Molecular Recognition

11. Reference deleted.

Received 30 March 1993

Molecular recognition in insulin assembly
G. G. Dodson, E. J. Dodson, J. P. Turkenburg and Xiao Bing
Department of Chemistry, University of York, Heslington, York Y01 5DD, U.K.

Introduction
The assembly of insulin into a zinc-containing hexamer is well understood both in biological and in structural terms [1]. The molecule is synthesized in the B-cells of the islets of Langerhans in the pancreas [2]. Initially it is manufactured as a single-chain molecule, pre-pro-insulin, which is directed across the endoplasmic-reticulum membrane. This molecule is rapidly cleaved to pro-insulin, which is still a single-chain molecule, which spontaneously folds into its correct conformation. The pro-insulin then assembles into dimers, and, in the presence of the high levels of zinc that are in the B-cell, into zinc-containing hexamers. The conversion of pro-insulin into insulin evidently occurs in the hexameric state in the storage vesicle [3].

The pro-insulin assembly process confers on the molecule properties that are highly advantageous for biosynthesis. Firstly, the pro-insulin hexamer is very soluble in the presence of zinc and calcium, both of which are at high levels in the B-cell. Secondly, the conversion of pro-insulin into insulin is associated with a profound change in solubility, since zinc ions precipitate insulin quantitatively. Thus, the concentrated solution of pro-insulin in the storage vesicle is efficiently converted into an insulin precipitate which, moreover, forms micro-crystals at the pH of the B-cell [3]. The assembly properties of insulin, which are illustrated in Figure 1, are therefore an inheritance of the biosynthetic machinery, and have nothing to do with the biological activity of the hormone, which is expressed only by the monomer.

The slow absorption of injected insulin in diabetic patients slows down the arrival of the hormone in the blood from the injection site [4]. A major factor in the sluggish delivery of injected insulin is its self-assembly. To improve insulin absorption, monomeric insulin was manufactured by protein-engineering methods. It was hoped that the knowledge of the active surfaces of the molecule would allow mutations to be made which would abolish self-assembly without compromising biological action [4].

Structures of the dimer-forming surfaces
Examination of the crystal structures of dimeric and of hexameric insulin molecules allows the dimer species to be characterized. In the presence of zinc ions the dimer assembles into the hexamer just as with the precursor pre-pro-insulin [5, 6]. The dimer-forming surface on the monomer is composed of a large non-polar and predominantly aromatic region of about 400 Å². This surface is roughly planar. Along one side lies the B chain C-terminal residues B22-B30, of which B24-B26 are in a β-strand conformation. In the dimer, these two C-terminal strands come together to form an antiparallel β-sheet. The hydrogen bonds give specificity and compactness to the interactions between the non-polar surfaces. The equilibrium constant for dimer...
The molecules are drawn with only Ca atoms shown. The B chain residues B1–B8 are drawn as thick lines, the positions of the zinc ions are indicated by circles and the His sidechains are shown. In the R6 hexamer, the phenol molecules can be seen packed between the dimers.

The contacts between the monomers and the dimers undergo some changes in different insulin-crystal forms and in different insulin oligomers. The structures of two dimeric insulin crystals that have been grown in the absence of zinc, one at acid pH, the other at neutral pH have been determined [7, 8]. The crystal that had been grown at neutral pH has one molecule in the asymmetric unit, the dimer two-fold axis is used by the crystal. The crystal grown at acid pH has two molecules in the asymmetric unit, here the local axis of the dimer is not precise. Comparison shows that the interactions along the dimer axis are quite well preserved, although there are some small but distinct movements away from symmetrical relationships in some of the residues. These variations in structure indicate that the non-polar dimer-forming surfaces are more adaptable than their close and specific interactions would suggest.

Within the dimers of the hexamer the same interactions are seen as in the dimers themselves; again there are deviations from symmetry that are similar in size to those that are seen in the low-pH dimer. In particular, the valine B12 sidechain is disordered in one molecule and is packed into two quite specific conformations [6]. The central B chain helix and the C-terminal extended segment are essentially identical in structure, and there is also close similarity between the A chains as well, though the A chains do move relative to the B chain in some structures (U. Derewenda and Xiao Bing, unpublished work) [5].

