Protein Cell 2010, 1(6): 537–551
DOI 10.1007/s13238-010-0069-z

Protein & Cell

REVIEW

Insulin: a small protein with a long journey

Qingxin Hua ✉

Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106-4935, USA
✉ Correspondence: qing-xin.hua@case.edu

ABSTRACT

Insulin is a hormone that is essential for regulating energy storage and glucose metabolism in the body. Insulin in liver, muscle, and fat tissues stimulates the cell to take up glucose from blood and store it as glycogen in liver and muscle. Failure of insulin control causes diabetes mellitus (DM). Insulin is the unique medicine to treat some forms of DM. The population of diabetics has dramatically increased over the past two decades, due to high absorption of carbohydrates (or fats and proteins), lack of physical exercise, and development of new diagnostic techniques. At present, the two largest developing countries (India and China) and the largest developed country (United States) represent the top three countries in terms of diabetic population. Insulin is a small protein, but contains almost all structural features typical of proteins: α-helix, β-sheet, β-turn, high order assembly, allosteric T→R-transition, and conformational changes in amyloidal fibrillation. More than ten years' efforts on studying insulin disulfide intermediates by NMR have enabled us to decipher the whole picture of insulin folding coupled to disulfide pairing, especially at the initial stage that forms the nascent peptide. Two structural switches are also known to regulate insulin binding to receptors and progress has been made to identify the residues involved in binding. However, resolving the complex structure of insulin and its receptor remains a challenge in insulin research. Nevertheless, the accumulated knowledge of insulin structure has allowed us to specifically design a new ultra-stable and active single-chain insulin analog (SCI-57), and provides a novel way to design super-stable, fast-acting and cheaper insulin formulations for DM patients. Continuing this long journey of insulin study will benefit basic research in proteins and in pharmaceutical therapy.

KEYWORDS insulin, diabetes, receptor binding, folding, disulfide pairing, NMR

INTRODUCTION

A sparrow may be a small bird, but it is fully equipped to act as a bird. Insulin, as a protein, can be viewed in a similar way: it is a small but fully functional protein. Insulin displays both key properties of naturally existing proteins: a stable structure and a physiologic function. On the one hand, insulin is one of smallest proteins of all known native functional proteins and it possesses most of the typical protein structural features: α-helix, β-sheet, β-turn, high order assembly, allosteric T→R-transitions and conformational changes by self-association, and amyloidal fibrillation (Blundell et al., 1971; Peking, 1971; Chothia et al., 1983; Baker et al., 1988; Derewenda et al., 1989; Hua et al., 1991; Hua and Weiss, 1991; Dobson, 1999; Hua and Weiss, 2004; Huang et al., 2005, 2006). As a small protein, insulin is a good model for basic structural research, especially for studying protein folding intermediates with a variety of disulfide pairings. On the other hand, insulin plays a central role in the regulation of vertebrate metabolism, especially carbohydrate metabolism. The initial step in hormonal control is the specific binding of insulin to the insulin receptor (IR) (Baker et al., 1988; De Meyts, 1994; De Meyts and Whittaker, 2002). Binding of insulin to its receptor triggers a series activation of tyrosine kinase. In the short term, insulin increases sugar transport and protein synthesis, inhibits lypolysis, activates enzymes (such as glycogen synthetase), and inactivates other antagonist-like enzymes (such as phosphorylase). In the long term, insulin activates glucokinase and suppresses gluconeogenesis enzymes. Although the hormone functions in the bloodstream as a Zn^{2+}-free monomer (Dodson and Steiner, 1998), insulin assembles as a Zn^{2+}-coordinated hexamer in the human β-cell (Blundell et al., 1971). The hexamer is proposed to function as a protective mechanism preventing toxic fibrillation of insulin during *in vivo* storage. This sequestration of the susceptible monomer within the zinc hexamer (Brange et al., 1997a; Brange and Langkjaer, 1997) is exploited in clinical formulations (Brange, 1987).

The transition of a globular domain to a well-organized cross-β fibril occurs in insulin fibrillation (Brange et al., 1997b). Various mutations causing conformational changes impair receptor binding and consequently induce diabetes mellitus (DM). Diabetes occurs either because of a lack of insulin or due to insulin resistance. Recently, the number of diabetes patients has increased dramatically. The statistics data from the National Diabetes Fact Sheet showed that nearly 24 million Americans (~8% of the population) have diabetes (http://www.diabetes.org/diabetes-basics/diabetes-statistics/). An additional 57 million people are pre-diabetic and are not included in this total. Of these patients, 90% are type-II diabetics. The World Health Organization (WHO) stated that people diagnosed with diabetes have increased from 30 million to 246 million over the past two decades. The complications of diabetes include heart disease and stroke, high blood pressure, blindness, kidney disease, neuropathy and death. Diabetes is the seventh leading cause of death listed in the US.

The rapid development of human genetics enables scientists to identify more gene mutations related to diabetes, especially to permanent neonatal-onset DM. The positions of mutations in the proinsulin sequence are marked as black circles in Fig. 1A and 1B. (Støy et al., 2007; Colombo et al., 2008; Edghill et al., 2008; Molven et al., 2008). The positions of adult-onset mutations ("classical mutations") are labeled as gray circles in Fig. 1A.

The insulin gene encodes a single-chain precursor polypeptide (preproinsulin) that contains a 24-residue signal peptide at the N-terminal (not shown in Fig. 1) and a 35-residue connecting domain (black in Fig. 1B) between the C-terminal residue of the B domain (B1) and the N-terminus of the A domain (A1) (Steiner, 1967). The C-peptide is cleaved by a specific prohormone convertase at conserved dibasic sites flanking the C domain (B-C and C-A junctions) (Steiner, 1998). Although insulin acts in blood in the absence of the C-peptide, newly identified mutations in neonatal-onset DM are found not only in insulin moiety, but also in the C-peptide or even in the signal peptide (24 residues) that is cleaved on translocation into the ER to yield proinsulin.

The Cys^{A7}→Tyr mutant can be viewed as a model to investigate the relationship between misfolding and DM. As shown in Fig. 1A, a mutation of Cys^{A7}→Tyr in permanent neonatal-onset DM is identical to that in the "Akita" mouse with native Type-II diabetes (Yoshioka et al., 1997; Wang et al., 1999; Oyadomari et al., 2002). This A7-Cys mutation and all other Cys-mutations are predicted to block folding of the precursor in the endoplasmic reticulum (ER) of pancreatic β-cells. Studies of the Akita mouse model have demonstrated that misfolding of the various proinsulin forms perturbs the biosynthesis of wild-type insulin, and in turn, the apoptosis of β-cells (Wang et al., 1999; Oyadomari et al., 2002; Ron, 2002; Izumi et al., 2003). These findings have restored interest again in the structural basis of disulfide pairing (Miller et al., 1993; Qiao et al., 2001; Hua et al., 2006a, b; Qiao et al., 2006; Guo et al., 2008). These models of impaired folding provide structural insight into a disease of toxic protein misfolding. However, they also raise questions, such as how insulin binds to its receptor, or why these mutations induce diabetes. The nature of the main underlying cause of diabetes—whether it is damage to the insulin-receptor specific binding or impairment of the proper conformation of insulin—is also unclear. Thus, to understand diabetes-related mutations, we have to interpret insulin structure in detail and reveal its structural-functional relationships.

Insulin is a globular protein containing two chains: the A-chain consists of 21 residues, and the B-chain has 30 residues. The molecule is linked by three disulfide bridges (two inter-chain: A7-B7 and A20-B19, and one intra-A-chain: A6-A11). A variety of dimeric and hexameric crystal structures of insulin have been determined. Structures of insulin in different crystal forms (Blundell et al., 1971; Peking, 1971; Bentley et al., 1976; Baker et al., 1988; Derewenda et al., 1989; Badger et al., 1991; Ciszak and Smith, 1994) have provided a pioneering example of long-range conformational change in a protein (Fig. 1D). Among these, Bake et al. (1988) described the insulin structure in monomer, dimer and hexamer forms in most detail—this paper could be called the "bible" of insulin structure. The T-state of the crystal structure of native insulin is shown in Fig. 1D. Allosteric T/R-transition occurs in various crystal insulin forms; a dimer is shown in Fig. 1C. The N-terminal extended arm of the B-chain (B1–B8, right promoter) in the T-state changes to an α-helix and forms a long helix from B2 to B19 in the R-state (left promoter). To circumvent the self-association of insulin in solution, an engineered monomeric DKP-insulin was designed and used as a template for further mutation. In this monomer, HisB10→Asp prevents the formation of a hexamer and a switch of the B28Pro/B29Lys residues prevents dimerization contact (Weiss et al., 1991; Hua et al., 1996a, 2001b, c, 2009). In addition to X-ray crystallographic approaches, nuclear magnetic resonance (NMR) can be used as a complimentary technique to determine the three-dimensional (3D) solution structures for insulin, its analogs and other flexible, molten-globule-like proteins (Wuthrich, 1986). Fig. 1E shows an ensemble of 20 NMR structures of DKP-insulin. All solution structures of insulin and its analogs in solution exhibit the T-state.

A large and ancestral family of insulin-related sequences has been discovered, including insulin, IGF-I (insulin-like growth factor-I), IGF-II, relaxin, bombyxin, *C. elegans* insulin-like peptide (ILP) and molluscan insulin-like peptide. In the insulin super-family, the sequence homology of Cys and the common "insulin fold" is shared (Humbel, 1990; Cooke et al., 1991; Eigenbrot et al., 1991; Murray-Rust et al., 1992; Sato et al., 1993; Torres et al., 1995; Nagata et al., 1995a, b; De Wolf et al., 1996; Gill et al., 1999; Sato et al., 2000; Hua et al., 2003). Study of the insulin structure is important for

