综述与专论

Insight Into Folding, Binding and Stability of Insulin by NMR

HUA Qing-Xin

(*Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio* 44106-4935, USA)

Abstract Insulin is one of the most important hormonal regulators of metabolism. Since the diabetes patients increase dramatically, the chemical properties, biological and physiological effects of insulin had been extensively studied. In last decade the development of NMR technique allowed us to determine the solution structures of insulin and its variety mutants in various conditions, so that the knowledge of folding, binding and stability of insulin in solution have been largely increased. The solution structure of insulin monomers is essentially identical to those of insulin monomers within the dimer and hexamer as determined by X-ray diffraction. The studies of insulin mutants at the putative residues for receptor binding explored the possible conformational change and fitting between insulin and its receptor. The systematical studies of disulfide paring coupled insulin folding intermediates revealed that in spite of the conformational variety of the intermediates, one structural feature is always remained: a "native-like B chain super-secondary structure", which consists of B9-B19 helix with adjoining B23-B26 segment folded back against the central segment of B chain, an internal cystine A20-B19 disulfide bridge and a short α-helix at C-terminal of A chain linked. The "super-secondary structure" might be the "folding nucleus" in insulin folding mechanism. Cystine A20-B19 is the most important one among three disulfides to stabilize the nascent polypeptide in early stage of the folding. The NMR structure of *C. elegans* insulin-like peptide resembles that of human insulin and the peptide interacts with human insulin receptor. Other members of insulin superfamily adopt the "insulin fold" mostly. The structural study of insulin-insulin receptor complex, that of *C. elegans* and other invertebrate insulin-like peptide, insulin fibril study and protein disulfide isomerase (PDI) assistant proinsulin folding study will be new topics in future to get insight into folding, binding, stability, evolution and fibrillation of insulin in detail.

Key words insulin, proinsulin, NMR, protein folding, insulin receptor binding and disulfide isomer

1 Introduction

1.1 Insulin and diabetes

Insulin is one of the most important hormonal regulators of metabolism. Insulin signals the fed state, stimulates the storage of fuels and the synthesis of proteins in a variety of ways. In short-term, insulin increases sugar transport and protein synthesis, inhibits lypolysis, activates enzymes (such as glycogen synthetase etc.), inactivates enzymes (such as phosphorylase etc.); and in long-term, insulin activates glucokinase and depresses controlling enzymes of gluconeogenesis. It acts on protein kinase cascades. Diabetes mellitus is an increasingly common disease, which occurs either because of lack of insulin or because of the presence of factors that oppose the action of the insulin. The result of insufficient action of insulin is an increase in blood glucose concentration (hyperglycaemia). Many other metabolic abnormalities occur, notably an increase in ketone bodies in the blood when there is a severe lack of insulin. Diabetes is the most common disease in the western countries and also increases quite fast in some developing country, such as in China.

New criteria for diagnosis of diabetes was proposed in 1997 by American Diabetes Association: (1) Symptoms of diabetes plus casual plasma glucose ≥11.1 mmol/L; (2) Fasting plasma glucose ≥7.0 mmol/L. Fasting is defined as no caloric intake for at least 8 h and (3) Two hours plasma glucose ≥11.1 mmol/L during oral glucose tolerance test using 75 g glucose load[1]. Followed after cancer and cardiac disease, diabetes now is the 3rd disease causing death of human being. Primary diabetes mellitus is of two types: insulin dependent diabetes (IDDM, type-Ⅰ diabetes) and non-insulin dependent diabetes

Brief bibliography of Dr. Hua: HUA Qing-Xin, Associate Professor of the Department of Biochemistry, Case Western Reserve University. Hua graduated from the Department of Biophysics, University of Science and Technology of China and then successively worked in the Institute of Biophysics, CAS (1963~1989); The National Institute of Medical Research, MRC, London (1985); The Institute of Biophysics and Molecular Biology, ETH-Honggerberg, Zurich (1987); The Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston (1989~1994); and The Department of Biochemistry and Molecular Biology, University of Chicago, Chicago (1994~1999). Hua's long-term research project is insight into folding, binding, evolution and stability of insulin and its analogues by use of Nuclear Magnetic Resonance (NMR) and other biophysical chemical methods. Hua published more than 40 research papers in world recognized leading periodicals, such as "Nature", "Nature: Structural Biology", "Proc Natl Acad Sci, USA", "Gene and Development" and "J Biol Chem", "J Mol Biol" etc. Hua also study transcriptional factors and chaperone, which is supported by NIH RO1 grant.

