Insulin and its receptor: structure, function and evolution

Pierre De Meyts

Summary

I present here a personal perspective on more than three decades of research into the structural biology of the insulin–receptor interaction. The solution of the three-dimensional structure of insulin in 1969 provided a detailed understanding of the insulin surfaces involved in self-assembly. In subsequent years, hundreds of insulin analogues were prepared by insulin chemists and molecular biologists, with the goal of relating the structure to the biological function of the molecule. The design of methods for direct receptor-binding studies in the 1970s, and the cloning of the receptor in the mid 1980s, provided the required tools for detailed structure–function studies. In the absence of a full three-dimensional structure of the insulin–receptor complex, I attempt to assemble the existing pieces of the puzzle generated by our and other laboratories, in order to generate a plausible mechanistic model of the insulin–receptor interaction that explains its kinetics and negative cooperativity. *BioEssays* 26:1351–1362, 2004.

The biography of insulin is being constantly amended and rewritten. Some parts of the story will persist longer than others. The more ephemeral portions serve an excellent purpose — they lead to work that results in a more permanent record.

Rachmiel Levine NEJM 277:1059–1064, 1967

Introduction and a bit of history

The successful isolation of insulin by Frederick Banting and Charles Best in John Macleod's laboratory at the University of Toronto in the summer of 1921 (beautifully reviewed in Michael Bliss's thorough account[1]) was nothing short of a miracle for people with diabetes. Until then, the only therapy that would somewhat prolong the life of the usually young patients was Frederick Allen's starvation diet, which gave parents the cruel choice between watching their child die slowly of cachexia or rather quickly of diabetic ketoacidosis. Within months of the discovery, the first patients were literally resurrected and, by 1923, insulin was readily available through Connaught Laboratories in Canada, Eli Lilly and Co in the USA and the Nordisk Insulin Laboratorium (founded by Hans Christian Hagedorn and Nobel laureate August Krogh in Denmark). The discovery won the Nobel prize in 1923 for Banting and Macleod, a choice that has remained controversial up to this day, although Banting shared his part of the prize with Best, and Macleod with Charles Collip, a chemist who played a critical role in the purification method.

Besides providing a life-saving treatment for insulin-dependent diabetics, insulin proved to be a bonanza for protein chemists and, on those accounts, may arguably deserve the title of "protein of the 20th century". Indeed many findings on the structure and properties of proteins were first made with insulin as a model. It was the first protein to have its primary amino acid structure sequenced in 1955, winning Fred Sanger his first Nobel prize in 1958.[2] It was also the first protein to be made by total chemical synthesis in the early sixties by Helmut Zahn's group in Germany,[3] Panayotis Katsoyannis's in the USA[4] and Yu Can Du and Zhang You Shang's in China.[5] John Abel in Baltimore crystallised it in 1929, but at the time did not realise the importance of zinc in stabilising the hexameric crystal and had trouble reproducing his finding. The thorough work of Jørgen Schlichtkrull at the Novo Terapeutisk Danish pharmaceutical company (the rival of the abovementioned Nordisk Insulin Laboratorium for over 60 years before they merged in 1989) established the conditions for reliable crystallisation and defined that two crystal forms existed with either two or four zinc atoms per six insulin molecules. Motivated by having a daughter with unstable diabetes, Schlichtkrull produced ultra-pure ("monocomponent") pig insulin, marketed in 1973. These findings were instrumental in enabling Dorothy Crowfoot Hodgkin and her colleagues in Oxford, after years of struggle, to solve the three-dimensional structure of insulin by X-ray crystallography in 1969 (having won the Nobel prize in 1964 for solving the structure of vitamin B12 and penicillin), a major achievement which is the stepping stone for much of the work discussed in this review.[6]

Insulin was also the first peptide hormone whose minute circulating amounts could be measured by radioimmunoassay by Solomon Berson and Rosalyn Yalow in 1960, winning Yalow the Nobel prize in 1977,[7] unfortunately after Sol's premature death. The biosynthetic pathway of insulin in the beta cells of the pancreas as a proinsulin precursor was elucidated by Don Steiner in 1967.[8] Insulin was the first

Funding agency: Danish Medical Research Council (Center for Growth and Regeneration); Øresund IT Academy; Medicon Valley Academy; Danish Ministry of Science, Technology and Development.
Receptor Biology Laboratory, Hagedorn Research Institute, Niels Steensens Vej 6, DK-2820 Gentofte, Denmark.
E-mail: pdm@novonordisk.com
DOI 10.1002/bies.20151
Published online in Wiley InterScience (www.interscience.wiley.com).

protein to be biosynthesised in microorganisms by recombinant DNA technology, paving the way for the design of insulin analogues attempting to optimise the pharmacodynamic profile of the molecule for diabetes therapy.[9]

