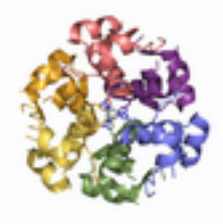
Insulin: An early structure with sweet success



As we don't eat constantly, the level of sugar in our bloodstream varies with time and needs to be controlled. For example, after the sugar rush of a meal, blood glucose levels rise too high and need to be reduced. This is achieved by the protein hormone insulin which is synthesised by beta cells in the pancreas. Once released into the bloodstream, insulin signals that glucose is in abundance and that liver, adipose (fat) and muscle cells need to take up the excess glucose and store it either as glycogen (liver) or as triglycerides (adipose).

Failure of this signalling results in high blood sugar and *diabetes mellitus*, which affects around 300 million people worldwide. This condition occurs if the body produces too little insulin, or if insulin which is produced fails to activate the receptors on the cells to which it should be transmitting its message. If left untreated, *diabetes mellitus* can have serious long-term complications which may be fatal. The discovery of insulin earned Banting and Macleod the Nobel Prize in physiology or medicine in 1923.

Small, but beautifully formed

Insulin is a small protein: its gene encodes only 110 amino acids. Actually, this is an immature form known as preproinsulin; the active form is even smaller and consists of two protein chains. To generate the active form, the protein must be processed (Figure 1). The first 24 amino acids form a signal peptide, indicating that the protein should be directed to the secretory apparatus of the cell. These residues are removed once preproinsulin is in the endoplasmic reticulum and the remaining 86 amino acid residues fold into proinsulin (PDB entry 2kqp) (view-1). This is further processed by protease enzymes (known as prohormone convertases) that remove the central section of the molecule leaving two discrete polypeptides of 21 and 30 amino acids in size (Figure 1). The amino-acid sequence of these two peptides was famously determined by Fred Sanger in 1955, for which he was awarded the Nobel Prize in chemistry in 1958. The three dimensional structure of insulin was one of the earliest protein structures to be solved. It was studied by Nobel Prize-winning crystallographer Dorothy Hodgkin. She and her team worked extensively on this from the 1930s, finally determining the structure of porcine insulin (which is only one amino acid different to human) in 1969.

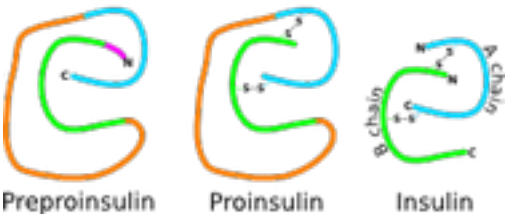


Figure 1. Insulin is produced as a preproprotein. The signal sequence (magenta) is cleaved in the endoplasmic reticulum to yield proinsulin. The central C-peptide (orange) is excised to give mature insulin.

The joy of six

The structure (e.g. PDB entry 2ins) revealed that the longer chain, called the B chain, forms an α-helix and a β-strand, whereas the A chain consists of two α-helices (view-1). These chains are linked by two disulfide bonds. This is the mature form of the insulin molecule, as secreted by the pancreas. The excised central portion, known as the C-peptide, is also secreted and has a biological activity of its own.

The two disulfide-linked polypeptides associate with another insulin molecule to form a dimer, and this aggregates with two additional dimers to form a hexamer (view-2). Within the pancreatic beta cells, proinsulin is stored in this hexameric form which is inactive and very stable. The hexamer is assembled around two zinc ions at its centre which are coordinated by a histidine residue from each monomer (PDB entry 1trz) (view-2). Insulin can bind more zinc and a four zinc hexamer (PDB entry 1zni) was crystallised soon after the two zinc form.

Stability? Precipitate is the solution

The proinsulin hexamer is soluble but once it is converted to the mature form in the storage granules the hexamer precipitates, actually forming microcrystals. This precipitation further increases the stability of the mature insulin, making it inaccessible to proteases. As only monomeric insulin is able to bind to its receptors in the liver and adipose tissue, insulin hexamers must dissociate once released from the beta cells to be biologically active. In the bloodstream, the concentration of zinc is much lower than in the storage vesicles of the beta cell, enabling dissociation by removal of the zinc. The pH of the blood is also higher than in the storage vesicles, probably contributing to the dissociation of the hexamer.