Tel: 216-368-4666, E-mail: qhua@biochemistry.cwru.edu Received: September 25, 2003 Accepted: October 25, 2003

(NIDDM, type-Ⅱ diabetes). IDDM is due to destruction of β-cell in the pancreatic islets with resulting loss of insulin production. Either genetic or environmental factors that trigger an autoimmune attack on the β-cell might be responsible. Insulin injection is a life-or-death issue for IDDM patients. There are numerous causes of NIDDM, which is now known to include a wide range of disorders with different progression and outlook. There are three major types of dominantly inherited NIDDM: MODY-1, MODY-2 and MODY-3 (MODY: maturity onset diabetes of young).

1.2 Diabetes patients increase dramatically

According to the criteria listed above, the global number of people with diabetes will rise from 150 million in year 2000 to 220 million in 2010, and 300 million in 2025[2]. Most cases will be of type Ⅱ diabetes due to recent sedentary lifestyle and obesity. Now 85 million people in Asia (exclude middle-Asia) is suffered from the diabetes and the patients will increase with as high as 57% to reach to more than 130 million in 2010. Even in North America and Europe, where the prevention, control and therapy of diabetes were well established, the diabetes patients will increase with rate of 24% to reach to 50 million in 2010. Among them diabetes in children and youth are now increasingly common. Within the past two decades diabetes patients were threefold increase in prevalence in certain areas of China[3]. The number of diabetes patients in China now reach to 40 millions. The World Health Organization predicts a doubling in diabetes's incidence in next two decades. Scientists now spend more efforts to study diabetes from its genetically programmed defects to insulin signaling, from mitochondrial function in diabetic β-cells to new drug targets for type Ⅱ diabetes. There are seven series papers of diabetes published in journal "Nature" (Insight) in December 2001[4].

1.3 History of insulin study and why does study of insulin continue?

Insulin is a small globular protein containing two chains (A chain, 21 residues and B chain, 30 residues). Two disulfide bridges (A7-B7 and A20-B19) link A- and B-chains, whereas the third disulfide (A6-A11) is intra-A-chain bond. Stored as a Zn^{2+}-stabilized hexamer in the human β-cell, the hormone functions as a Zn^{2+}-free monomer[5]. The history of insulin occupies a unique position in the evolution of understanding of protein and peptide hormones. The determination of the amino acid sequence of insulin by F. Sanger in 1953 is a landmark in biochemistry because it is the first study to show that a protein has a precisely defined amino acid sequence. Table 1 summarized a series of notable "firsts" in variety fields.

Table 1 History of insulin study

Year	Event	Investigator(s)
1921	Discovery and isolation of insulin	F. G. Banting[1] and C. Best
1955	Determination of primary sequence	F. Sanger[1]
1965	Organic synthesis of insulin	Shanghai research group
1967	Discovery of proinsulin	D. F. Steiner and R. Chance
1969	Determination of crystal structure	D. Crowfoot-Hodgkin[1]
1969	Determination of crystal structure	Beijing research group
1979	Cloning of insulin gene	W. Rutter and H. Goodman

[1] Received Nobel Prize.

A variety of dimeric and hexameric crystal structures of insulin have been determined. Structures of insulin in different crystal forms[6~12] provided a pioneering example of long-range conformational change in a protein as shown in Figure 1a. Among them, E. N. Baker *et al.*[6] described the insulin structure in monomer, dimmer and hexamer in detail, the paper can be called the "bible" of the insulin structure. Crystal structure of native insulin contains three helices (B9-B19, A2-A8 and A13-A19, designated to be helix-1, -2 and -3 respectively), one β-turn (B20-B23) and one β-strand (B24-B28). Two distinct crystallographic protomers are observed, designated T and R state. Residues B1-B8 exhibit variable secondary structure, either extended (T-state) or helical (R-state). The insulin receptor provided the first example of metabolic regulation *via* a receptor tyrosine kinase[13]. Despite these landmarks, several central questions remain to be answered: How does insulin bind to its receptor? What is the relationship between insulin structure, stability and function? What is the solution structure of insulin and its diabetes-mellitus mutants? Why are some residues important for receptor binding and the others don't? What is the structure of insulin folding intermediate, which coupled with disulfide pairing during the synthesis of nascent polypeptide? To answer these questions the study of insulin continues that thereby will fill a critical gap in our knowledge of insulin.