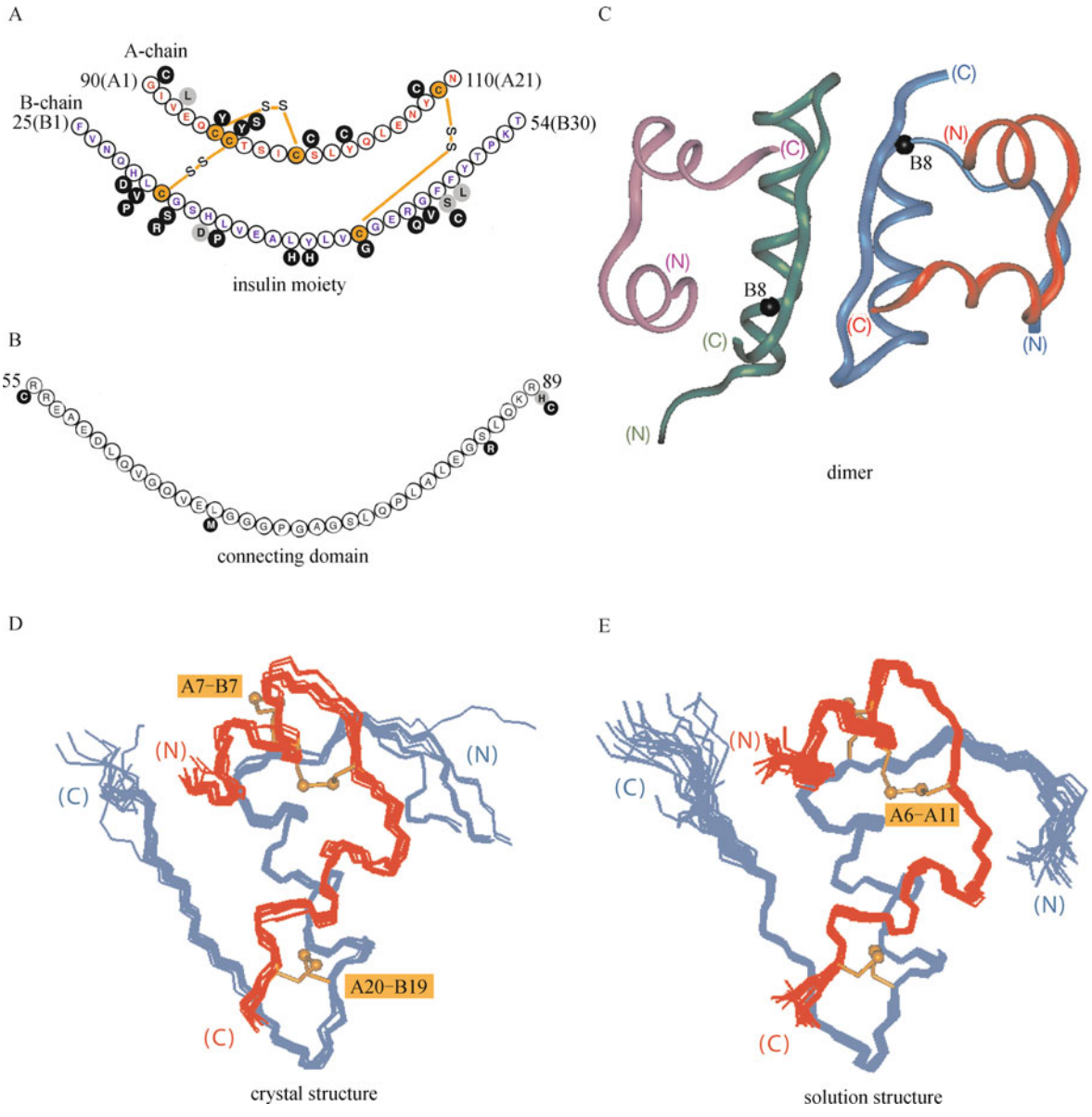

Figure 1. Models for insulin structure and diabetes-associated mutations. (A) Proinsulin sequence showing diabetes-associated amino-acid substitutions in insulin or the insulin moiety of proinsulin. The A- and B-domain of proinsulin, which mediates disulfide pairing, contain multiple sites of mutations causing neonatal DM (filled black circles). The circle at B22 indicates a Maturity-Onset Diabetes of the Young (MODY) phenotype. Substitutions permitting native disulfide pairing (gray) are associated with an adult-onset (classical) phenotype of variable penetrance; these mutations lead to secretion of mature mutant insulin (Leu^{A3}, Ser^{B24}, and Leu^{B25}) or constitutive secretion of a mutant proinsulin (Asp^{B10}). Labeled number: Signal peptide (1–24, not shown), B-chain (25–54, or B1–B30), C-domain (55–89) and A-chain (90–110 or A1–A21). (B) Mutations in the connecting domain are associated with either neonatal DM (black circle) or a syndrome of impaired prohormone processing permissive of folding (gray). (C) Ribbon representation of the T_3R_{f3} crystal structure of insulin in the dimer (PDBID: 1TRZ). Right panel (T-state) showing the A-chain in red and the B-chain in blue, whereas purple and green in the left (R-state) are the A- and B-chain, respectively. The black circle in B8 is a structural switch for a T→R transition. (D) Crystal structure of native insulin in different crystal forms (1MSO, 4INS, 1APH, 1BPH, 1CPH, 1DPH, 1TRZ, 1TYL, 1TYM, 2INS, 1ZNI, 1LPH, 1G7A) and (E) Solution structure of a 20 NMR ensemble of DKP-insulin (PDB ID: 2JMN) with the same alignment according to the backbone atom from A2–A8, A13–A19 and B9–B19.

interpreting insulin and DM as well as for understanding other members of the insulin superfamily. This review primarily focuses on the structure of insulin and its analogs.

STRUCTURAL FEATURES

Hydrophobic core and ordered secondary structure

Despite its small size, insulin contains representative structural features of larger proteins, including a well-ordered hydrophobic core and canonical elements of secondary structure (Baker et al., 1988). As shown in Fig. 1, the solution structure of insulin (E) resembles the crystal T-state structures of insulin in various crystal forms (D): three helices (B9–B19, A2–A8 and A13–A19, designated as helix-1, -2 and -3), two β-turns (B7–B10 and B20–B23) and one β-strand (B24–B28).

The hydrophobic core of insulin is composed of four structural elements: the central helix of B-chain (B9–B19), a mostly buried disulfide bridge (A20–B19), the C-terminal helix of the A-chain (A13–A19) and a β-turn at B20–B23. This β-turn allows the C-terminal β-strand segment to pack against the central helix of the B-chain (Bao et al., 1997; Hua et al., 2001, 2006b). As described above, three helices form the major body in the insulin structure and make insulin a helical rich protein. Of these helices, the A2–A8 helix is not well folded like other two helics (Baker et al., 1988; Hua et al., 1996a, b, 2002a). Insulin is a molten-globule like small protein in some situations, as observed by NMR and other biophysical chemical methods (Hua et al., 1992a).

Three disulfide bridges

Three disulfide bridges (A6-A11, A7-B7, and A20-B19) are "framework" in all members of the insulin family. These bridges (shown in gold in Figs. 1D and 1E) play a critical role for protein stability and biological activity (Narhi et al., 1993; Dai and Tang, 1996; Hua et al., 1996a; Hober et al., 1997; Weiss et al., 2000; Guo and Feng, 2001, 2003; Jia et al., 2003; Yan et al., 2003; Hua et al., 2006b). The structure, stability, and receptor binding activity of insulin require appropriate maintenance of each disulfide bridge (Narhi et al., 1993; Dai and Tang, 1996; Hua et al., 1996a; Weiss et al., 2000; Guo and Feng, 2001; Feng et al., 2003; Jia et al., 2003; Yan et al., 2003; Hua et al., 2006b). The cystine provides interior struts in the hydrophobic core (A19-B20 and A6-A11) and an external staple between chains (A7-B7). Although the (A6-A11) and (A20-B19) disulfide bridges are both buried, the crystal and NMR structural studies indicate that the (A6-A11) bridge is relatively flexible and poorly ordered, whereas the (A20-B19) bridge presumably stabilizes the entire molecule. The exposed (A7-B7) bridge restrains the two N-terminals of the A- and B-chain like a special staple. Mutation at any Cys dramatically impairs receptor binding and destroys the conformation, so that the yield of biosynthesis and chain combination is normally quite low or even undetectable (such as deletion of A20-B19 bond).

Two structural switches and a Gly^{B8} related allosteric T→R transition

When insulin binds to its receptor, a conformational change of insulin occurs at the positions of two structural switches. Analysis of a single insulin chain linked by a B29-A1 covalent bond reveals that it is completely inactive (Derewenda et al., 1989, 1991). However, Fig. 2A shows that the structure of mini-proinsulin is nearly identical to the native T-state insulin (Fig. 1D), which implies that when the C-terminal of the B-chain is "closed" against the central helix of the B-chain,

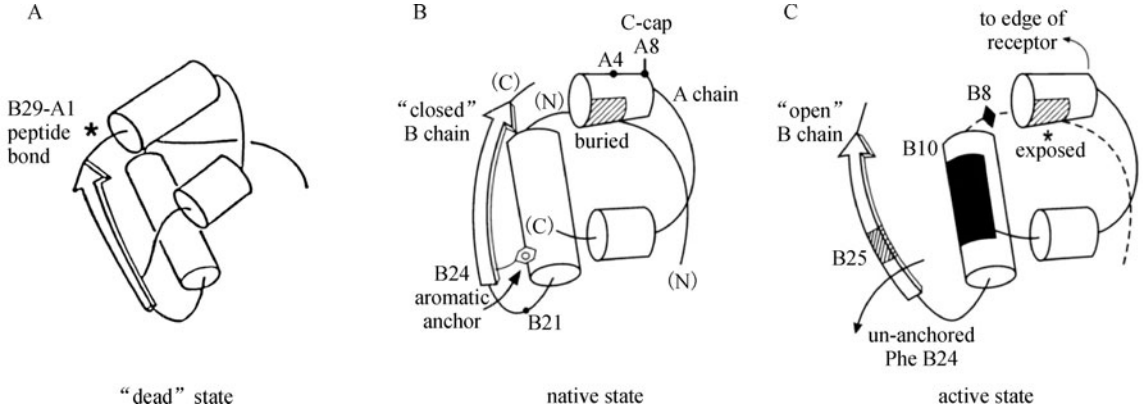

Figure 2. Cylinder models of insulin. (A) Structure of an inactive single-chain insulin: Cylinder model of a 50-residue single-chain insulin analog in which a peptide bond tethers Lys^{B29} to Gly^{A1}. Although the crystal structure of this analog (as a $T_3R^f_3$ zinc hexamer) is essentially identical to that of wild-type insulin, its receptor binding activity is reduced by at least 1000-fold. (B and C) Cylinder representations of insulin: The "closed" model of insulin (B) and the proposed model of the active "open" conformation (C) show that partial detachment of the C-terminal β-strand of the B-chain exposes the side chains of Ile^{A2} and Val^{A3} (shaded; asterisk in C) in a recognition α-helix.

this represents the "inactive" conformation. The crystal form of insulin seems to be a "dead" structure, suggesting that insulin conformation must change upon receptor binding (Derewenda et al., 1991; Liang et al., 1992). The NMR structure of Gly^{B24}-insulin provided the first experimental evidence for this hypothesis (Hua et al., 1991, 1992b). A Phe to Gly substitution at position B24 is tolerant to receptor binding, and it retains as high as 80% of the binding activity of native insulin. Its structure is "opened" at the C-terminal of the B-chain (Fig. 2C). Fig. 2 shows that the "closed" B-chain at the native state is "inactive", whereas the "opened" B-chain may represent an active state. The B24 Phe plays an important role in the structural switch. The "open" state, in which the side-chain of the A2 Ile and A3 Val are exposed from their hidden position, enables key residues of the receptor to bind to the A2 and A3 residues (Nanjo et al., 1986; Steiner et al., 1990; Nakagawa and Tager, 1992; Hua et al., 2002a; Huang et al., 2007).

Chiral mutagenesis at B24 by substitution of D-Ala or L-Ala further demonstrated that the conserved side chain of Phe^{B24} is an anchor for the B24–B28 strand; its substitution by Ala (Fig. 3A) leads to an unstable but native-like analog of low activity (~4% of native insulin). However, D-Ala substitution is equally unstable but exhibits enhanced activity (2-fold affinity) relative to insulin, with segmental unfolding of the β-strand as shown in Fig. 3B (Hua et al., 2009). The Phe^{B24} and adjoining B20–B23 β-turn is required for functional activity (Bao et al., 1997). A bifunctional β-strand B24–B28 mediates both self-assembly (within β-cell storage vesicles) and receptor binding (in the bloodstream).