My own wandering into the insulin field was rather serendipitous. I finished medical school at the University of Liège in Belgium the very summer of 1969 when Dorothy Hodgkin solved the insulin structure. I was accepted for specialisation in internal medicine (with emphasis on endocrinology and diabetes, under the supervision of Pierre Lefèbvre) and started research in the laboratory of Paul Franchimont, who pioneered radioimmunoassay in Europe. I was attempting to establish a radioimmunoassay for ACTH.[10]

The publication of the first radioreceptor assay for a peptide hormone (in fact ACTH) in 1970 in Science by Bob Lefkowitz, Jesse Roth and Ira Pastan at the NIH[11] radically changed my perspective: I decided that I was more interested in learning how hormones activate their receptors than in measuring their circulating levels, especially since I could continue using my newly acquired experience in radioligand-binding assays. Jesse Roth also came to receptors from the radioimmunoassay field, having trained with Yalow and Berson, and established with Seymour Glick the first growth hormone radioimmunoassay. Luckily, our department was very involved in organising the second International Symposium on Protein and Polypeptide Hormones, held in Liège in 1971, attended by the Who's Who in the field including Jesse Roth, Martin Rodbell, David Rodbard, Roger Ekins, Sol Berson and Rosalyn Yalow. It was a thrilling experience to mingle with those giants. Jesse presented the first insulin-binding studies using IM-9 human cultured lymphocytes (to this day still my golden standard for structure–function relationships of insulin analogues). Paul Franchimont introduced me to Jesse, who accepted me as a post-doc to work on ACTH receptors. I obtained a Fogarty NIH fellowship, and landed in Bethesda in January 1973. It turned out that the laboratory was now focused on insulin and growth hormone; so Jesse suggested I work on insulin receptors instead. So started my more than 30-year love affair with my favorite molecules: insulin and its receptor.

The structure of insulin

The circulating (and biologically active) form of insulin is a monomer (Fig. 1A) consisting of two chains, an A chain of 21 amino acids and a B chain of 30 amino acids (in man), linked by two disulfide bridges, A7–B7 and A20–B19. The A chain contains an intra-chain disulfide bridge between A7 and A11. At micromolar concentrations, insulin dimerises and, in the presence of zinc ions, further associates into hexamers, a stunningly beautiful symmetrical structure (Fig. 2). In the 2-Zn hexamer solved by Dorothy Hodgkin and her colleagues,[12,13] the A chain has an N-terminal helix, A1–A8, linked to an antiparallel C-terminal helix, A12–A20. The B chain has a central helix B8–B19, extended by N- and C-terminal strands. This crystallographic conformation is referred to as the T conformation (Fig. 1A). In the 2-Zn crystal form, all six monomers are in the T conformation (T6). An alternative conformation exists in which the B-chain helix extends all the way to the N-terminal (B1–B19, Fig. 1B), referred to as the R conformation. In the 4-Zn hexamer generated by high chloride concentrations (discovered by Schlichtkrull), three of the monomers are in the R form and three in the T form (R3T3),

Figure 1. The insulin peptide family protein folds. **A:** Pig insulin: T form. The A-chain is red, the B-chain yellow. PDB accession number: 9INS. **B:** Human insulin: R form. The A-chain is red, the B-chain yellow. PDB accession number: 1ZNJ. **C:** Human IGF-I. The A domain is red, the B domain yellow, the C domain blue and the D domain green. PDB accession number: 1GZR. **D:** Human relaxin 2. The A domain is red, the B domain yellow. PDB accession number: 6RLX. **E:** *Bombyx Muri* Bombyxin-II. The A domain is red, the B domain yellow. PDB accession .number: 1BOM. Graphics program: DSViewerPro from Accelrys

Figure 2. The insulin hexamer. On the right, the human 2-Zn insulin R6 hexamer. NMR coordinates from Chang et al.[15] PDB accession raphicsGnumber 1AIY. One monomer of each dimer is colored in dark blue, the second monomer in light blue. Zinc in yellow is in the middle. heTprogram: DSViewerPro from Accelrys. On the left, the rose window of the Sanctuary of Notre-Dame de la Prie`re in Pontmain, France. ructure.tKeats citation was mentioned in the legend to the insulin hexamer in Tom Blundell's landmark 1972 comprehensive review of insulin s ((12

Figure 3. Dimer- and hexamer-forming surfaces of insulin. The dimer-forming surface is shown in yellow. It is made or amino acids B8 Gly, B9 Ser, B12 Val, B13 Glu, B16 Tyr, B23 Gly, B24 Phe, B25 Phe, B26 Tyr and B27 Thr. The hexamer-forming surface is shown in red. It is made of amino acids B1 Phe, B2 Val, B4 Gln, B13 Glu, B14 Ala, B17 Leu, B18 Val, B19 Cys, B20 Gly, A13 lau, A14 Tyr and A17 Glu. Backbone is .shown in blue. PDB accession number: 9INS. Graphics program: DSViewerPro from Accelrys

a structure solved in 1976 by Graham Bentley, Dan Mercola and Guy Dodson in Dorothy Hodgkin's group.[14] In phenol-containing crystals, solved by Dodson's group in 1989, all six monomers are in the R form (R6).[14] The solution structure of the R6 hexamer has also been solved by NMR[15] (Fig. 2). The T–R transition, a true allosteric equilibrium, has been extensively studied in solution by Axel Wollmer,[16] Niels Kaarsholm, Michael Dunn and their colleagues.[17] It plays an important role in the pharmaceutical formulations of insulin where phenol is used as an antimicrobial agent and chloride as an isotonic agent.

