (pp 851–857) discuss *O*-linked GlcNAc transferase (OGT) and the possible cellular functions of O-GlcNAcylation, an important post-translational modification that is only beginning to be understood.

Imperialia and Dempski describe oligosaccharyl transferase as the “gatekeeper to the secretory pathway” because all proteins that are destined to be released from the cell or integrated into cellular membranes must first be tagged by this enzyme, which thus plays a huge role in the glycobiology of the organism. Oligosaccharyl transferase, which is highly conserved in eukaryotes, is a multi-protein complex that attaches a 14-residue oligosaccharide to the asparagine amides of selected Asn-Xxx-Ser sequences as polypeptides are extruded into the lumen of the ER during translation. Oligosaccharyl transferase differs from more typical NDP-utilizing glycosyltransferases in two key ways. The first is in the nature of the glycosyl donor, which consists of a huge oligosaccharide attached to a dolichol pyrophosphate leaving group. The second is in the structure of the enzyme itself, which consists of not one, but many different proteins. In yeast, for example, nine different membrane-bound subunits have been identified, of which five are essential for cell viability. The other four, while not absolutely essential for survival, appear to play an important role in determining the extent of glycosylation of proteins that enter the secretory pathway. Because there are so many interdependent components of oligosaccharyl transferase, and because these components are membrane-bound, it has been an enormous challenge simply to characterize the multi-enzyme complex. Imperiali and Dempski discuss the work that has led to the identification of the components of oligosaccharyl transferase and summarize what is known about the role of each component in *N*-linked glycosylation. The authors also mention recent work linking certain congenital disorders of glycosylation to mutations that affect the assembly of the tetradecasaccharide on the dolichol pyrophosphate donor. Because OT is promiscuous

with respect to the length and composition of the carbohydrate on the dolichol pyrophospahte donor, the truncated oligosaccharides that are formed can still be transferred onto a polypeptide chain. Therefore, as the authors imply, one way to learn more about the role of oligosaccharides on serum glycoproteins and/or on cell surfaces is to mutate enzymes responsible for assembling the oligosaccharide substrate of oligosaccharyl transferase. Imperiali and Dempski conclude their review by noting that, “with the identities of the molecular players in hand, future studies can focus on the structure and molecular mechanism of OT”. We look forward to learning more about this unusual multi-enzyme complex.

The review by Hart and co-workers is an excellent complement to Imperiali’s review. *O*-linked GlcNAc transferase (OGT) is an enzyme that transfers *N*-acetyl glucosamine to serines and threonines of nuclear pore proteins and transcription factors. Like OT, OGT is essential for cell viability and is highly conserved in eukaryotes. Both OT and OGT glycosylate a wide range of different kinds of proteins and thus play roles in a wide range of different biological processes. However, the differences between these enzymes end there. Whereas *N*-linked glycosylation takes place co-translationally in the ER and is catalysed by a membrane-bound, multiprotein complex, O-GlcNAcylation is a post-translational modification mediated by a soluble enzyme found in the cytoplasm and nucleus of eukaryotic cells. OT marks proteins for secretion from the cell and/or insertion into cellular membranes. Whereas the composition of the oligosaccharide attached by OT changes during processing through the secretory pathway, the asparagine glycosidic linkage is a permanent modification that remains for the lifetime of the protein. In contrast, OGT temporarily tags serine or threonine residues in proteins with a single *N*-acetyl glucosamine residue. Glycosylation evidently occurs in response to changing conditions within the cell, and the GlcNAc moieties can be removed under appropriate conditions by a glycosidase that is the counterpart to OGT. Whereas OT is aptly described as the “gatekeeper” to the secretory pathway, OGT functions as a sensor that glycosylates proteins to regulate their activities.