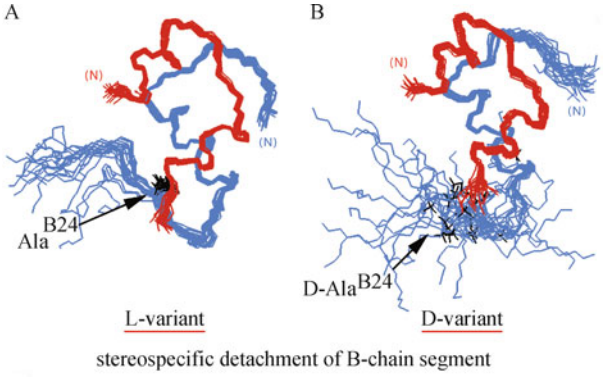

Figure 3. Structural models. (A) NMR ensemble of L-Ala^{B24}-DKP-insulin. (B) NMR ensemble of D-Ala^{B24}-DKP-insulin. In each panel, the A-chain is shown in red and B-chain in blue.

Another structural switch is Gly^{B8}, which is a conserved residue in all species with a positive ϕ dihedral angle. While any L-substitution at B8 impairs chain combination, the corresponding D-substitution enhances yield (Nakagawa et al., 2005; Hua et al., 2006c). A stereospecific modification of the structural switch B8 implies that this type of "chiral mutagenesis" blocks receptor binding of the D-analog, but retains the "native state" structure of insulin. The D-Ser substitution enhances stability, but impairs receptor binding by 100-fold; whereas the L-Ser substitution remarkably impairs stability with only 2-fold reduction in receptor binding. Similarly, while D-Ala at B8 "locks" insulin conformation into the T-state and causes only 0.1% affinity, L-Ala-substitution "rescues" receptor binding as high as 20 folds (Nakagawa et al., 2005; Hua et al., 2006c). The sign of the B8 ϕ angle possibly contributes to the alignment or misalignment of the neighboring Cys^{B7} for disulfide pairing (Baker et al., 1988; Adams et al., 2000).

About 20 years ago, B6 Leu was reported to possibly participate in receptor binding (Nakagawa and Tager, 1991). Replacement of Leu^{B6} by Gly has a nearly equivalent effect as deletion of residues B1–B6 in decreasing receptor binding potency (to only 0.05% of insulin). Ala scanning also revealed that replacement impairs binding (Kristensen et al., 1997). This suggests that the position of B6 might undergo some movement during receptor binding. The T→R transition has been proposed as a model for insulin's active conformation (Nakagawa et al., 2005). D-amino acids at B8 would thus favor the T state whereas L substitutions would favor the R state. The T→R transition is an allosteric feature of the zinc-hexamer assembly. Gly^{B8} related T→R transition exploits the intrinsic flexibility of Gly^{B8}. Therefore, both N- and C-terminal segments of the B-chain would be expected to reorganize receptor binding and the crystallographic T→R transition in part represents how the insulin monomer changes conformation upon receptor binding. It is reasonable to speculate that a B8 switch related T→R transition is required for receptor binding.

Inter-chain hydrogen bond

As a small protein, insulin is stabilized by bulk and subtle weak interactions, such as hydrogen bonding and Van der Waals contact. Among these, inter-chain hydrogen bonds in insulin are important for proper folding, such as those at A6 CO-B6 HN, A7 CO-B5HϵN, A11 HN-B4 CO, A19 CO-B25 HN and A21 HN-B23 CO. The NMR structure of the depsi-insulin (B24–B25 CO-O) with native sequence and a B24-B25 ester bond (Kurapkat et al., 1997) maintains a structure similar to that of native insulin, but it retains only 3%–4% potency. This modification breaks the H-bond between the amide proton of B25 and the carbonyl oxygen of A19. The crystal structure studies of Asn^{A21} substituted by Gly, Ala, Asp and Gln revealed the conformational correlation and the coupled motion between residues in the A21 and B25 side chains, respectively (Zeng et al., 2000). Invariant residue A21 may affect receptor binding. The lack of a hydrogen bond between B25 and A19 or the replacement of A21 both affect the position of the B25 Phe side chain and consequently impair

receptor binding. The hydrogen bond A21-B23 and the local B25 environment are sensitive for receptor binding. Actually, a few invertebrate insulin-like peptides (such as β-subfamily in *C. elegans* ILP) contain a fourth disulfide bond (A21-B23), which is interesting in terms of the evolution of insulin (Duret et al., 1998; Pierce et al., 2001; Hua et al., 2003).

Mutational sites His^{B5} and Gly^{B8} (both broadly conserved among vertebrate insulins) have been well characterized (Nakagawa et al., 2005; Hua et al., 2006a, c; Wan et al., 2008). His^{B5} is an important neighbor residue of B8. The solution structure of the Ala^{B5} insulin analog retains native features and exhibits substantial biological activity (129% relative to native insulin) despite its reduced foldability (< 10% yield relative to wild-type insulin), which implies that His^{B5} is not directly involved in receptor binding but significantly affects the yield of insulin (Hua et al., 2006a). Explanation of these properties of Ala^{B5}-DKP-insulin might be as follows. Sequence homology analysis of insulin indicates that a His^{B5} residue exists in almost all vertebrate insulins and that Arg^{B5} exists in some non-mammalian such as coypu or rabbit fish. Both these His and Arg residues contain side-chain NH groups. An important difference between structures of Ala^{B5}-DKP-insulin and its template DKP-insulin is the existence of an H-bond between the B5 His imidazole proton and the A7 (or A9) carbonyl oxygen in DKP-insulin. It seems that formation of a disulfide bridge (A7-B7) in insulin is difficult without this hydrogen bond. This hypothesis has lately been proved by the observation that the crystal structure of Arg^{B5}-insulin contains a B5-A7 hydrogen bond (Wan et al., 2008). Thus, in the native state, His^{B5} can be proposed to pack within an inter-chain crevice, changing one or more hydrogen bonds to carbonyl oxygens in the A chain. Substitution of His^{B5} by diverse side chains impairs disulfide pairing in chain combination (Hua et al., 2006a),whereas Arg^{B5} is well tolerated, presumably because of formation of analogous inter-chain hydrogen bonds (B5-A7) (Wan et al., 2008). Substitution of His^{B5} by Asp in mammalian cell culture blocks the folding and secretion of human proinsulin (Hua et al., 2006a). Other inter-chain H-bonds (B6-A6 and A11-B4) may also contribute to the formation of the A7-B7 disulfide bond.

FOLDING: DISULFIDE PAIRING VS. CORRECT FOLDING

From the chemical aspect, a protein is a linear heteropolymer. Protein folding has been intensively studied over the last three decades, since Anfinsen's famous assumption proposed that the native structure of a protein is uniquely determined by its linear amino acid sequence (Anfinsen, 1973). Protein folding study is important for both basic research and in the pharmaceutical industry. The structure of folding intermediates is difficult to solve due to their short "life time" and easy aggregation. Furthermore, for Cys-containing proteins, the appropriate folding is always coupled with correct disulfide pairing. Air oxidation of isolated A- and B-chains of insulin preferentially yields insulin with correct disulfide pairing (Du et al., 1965; Katsoyannis and Tometsko, 1966; Zahn and Schmidt, 1967). In each case, oxidative refolding of proinsulin and insulin chain combinations utilizes similar native disulfide pairing (Tang and Tsou, 1990; Wang and Tsou, 1991). The folding pathway of insulin or its single chain analog PIP (Porcine Insulin Precursor with Ala-Lys di-peptide linker between B30 and A1) reveals certain folding intermediates with different disulfide pairings (Milner et al., 1999; Qiao et al., 2001, 2003; Yan et al., 2003). Oxidative refolding of proinsulin or chain combination of insulin yields only the native disulfide pairing scheme (Hua et al., 2006b); however, human IGF-I refolds to form two products: one native IGF-I (~60%) and the other with a specific disulfide-swap (~40%). The two isomers are in thermodynamic equilibrium (Miller et al., 1993; Narhi et al., 1993; Hua et al., 1995, 1996b). Because these folding intermediates are difficult to crystallize, NMR is an alternative approach to solve or mimic the solution structure of these intermediates. Insulin and IGF-I exhibit stepwise acquisition of structures with successive formation of one, two and three disulfide bridges (Narhi et al., 1993; Hua et al., 2001). The model disulfide intermediates of insulin (or IGF-1) with different disulfide bonds are listed in Table 1.

Table 1 List of disulfide intermediate models of insulin

	insulin	IGF-1
one disulfide bond isomer		$Ala^{47, 52}$, $Ser^{6, 48}$-IGF-1 IGF-I-peptide
two disulfide bond isomer	$Ser^{A6, A11}$-DKP, $Ala^{A6, A11}$-DKP, $Ser^{A7, B7}$-DKP	$Ala^{47, 52}$-IGF-1
three disulfide bond isomer	[A6-B7, A7-A11, A20-B19]-HI	[48-52, 6-47, 18-61]-IGF-1
	A6-A7, A11-B7, A20-B19]-HI	[47-48, 6-48, 18-61]-IGF-1

* Disulfide bonds with underlining are abnormal swapped pairings.

Three disulfide insulin isomers: insulin disulfide isomer-I and -II

The formation of insulin folding isomers with three disulfides can be induced by disulfide exchange in dithiothreitol (DTT) with Guanidine-HCl (Hua et al., 1995, 2002b). The products of disulfide reassortment in the denatured condition are isomer-I (insulin-swap-1, Fig. 4A) and isomer-II (insulin-swap-2, Fig. 4B), both of which retain the native A20-B19 disulfide. Swap-1 contains (A6-B7 and A7-A11) disulfides while swap-2 contains (A6-A7 and A11-B7) disulfides, as listed in Table 1. While the main structural feature of native insulin is retained in swap-1 with movement of N-terminal of B-chain, the structure is much less ordered in swap-2, which coincides