The basic insulin fold described above (three helices, three conserved disulfide bridges) is present in all members of the insulin peptide family (Fig. 1) throughout evolution (see below) despite sometimes completely divergent sequences as in *C. elegans*. Around its hydrophobic core, the insulin monomer has two extensive nonpolar surfaces (Fig. 3).[12,13] One is flat and mainly aromatic, and buried upon dimer formation in an antiparallel beta sheet structure. The other is more extensive and is buried when the dimers assemble to form hexamers. As discussed below, it is now clear that, not surprisingly for a small globular protein, insulin uses the same surfaces that it uses for self-assembly, for binding to its receptor.

The insulin receptor: early studies

The existence of a membrane receptor mediating the cellular effects of insulin had been postulated among others by Rachmiel Levine, a pioneer of diabetes research who had shown in 1949 that insulin stimulates glucose entry into cells. The use of radiolabelled insulin to study cell membrane binding (e.g. to diaphragm muscle) had been tried, but the binding observed was mostly non-specific. The Chloramin T method of Hunter and Greenwood for radioiodination of peptide hormones,[18] designed to obtain maximal specific activity, was a stepping stone in the prodigious expansion of radio-immunoassay. Jesse Roth (and independently Lin and Good-friend in Madison working on angiotensin) realised that the stoichiometry used in this method was deleterious to the biological activity of the fragile peptides, and designed a gentler protocol that resulted in monoiodinated ligands, as applied to ACTH in Lefkowitz's 1970 paper,[11] In the case of insulin, the molecule is labelled at either of the four tyrosyle residues A14, A19, B16 and B26, and later refinements included purification of the labelled A-14 insulin isomer (which binds best) by HPLC. House and Weidemann in Canberra were the first to publish binding of a low-specific-activity insulin (although the labelling method was not described) to liver plasma membrane fractions in 1970,[19] P.D.R House wrote a thesis on insulin receptors that contained extensive biochemical studies, but unfortunately died of diabetes complications before much was published. Pierre Freychet and Jesse Roth,[20] and Pedro Cuatrecasas' labs[21] published their papers on binding of monoiodinated insulin to rat liver plasma membranes in 1971, starting a sometimes fierce competition between the two laboratories that lasted several years. Steen Gammeltoft and Jørgen Gliemann independently developed insulin-binding assays to isolated fat cells in Copenhagen.[22]

This is how I came to think that the curved Scatchard plot therafor insulin binding may be due to negative cooperativity r than binding sites heterogeneity. After playing with the tbinding data in a number of ways, I convinced Jesse that i was worth investigating. The problem is that equilibrium binding data do not allow to distinguish between the two hhypotheses. During one of those brainstorming sessions wit Jesse, we came up with the concept of comparing the ''edissociation rate of prebound ^{125}I-insulin in an ''infinit edtdilution in buffer alone (to prevent rebinding of dissocia ofntracer) or in buffer con-taining a saturating concentratio cold insulin. This retro-spectively simple comparison had edcnever been done. It turned out that unlabeled insulin indu a marked acceleration of the dissociation of ^{125}I-insulin, eceptorrinexplicable on the basis of The hot topic of the day was lpingedownregulation, and I learned the tricks of the trade by h Jim Gavin finish some experiments for his study of insulin receptor downregulation in IM9 lymphocytes for which I got a generous coauthorship on what became , a ''citation classic''.[23] Andy Soll and Ron Kahn had shown that hyper-insulinemic ob/ob mice had a marked decrease in their liver membrane receptor concentration. Ron had also analysed in detail the quantitative aspects of insulin binding to liver plasma membranes, and demonstrated curvilinear Scatchard plots (a plot of Bound/Free versus Bound ligand, which is a straight line in case of a simple binding reaction).[24] I was familiar with Scatchard plots, which had been used by Berson and Yalow in radioimmunoassay, and curvilinear ones were usually inter-preted as indicating the coexistence of populations of anti-bodies or receptors with distinct .affinities

On Jesse's advice, I took a tutorial course in protein chemistry run by Alan Schechter, Bob Adelstein and Irvin Chaiken, and was given the task to explain to the class the mechanism of cooperativity in haemoglobin, including the classic models of Monod, Wyman and Changeux and Koshland, Nemethy and Filmer. The MWC (''allosteric'' or concerted) model can only explain positive cooperativity, while the KNF sequential model can generate both positive and negative cooperativity. I had been fascinated by the allosteric concept while studying pharmacology in medical school.