Mutations to reduce dimer stability
The most direct and least disturbing approach to reduce dimer formation appeared to be to perturb the specific fit made by the dimer-forming surfaces.
Unfortunately, this philosophy failed. The repulsion between charged groups that are brought together by dimer formation, however, proved much more successful. In these experiments, moreover, most of the residues selected were known to be unimportant for binding of insulin to the receptor and for biological activity, and thus the therapeutic efficacy of the mutant insulin should be unaffected.

**B12 Val → Ile**

The mutation B12 Val → Ile inserts two methyl groups into the monomer–monomer interface. In response to the arrival of these groups there are adjustments in sidechains and in main-chain structure with magnitudes of up to 1 Å; these movements extend via a series of linked contacts over a considerable distance from the mutation sites (Xiao Bing, unpublished work) [9]. The ability of this mutant to dimerize is reduced, but only partially, suggesting that the energy penalty that is paid to restructure the interface between the molecules is only of the order of 1–2 kcal/mol (4.2–8.4 kJ/mol). The variations between different crystal forms suggest that the non-polar regions of the insulin surface are relatively plastic. It may well be that this adaptability is a general phenomenon in proteins.

**B9 Ser → Asp**

The stability of the dimer has been reduced successfully by electrostatic repulsions that have been introduced by mutations at groups that are brought together when the molecule dimerizes [4]. Two such mutants have been investigated. These are B9 Ser → Asp (with the additional mutations, B27 Thr → Glu and A8 Thr → His) and B10 His → Asp (and B28 Pro → Asp). The substitution of aspartic acid at B9 brings four carboxylic acids (two B9 aspartic acids and two B13 glutamic acids) into close proximity on the dimer surface. This mutation generates a molecule with strongly monomeric properties and about 30% potency [4]. In spite of its monomeric properties, crystals of a dimeric species of this mutant of insulin have been grown. The existence of the dimer and the close approach of the carboxylic acids is interesting. In the crystal it can be seen that the B13 glutamates are separated by 3.2 Å, suggesting that they are protonated and H-bonded. In the native insulin dimer crystal structures, the B13 residue separation is 6.1 Å at neutral pH and is at H-bonding distance at acid pH. Studies in solution by n.m.r. have demonstrated that, in the B9 Asp mutant insulin, the pK of a carboxylic acid, probably B13, is elevated at very high concentrations when the dimeric species can be formed [10]. This phenomenon is accounted for by the formation of the hydrogen bond seen in the crystal structure.

The crystal structure of the B10 Asp mutant insulin has also been solved recently (J. Turkenburg and Xiao Bing, unpublished work). This molecule, as with the B9 Asp, B27 Glu, A8 His mutant, forms a dimer in the crystal. Analysis shows that the B10 and B13 carboxylate groups are directed away from each other. There is a H-bond between the B13 Glu and the B10 His that will reduce the charge repulsion. Interestingly, this interaction is not seen in the native insulin crystals, although it appears energetically possible.

**Hexamer-forming surfaces**

The insulin hexamer is formed when three dimers are assembled together around central zinc ions that are co-ordinated through the B10, and sometimes through the B5 histidines [5, 6]. The zinc ions are essential for hexamer formation by native insulin. Three distinct hexamers have been crystallized and their structures solved [5, 6, 11, 12]. Firstly, there is 2Zn insulin. The co-ordination of the two Zn axial ions here is octahedral, and the overall hexamer point-group symmetry is 32. All the monomers in the three dimers of the hexamer have essentially the same structure, with their B chain N-terminal residues extended. This conformation in the monomer is referred to as T, and in the 2Zn insulin hexamer as T6 [13]. In this hexamer, these N-terminal segments cross over and are buried in adjacent dimers.