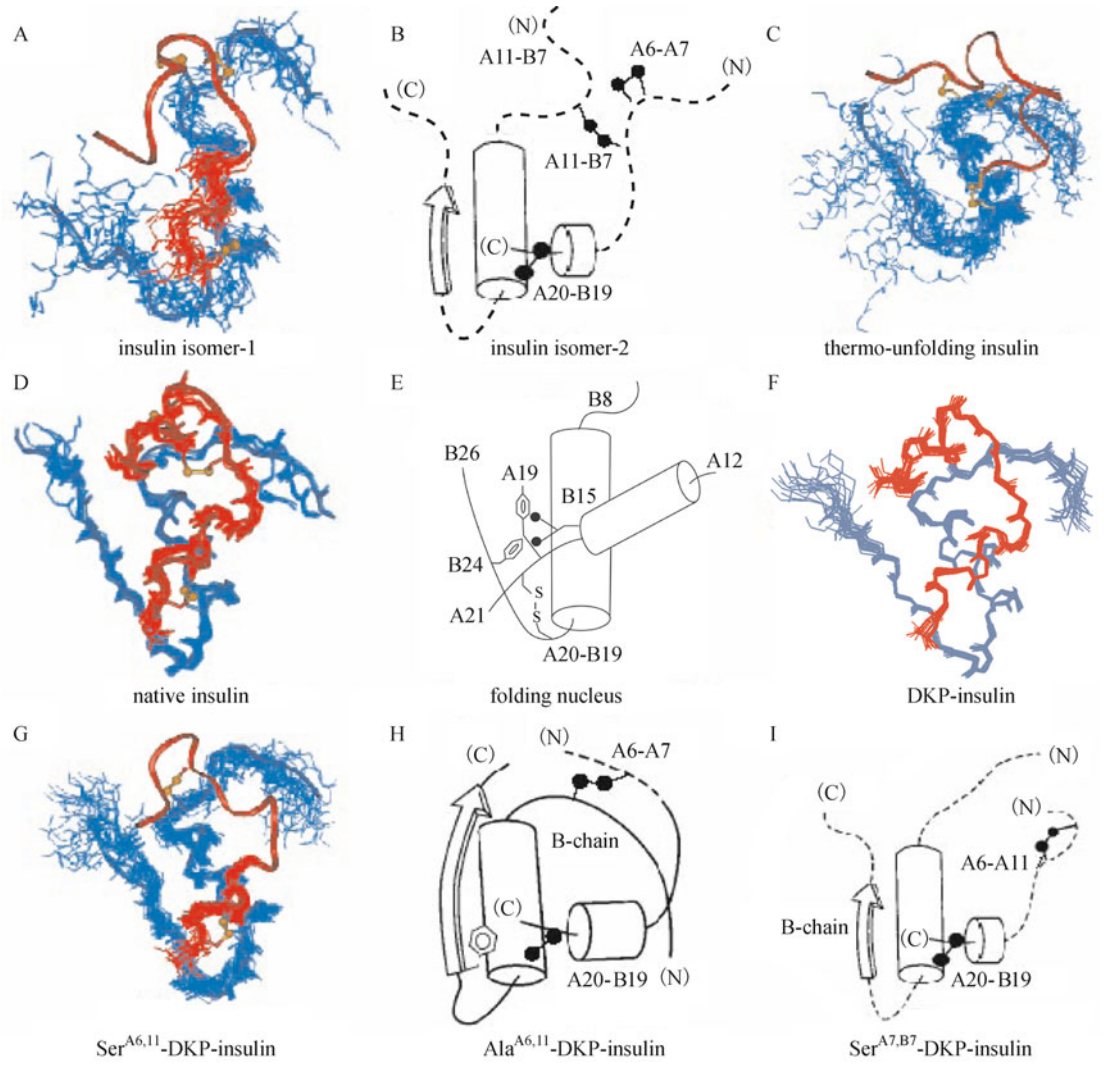

Figure 4. Summary of structures of insulin folding intermediates coupled with disulfide pairing. (A) An ensemble of 20 NMR structures of insulin disulfide intermediates isomer-1 (PDB ID: 1SGL). (B) Proposed model of insulin isomer-2. (C) Thermo-unfolded insulin at 60°C (PDB ID: 1SF1). (D) An ensemble of 10 NMR structures of native insulin (PDB ID: 2HIU). (E) Cylinder model of folding nucleus representing the structural motif, the super-secondary structure. (F) An ensemble of 20 NMR structures of DKP-insulin (PDB ID: 2JMN). (G) An ensemble of 20 NMR structures of Ser$^{A6,\ A11}$-DKP-insulin (PDB ID: 1VKT). (H) Proposed model of Ala$^{A6,\ A11}$-DKP-insulin. (I) Proposed model of Ser$^{A7,\ B7}$-DKP-insulin. The A-chain backbone and ribbon are shown in red and the B-chain in blue. The structures in (A, C, D, F and G) were aligned in respect to the main chain atoms of B9–B19. All of the intermediates retained the super-secondary structure shown in (E).

with a 3.9 kcal/mol reduction in denatured stability relative to native insulin (Hua et al., 2002b). The two N-terminal "arms" of insulin exhibit high flexibility, with only one "staple" existing between A11 and B7 (Fig. 4B). The NMR structure of IGF-I-*swap* exhibits a similar pattern to that of insulin-swap-I (Sato et al., 2000).

The pre-fibril solution structure of insulin (Fig. 4C) maintaining three disulfide bonds was obtained under amyloidogenic conditions (pH 2.4 and 60°C). A novel partial fold is observed similar to swap-2 (Fig. 4B). The N-terminal segments of both chains of insulin detach from the core (Hua and Weiss, 2004). Unfolding of the N-terminal α-helix of the A-chain exposes a hydrophobic surface but maintains native-like packing, including the remaining parts of the α-helices. The C-terminal segment of the B-chain, although not well ordered, remains partially tethered to this "shrunken" helical core.

Two disulfide insulin isomers: pair-substitution of Cys→Ala (or Ser)

Two disulfide isomers of insulin exhibit molten-globule

structures of marginal stability and low biological activity (Sieber et al., 1978; Hua et al., 1995, 2002b). As listed in Table 1, four insulin analogs exist containing two disulfides are listed: $Ser^{A6, A11}$-DKP, $Ala^{A6, A11}$-DKP, $Ala^{47, 52}$-IGF-I and $Ser^{A7, B7}$-DKP (Hua et al., 1996a, b, 2001; Weiss et al., 2000). The solution structure of Ser substitution at A6 and A11 in insulin (Fig. 4G) and in IGF-I is remarkable for segmental unfolding of the N-terminal A-chain α-helix (A2–A8) in an otherwise native sub-domain (Hua et al., 1996a, b). The A6-A11 disulfide bridge is buried and surrounded by hydrophobic residues in native insulin. The preference to form a helix in the N-terminal segment makes it possible to dock into the "internal template" by a hydrophobic interaction, and consequently, the A2–A8 helix formation may help to link the disulfide bridge at A6-A11. The Ala-substituted analog retains the same structure as the Ser analog (Fig. 4H) (Weiss et al., 2000).

Intriguingly, the structure of $Ser^{A7, B7}$-DKP-insulin changed dramatically (Fig. 4I), and the most exposed disulfide A7-B7 in native insulin was broken (Hua et al., 2001). Circular dichroism (CD) and NMR results indicate that this intermediate model retains a much less well-defined secondary structure than that of the above models. In comparison to insulin swap-2, its two N-terminal "arms", without the "staple" between A7 and B7, (Fig. 4I) exhibit even higher flexibility (Fig. 4B).

A one-disulfide IGF-I analog: a folding nucleus

An insulin analog with only one disulfide bond (A20-B19) is rarely synthesized because of its easy aggregation and extremely low yield during chain combination. To mimic a one-disulfide bond model, a corresponding one-disulfide IGF-analog was prepared, which contains only 37 residues with an 18-61 disulfide bridge (corresponding to A20-B19 in insulin) (Hua et al., 2006b). This one-disulfide-linked IGF-I analog is not well ordered. However, NMR spectra illustrate the existence of nascent helical elements corresponding to native α-helices and establishes transient formation of a native-like partial core, suggesting that early events in the folding of insulin-related polypeptides are nucleated by a native-like molten subdomain containing a (A20-B19) bridge.

The intermediate studies of insulin (and IGF-I) are summarized by structural models in Fig. 4, in which a folding nucleus is shown in the middle. In spite of the conformational differences among a number of the intermediates in Fig. 4, one structural feature always remains: the structural motif (also referred to as the super-secondary structure) as shown in Fig. 4E, which consists of two helices: B9–B19 and A13–A19 helix (or even a short helix of A17–A20), one disulfide bridge (A20-B19), and an adjoining B23–B26 segment that folds back against the central B-chain. We speculate that this super-secondary structure is the "folding nucleus". Cystine A20-B19 is the unique one-disulfide species to accumulate in the refolding of mini-proinsulin and IGF-I (Miller et al., 1993, 1999; Qiao et al., 2001), and thus, it is proposed to contribute to the stabilization of a specific folding nucleus (Hua et al., 1996b, 2001, 2006b). The intrinsic conformational characteristics of the flexible disulfide-linked peptide define the folding nucleus, foreshadowing the structure of the native state. Folding can proceed after A20-B19 pairing through multiple alternative channels, including on-pathway intermediates and off-pathway kinetic trapping (see reference (Hua et al., 2001; Qiao et al., 2001; Yan et al., 2003; Hua et al., 2006b) for a detailed discussion). This disulfide bridge packs within a cluster of conserved aliphatic and aromatic side chains in the hydrophobic core (Hua et al., 1995, 2002a, b). As successive disulfide bridges are introduced into the above models, NMR spectra exhibit progressively increased chemical-shift dispersion, suggesting stepwise stabilization of structure.

BINDING WITH THE INSULIN RECEPTOR

How insulin binds to its receptor is still unknown, despite decades of investigation. Insulin's active conformation and how it relates to the classical crystal structures of insulin is not yet clear, nor are the points of contact between insulin and the insulin receptor (IR). Whether insulin binding triggers the reorganization of the ectodomain of the IR also remains to be established. The best answer to these questions would come from determining the structure of the complex of insulin and its receptor. However, this complex has not yet been characterized, either by X-diffraction or by NMR.

Structure of the insulin receptor

The insulin receptor and the IGF-I receptor belong to the superfamily of receptor tyrosine kinases. To date, the crystal structure of the IR ectodomain in Fig. 5 provides the highest resolution for the IGF-I structure, which seems different from that of the EGFR, especially in the regions governing ligand specificity (See details in references (Lou et al., 2006; McKern et al., 2006; Lawrence et al., 2007; Ward et al., 2008)).

