

# Cell Macromolecules

Jeremy Griggs, *University of Cambridge, Cambridge, UK*

Dennis Bray, *University of Cambridge, Cambridge, UK*

Macromolecules: nucleic acids, proteins and polysaccharides, form the molecular basis of all living organisms. Molecular recognition between macromolecules governs all of the most sophisticated processes in cells.

## Macromolecules

Macromolecules are very large molecules (greater than about 5000 daltons) made as polymers of subunits in an exact sequence. Living cells contain three major classes of macromolecules: (1) nucleic acids, polymers of nucleotides; (2) proteins, polymers of amino acids; and (3) polysaccharides, polymers of sugars. Macromolecules constitute a major portion of the total dry weight of both eukaryotic and prokaryotic cells. They are responsible for all of the most sophisticated functions of living cells, such as the catalysis of chemical reactions, the building of complex structures, the generation of movement and the transmission of hereditary information.

Macromolecules individually fulfil many distinct roles in the cell and, in addition, work together in large multi-molecular complexes. Nucleic acids alone encode the inherited genetic information of the cell, whereas proteins interact with DNA to regulate the expression of genes, which in turn direct the production of further proteins. Polysaccharides are the principal source of stored energy required to drive the cells' countless chemical reactions, but they are also involved in maintenance of cell structure integrity. Thus, the ability of macromolecules to recognize other molecules with great specificity is essential for their function in the cell.

## How Macromolecules are Made

All macromolecules are formed as a result of additive polymerization of their constituent subunits. In every case, the polymer-forming reactions are driven by metabolic energy, and coupled to nucleoside triphosphate hydrolysis. However, the way in which macromolecular subunits are linked is distinct for each class of macromolecule.

### Nucleic acids

The nucleic acids deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are polymers of nucleotides, arranged in a specific sequence. To form macromolecular polymers, nucleotides are joined between the 3' and 5' carbon atoms in their sugar moiety by a phosphodiester

bond, giving rise to a nucleic acid with a sugar–phosphate 'backbone' to which is attached a series of bases in a specific order. Hydrogen bonding between pairs of bases can occur, leading to the formation of double-stranded polymers if the sequences are complementary (see below).

### Proteins

Proteins are polymers of amino acids arranged in a specific sequence. Adjacent amino acids are linked by a peptide bond between the carboxylic group of one amino acid and the amino group of the next, so that a protein molecule has an amino, or N-terminus, and a carboxyl, or C-terminus. The polymerization event occurs during translation within a multimeric complex composed of proteins and ribosomal RNA (rRNA) called the ribosome. The specific amino acid sequence is determined by the sequence of bases in a messenger RNA (mRNA) molecule, produced by gene transcription, that is also transiently associated with the ribosome.

### Polysaccharides

The sugars that comprise polysaccharides are typically glucose or *N*-acetyl galactosamine. The subunit sugars may be linked in straight or branched chains, and may be repeating units of the same or different monosaccharides. Glycogen, for example, is a polysaccharide made only of glucose units. Subunits of polysaccharides are bonded by glycosidic bonds, formed in a condensation reaction that requires energy, provided by nucleoside triphosphate hydrolysis.

## Forces, Interactions and Molecular Recognition

The shapes of macromolecules and their functions all depend on chemical interactions. In broad terms, these are of two kinds. Covalent bonds, stable chemical bonds between atoms which share one or more pairs of electrons, link atoms within a macromolecule. Weaker, noncovalent forces act both within a macromolecule, between different

### Introductory article

#### Article Contents

- Macromolecules
- How Macromolecules are Made
- Forces, Interactions and Molecular Recognition
- Nucleic Acids
- Proteins

macromolecules, and between macromolecules and other molecules such as water. Although they are individually much weaker than covalent bonds, noncovalent forces have, collectively, a major influence on macromolecular conformation and interaction.

## Hydrogen bonds

Hydrogen bonds are noncovalent interactions in which a hydrogen atom is sandwiched between two electronegative atoms, usually oxygen or nitrogen. The most common example is in liquid water, where hydrogen bonds form between oxygen atoms in neighbouring molecules. However, hydrogen bonds are also a universal feature of macromolecules and account for many of their remarkable properties. The pairing of bases in a DNA double helix for example, arises from hydrogen bonds, three in the case of cytosine:guanine pairing, and two between adenine and thymine. Hydrogen bonds are also crucial in the folding of a polypeptide chain into the correct three-dimensional conformation, where hydrogen bonds form between neighbouring amino acids.