This is how I came to think that the curved Scatchard plot for insulin binding may be due to negative cooperativity rather than binding sites heterogeneity. After playing with the binding data in a number of ways, I convinced Jesse that it was worth investigating. The problem is that equilibrium binding data do not allow to distinguish between the two hypotheses. During one of those brainstorming sessions with Jesse, we came up with the concept of comparing the dissociation rate of prebound ^{125}I-insulin in an ''infinite'' dilution in buffer alone (to prevent rebinding of dissociated tracer) or in buffer con-taining a saturating concentration of cold insulin. This retro-spectively simple comparison had never been done. It turned out that unlabeled insulin induced a marked acceleration of the dissociation of ^{125}I-insulin, inexplicable on the basis of

site heterogeneity. The concept of negative cooperativity in the insulin receptor was born. It became clear that both the concentration and affinity of insulin receptors were subject to regulation. The cooperative effect was seen at physiologic insulin concentrations and mirrored receptor occupancy, but was reverted at supraphysiological concentrations above 100 nM ("bell-shaped curve"). Our initial paper in BBRC[25] generated considerable interest (it was quoted over 600 times), as well as lasting controversy.[26–28] I did not realise then that it would define the rest of my research career. Today, with the progress made over the last two decades in understanding the structure of the insulin receptor, but still short of a structure of the insulin–receptor complex, we and others have at last found plausible mechanistic models compatible with these findings, as explained below.

Towards a structure of the insulin receptor

The subunit structure of the insulin receptor (see Ref. 29 for review) was investigated in the early eighties among others by Cecil Yip, Steve Jacobs and Pedro Cuatrecasas, and Joan Massague and Mike Czech, who showed that it is made of a covalent dimer of two α subunits and two β subunits. Following the demonstration in 1982 by Ora Rosen's group[30] that a tyrosine kinase was closely associated with the insulin receptor, several groups showed that the insulin receptor itself is a tyrosine kinase, which catalyses the transfer of the γ phosphate of ATP to tyrosine residues on protein substrates, the first being the receptor itself.[31] It is a member of the RTK superfamily, which, in the human genome, comprises $\sim$60 members distributed in $\sim$20 subfamilies depending on the modular architecture of their extracellular domains and the degree of identity in their intracellular kinase domains.

Cloning of the insulin receptor cDNA was accomplished in 1985 by the groups of Axel Ullrich at Genentech[32] and Bill Rutter at UCSF,[33] soon followed by that of the related IGF-I receptor (IGF-IR) and by the sequencing of both genes, giving valuable insight into the receptor structure and organisation.

The gene for a mammalian related receptor (insulin-receptor-related receptor or IRR) was identified in 1989. No ligand for this receptor has been identified. The mouse IRR knockout has no phenotype, but the triple IR/IGF-IR/IRR knockout results in gonadal dysgenesis and male-to-female somatic sex reversal, providing the first glimpse of a function for this orphan receptor.[34]

The insulin and IGF-I receptors differ from the other RTKs in that they are covalent dimers in the absence of ligand, although all the RTKs dimerise or oligomerise upon ligand binding, resulting in activation of the kinase by transphosphorylation.

We and others have recently reviewed extensively the structural biology of insulin and IGF-I receptors,[29,35] and I will only summarise here the most salient features of the insulin receptor.

The insulin receptor has a modular structure encoded by a gene with 22 exons and 21 introns (Fig. 4). The short exon 11 is alternatively spliced, resulting in two isoforms (A and B) that differ slightly in affinity for insulin. The B isoform binds the IGFs with at least 100 times lower affinity than insulin, while the A

Figure 4. Modular structure of the insulin receptor. Cartoon of the $\alpha 2\beta 2$ structure of the insulin receptor, drawn to scale. On the left half of the receptor, spans of the 22 exon-encoded sequences. On the right half, spans of predicted protein modules. Orange arrowheads: N-glycosylation sites. Green arrowheads: ligand-binding "hotspots" identified by single amino acids site-directed mutagenesis. L1: large domain 1. CR: cystein-rich domain. L2: large domain 2. Fn: fibronectin type III domains. Ins: Insert. TM: transmembrane domain. JM: juxtamembrane domain. TK: tyrosine kinase domain. CT: C-terminal domain. Black bar: immunogenic domain. From De Meyts and Whittaker. Structural biology of insulin and IGF-I receptors: implications from drug design. Nat Rev Struct Biol 2002 1:769–783.