Hart and co-workers discuss what is currently known about OGT, which is an important but still somewhat mysterious enzyme. There is growing evidence that O-GlcNAcylation of proteins, mediated by OGT, is involved in the development of insulin resistance, which is the hallmark of type II diabetes. Evidently, increased production of UDP-GlcNAc, resulting from excess flux through the hexosamine synthesis pathway, leads to increased O-GlcNAcylation of proteins. OGT thus appears to function as a sensor for flux through a key nutrient-sensing pathway. The connection between increased flux through the HSP and increased GlcNAcylation of proteins seems clear, but it is not clear how the increased O-GlcNAcylation leads to insulin resistance. One possibility is that some of the O-GlcNAcylated proteins regulate the transcription of a key set of genes that control insulin sensitivity/resistance. The authors discuss mounting evidence that O-GlcNAcylation can regulate gene transcription by glycosylating both transcription factors and proteins that associate with transcription factors. In this regard, OGT functions rather like a kinase. In fact, in some proteins, including RNA Pol II, glycosylation and phosphorylation occur at the same sites but appear to have very different effects on activity. Unfortunately, OGT function has proven to be more difficult to classify than kinase function because (as far as is known) there is only one glycosyltransferase responsible for all post-translational glycosylation of nuclear proteins. In contrast, there are many kinases, each responsible for phosphorylating a subset of proteins. While crosstalk between pathways involving kinases complicates the analysis of what these enzymes do, they can be assigned to signalling pathways/regulatory networks. OGT glycosylates an enormous range of different proteins and only a fraction has even been identified. While it has been suggested that OGT plays a general role in gene silencing, Hart and co-workers note that “it is unlikely that all O-GlcNAc found on transcriptional machinery will have a common negative regulatory function”. Clearly, a lot remains to be done to understand OGT and its role in the cell. Hart and co-workers describe the need for better tools to analyze the proteins that are glycosylated by OGT as well as small molecules that can be used to specifically inhibit OGT to probe its function in various cell types and under various conditions.

Peptide biopolymers

The proteins sub-section of this issue focuses on several rapidly advancing areas of research. Tao and Cornish (pp 858–864) discuss new methods for the directed evolution of enzymes, and the use of these methods to increase the robustness and substrate selectivity of natural enzymes. The problem can be reduced to two basic problems: first, generating variants of a given protein; and second, selecting or screening for enzymes with improved properties. With increasing frequency, structure-based approaches are being combined with more random combinatorial approaches to increase the likelihood of success in the first step. Also, increasingly ingenious methods are being devised to select for proteins with the desired activity.

Protein and peptide design has provided a valuable tool for the design of nanoscale molecular assemblies. The overview of Zhang and co-workers (pp 865–871) provides a brief primer on the forces and principles involved in molecular assembly. Peptides and peptide–lipid conjugates have been fashioned that specifically self-assemble to form fibers, tubes and vesicles. Such assemblies have potential for biomedical applications, and might also be used for the fabrication of nanowires and other miniaturized electronic components. Also, the study of the assembly of certain classes of self-assembling peptides can provide insight into the mechanism of protein conformational diseases.

As our understanding of protein folding and design matures, it should be relatively straight-forward to translate this molecular knowledge into the design of oligomers that are not based on the standard backbone observed in proteins. Patch and Barron (pp 872–877) review recent advances in this area of research. To date, a number of backbones have been investigated including β- and γ-peptides, peptoids, hydrazino peptides, oligocarbamates, oligoureas and arylamides. The secondary structures formed by some of these oligomers have been experimentally or theoretically determined, and are now being used as frameworks for the construction of mimics of antimicrobial peptides and HIV-Tat. A major future challenge will be to design mimics of proteins with well-defined tertiary structures.

Until recently, our understanding of the determinants of the folding and function of membrane proteins has lagged very significantly behind our understanding of water-soluble proteins. However, as the number of three-dimensional structures of membrane proteins increases, we are beginning to obtain a glimpse of the features that stabilize membrane protein structures and how these structures dictate their functions. Liang (pp 878–884) summarizes recent analyses of membrane protein structures. These studies have shown that polar–polar interactions play an important role in driving helix–helix associations in membranes. A number of sequential and spatial motifs have been identified as being particularly important in mediating interhelical associations.