## Ionic bonds

Ionic bonds are most familiar in substances such as solid sodium chloride, but also play an important role in many macromolecules. For example, ionic bonds often facilitate the interaction between an enzyme and its substrate, allowing charged amino acid residues in the enzyme's active site to be precisely positioned, facilitating interaction with oppositely charged groups on the substrate.

## Van der Waals forces

Van der Waals forces occur between atoms at very short distances. Fluctuations in the electrical charges of two atoms cause a weak bonding between them, a force which is abolished and changes to repulsion if the two atoms are brought closer together. The distance at which an atom shows the strongest bonding to a neighbouring atom is called its van der Waals radius.

## Hydrophobic forces

Water constitutes approximately 70% of eukaryotic cells, therefore interactions between water and other molecules in the cell are extremely important. Macromolecules typically contain both hydrophilic and hydrophobic chemical groups, and in water tend to adopt conformations in which hydrophobic groups (such as hydrocarbon side-chains) are clustered on the interior of the molecule, and hydrophilic groups are exposed on the surface of the molecule. In contrast, proteins that are in a hydrophobic environment, such as a lipid membrane, often take up the

opposite configuration, with their hydrophobic regions arranged around a hydrophilic core. This segregation of groups, which can be thought of as another noncovalent force, is of major importance in the folding of proteins.

## Nucleic Acids

### Molecular recognition and the DNA double helix

Complementary strands of DNA form a double helix. Hydrogen bonds link complementary base pairs (three between guanine and cytosine, two between adenine and thymine). The overall geometry of the DNA double helix has important implications for macromolecular recognition of specific DNA sequences by factors which regulate the expression of individual genes.

Noncovalent bonds again play a crucial role. The edge of each base pair is exposed at the surface of the double helix, each with a distinct configuration of hydrogen bond donors and acceptors. Sequences of base pairs then permit the specific binding of gene regulatory proteins to DNA sequences such that precise control of gene expression can be achieved.

### Importance of sequence integrity

The fundamental importance of sequence specificity in a macromolecular polymer is most apparent when we consider the nucleic acids, particularly DNA. Nucleic acids are polymers of nucleotide subunits in a precise sequence. Every time a cell divides, its DNA must be faithfully duplicated. This process must be extremely accurate, to ensure that the hereditary information is passed on in a correct form to the daughter cells. In fact, fewer than one mistake in  $10^9$  nucleotides arises during DNA replication. Mistakes that do arise, where the nucleotide sequence is changed, are termed mutations – permanent changes in the DNA sequence. The information contained within a gene must also be faithfully decoded through transcription and translation each time a protein is produced, to ensure the correct amino acid sequence.

Why is the sequence of bases in a DNA molecule important? To answer this question we need to examine how the cell's genetic information is stored and used, and the implications for the cell, and indeed the organism, if either of these mechanisms fails.

DNA encodes the genetic information inherited from the parent cell. Genes are broadly defined as sequences of DNA that are transcribed as a single unit of RNA, which either (a) encodes a protein or one of a series of related polypeptides, or (b) is a functional RNA molecule: rRNA or transfer RNA (tRNA). The production of proteins is achieved by transcription of the DNA sequence into an

mRNA message, and the subsequent translation of that information by the ribosome to generate polypeptides: DNA makes RNA makes protein – the central dogma of molecular biology.

Genes are expressed (usually by the synthesis of a specific protein) as a result of activation of gene transcription, in response to a vast variety of factors. Since the information at the genetic level is faithfully transcribed and translated, it is the downstream implications for RNA and hence, protein subunit sequence specificity that are of most importance when considering errors in sequence at the DNA level.

## Mutations: an illustration

The gene for human  $\beta$ -globin, part of the oxygen-carrying haemoglobin molecule, consists of a sequence of bases which are translated into amino acids. The sixth codon of beta globin reads 'GAA', which is translated into the amino acid glutamic acid. At some point in the past, a random mutation occurred that changed this triplet from GAA to GTA, causing it to encode valine in the place of glutamic acid (**Figure 1**). Because valine is more hydrophobic than glutamic acid, it changes the properties of the  $\beta$ -globin, so that it aggregates into long polymeric rods. The protein rods distend the plasma membrane of red blood cells carrying the mutant haemoglobin, making them fragile and liable to clog fine capillaries. People carrying this mutation suffer from a form of chronic anaemia: 'sickle cell disease' and often die before adulthood. Interestingly, the birth frequency of homozygotes in parts of equatorial Africa is 1 in 40, and 1 in 3 people carry the mutation, an unexpectedly high incidence for such a damaging phenotype. In fact, this mutation has a selective advantage in areas where malaria caused by *Plasmodium*

*falciparum* is a major cause of death, as infected red cells in individuals heterozygous for the mutation are cleared more rapidly and the chance of recovery from a malarial infection is increased.