While examining the structure of the hexamer, attention of the crystallographers was drawn to glutamate 13 in the B chain. This residue is located at the centre of the hexamer, an unusual position as it places three negative charges together, all repelling each other (view-3). The presence of the histidine-coordinated zinc is crucial to overcoming the charge repulsion and stabilising the hexamer. This was demonstrated when the glutamate was replaced by glutamine (PDB entry 1izb), which doesn't have a charged sidechain; the resulting mutant protein formed hexamers even in the absence of zinc. Why does insulin have a glutamate if it is potentially deleterious to hexamer formation? It is likely that this residue promotes the dissociation of the hexamer into the active monomeric form once in the bloodstream.

The amino acid sequence of mature insulin is highly conserved across the animal kingdom. Porcine insulin differs from human insulin in only one amino acid at the C-terminus of the B chain, and bovine insulin differs from human in only three residues. This conservation meant

that, before the advent of recombinant technology, animal insulin could be purified and used to treat diabetes.

A spoonful of sugar helps the medicine go down...

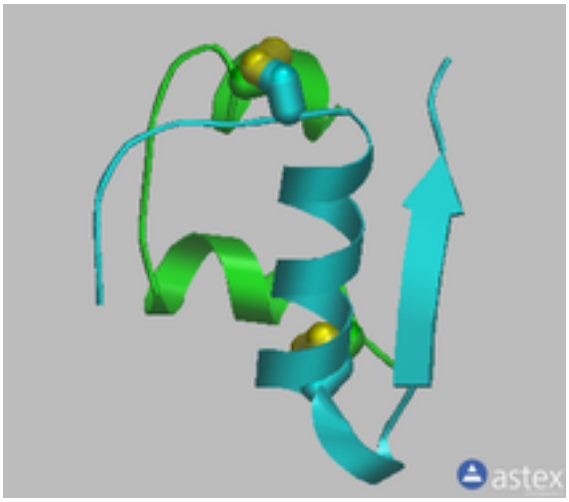
...But not if the medicine is insulin as this is administered by injection under the skin! With this method, the rapid dilution of zinc and the pH change that take place when insulin is secreted by the pancreas do not occur and stable hexamers remain, which leads to suboptimal drug delivery. With the advent of recombinant technology, modified forms of insulin have been produced which are better suited as drugs compared to the wildtype human protein. A form in which residue 28 at the C-terminal end of the B chain is mutated from proline to aspartate (insulin aspart - NovoRapid) (PDB entries 1zeg, 1zeh or 1kei) (view-4) displays greater flexibility in the C-terminal region, disrupting the assembly of the insulin dimer and favouring the monomeric form. This leads to a more rapid transfer of the molecule from the subcutaneous injection site into the bloodstream. Similarly, the drug Humalog (PDB entries 1lph or 2kjj) (view-4), sold by Eli Lily, has two amino acids swapped at the C- terminus of the B chain (Pro28 and Lys29 are changed to Lys28 and Pro29). This change also favours the monomeric form whilst not affecting the activity of the hormone. In both cases, the hexameric form prevails in the presence of zinc and the high concentrations used for crystallization and drug storage, but compared to wildtype insulin, monomerisation is much more rapid *in vivo*.

In some cases, a fast acting insulin may not be what is required. A synthetic form has been engineered and is marketed under the name Levemir which has a fatty acid covalently bound (PDB entry 1xda) (view-4). This causes the molecule to bind to albumin in the blood serum, which effectively competes with the insulin receptor and thereby prolongs the action of the drug. A further form, insulin degludec, (PDB entry 4ajx), has a different fatty-acid modification and is being developed by Novo Nordisk. It has an even longer action because, in addition to the albumin binding, its hexamers are very stable. If insulin degludec should come to the market, diabetic patients would possibly only need to inject insulin once every 2-3 days rather than several times per day.