Figure 5. Natural mutations in the insulin and receptor insulindmolecules in patients with genetic forms of diabetes an tions inaresistance syndromes. Left. The location of natural mut n aswthe insulin molecule that alter receptor binding is sho ueyellow Van der Waals spheres. Mutations of B25 Phe to L and Val,(Insulin Chicago), B24 Phe to Ser (Insulin Los Angeles) A3 to Leu (Insulin Wakayama) markedly decrease insulin-ulinsbinding affinity. Mutation of B10 His to Asp impairs in ssedeprocessing and results in hyperproinsulinemia; the proc . ThetB10 Asp insulin has a 3-fold increase in affinity. Righ lin bindingulocations of insulin receptor mutations that impair ins yellowyin patients with extreme insulin resistance are shown b rminaleVan der Waals spheres. All are located in the L1CL2 N-t anmdomain, which has been modelled on the corresponding hu :IGF-I receptor domain coordinates (PDB accession number .I1IGR) by Jonathan Whittaker using the program SwissMode .Graphics program: DSViewerPro from Accelrys

Figure 6. Schematic illustration of the two-step bivalent 1Lreceptor cross-linking model. The two receptor α-subunit l(blue)-CR (green)-L2 (red) domains (modelled on the actua ndaIGF-I receptor coordinates using the program SwissModel)),IFn$_0$ domains (simulated module kindly given by Tom Blundel l aslwhich contain the insulin-binding sites, are shown, as we ssicalathe disulfide bridges. Insulin, in the middle, has the cl dbinding surface shown as yellow Van der Waals spheres, an the newly mapped binding surface shown in red. The two the arrows indicate the stepwise binding process and point to alsogtwo respective putative receptor-binding epitopes. Bindin edtrequires sequence 704–715 in the insert. Adapted and upda ulinsfrom De Meyts P and Whittaker J. Structural biology of in veand IGF-I receptors: implications from drug design. Nat R .Drug Disc 2002 1: 769–783

isoform has significantly higher affinity than the B isoform for IGF-I and especially IGF-II. The IGF-I receptor binds IGF-II with a lower affinity than IGF-I and insulin with a 500-fold lower affinity.

The receptors are synthesised as single chain preproreceptors that are processed, glycosylated, folded and dimerised to yield the mature $\alpha_2\beta_2$ receptor.

In cells expressing both insulin and IGF-I receptors, hybrid receptors are formed consisting of one half of each. Their physiological role is unknown.

Comparative sequence analysis of the insulin/IGF-I receptors and the related EGF receptor had led Mona Bajaj in Tom Blundell's group[36] to suggest that the N-terminal half consists of two large homologous globular domains, L1 and L2, separated by a cysteine-rich region later predicted to consist of a series of disulfide-linked modules similar to those found in the tumor-necrosis factor (TNF) receptor[37] and laminin. These predictions were confirmed by the determination of the crystal structure of this domain (see below). The C-terminal half of the receptors was predicted to consist of three fibronectin type III (FnIII) domains. The second FnIII domain contains a large insert domain of unknown structure containing the site of cleavage between α- and β-subunits. The intracellular portion of the β-subunit contains the kinase domain flanked by two regulatory regions, a juxtamembrane region involved in docking insulin receptor substrates (IRS) 1–4 and Shc as well as in receptor internalisation, and a C-terminal tail.

The structure of the L1/Cys-rich/L2 (L1CL2) domain fragment of the IGF-I receptor has been solved by Tom Garrett et al. in Colin Ward's laboratory in Melbourne.[35] A bilobed structure comprises the two globular L domains with a new type of right-handed β-helix fold flanking the Cys-rich domain. Fig. 5 shows the equivalent domains of the insulin receptor modelled on the IGF-I receptor structure. A cavity occupies the centre of the molecule and represents a potential binding pocket, although this construct does not bind the ligand. The position of L2 is probably more parallel to L1 in the native structure, as shown in the recently determined structures of the structurally related EGF receptor complex with EGF or TGFα (for review see Ref. 38).

The structures of the tyrosine kinase domain of the insulin and IGF-I receptors, both in the inactive and in the active states have been determined by Stevan Hubbard.[39]

Peter Ottensmeyer, Cecil Yip and co-workers have proposed a quaternary structure of the insulin–receptor complex based on elegant cryoelectron microscopic studies.[40] We have expressed reservations regarding the details of their proposed arrangement of the insulin–receptor interface.[29] Nevertheless, these studies make it likely that the insulin–receptor complex is a globular molecule with a diameter of 150 Å. In the extracellular component, the L1CL2 domains form the membrane distal core and the FnIII domains form the membrane proximal part.

Molecular evolution of the insulin/IGF system

This topic has been reviewed in more detail elsewhere.[41–43] In the human genome, the insulin-like peptide family comprises ten members, the closely related insulin and insulin-like growth factors (IGF)-I and II, and seven peptides related to relaxin (Fig.1D) (INSL/RLFs). While insulin and the IGFs bind to RTKs, the relaxin-like peptides unexpectedly bind to leucine-rich repeat-containing G-protein-coupled receptors (LGRs), involved in the development and physiology of the reproductive tract.

The sequence of the zebrafish and fugu genomes has revealed the existence of two insulin genes; evidence suggests that genome duplication occurred on the lineage leading to teleost fish.

Several insulin-like family members have been identified in invertebrates, including 38 bombyxins (Fig. 1E) in the silkworm *Bombyx mori* and seven insulin-like genes in the fruit fly *Drosophila melanogaster*. In addition, three families of insulin-like peptides (38 in all) with atypical disulphide bond pattern have been identified in the nematode *C. elegans* that may be ligands (both agonists and antagonists) for the *daf-2* insulin-like receptor.