## Proteins

Proteins, the final products of gene expression, are polymers of amino acids, arranged in a specific sequence, encoded at the genetic level. Proteins govern the behaviour of the cell, and are necessarily involved in many processes. Numerous 'housekeeping' proteins are present in the cell at relatively constant levels, and are involved in pathways that are ongoing in a living cell. Others must be present at an abundant level only at specific times, for example in the cell cycle. The production of many proteins, via the expression of the relevant genes, dramatically increases in response to changes in the cells microenvironment, or as the final step of a signal cascade in response to an external signal such as a growth factor binding to its receptor on the cell surface. Each aspect of protein biochemistry is a complex area of expertise. An underlying theme in all areas of protein function, however, is that of molecular recognition, which in turn depends on the structure of the protein molecule.

## Protein structure

To understand the implications of amino acid subunit sequence in protein folding from a macromolecular perspective requires an understanding of the properties of the 20 different amino acids. Amino acids share a common molecular structure comprising an amino group and a carboxylic acid group linked by a central carbon atom,  $C_{\alpha}$ . A side-chain group, also linked to  $C_{\alpha}$ , confers the individual chemistry of each amino acid. The properties of these side-chain groups determine how each amino acid will behave as part of a polypeptide structure and how, collectively, the amino acid sequence dictates the overall structure and properties of the whole protein.

Amino acid side-chains can be classed as acidic, basic, uncharged polar or nonpolar. These characteristics play a major role in determining the structure of proteins, as a consequence of the various noncovalent interactions discussed above. For example, nonpolar amino acids tend to cluster on the inside of proteins, whereas hydrophilic uncharged polar amino acids position themselves predominantly on the outside of proteins. Proteins adopt the most stable conformation that results from interactions between side-chains of amino acids with each other and with the cellular environment and principally involve noncovalent interactions such as hydrogen and ionic bonds. Additionally, cysteine residues, which contain a sulfur atom in the side-chain, can pair to form covalent

Amino acid number	1	2	3	4	5	6	7	
Amino acid residue	Val	His	Leu	Thr	Pro	Glu	Glu	$\beta$ chain
	Val	His	Leu	Thr	Pro	Val	Glu	mutant $\beta$ chain
mRNA sequence						G A G		$\beta$ chain
						G U G		mutant $\beta$ chain
DNA sequence						C T C		$\beta$ chain
						C A C		mutant $\beta$ chain

**Figure 1** A single point mutation in the gene for the  $\beta$  chain of human haemoglobin changes the amino acid encoded by codon 6 from glutamine to valine. This single subunit dramatically alters the properties of the protein product, which aggregates within red blood cells, causing them to be fragile and liable to clog fine capillaries and producing a form of anaemia: sickle cell disease.

disulfide bonds, which pin portions of the folded chain together.

The sequence of amino acids itself is termed the primary structure. Local folding of the polymeric chain, governed largely by hydrogen bonds, between adjacent amino acids, give rise to the secondary structure of the protein. Tertiary structure is derived from globular units, formed by combinations of secondary structure components. Multimeric protein complexes are also frequently produced from multiple subunits, giving rise to structures such as ribosomes and enzyme complexes.

Noncovalent interactions, principally hydrogen bonding, are crucial in ensuring that a polypeptide chain folds into the correct three-dimensional conformation. Two common folding patterns produced in this way are  $\alpha$  helices and  $\beta$  sheets. Both play an important role in the secondary structure of a protein, enabling it to adopt its correct structural conformation. An  $\alpha$  helix results from the regular folding of a single polypeptide and the formation of hydrogen bonds between adjacent peptide bonds. Although individual  $\alpha$  helices are unstable by themselves they are often stabilized by wrapping around other  $\alpha$  helices to produce a coiled coil. Beta sheets result from the polypeptide folding back and forth to form 'layers' between which hydrogen bonds form to give a stable structure. The polypeptide chains in adjacent layers may run in the same or in opposite directions, forming parallel or antiparallel  $\beta$  sheets respectively. Combinations of  $\alpha$  helices and/or  $\beta$  sheets form the core of protein domains and confer stability on the protein. Certain combinations occur frequently in diverse protein types and are known as 'motifs'.

The complicated folding of a polypeptide chain brings many amino acids at distant parts of the linear polymer into close proximity, usually held in precise geometrical relationships. The atomic contours and chemical properties of such juxtaposed amino acids are crucial for the distinctive properties of a protein, for example, its ability to catalyse a specific reaction.