The second hexamer is the so-called 4Zn insulin, which is grown from high chloride-ion concentrations [11, 12]. Here, the monomer structures in the dimers are distinctly different at the B chain N-terminus. In one molecule the B1–B8 residues are extended, roughly as in the T structure; in the other molecule of the dimer they are helical. The molecule with the helical structure at B1–B8 is referred to as being the R state [13]. The helix at B1–B8 forms a continuation of the central helix at B9–B20 that is a feature of all insulin structures. Associated with this large rearrangement of B1–B8 is a marked separation of the B chain C-terminus and the A chain [5, 11]. This hexamer is described as T3R3 and it has point-group symmetry 3. It has one axial site of octahedral zinc co-ordination, one axial tetrahedral site and one off-axial tetrahedral site. The tetrahedral sites are both associated with the R molecule.

The third hexamer, which is grown in the presence of phenol, has a helical structure at B1–B8 (that is, B1–B20) in both monomers in the dimers,
giving it an R6 organization. This equivalence in the monomers restores the hexamer to 32-point symmetry [14]. It contains two axial zinc sites, both of which are tetrahedral.

Clearly the contacts between the dimers in these three hexamer types are very different. The T6 hexamer dimer–dimer contacts are relatively loose compared with the monomer–monomer contacts. There is a buried non-polar, entirely aliphatic core, while on the external surface the B1 Phe and A14 Tyr pack together, apparently closing the envelope of contact [6]. In the T3R3 hexamer, many of these contacts are broken. The residues A13 and B14, and B17 and B18 are generally unaffected. However, the loss of B1–B8 particularly affects B1 and B2, which make new and weak interactions, and B6 and A14, which now make no dimer–dimer contacts. There are new compensating hexamer contacts that are generated by the helix at B1–B8, notably those between three axially related B6 leucines. Incidentally, this packing of the leucines in the R molecule buries the zinc site, and significantly increases the stability of the hexamer. In the R6 hexamer, the contacts along the envelope between the dimers are further reduced; there are only contacts between A13, B14, B17 and B18. The contacts between the helices that are made by B6 along the axis are, however, strengthened, and both zinc sites are now buried.

An additional factor that stabilizes the R6 hexamer is the contacts that are made by the six phenol molecules that are specifically bound between the dimers. Examination of the environment of the phenol shows it makes two H-bonds to main-chain groups, it has displaced water molecules and that there are some close non-polar contacts between the dimers. All of these phenomena will favour the hexamer [14].

Figure 2
Structure of the B10 Asp insulin dodecamer

The Ca are shown with one set of three dimers as thick lines and the other as thin lines. Sidechains for the co-ordinating B5 histidines and for the mutated B10 Asp are shown. The two zinc ions on the central three-fold axis are represented by overlapping triangles.
Mutations at the hexamer-forming surface

B10 His → Asp

Zinc ions are essential for hexamer formation by native insulin, and, therefore, one obvious way to alter hexamer assembly is to modify the zinc-binding residues [15]. The zincs are co-ordinated by the imidazole groups of the B10 His, although sometimes there is also co-ordination by the B5 His in the T3R3 hexamer. Not surprisingly, the mutation B10 His → Asp abolishes the capacity of the molecule to hexamerize and produces a molecule that has a reduced ability to dimerize [4]. In the presence of zinc, however, the B10 Asp mutant of insulin forms a dodecamer, in which two sets of three dimers are co-ordinated to two zinc ions on the central three-fold axis (Figure 2). The zinc co-ordination is through the B5 His imidazole ring, and is reminiscent of the tetrahedral co-ordination that is seen in the R6 insulin hexamer. The monomers in the structure have the T conformation, and the dimer–dimer contacts include many of the same residues that are used in the T6 insulin hexamer. The contacts are, however, very different. In particular, a critical salt bridge is made by the mutated B10 Asp to the B chain α-NH₂ group. The B13 Glu sidechains are not packed together, in contrast with the native insulin hexamers, but H-bond to the A9 Ser in one dimer interaction, and to the A16 Tyr in the other. Another essential interaction is between the two B4 Gln residues, which form three pairs of complementary H-bonds about the three-fold axis.

B9 Ser → Asp

An alternative approach to perturbing hexamer formation is to exploit the concentration of charge at the hexamer centre that is generated by the six central and closely packed B13 Glu residues. The residue B9, which is adjacent to the B13 Glu, when mutated from serine to aspartic acid, greatly reduces dimerization and generally abolishes hexamer formation, even in the presence of zinc ions [4]. This is expected since such a mutant hexamer would pack 12 negatively charged groups together at the centre. Hexamers of this insulin mutated to aspartic acid at B9 have, however, been crystallized as hexamers in the presence of zinc, calcium and phenol. It turns out that the six B9 Asp–B13 Glu pairs bind calcium and hence balance the charges and remove the repulsive forces. The phenols, by increasing the extra contacts between the dimers and by burying the zinc ions, further stabilize this hexamer [15].