Insulin-receptor binding surface

There are two ways to identify the residue proposed for receptor binding: an alanine scan (and other site-directed mutagenesis by other residues) (Kristensen et al., 1997) and photo cross-linking (Kurose et al., 1994). While alanine scanning mutagenesis has been systematically used to identify specific side chain of insulin (Kristensen et al., 1997), the photo cross-linking method directly reveals the contacts between specific residues of insulin and its IR binding points (Huang et al., 2004; Xu et al., 2004a, b, 2009). To determine the contact between insulin and IR by a photo cross-linking method, insulin analogs containing a

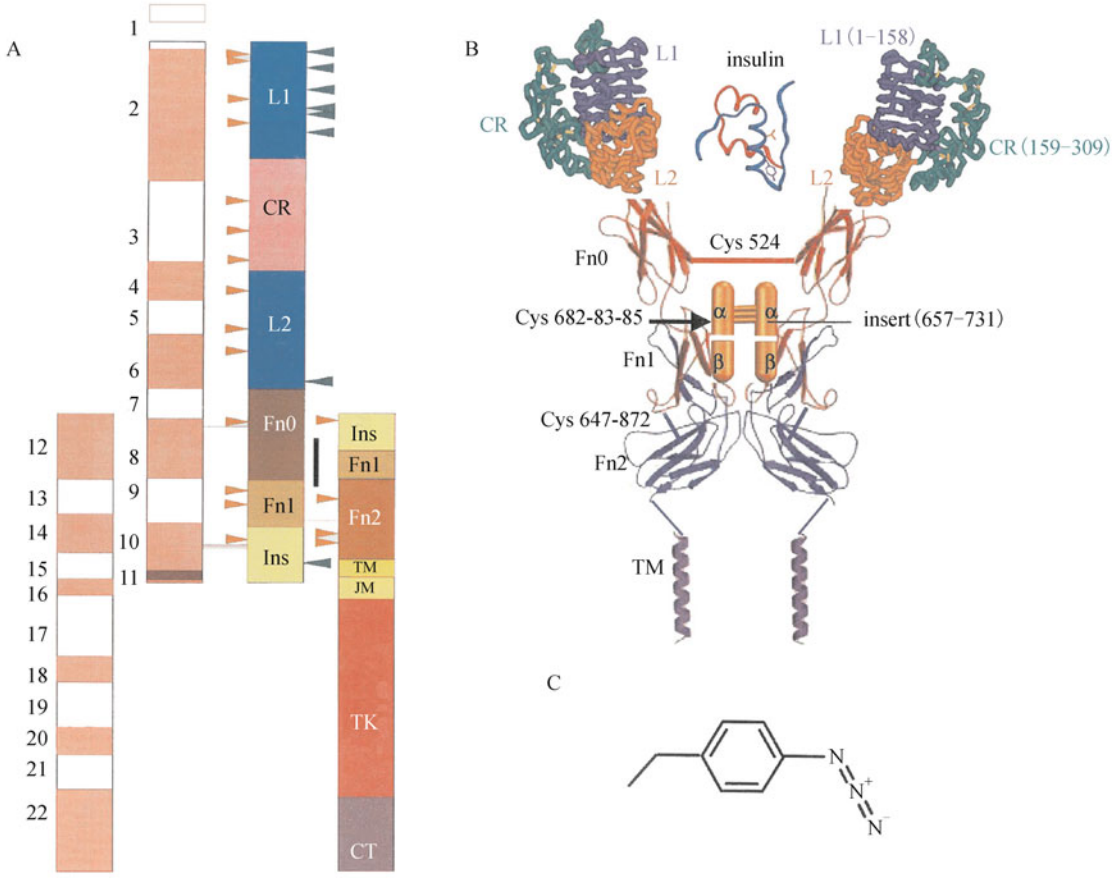

Figure 5. Structural models of the insulin receptor. (A) Cartoon of a modular structure of the $\alpha_2\beta_2$ insulin receptor (IR). On the left half of the receptor, spans of the sequences encoded by the 22 exons of the IR gene are shown in alternating white and pink boxes. The alternatively spliced exon 11 is highlighted. On the right half, modules are predicted from secondary-structure. Module boundaries mostly correspond to exon boundaries. L1 and L2, large domains 1 and 2, containing leucine-rich repeats; CR, cysteine-rich domain; Fn, fibronectin type III domains; Ins, insert domain; TM, trans-membrane domain; JM, juxtamembrane domain; TK, tyrosine kinase domain; CT, C-terminal tail. All modules are drawn to scale. (B) Corresponding ribbon representations. The tyrosine kinase domain is cut off. Structure of insulin is in the middle of the IR dimer. The A-chain is in red and the B-chain in blue. (C) Chemical structure of para-azide-phenylalanine; only the side-chain is shown.

photo-activatable derivative of phenylalanine, *para*-azido-Phe (Pap, Fig. 5C), were first synthesized (Kurose et al., 1994). The α-amino group of the B-chain is biotinylated to enable efficient detection of cross-linked peptides (Kurose et al., 1994; Huang et al., 2004; Xu et al., 2009). In addition to the identification of insulin residues for binding, the resulting cross-linked peptide fragments can be separated by affinity chromatography and sequenced; consequently, some of the smallest identified fragments in the α-subunit of IR have been analyzed (Kurose et al., 1994; Xu et al., 2009). Interestingly, a recent report shows that insulin inserts a detachable arm between receptor domains. For residues in the insulin B-chain β-strand segment, even-numbered probes (at positions B24 and B26) contact the N-terminal L1 domain of the α-subunit, while odd-numbered probes (at positions B25 and B27) contact its C-terminal insert domain (Xu et al., 2009).

Models of the insulin receptor binding surface have been extended to define two distinct clusters of contact residues. The major (or classical) receptor binding surface (Pullen et al., 1976) is proposed to contain residues Gly^{A1}, Ile^{A2}, Val^{A3}, Tyr^{A19}, Val^{B12}, Tyr^{B16}, Gly^{B23}, Phe^{B24}, Phe^{B25} and Tyr^{B26} (De Meyts and Whittaker, 2002), whereas the accessory surface is proposed to contain Leu^{A13} and Leu^{B17} (De Meyts, 1994; Schäffer, 1994; De Meyts and Whittaker, 2002). One ligand would appear to serve as a cross-link between two subunits within the $\alpha_2\beta_2$ heterotetramer (Fig. 5B).

Induced fit of insulin on receptor binding

As discussed above, two structural switches exist in the insulin structure. “Chiral mutagenesis”, which compares matched D- and L-amino-acid substitutions at sites, reveals

induced fitting at B8 or B24. A conserved residue Phe^{B24} is involved in receptor binding since its substitution by D-Phe or D-Ala enhances insulin activity (Kobayashi et al., 1982; Shoelson et al., 1992), whereas GlyB8 is not directly involved in binding, but seems indirectly involved in binding via crystallographic T→R transition. The side chain of Leu^{B6} is buried in the T state, but might be exposed on a change in configuration of Gly^{B8}. The activity of D-Ala^{B8} analogs is the lowest among known analogs of native-like structure. The minimal photo-cross-linking properties of a low-affinity D-*para*-azido-Phe^{B8} analog suggests that the impaired activity of D analogs is due to "locking" of the B8 ϕ dihedral angle and prevention of binding of the D-side chain to the receptor.

At the ostensibly unrelated processes of disulfide pairing, allosteric assembly and receptor binding appear to utilize the same residue as a structural switch: an "ambiguous" B8 glycine unhindered by the chiral restrictions of the Ramachandran plane.

However, it is still not clear which state insulin adopts when it binds to its receptor. The NMR structure and stability analysis of D-Ala^{B8}- and D-Ser^{B8}-DKP-insulin imply at least that insulin is not in the T-state when binding.

STABILITY AND FLEXIBILITY

Fibrillation and misfolding

Stability is required for correct folding, whereas flexibility is required for receptor binding, especially for structural switches. However, misfolding is an unstable state with elevated flexibility that favors fibril formation. Protein fibrillation (Anfinsen, 1973; Cohen, 1999; Dobson, 1999) is a misfolding process of both fundamental and medical interest. As described above, folding intermediates easily undergo misfolding-coupled aggregation (Fig. 4) and insulin will be in the pre-fibril state (Fig. 4C) in some conditions, such as in the presence of urea, guanidine, co-solvents (such as ethanol), hydrophobic surfaces, stirring and high temperature (Nielsen et al., 2001; Ahmad et al., 2003). Protein fibrillation, which is a cytotoxic process of broad structural interest, is recognized as a mechanism of disease, including diverse amyloidoses and prion-associated encephalopathies (Dobson, 1999; Bucciantini et al., 2002). As a classical amyloidogenic protein, insulin provides a good model (Waugh, 1941, 1944, 1946a, 1946b). Fibrillation is promoted by factors that destabilize the classical self-association pathway from monomer → dimer → tetramer → hexamer → higher-order native assemblies, presumably by increasing the availability of the susceptible monomer (Brange and Langkjaer, 1997). Insulin is a helix-rich protein; therefore, it is a good example of the canonical transition of an α-helix to a cross-β-structure, as observed in amyloid diseases and prions (Laws et al., 2001). Under conditions that promote the rapid fibrillation of zinc-free insulin, proinsulin is refractory to fibrillation. The connecting peptide, although not well ordered in native proinsulin, participates in a fibril-specific β-sheet (Huang et al., 2005). The B30-A1 tethered mini-proinsulin impedes the fiber-specific α→β transition, leading to slow formation of a structurally non-uniform amorphous precipitate. Whereas native insulin exhibits pairings (A6-A11, A7-B7 and A20-B19), metastable isomers with alternative pairings (A6-B7, A7-A11 and A20-B19, or A6-A7, A11-B7 and A20-B19) readily undergo fibrillation with essentially identical α→β transitions (Huang et al., 2006). Acid-induced "molten" features might also possibly increase the concentration of a putative fibrillogenic intermediate.

Design of new active ultra-stable insulin analog

Development of commercially used insulin has to consider the compensation between structural stability and receptor binding. Two commercially used insulin analogs are Humalog (Lys^{B28} and Pro^{B29}-insulin, from Eli Lilly) and Novalog (Asp^{B28}-insulin, from Novo-Nordisk) (Vajo and Duckworth, 2000); however, both of these were delayed from reaching the market for several years due to enhanced propensity for fibrillation. Recently, an active ultra-stable single-chain insulin analog (SCI-57) has been successfully designed (Hua et al., 2008). SCI-57 contains a total of 57 residues with a six-residue linker (GGGPRR). Its native receptor binding affinity (129% relative to wildtype) is achieved, as hindrance by the linker is offset by favorable substitutions in the insulin moiety with a Thr^{A8}→His substitution. The thermodynamic stability of SCI-57 is markedly increased ($\Delta\Delta G_u$ 0.7 kcal/mol relative to the corresponding 57 residues two-chain analog and $\Delta\Delta G_u$ 1.9 kcal/mol relative to wild-type insulin). The design and structural model are shown in Fig. 6. A new six-residue (GGGPRR) linker is favored for suitable balance between stability and flexibility in C-A junction of proinsulin: three Gly residues are quite flexible and Pro34 enables the linker twist, and the following two Arg are similar to conserved dibasic sites in the C-A junction of proinsulin. Neither longer nor shorter linkers are favored for proper stability and flexibility. In addition to their biophysical interest, ultra-stable SCIs may enhance the safety and efficacy of insulin replacement therapy in the developing world, where no refrigerator may be available to store insulin (Hua et al., 2008). Based on the results of insulin study, the design of a new insulin analogs for commercial use is desired by numerous diabetes patients.