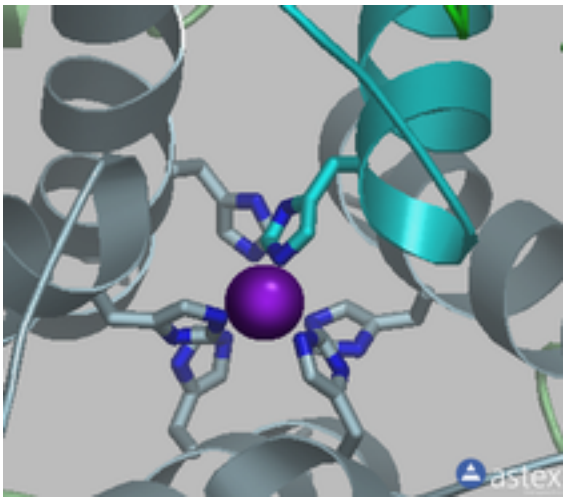
Further exploration

More than forty years after the structure of insulin was first determined, and after four Nobel prizes, you might think that we know all there is to know about insulin. However, there are still structures of insulin being deposited in the PDB, and amazingly, the structure of insulin bound to its receptor was only published in early 2013 (PDB entries 3w11, 3w12, 3w13 and 3w14). So for all young structural biologists out there, there is still much to be done.

Views

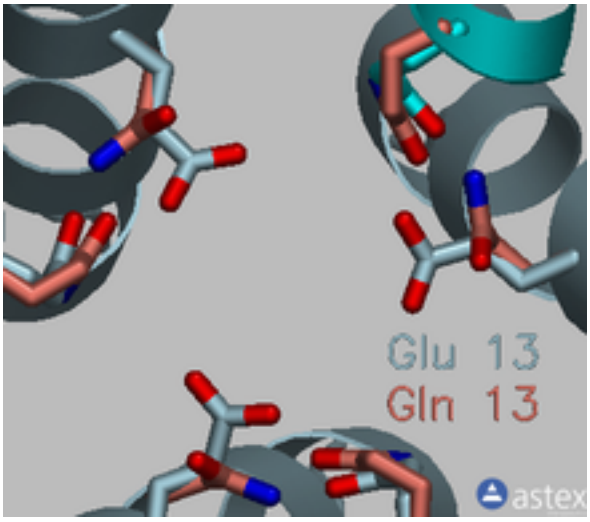


View 1: Proinsulin is converted to insulin. Proinsulin is shown in cartoon, highlighting the A chain (**green**), the B chain (**cyan**) and the C-peptide (**orange**) which is excised to yield mature insulin. The two intermolecular disulfide bonds are shown as sticks.

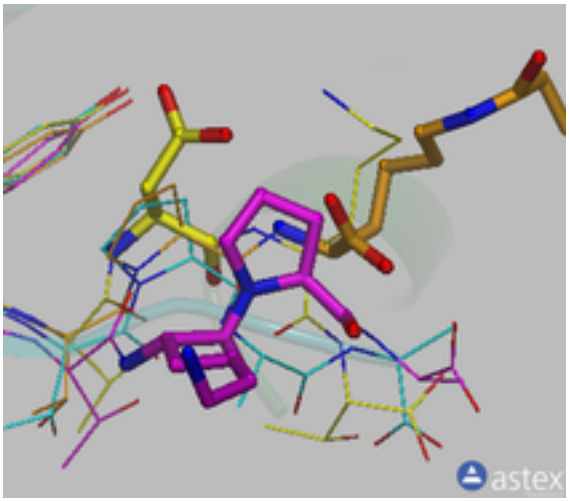


View 2: Insulin forms a hexamer. A monomer, consisting of an A chain (**green**) and B chain (**cyan**), forms a dimer with another monomer. Three dimers then form a hexamer. Two zinc ions on the central threefold axis (**purple** spheres) are coordinated by histidine residues (shown as

sticks).



View 3: Repulsive glutamates. Glutamate 13 (shown as **light blue** sticks) is at the centre of the hexamer. Overlayed in **salmon** is the Glu13Gln mutant in which the glutamine side chains only interact within the insulin dimers.



View 4: New insulin forms to use as drugs. Three commercially available engineered forms of insulin are compared to wildtype insulin. The C-terminal region of their B chains, where the differences lie, are shown as stick models: wildtype (**cyan** and **green** carbons), Novorapid (**yellow** carbons; Pro28Asp mutant), Humalog (**magenta** carbons; Pro28Lys and Lys29Pro double mutant), Levemir (**orange** carbons; myristoylated Lys29).