## Molecular recognition

The regular secondary structures of a protein, such as  $\alpha$  helices and  $\beta$  sheets, stabilize the interior of the protein and control its overall shape, or conformation. The exposed amino acid residues dictate the chemical properties of the protein, including the capacity for molecular recognition.

Molecules to which individual proteins can bind are termed ligands. Ligand:protein interactions are noncovalent and occur as a result of the noncovalent forces and interactions already discussed. As these forces are individually very weak, many are required to enable a high degree of affinity between protein and ligand. Only ligands that 'fit' precisely onto the binding site on the surface of the protein can bind tightly to it.

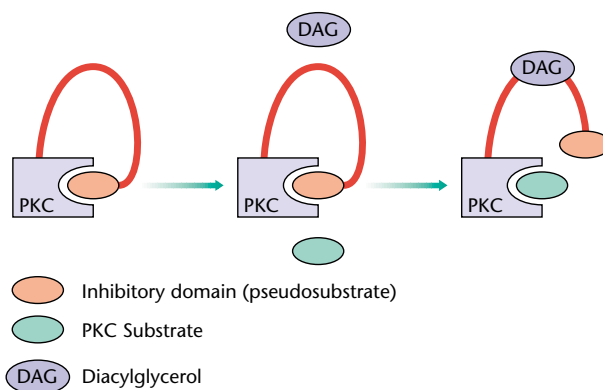
A binding site on the surface of a protein is formed from the specific position of surface amino acid residues – an aspect of the protein's conformation. The affinity of this binding site for its ligand is often increased by the exclusion of water molecules from the site, achieved because it is energetically more favourable for water molecules to form a hydrogen-bonded network than to enter into protein surface crevices. Additionally, the clustering of charged amino acid residues in a binding site (forced together as a result of protein folding, despite their mutual repulsion) often dramatically increases the affinity of the binding site for a ligand with charges of the opposite sign.

## Molecular recognition: an illustration

Proteins direct the activities of the cell, govern and regulate gene expression, and participate in many regulatory and signalling events. The mechanisms by which they achieve these wonders are a subject by themselves, and outside the scope of this short article. But we can catch a glimpse of how real proteins work in the cell with a specific example.

Protein kinase C is an enzyme that catalyses the transfer of a phosphate group from adenosine triphosphate (ATP) to specific amino acids in target proteins. This reaction is tightly controlled and takes place only when the enzyme is 'switched on' in response to chemical signals coming from other parts of the cell. The mechanism of this switch provides an insight into the way macromolecules work.

Protein kinase C contains a catalytic domain, which is the binding site for its target protein ligand (Figure 2). However, in the inactive state of the enzyme, this binding site is occupied by another domain of the same protein, which folds back on itself. So long as the catalytic domain is occupied by the inhibiting domain, the actual ligand for the catalytic binding site cannot enter and protein phosphor-



**Figure 2** The catalytic activity of the enzyme protein kinase C (PKC) is regulated by diacylglycerol (DAG). In the absence of DAG, the active site of PKC is occupied by another part of the macromolecule (the pseudosubstrate), rendering the enzyme inactive. PKC is activated when DAG binds, causing a conformational change such that the active site is no longer occupied by the pseudosubstrate and can interact with its substrate.

ylation is prevented. However, if the protein becomes exposed to the small molecule, called diacylglycerol, which is produced in response to specific environmental signals, kinase C can become active. The inhibitory domain (the 'pseudosubstrate') dissociates from the catalytic site, which is thereby able to phosphorylate target proteins.

Changes similar to those just described for protein kinase C are seen over and over again in many thousands of other proteins inside a typical animal cell. Many protein sequences have the capacity to fold in different ways, giving rise to multiple ligand-binding sites for different ligands on the same protein. Frequently, as in protein kinase C, the different conformations have a regulatory effect, allowing the process to be controlled by another unrelated substance in the cell, and there are other enzymes that use a similar

'foot-in-the-mouth' foldback inhibition. But the possibilities do not end there, and there are also proteins that build precise structures, pump ions, walk along filaments, generate light, and perform many other sophisticated functions, all based on the same general principles.

## Further Reading

- Alberts B, Bray D, Lewis J *et al.* (1994) *Molecular Biology of the Cell*, 3rd edn. New York: Garland.
- Branden C and Tooze J (1999) *Introduction to Protein Structure*, 2nd edn. New York: Garland.
- Connor M and Ferguson-Smith M (1997) *Essential Medical Genetics*, 5th edn. Boston, MA: Blackwell Science.
- Stryer L (1988) *Biochemistry*, 3rd edn. New York: WH Freeman.