The existence of a single insulin-like receptor has been reported in various invertebrates such as *D. melanogaster*, *B. Mori*, other insects, molluscs, *C. elegans* and cnidarians.

Interestingly, the protochordate amphioxus (*Branchiostoma californiensis*) contains a single gene for an insulin-like peptide and a related receptor, both of which have hybrid characterictics of the insulin and IGF systems and may therefore represent the ancestral vertebrate genes.

Genetic evidence suggests that the signalling pathways and biological effects mediated by these receptors in mammals, nematodes and fruit flies are highly conserved. Insulin and IGF-I receptor knockout studies in mice as well as mutational studies of their orthologues in *Drosophila* and *C. elegans* indicate that, in addition to intermediary metabolism and growth, these receptors are involved in regulating energy homeostasis, reproductive function and longevity.

Structure–function relationships of insulin binding

As soon as the structure of insulin was solved, a major focus of investigation was to determine which surface residues were involved in receptor binding and biological activity, questions pioneered initially by Tom Blundell,[12] Guy Dodson and their colleagues.[13] In vivo insulin assays are not appropriate to answer such questions due to the lack of correlation between in vitro binding and in vivo activity, as shown by Richard Jones and Peter Sönksen, and Ulla Ribel and colleagues at Novo

Nordisk. The availability of receptor-binding assays in the early 1970s provided the much-needed biochemical basis for answering such questions, which became also a major focus of my work. Pierre Freychet and Jesse Roth had used insulin analogues such as proinsulin and chemically modified insulins to evaluate binding selectivity,[20] followed by a similar approach by Gliemann and Gammeltoft in 1974.[44] In those days, available modified insulins were insulins purified from the pancreas of various animal species, chemically modified insulins, modified insulins made by total synthesis (Panayotis Katsoyannis in New York studied a countless number of these, and was the first to use the term insulin analog(ue) in 1973), or analogues made by enzymatic semi-synthesis (an approach pioneered by Ken Inouye in Japan and Hans Gattner, Dietrich Brandenburg and Helmut Zahn at the German Wool Research Institute in Aachen, Germany[45]). Other active groups were Rolf Geiger's at Hoechst and Franz Märki and Werner Rittel's at Ciba Geigy.

From those early studies, epitomised in the classical Nature paper by Pullen et al. in 1976,[46] the consensus was that a number of surface residues that have been widely conserved during vertebrate evolution are probably involved in receptor binding: Gly A1, Gln A5, Tyr A19, Asn 21, Val B12, Tyr B16, Gly B23, Phe B24, Phe B25, Tyr B26 (referred to here as the "classical binding surface").

We showed that a subset of residues from this domain (A21, B23–26) was essential for the negative cooperativity in binding, defining a "cooperative" site on the insulin surface. Our article made the cover of Nature in 1978[47] with the first published color space-filling model of a protein, thanks to my collaboration with computer wizards Richard Feldmann and Tom Porter at the NIH. Tom had just solved the so-called hidden surface algorithm,[48] allowing virtual space-filling models; he went on to Lucasfilm and Pixar, and became a pioneer in 3D computer animation, winning three Oscars!

The critical importance of the B23–26 region of insulin for receptor binding was confirmed and extensively investigated by Howard Tager, Satoe Nakagawa in Don Steiner's group in Chicago,[49] and by Mike Weiss (now in Cleveland).[50] Interestingly, two of the six known mutations in the insulin molecule in diabetic patients are in this region (Fig. 5), insulin Chicago (Phe B25 to Leu) and insulin Los Angeles (B24 Phe to Ser).[50,51]

In addition, Ile A2 and Val A3 (a residue also mutated in a diabetic patient: insulin Wakayama, Fig. 5),[50,51] which are not on the surface of the molecule, probably become exposed and interact with the receptor upon displacement of the B-chain C terminus during the receptor-binding process.[52,53] This requirement for rearrangement is compatible with the very low activity of a single chain B29-A1 peptide-linked insulin, despite close structural similarities with native insulin.[54]

The concept that the classical binding surface was the only one involved in receptor binding was subsequently challenged in studies by us and others of the properties of the phylogenetically ancient insulin from the Atlantic hagfish (*Myxine glutinosa*).[55] Despite absolute conservation of the classical binding surface, and a crystallographic structure nearly identical to that of pig insulin, hagfish insulin displays different binding behaviour from most mammalian insulins: low affinity, low metabolic potency, slow association kinetics, decreased negative cooperativity with a sigmoid (not bell-shaped) dose-response curve. This suggested that certain residues outside the classical binding surface must contact the receptor and their substitution in hagfish insulin explains the abnormal behaviour.[13,56] The evolutionarily divergent hystricomorph insulins (guinea-pig and cousins) exhibit similar binding behaviour.[56] The aberrant behaviour of hagfish and hystricomorph insulins remained unexplained for nearly three decades.