FOOTNOTE

Chinese scholars have made great contributions to the study of insulin at different stages. Recently, Prof. Youshang Zhang reminisced about the total synthesis of crystalline insulin by a Chinese group 50 years ago (Zhang, 2010). Prof. Dacheng Wang provided a retrospective of the determination of insulin's crystal structure by Chinese crystallographers 40 years ago (Wang, 2010) and Prof. Youming Feng summarized recent important contributions on the

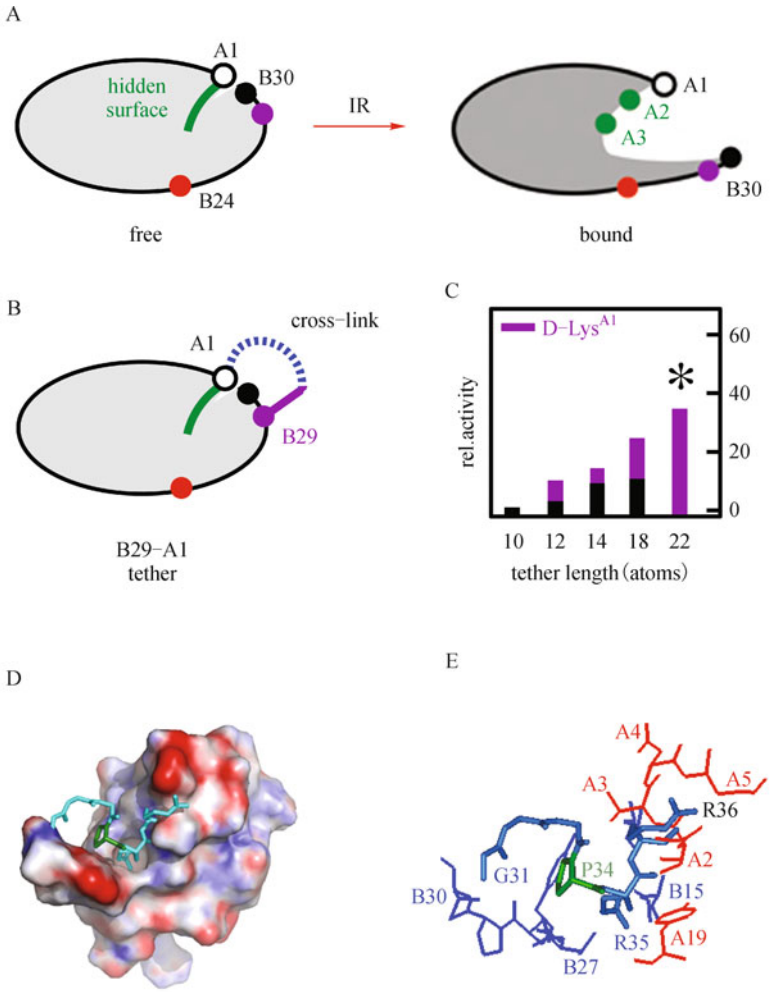

Figure 6. Model of induced fit and a structural model showing the electrostatic surface of a new analog with a linker. (A) Whereas the free conformation of insulin is closed (left), upon receptor binding, detachment of the C-terminal B-chain β-strand is proposed to expose the hidden non-polar surface of the A-chain (right), including Ile^{A2} and Val^{A3} (green). The aromatic side chain of Phe^{B24} (red) may function as a structural switch. Gly^{A1} is shown as an open circle, and residues B29 and B30 as purple and black spheres, respectively. (B) Tethering the A- and B-chains was affected by bi-functional cross-linking (blue dashed line) between the ε-amino group of Lys^{B29} (purple) and the α-amino group of Gly^{A1} (open circle). (C) Histogram showing relative activity of tethered insulin analogs is plotted as a function of the number of atoms between the C_a carbons of residues B29 and A1 (horizontal axis), including the B29 side chain and, when present, the side chain of D-Lys^{A1} (magenta). Bar heights indicate receptor binding affinities relative to wild-type human insulin; black bars indicate linkage to the α-amino group of Gly^{A1} whereas magenta bars indicate linkage to the ε-amino group of D-Lys^{A1}. Asterisk indicates substantial relative activity (35%) exhibited by $N^{\varepsilon A1}$, $N^{\varepsilon B29}$-ethyleneglycol-*bis*-succinoyl-D-Lys^{A1}-insulin. (D) Model showing packing of linker (cyan) related to electrostatic surface of SCI-57 analog: negative (red) and positive (blue). (E) Stick view shown linker (G31–R36) in middle and surrounding residues of A- and B-chain. Same orientation in (D) and (E) was adopted.

structure-function and folding study of insulin (Feng, 2010). A review paper of folding, binding, and stability of insulin can be seen in reference (Hua, 2004).

ACKNOWLEDGEMENTS

The author is grateful to Prof. M. A. Weiss of Case Western Reserve University for long-term support of insulin study and for grant support from the National Institute of Health (GM-61446).

REFERENCES

Adams, T.E., Epa, V.C., Garrett, T.P.J., and Ward, C.W. (2000). Structure and function of the type 1 insulin-like growth factor receptor. Cell Mol Life Sci 57, 1050–1093.

Ahmad, A., Millett, I.S., Doniach, S., Uversky, V.N., and Fink, A.L. (2003). Partially folded intermediates in insulin fibrillation. Biochemistry 42, 11404–11416.

Anfinsen, C.B. (1973). Principles that govern the folding of protein

chains. Science 181, 223–230.

Badger, J., Harris, M.R., Reynolds, C.D., Evans, A.C., Dodson, E.J., Dodson, G.G., and North, A.C. (1991). Structure of the pig insulin dimer in the cubic crystal. Acta Crystallogr B 47, 127–136.

Baker, E.N., Blundell, T.L., Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Hodgkin, D.M., Hubbard, R.E., Isaacs, N.W., Reynolds, C.D., *et al*. (1988). The structure of 2Zn pig insulin crystals at 1.5 A resolution. Philos Trans R Soc Lond B Biol Sci 319, 369–456.

Bao, S.J., Xie, D.L., Zhang, J.P., Chang, W.R., and Liang, D.C. (1997). Crystal structure of *des*heptapeptide(B24–B30)insulin at 1.6 A resolution: implications for receptor binding. Proc Natl Acad Sci U S A 94, 2975–2980.

Bentley, G., Dodson, E., Dodson, G., Hodgkin, D., and Mercola, D. (1976). Structure of insulin in 4-zinc insulin. Nature 261, 166–168.

Blundell, T.L., Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G. G., Hodgkin, D.C., Mercola, D.A., and Vijayan, M. (1971). Atomic positions in rhombohedral 2-zinc insulin crystals. Nature 231, 506–511.

Brange, J. (1987). Galenics of insulin: the physico-chemical and pharmaceutical aspects of insulin and insulin preparations (Berlin: Higher Education Press and Springer-Verlag).

Brange, J., Andersen, L., Laursen, E.D., Meyn, G., and Rasmussen, E. (1997a). Toward understanding insulin fibrillation. J Pharm Sci 86, 517–525.

Brange, J., Dodson, G.G., Edwards, D.J., Holden, P.H., and Whittingham, J.L. (1997b). A model of insulin fibrils derived from the x-ray crystal structure of a monomeric insulin (*des*pentapeptide insulin). Proteins 27, 507–516.

Brange, J., and Langkjaer, L. (1997). Insulin formulation and delivery. Pharm Biotechnol 10, 343–409.

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C.M., and Stefani, M. (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 416, 507–511.

Chothia, C., Lesk, A.M., Dodson, G.G., and Hodgkin, D.C. (1983). Transmission of conformational change in insulin. Nature 302, 500–505.

Ciszak, E., and Smith, G.D. (1994). Crystallographic evidence for dual coordination around zinc in the T3R3 human insulin hexamer. Biochemistry 33, 1512–1517.

Cohen, F.E. (1999). Protein misfolding and prion diseases. J Mol Biol 293, 313–320.

Colombo, C., Porzio, O., Liu, M., Massa, O., Vasta, M., Salardi, S., Beccaria, L., Monciotti, C., Toni, S., Pedersen, O., *et al*., and the Early Onset Diabetes Study Group of the Italian Society of Pediatric Endocrinology and Diabetes (SIEDP). (2008). Seven mutations in the human insulin gene linked to permanent neonatal/infancy-onset diabetes mellitus. J Clin Invest 118, 2148–2156.

Cooke, R.M., Harvey, T.S., and Campbell, I.D. (1991). Solution structure of human insulin-like growth factor 1: a nuclear magnetic resonance and restrained molecular dynamics study. Biochemistry 30, 5484–5491.

Dai, Y., and Tang, J.G. (1996). Characteristic, activity and conformational studies of [A6-Ser, A11-Ser]-insulin. Biochim Biophys Acta 1296, 63–68.

De Meyts, P. (1994). The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling. Diabetologia 37, S135–S148.

De Meyts, P., and Whittaker, J. (2002). Structural biology of insulin and IGF1 receptors: implications for drug design. Nat Rev Drug Discov 1, 769–783.

De Wolf, E., Gill, R., Geddes, S., Pitts, J., Wollmer, A., and Grötzinger, J. (1996). Solution structure of a mini IGF-1. Protein Sci 5, 2193–2202.

Derewenda, U., Derewenda, Z., Dodson, E.J., Dodson, G.G., Bing, X., and Markussen, J. (1991). X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue. J Mol Biol 220, 425–433.

Derewenda, U., Derewenda, Z., Dodson, E.J., Dodson, G.G., Reynolds, C.D., Smith, G.D., Sparks, C., and Swenson, D. (1989). Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer. Nature 338, 594–596.

Dobson, C.M. (1999). Protein misfolding, evolution and disease. Trends Biochem Sci 24, 329–332.

Dodson, G., and Steiner, D. (1998). The role of assembly in insulin's biosynthesis. Curr Opin Struct Biol 8, 189–194.

Du, Y.C., Jiang, R.Q., and Tsou, C.L. (1965). Conditions for successful resynthesis of insulin from its glycyl and phenylalanyl chains. Sci Sin 14, 229–236.

Duret, L., Guex, N., Peitsch, M.C., and Bairoch, A. (1998). New insulin-like proteins with atypical disulfide bond pattern characterized in *Caenorhabditis elegans* by comparative sequence analysis and homology modeling. Genome Res 8, 348–353.

Edghill, E.L., Flanagan, S.E., Patch, A.M., Boustred, C., Parrish, A., Shields, B., Shepherd, M.H., Hussain, K., Kapoor, R.R., Malecki, M., *et al*., and the Neonatal Diabetes International Collaborative Group. (2008). Insulin mutation screening in 1,044 patients with diabetes: mutations in the INS gene are a common cause of neonatal diabetes but a rare cause of diabetes diagnosed in childhood or adulthood. Diabetes 57, 1034–1042.

Eigenbrot, C., Randal, M., Quan, C., Burnier, J., O"Connell, L., Rinderknecht, E., and Kossiakoff, A.A. (1991). X-ray structure of human relaxin at 1.5 A. Comparison to insulin and implications for receptor binding determinants. J Mol Biol 221, 15–21.

Feng, Y., Liu, D., and Wang, J. (2003). Native-like partially folded conformations and folding process revealed in the N-terminal large fragments of staphylococcal nuclease: a study by NMR spectroscopy. J Mol Biol 330, 821–837.

Feng, Y.M. (2010). Chinese work on insulin (Kerala, India: Transworld Research Network).

Gill, R., Verma, C., Wallach, B., Ursø, B., Pitts, J., Wollmer, A., De Meyts, P., and Wood, S. (1999). Modelling of the disulphide-swapped isomer of human insulin-like growth factor-1: implications for receptor binding. Protein Eng 12, 297–303.

Guo, Z.Y., and Feng, Y.M. (2001). Effects of cysteine to serine substitutions in the two inter-chain disulfide bonds of insulin. Biol Chem 382, 443–448.

Guo, Z.Y., Qiao, Z.S., and Feng, Y.M. (2008). The *in vitro* oxidative folding of the insulin superfamily. Antioxid Redox Signal 10, 127–139.

Hober, S., Uhlén, M., and Nilsson, B. (1997). Disulfide exchange folding of disulfide mutants of insulin-like growth factor I in vitro. Biochemistry 36, 4616–4622.

Hua, Q.X. (2004). Insight into folding, binding and stability of insulin by

NMR. Prog Biochem Biophys 31, 1–26.