The observations above illustrate the limitations in predicting the consequences of protein engineering, and remind us of how complex protein systems are and how much there is still to learn about their recognition properties. However, it is encouraging to see how successfully the assembly of the insulin molecule can be manipulated, and to see that there are many possibilities for further experimentation that involves both protein and chemical thinking.

We are most grateful to Drs. J. Brange, N. Kaarsholm, J. Markussen and their colleagues at the Novo-Nordisk research laboratories for providing the mutant insulins and for informed discussions. We are also grateful to Drs. U. and Z. Derewenda for lively discussions and also for crystallographic contributions. This research has been supported by the Juvenile Diabetes Foundation, Novo-Nordisk a/s, the Medical Research Council and the Science and Engineering Research Council.

Molecular recognition in biological systems: From activation to inhibition
Anthony A. Kossiakoff, Thomas Hynes and Abraham de Vos
Department of Protein Engineering, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, U.S.A.

A fine-tuned system of molecular recognition is a fundamental feature of all biological processes. Although the basic principles of how molecular surfaces fit together in a complementary fashion have been established to a first approximation, the details of the interplay between electrostatic, steric, entropic and solvation effects are not well understood. To provide insight into the stereochemical factors that govern recognition and packing at protein interfaces, a host of biochemical and biophysical techniques have been applied. Perhaps the most widely used of these techniques are mutagenesis and crystallography, and when used together they have proved to be extremely powerful in defining the functional elements that are important in protein interactions.

Using these protein-engineering techniques we have focused our studies on the role of molecular recognition in two important, but distinctly different, biological processes: (1) the role of molecular recognition in receptor activation, and (2) the stereochemical factors that define specificity and provide binding energy to inhibitors that are associated with their target proteinases. Our findings, discussed below, suggest that the conventional wisdom about specificity and binding needs to be rethought.

Role of molecular recognition in the activation of the human-growth-hormone receptor

Human growth hormone (hGH) is a 22 kDa protein that stimulates the growth of bone, muscle and cartilage, by binding to a specific cell-surface receptor [1]. Two specific receptors that bind to hGH have been identified, the activation of which results in distinct pharmacologic effects: the growth hormone (GH) receptor and the prolactin (PRL) receptor. These receptors are classified, based on sequence similarities between their extracellular domains (ECDs), as belonging to the haematopoietic-receptor superfamily. This extensive receptor family includes the endocrine receptors, such as the GH and PRL receptors, and a number of other cytokine receptors [2, 3]. These receptors have a three-domain organization: the extracellular portion that binds the activating ligand; a transmembrane segment, which is usually hydrophobic in nature and 20–30 residues long; and a cytoplasmic domain that is involved in producing the response, the so-called 'second message', within the cell.

How does hormone binding trigger the activation of these receptors? Several mechanisms for the activation process had been proposed [4]. The most plausible of the mechanisms is based on some type of receptor aggregation, which assumed the formation of complexes of hormone–receptor pairs; that is, aggregates of some multiple of hormone–receptor dimers. However, the biochemical data assigning stoichiometry were equivocal and there was no structural information about the nature of the aggregation.

The growth hormone–receptor structural studies were undertaken using hGH and the ECD of its receptor. Crystals that were grown of the complex diffracted to 2.7 Å [5]. During the initial stages of the structure determination, the 5 Å electron-density map clearly showed that the complex contains three molecular components, not the two or four that are expected if the complex was a multiple of one hormone:one receptor ECD. It was determined that the stoichiometry found in the crystal, and subsequently verified in solution [6], is two ECD molecules to 1 hGH. This finding presented a conundrum; how do two receptor ECDs bind to one hGH molecule, considering that hGH has no semblance of symmetry (an asymmetric four-helix bundle protein; Figure 1) that could support the