In the early 1980s, human insulin prepared by recombinant DNA technology became available (see Hall, 1987[57] for a thorough account of this important milestone). We showed, in collaboration with Bruce Frank and Ron Chance at Eli Lilly, that its receptor-binding properties were identical to those of native human insulin.[58] The DNA recombinant technology opened the door to further exploration of the structure–activity of the molecule by site-directed mutagenesis. In the eighties, Jens Brange and colleagues at Novo embarked on an extensive program of genetically engineered insulin analogues in order to design variants with impaired aggregation properties that would have a faster absorption from the skin in diabetic patients.[59] With this new generation of insulin analogues kindly made available by Jens, I went back to studying the impact of such mutations on receptor-binding affinity and kinetics.

An important finding was that insulin analogues with mutations at two residues in the hexamer-forming surface, Leu A13 and Leu B17, bound to the receptor with characteristics similar to those of hagfish insulin, including a 20-fold lower association rate than native human insulin.[56] Lauge Schäffer at Novo Nordisk observed the slow kinetics independently using a purified receptor preparation.[60] These data supported the hypothesis that residues outside the classical binding surface, e.g. A13 and B17, are involved in receptor binding.

These findings prompted a systematic reexamination of the structure–function relationship of insulin by alanine scanning mutagenesis (an approach pioneered for the study of the growth hormone–receptor interaction by Jim Wells and colleagues at Genentech[61]). Claus Kristensen tested 21 new insulin analogues with single alanine substitutions for low-affinity binding to the soluble insulin receptor ectodomain fused to IgG[62] and showed that mutation of residues A2, A19, B23 and B24 from the classical binding surface and of B13 were most disruptive of binding. Mutations of B6 and B8 were also disruptive, probably due to conformational alterations.

Since high-affinity binding is more amenable to precise determination in the context of the cell-bound receptor, we tested Claus' analogues using IM-9 human lymphocytes. We confirmed the importance of residues from the classical binding surface. In addition, we found that a cluster of residues outside the classical surface disrupted binding:[63] Ser A12, Leu A13, Glu A17, His B10, Glu B13, Leu B17 (Fig. 6). Interestingly, all of these are involved in the formation of the hexamer. Moreover, we showed by introducing substitutions from hystricomorph[64] or hagfish[65] insulins into human insulin, either alone or in combination, that the aberrant behaviour of these species is largely explained by a small number of deleterious substitutions in this novel binding surface, thus solving a 30-year-old riddle. I believe that we now have a more or less complete map of the bioactive surface of the insulin molecule (Fig. 6).

Mapping of insulin-binding domains on the insulin receptor

A variety of approaches have been used to map putative insulin contact sites on the insulin receptor,[29] including covalent crosslinking of the ligand (most recently by Mike Weiss and colleagues[66]), chimeric insulin and IGF-I receptors, minimised receptor constructs, as well as alanine scanning mutagenesis by Jonathan Whittaker and colleagues.[67,68] We have reviewed these data in detail recently[29] and I will only briefly summarise them here. Photo-affinity labelling data indicate that the classical binding surface of insulin binds to a receptor site that is composed of the L1 domain, and an α-subunit C-terminal peptide sequence (amino acids 704–715), the functional epitopes of which have been identified by alanine scanning. Several lines of evidence suggest that a second binding site exists within the C-terminal part of L2 and the first FnIII domain (Fn_0); this site has not yet been mapped.

The location of mutations that impair insulin binding found in patients with extreme resistance syndromes also provides clues to the mapping of the ligand contact sites (Fig. 5).

Mechanistic models of the binding interaction

Using purified recombinant insulin receptors, Lauge Schäffer and Jan Markussen showed that only one insulin molecule binds to the dimeric receptor with high affinity.[60] However, a soluble extracellular ectodomain binds two insulin molecules, but with low affinity and lack of negative cooperativity. Likewise, reduced half receptors bind insulin with low affinity. These data suggested that each half receptor contains components of the high-affinity binding site, but that cooperation between the two half-receptors is required to create high-affinity binding,[56,60] maybe as observed for the bivalent growth hormone–receptor-binding mechanism.[61]

25 years ago, Martin Raff was the first to speculate that the complex insulin-binding kinetics may be the result of insulin having more than one binding site for the receptor and its consequent ability to crosslink the receptor.[56] Later, Paul Pilch and Cecil Yip's laboratories also proposed that insulin may contact both halves of the receptor. Lauge Schäffer proposed in 1994 an elegant bivalent crosslinking model whereby two different receptor sites on each half-receptor contact an insulin molecule.[60] I proposed that all parameters of the negative cooperativity (high- and low-affinity sites, accelerated dissociation and bell-shaped dose–response curve) could be best explained if the two receptor-binding sites were disposed in an antiparallel symmetry.[56] A second insulin molecule bridging both leftover sites 1 and 2 upon partial dissociation of the first crosslink would prevent its rebinding and accelerate dissociation of the first bound insulin. The bell-shaped curve observed for negative cooperativity is the hallmark of an effect that requires crosslinking (as demonstrated by the dose–response curve for biological effects of GH[69,70]).