Hua, Q.X., Chu, Y.C., Jia, W., Phillips, N.F., Wang, R.Y., Katsoyannis, P.G., and Weiss, M.A. (2002a). Mechanism of insulin chain combination. Asymmetric roles of A-chain alpha-helices in disulfide pairing. J Biol Chem 277, 43443–43453.

Hua, Q.X., Gozani, S.N., Chance, R.E., Hoffmann, J.A., Frank, B.H., and Weiss, M.A. (1995). Structure of a protein in a kinetic trap. Nat Struct Biol 2, 129–138.

Hua, Q.X., Hu, S.Q., Frank, B.H., Jia, W., Chu, Y.C., Wang, S.H., Burke, G.T., Katsoyannis, P.G., and Weiss, M.A. (1996a). Mapping the functional surface of insulin by design: structure and function of a novel A-chain analogue. J Mol Biol 264, 390–403.

Hua, Q.X., Jia, W., Frank, B.H., Phillips, N.F., and Weiss, M.A. (2002b). A protein caught in a kinetic trap: structures and stabilities of insulin disulfide isomers. Biochemistry 41, 14700–14715.

Hua, Q.X., Kochoyan, M., and Weiss, M.A. (1992a). Structure and dynamics of *des*-pentapeptide-insulin in solution: the molten-globule hypothesis. Proc Natl Acad Sci U S A 89, 2379–2383.

Hua, Q.X., Liu, M., Hu, S.Q., Jia, W., Arvan, P., and Weiss, M.A. (2006a). A conserved histidine in insulin is required for the foldability of human proinsulin: structure and function of an ALA^{B5} analog. J Biol Chem 281, 24889–24899.

Hua, Q.X., Mayer, J.P., Jia, W., Zhang, J., and Weiss, M.A. (2006b). The folding nucleus of the insulin superfamily: a flexible peptide model foreshadows the native state. J Biol Chem 281, 28131–28142.

Hua, Q.X., Nakagawa, S.H., Hu, S.Q., Jia, W., Wang, S., and Weiss, M.A. (2006c). Toward the active conformation of insulin: stereo-specific modulation of a structural switch in the B-chain. J Biol Chem 281, 24900–24909.

Hua, Q.X., Nakagawa, S.H., Jia, W., Hu, S.Q., Chu, Y.C., Katsoyannis, P.G., and Weiss, M.A. (2001). Hierarchical protein folding: asymmetric unfolding of an insulin analogue lacking the A7-B7 interchain disulfide bridge. Biochemistry 40, 12299–12311.

Hua, Q.X., Nakagawa, S.H., Jia, W., Huang, K., Phillips, N.B., Hu, S. Q., and Weiss, M.A. (2008). Design of an active ultrastable single-chain insulin analog: synthesis, structure, and therapeutic implications. J Biol Chem 283, 14703–14716.

Hua, Q.X., Nakagawa, S.H., Wilken, J., Ramos, R.R., Jia, W., Bass, J., and Weiss, M.A. (2003). A divergent INS protein in *Caenorhabditis elegans* structurally resembles human insulin and activates the human insulin receptor. Genes Dev 17, 826–831.

Hua, Q.X., Narhi, L., Jia, W., Arakawa, T., Rosenfeld, R., Hawkins, N., Miller, J.A., and Weiss, M.A. (1996b). Native and non-native structure in a protein-folding intermediate: spectroscopic studies of partially reduced IGF-I and an engineered alanine model. J Mol Biol 259, 297–313.

Hua, Q.X., Shoelson, S.E., Kochoyan, M., and Weiss, M.A. (1991). Receptor binding redefined by a structural switch in a mutant human insulin. Nature 354, 238–241.

Hua, Q.X., Shoelson, S.E., and Weiss, M.A. (1992b). Nonlocal structural perturbations in a mutant human insulin: sequential resonance assignment and 13C-isotope-aided 2D-NMR studies of [PheB24— > Gly]insulin with implications for receptor recognition. Biochemistry 31, 11940–11951.

Hua, Q.X., and Weiss, M.A. (1991). Comparative 2D NMR studies of human insulin and *des*-pentapeptide insulin: sequential resonance assignment and implications for protein dynamics and receptor recognition. Biochemistry 30, 5505–5515.

Hua, Q.X., and Weiss, M.A. (2004). Mechanism of insulin fibrillation: the structure of insulin under amyloidogenic conditions resembles a protein-folding intermediate. J Biol Chem 279, 21449–21460.

Hua, Q.X., Xu, B., Huang, K., Hu, S.Q., Nakagawa, S., Jia, W., Wang, S., Whittaker, J., Katsoyannis, P.G., and Weiss, M.A. (2009). Enhancing the activity of a protein by stereospecific unfolding: conformational life cycle of insulin and its evolutionary origins. J Biol Chem 284, 14586–14596.

Huang, K., Chan, S.J., Hua, Q.X., Chu, Y.C., Wang, R., Klaproth, B., Jia, W., Whittaker, J., De Meyts, P., Nakagawa, S.H., *et al*. (2007). The A-chain of insulin contacts the insert domain of the receptor. Photo-cross-linking and nonstandard mutagenesis of the diabetes-related A3 crevice. J Biol Chem 282, 35337–35349.

Huang, K., Dong, J., Phillips, N.B., Carey, P.R., and Weiss, M.A. (2005). Proinsulin is refractory to protein fibrillation: topological protection of a precursor protein from cross-beta assembly. J Biol Chem 280, 42345–42355.

Huang, K., Maiti, N.C., Phillips, N.B., Carey, P.R., and Weiss, M.A. (2006). Structure-specific effects of protein topology on cross-beta assembly: studies of insulin fibrillation. Biochemistry 45, 10278–10293.

Huang, K., Xu, B., Hu, S.Q., Chu, Y.C., Hua, Q.X., Qu, Y., Li, B., Wang, S., Wang, R.Y., Nakagawa, S.H., *et al*. (2004). How insulin binds: the B-chain alpha-helix contacts the L1 beta-helix of the insulin receptor. J Mol Biol 341, 529–550.

Humbel, R.E. (1990). Insulin-like growth factors I and II. Eur J Biol 277, 103–118.

Izumi, T., Yokota-Hashimoto, H., Zhao, S., Wang, J., Halban, P.A., and Takeuchi, T. (2003). Dominant negative pathogenesis by mutant proinsulin in the Akita diabetic mouse. Diabetes 52, 409–416.

Jia, X.Y., Guo, Z.Y., Wang, Y., Xu, Y., Duan, S.S., and Feng, Y.M. (2003). Peptide models of four possible insulin folding intermediates with two disulfides. Protein Sci 12, 2412–2419.

Katsoyannis, P.G., and Tometsko, A. (1966). Insulin synthesis by recombination of A and B-chains: a highly efficient method. Proc Natl Acad Sci U S A 55, 1554–1561.

Kobayashi, M., Ohgaku, S., Iwasaki, M., Maegawa, H., Shigeta, Y., and Inouye, K. (1982). Supernormal insulin: [D-PheB24]-insulin with increased affinity for insulin receptors. Biochem Biophys Res Commun 107, 329–336.

Kristensen, C., Kjeldsen, T., Wiberg, F.C., Schäffer, L., Hach, M., Havelund, S., Bass, J., Steiner, D.F., and Andersen, A.S. (1997). Alanine scanning mutagenesis of insulin. J Biol Chem 272, 12978–12983.

Kurapkat, G., De Wolf, E., Grötzinger, J., and Wollmer, A. (1997). Inactive conformation of an insulin despite its wild-type sequence. Protein Sci 6, 580–587.

Kurose, T., Pashmforoush, M., Yoshimasa, Y., Carroll, R., Schwartz, G.P., Burke, G.T., Katsoyannis, P.G., and Steiner, D.F. (1994). Cross-linking of a B25 azidophenylalanine insulin derivative to the carboxyl-terminal region of the alpha-subunit of the insulin receptor. Identification of a new insulin-binding domain in the insulin receptor. J Biol Chem 269, 29190–29197.

Lawrence, M.C., McKern, N.M., and Ward, C.W. (2007). Insulin receptor structure and its implications for the IGF-1 receptor. Curr Opin Struct Biol 17, 699–705.

Laws, D.D., Bitter, H.M., Liu, K., Ball, H.L., Kaneko, K., Wille, H., Cohen, F.E., Prusiner, S.B., Pines, A., and Wemmer, D.E. (2001). Solid-state NMR studies of the secondary structure of a mutant prion protein fragment of 55 residues that induces neurodegeneration. Proc Natl Acad Sci USA 98, 11686–11690.

Liang, D.C., Chang, W.R., Zhang, J.P., and Wan, Z.L. (1992). The possible mechanism of binding interaction of insulin molecule with its receptor. Sci China B 35, 547–557.

Lou, M., Garrett, T.P., McKern, N.M., Hoyne, P.A., Epa, V.C., Bentley, J.D., Lovrecz, G.O., Cosgrove, L.J., Frenkel, M.J., and Ward, C.W. (2006). The first three domains of the insulin receptor differ structurally from the insulin-like growth factor 1 receptor in the regions governing ligand specificity. Proc Natl Acad Sci U S A 103, 12429–12434.

McKern, N.M., Lawrence, M.C., Streltsov, V.A., Lou, M.Z., Adams, T. E., Lovrecz, G.O., Elleman, T.C., Richards, K.M., Bentley, J.D., Pilling, P.A., *et al*. (2006). Structure of the insulin receptor ectodomain reveals a folded-over conformation. Nature 443, 218–221.

Miller, J.A., Narhi, L.O., Hua, Q.X., Rosenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., Stoney, K.S., Tsai, L., *et al*. (1993). Oxidative refolding of insulin-like growth factor 1 yields two products of similar thermodynamic stability: a bifurcating protein-folding pathway. Biochemistry 32, 5203–5213.

Milner, S.J., Carver, J.A., Ballard, F.J., and Francis, G.L. (1999). Probing the disulfide folding pathway of insulin-like growth factor-I. Biotechnol Bioeng 62, 693–703.

Molven, A., Ringdal, M., Nordbø, A.M., Raeder, H., Støy, J., Lipkind, G.M., Steiner, D.F., Philipson, L.H., Bergmann, I., Aarskog, D., *et al*., and the Norwegian Childhood Diabetes Study Group. (2008). Mutations in the insulin gene can cause MODY and autoantibody-negative type 1 diabetes. Diabetes 57, 1131–1135.

Murray-Rust, J., McLeod, A.N., Blundell, T.L., and Wood, S.P. (1992). Structure and evolution of insulins: implications for receptor binding. Bioessays 14, 325–331.

Nagata, K., Hatanaka, H., Kohda, D., Kataoka, H., Nagasawa, H., Isogai, A., Ishizaki, H., Suzuki, A., and Inagaki, F. (1995a). Identification of the receptor-recognition surface of bombyxin-II, an insulin-like peptide of the silkmoth Bombyx mori: critical importance of the B-chain central part. J Mol Biol 253, 759–770.

Nagata, K., Hatanaka, H., Kohda, D., Kataoka, H., Nagasawa, H., Isogai, A., Ishizaki, H., Suzuki, A., and Inagaki, F. (1995b). Three-dimensional solution structure of bombyxin-II an insulin-like peptide of the silkmoth Bombyx mori: structural comparison with insulin and relaxin. J Mol Biol 253, 749–758.

Nakagawa, S.H., and Tager, H.S. (1991). Implications of invariant residue LeuB6 in insulin-receptor interactions. J Biol Chem 266, 11502–11509.

Nakagawa, S.H., and Tager, H.S. (1992). Importance of aliphatic side-chain structure at positions 2 and 3 of the insulin A chain in insulin-receptor interactions. Biochemistry 31, 3204–3214.

Nakagawa, S.H., Zhao, M., Hua, Q.X., Hu, S.Q., Wan, Z.L., Jia, W., and Weiss, M.A. (2005). Chiral mutagenesis of insulin. Foldability and function are inversely regulated by a stereospecific switch in the B-chain. Biochemistry 44, 4984–4999.

Nanjo, K., Sanke, T., Miyano, M., Okai, K., Sowa, R., Kondo, M., Nishimura, S., Iwo, K., Miyamura, K., Given, B.D., *et al*. (1986). Diabetes due to secretion of a structurally abnormal insulin (insulin Wakayama). Clinical and functional characteristics of [LeuA3] insulin. J Clin Invest 77, 514–519.

Narhi, L.O., Hua, Q.X., Arakawa, T., Fox, G.M., Tsai, L., Rosenfeld, R., Holst, P., Miller, J.A., and Weiss, M.A. (1993). Role of native disulfide bonds in the structure and activity of insulin-like growth factor 1: genetic models of protein-folding intermediates. Biochemistry 32, 5214–5221.

Nielsen, L., Khurana, R., Coats, A., Frokjaer, S., Brange, J., Vyas, S., Uversky, V.N., and Fink, A.L. (2001). Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. Biochemistry 40, 6036–6046.

Oyadomari, S., Koizumi, A., Takeda, K., Gotoh, T., Akira, S., Araki, E., and Mori, M. (2002). Targeted disruption of the Chop gene delays endoplasmic reticulum stress-mediated diabetes. J Clin Invest 109, 525–532.

Peking, I.R.G. (1971). Insulin's crystal structure at 2.5A resolution. Peking Rev 40, 11–16.

Pierce, S.B., Costa, M., Wisotzkey, R., Devadhar, S., Homburger, S. A., Buchman, A.R., Ferguson, K.C., Heller, J., Platt, D.M., Pasquinelli, A.A., *et al*. (2001). Regulation of DAF-2 receptor signaling by human insulin and ins-1, a member of the unusually large and diverse *C. elegans* insulin gene family. Genes Dev 15, 672–686.

Pullen, R.A., Lindsay, D.G., Wood, S.P., Tickle, I.J., Blundell, T.L., Wollmer, A., Krail, G., Brandenburg, D., Zahn, H., Gliemann, J., *et al*. (1976). Receptor-binding region of insulin. Nature 259, 369–373.

Qiao, Z.S., Guo, Z.Y., and Feng, Y.M. (2001). Putative disulfide-forming pathway of porcine insulin precursor during its refolding in vitro. Biochemistry 40, 2662–2668.

Qiao, Z.S., Guo, Z.Y., and Feng, Y.M. (2006). *In vitro* folding/unfolding of insulin/single-chain insulin. Protein Pept Lett 13, 423–429.

Qiao, Z.S., Min, C.Y., Hua, Q.X., Weiss, M.A., and Feng, Y.M. (2003). *In vitro* refolding of human proinsulin. Kinetic intermediates, putative disulfide-forming pathway folding initiation site, and potential role of C-peptide in folding process. J Biol Chem 278, 17800–17809.

Ron, D. (2002). Proteotoxicity in the endoplasmic reticulum: lessons from the Akita diabetic mouse. J Clin Invest 109, 443–445.

Sato, A., Koyama, S., Yamada, H., Suzuki, S., Tamura, K., Kobayashi, M., Niwa, M., Yasuda, T., Kyogoku, Y., and Kobayashi, Y. (2000). Three-dimensional solution structure of a disulfide bond isomer of the human insulin-like growth factor-I. J Pept Res 56, 218–230.

Sato, A., Nishimura, S., Ohkubo, T., Kyogoku, Y., Koyama, S., Kobayashi, M., Yasuda, T., and Kobayashi, Y. (1993). Three-dimensional structure of human insulin-like growth factor-I (IGF-I) determined by 1H-NMR and distance geometry. Int J Pept Protein Res 41, 433–440.

Schäffer, L. (1994). A model for insulin binding to the insulin receptor. Eur J Biochem 221, 1127–1132.

Shoelson, S.E., Lu, Z.X., Parlautan, L., Lynch, C.S., and Weiss, M.A. (1992). Mutations at the dimer, hexamer, and receptor-binding surfaces of insulin independently affect insulin-insulin and insulin-receptor interactions. Biochemistry 31, 1757–1767.

Sieber, P.S., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Märki, F., and de Gasparo, M. (1978). Synthesis and biological activity of two disulphide bond isomers of human insulin: [A7–A11,

A6–B7-cystine]- and [A6–A7,A11–B7-cystine]insulin (human). Hoppe Seylers Z Physiol Chem 359, 113–123.

Steiner, D.F. (1967). Evidence for a precursor in the biosynthesis of insulin. Trans N Y Acad Sci 30, 60–68.

Steiner, D.F. (1998). The proprotein convertases. Curr Opin Chem Biol 2, 31–39.

Steiner, D.F., Tager, H.S., Chan, S.J., Nanjo, K., Sanke, T., and Rubenstein, A.H. (1990). Lessons learned from molecular biology of insulin-gene mutations. Diabetes Care 13, 600–609.

Støy, J., Edghill, E.L., Flanagan, S.E., Ye, H., Paz, V.P., Pluzhnikov, A., Below, J.E., Hayes, M.G., Cox, N.J., Lipkind, G.M., *et al*., and the Neonatal Diabetes International Collaborative Group. (2007). Insulin gene mutations as a cause of permanent neonatal diabetes. Proc Natl Acad Sci U S A 104, 15040–15044.

Tang, J.G., and Tsou, C.L. (1990). The insulin A and B-chains contain structural information for the formation of the native molecule. Studies with protein disulphide-isomerase. Biochem J 268, 429–435.

Torres, A.M., Forbes, B.E., Aplin, S.E., Wallace, J.C., Francis, G.L., and Norton, R.S. (1995). Solution structure of human insulin-like growth factor II. Relationship to receptor and binding protein interactions. J Mol Biol 248, 385–401.

Vajo, Z., and Duckworth, W.C. (2000). Genetically engineered insulin analogs: diabetes in the new millenium. Pharmacol Rev 52, 1–9.

Wan, Z.L., Huang, K., Hu, S.Q., Whittaker, J., and Weiss, M.A. (2008). The structure of a mutant insulin uncouples receptor binding from protein allostery. An electrostatic block to the TR transition. J Biol Chem 283, 21198–21210.

Wang, C.C., and Tsou, C.L. (1991). The insulin A and B-chains contain sufficient structural information to form the native molecule. Trends Biochem Sci 16, 279–281.

Wang, D.C.G., and Gu, X.C. (2010). A brief account on the study of the insulin crystal structure. Sin China Series C 53, 13–15.

Wang, J., Takeuchi, T., Tanaka, S., Kubo, S.K., Kayo, T., Lu, D., Takata, K., Koizumi, A., and Izumi, T. (1999). A mutation in the insulin 2 gene induces diabetes with severe pancreatic beta-cell dysfunction in the Mody mouse. J Clin Invest 103, 27–37.

Ward, C., Lawrence, M.C., Streltsov, V., Garrett, T., McKern, N., Lou, M.Z., Lovrecz, G., and Adams, T. (2008). Structural insights into ligand-induced activation of the insulin receptor. Acta Physiol (Oxf) 192, 3–9.

Waugh, D.F. (1941). The properties of protein fibers produced. Reversibly from soluble protein molecules. Am J Physiol 133, 484–P485.

Waugh, D.F. (1944). The linkage of corpuscular protein molecules. I. A fibrous modification on insulin. J Am Chem Soc 66, 663.

Waugh, D.F. (1946a). Fed Proc 5, 111.

Waugh, D.F. (1946b). A fibrous modification of insulin. I. The heat precipitate of insulin. J Am Chem Soc 68, 247–250.

Weiss, M.A., Hua, Q.X., Jia, W., Chu, Y.C., Wang, R.Y., and Katsoyannis, P.G. (2000). Hierarchical protein “un-design”: insulin’s intrachain disulfide bridge tethers a recognition alpha-helix. Biochemistry 39, 15429–15440.

Weiss, M.A., Hua, Q.X., Lynch, C.S., Frank, B.H., and Shoelson, S.E. (1991). Heteronuclear 2D NMR studies of an engineered insulin monomer: assignment and characterization of the receptor-binding surface by selective ^{2}H and ^{13}C labeling with application to protein design. Biochemistry 30, 7373–7389.

Wuthrich, K. (1986). NMR of Proteins and Nucleic Acids (New York, NY: Higher Education Press and Springer-Verlag).

Xu, B., Hu, S.Q., Chu, Y.C., Huang, K., Nakagawa, S.H., Whittaker, J., Katsoyannis, P.G., and Weiss, M.A. (2004a). Diabetes-associated mutations in insulin: consecutive residues in the B-chain contact distinct domains of the insulin receptor. Biochemistry 43, 8356–8372.

Xu, B., Hu, S.Q., Chu, Y.C., Wang, S., Wang, R.Y., Nakagawa, S.H., Katsoyannis, P.G., and Weiss, M.A. (2004b). Diabetes-associated mutations in insulin identify invariant receptor contacts. Diabetes 53, 1599–1602.

Xu, B., Huang, K., Chu, Y.C., Hu, S.Q., Nakagawa, S., Wang, S., Wang, R.Y., Whittaker, J., Katsoyannis, P.G., and Weiss, M.A. (2009). Decoding the cryptic active conformation of a protein by synthetic photoscanning: insulin inserts a detachable arm between receptor domains. J Biol Chem 284, 14597–14608.

Yan, H., Guo, Z.Y., Gong, X.W., Xi, D., and Feng, Y.M. (2003). A peptide model of insulin folding intermediate with one disulfide. Protein Sci 12, 768–775.

Yoshioka, M., Kayo, T., Ikeda, T., and Koizumi, A. (1997). A novel locus, Mody4, distal to D7Mit189 on chromosome 7 determines early-onset NIDDM in nonobese C57BL/6 (Akita) mutant mice. Diabetes 46, 887–894.

Zahn, H., and Schmidt, G. (1967). Synthesis of an insulin B-chain disulfide polymer. Tetrahedron Lett 50, 5095–5098.

Zeng, Z.H., Liu, Y.S., Jin, L., Zhang, Y., Havelund, S., Markussen, J., and Wang, D.C. (2000). Conformational correlation and coupled motion between residue A21 and B25 side chain observed in crystal structures of insulin mutants at position A21. Biochim Biophys Acta 1479, 225–236.

Zhang, Y.S. (2010). The first protein ever synthesized in vitro. Sin China Series C 53, 16–18.