These bivalent crosslinking models not only explain binding kinetics but provide a molecular basis for activation of the receptor tyrosine kinase and signalling pathways, by approximating the two kinase domains and permitting transphosphorylation. Since mutations at the C terminus of the B-chain create antagonists for negative cooperativity,[56] I postulate that the classical binding surface binds last, and that the novel binding surface containing Leu A13 and Leu B17 binds first (Fig. 6), in agreement with the fact that mutations of these two residues slow down the initial association 20-fold.

In summary, the studies described above support the concept that insulin binds asymmetrically to two discrete sites in the receptor dimer, crosslinking the constituent monomers. One of these sites has been characterised in detail and is partly located in the L1 domain and also contains a peptide from the C terminus of the α-subunit. The second site is less well studied but is located in the L2-$FnIII_0$ region (Fig. 6).[29]

Biological significance of negative cooperativity

An interesting question has always been whether the negative cooperativity observed in the insulin receptor is only an epiphenomenon of the binding mechanism, or whether the acceleration of the dissociation rate of the insulin–receptor complex by increased insulin concentration has a regulatory role on insulin actions. The observation by us and others that the mitogenic effects of insulin are disproportionally enhanced for analogues that have slower dissociation rates than native human insulin suggests that the negative cooperativity may play an important role in limiting the mitogenic effects of insulin.[71] Theoretical studies with Ron Shymko (my accomplice for nearly 20 years in computer modelling of receptor binding) based on logical (Boolean) analysis of signalling pathways have demonstrated the plausibility of the model whereby regulation of residence time on the insulin receptor may indeed determine choices in branching signalling pathways.[72] We are currently addressing this issue by attempting to measure real-time signalling for analogues with different binding kinetics.

Future directions

With the huge progress in genomics (including gene invalidation approaches) and proteomics over the last decade, much of biological science has turned from a hypothesis-driven endeavour into a giant fishing expedition (''data mining''). While the amount of relevant information regarding living systems generated by such approaches is invaluable, information is not knowledge. Ron Kohanski (who has made important contributions to our understanding of the insulin receptor kinase enzyme kinetics[73,74]) and I recently reflected that we are in danger of being regarded as dinosaurs for persisiting with our biochemical approaches to the insulin–receptor interaction. I do hope that our kind of dinosaurs will not become extinct. I believe there is more need than ever for good biochemistry and biophysics, structural biology, kinetics and mathematical modelling (topics rarely found in diabetes, endocrinology or signal transduction meetings nowadays) to make sense of the expanding networks of protein interactions. I hope that a new generation of young dinosaurs will train in those disciplines and take up the challenge of understanding structure–function relationships, as well as the specificity of signal transduction mechanisms.

As far as the insulin receptor system is concerned, I hope that the efforts at obtaining a crystallographic structure of the insulin–receptor complex currently under way, will succeed in a not too distant future.

My laboratory continues to pursue the site-directed mutagenesis approach in order to validate and refine our mapping of the insulin–receptor-binding interface, and extend such studies to IGF-I and II. With help from our Novo Nordisk colleagues, we are scaling up our production system from the minute amounts of analogues required for binding analysis to the production of amounts suitable for structure analysis.

Of particular interest to biologists is the exciting discovery of the expanding role of the insulin signalling system. Recent work on the genetics of the insulin/IGF system in invertebrates, rodents and humans suggest that it plays a major role not only in metabolism, growth and development, but also in reproduction, survival and lifespan[75] and central nervous system function including appetite regulation and memory. This is expected to be a source of innovative research in the years to come, with important outcomes in drug discovery.

We plan to address more closely the molecular evolution of the insulin system by extending the biochemical work that we have done on mammalian, hystricomorph and hagfish insulins to some of the ligands of *C. elegans*, *Drosophila* and other phylogenetically ancient systems, and their receptors.

Another aspect of the work of our laboratory, which is outside the scope of this essay, concerns the specificity of signal transduction by insulin and the IGFs. In order to understand better the combinatorial nature of these systems, we are attempting to combine real-time kinetics of early signal transduction steps using e.g. FRET, cellular localisation of signals by confocal microscopy, and microarray gene profiling as end points. A future goal is to combine these approaches with mathematical modelling and ''reverse engineering'', in order to build a systems biology of signal transduction, and ultimately of diabetes. While still very much a buzzword, systems biology may be the way of the future to reconciliate the current trendy but phenomenologic biological science with hypothesis-driven science.[76] Hopefully within the lifespan of the dinosaur who signs these pages.

Acknowledgments

This essay is dedicated to Jesse Roth on an important birthday. The Receptor Biology Laboratory and the Hagedorn Research Institute are independent basic research components of Novo Nordisk A/S. I want to acknowledge my long-standing friendships and collaborations with many of the scientists quoted in this review.

I am grateful to Ron Kohanski and Ron Shymko, as well as the anonymous referees, for critical review of the manuscript.