1.4 NMR is the unique technique to determine the three dimensional structure of macromolecule in solution

Determination of structure of macromolecules is a key to open the door for exploring the exquisite functions of the molecules. After more than one century's development, X-ray diffraction now is a mature method to determine the structure for single crystal at atomic level. Using X-ray diffraction one can determine large molecule with more precise detail, but lots molecules can't be crystallized, especially for the proteins with flexible segments. Nuclear magnetic resonance spectroscopy (NMR) is a powerful technique, which can determine the solution structure

of molecule at atomic level (Å). Now 3 172 (or ~ 16%) atomic coordinates of total 20 800 in the Protein Data Bank had been determined by NMR (http://www.rcsb.org/pdb/, May 2003). X-diffraction and NMR are two most important complementary methods to determine the structure of macromolecules either in crystal or in solution. NMR provides the knowledge of structures in aqueous solutions, which mimic the physiological conditions. It can be applied to those proteins that single crystals are difficult to be obtained, typically of molecules with disordered and highly flexible parts. NMR might be a unique technique, for which Nobel Prize awarded four times to NMR scientists due to their great contributions. In 1946 Purcell and Bloch discovered NMR as a physical phenomena and they won Nobel Prize in 1952. In 1991 Nobel Prize in Chemistry awarded to Swiss scientist Ernst for his contribution of development of multi-dimensional NMR in theoretically and practically. In 2002 Nobel Prize of chemistry awarded to Wuthrich in ETH, Switzerland for his contribution to use of NMR in structural genome. Now (2003) The Nobel Prize in Physiology and Medicine gave to Dr. Lauterbur at The University of Illinois at Urbana, US and Sir Mansfield at The University of Nottingham, UK for their discoveries concerning Magnetic Resonance Imaging (MRI). That is very rare for one technique shown such an importance and rapid development. NMR signals (chemical shifts, coupling constants etc.) contain information about the chemical environment of the nucleus studied. The nuclear overhauser enhancement effect (NOE) provides structural information about inter-atomic distances between nuclei close in space, which is proportion to the minus sixth power of the inter-proton distance, while the coupling constant is related to dihedral angle, so that NMR became an important technique to determine the conformation of bio-molecules in solution. Now NMR can be used for protein with molecular mass normally around 300 ku and sometimes up to 800 ku for GroEL[14] by use of transverse relaxation-optimized spectroscopy (TROSY) and cross-correlated relaxation-enhanced polarization transfer (CRINEPT)[15].

2 Insulin and its receptor binding

2.1 DPI and HI exhibit a similar structure of crystal insulin

Des-pentapeptide (B26-B30) insulin (DPI) is the first insulin analogue, whose structure was determined by NMR in solution. The difficulty to solve the solution structure of insulin is the complicated aggregation behavior. Insulin is easy to form dimer by anti-parallel β-sheet in between C-terminal of B-chain of two molecules. Lacking of five residues at C-terminal of B chain, DPI is a monomeric insulin analogue that gives nice dispersion of amide peaks among a variety insulin derivatives. Its receptor binding affinity retains ~20% relative to native insulin. The sequential resonance assignments were completely finished both in 20% deuteric acetic acid and in 10% DMSO at pH 2[16,17]; consequently, the assignments of native human insulin (HI) were also solved[18]. Figure 1a shows 14 crystal structures of T-state insulin at different crystal forms, which exhibit very similar "insulin fold" with slightly conformational differences, while NMR ensemble of 10 native insulin structures in Figure 1b show very similar fold with flexible terminals of B chain.

2.2 DKP- HI and Glu^{B16}-HI are engineered monomeric insulin

To circumvent the self-association of insulin in solution, which precluded 2D NMR study under physiological conditions, an engineered insulin monomer Asp^{B10}, Lys^{B28}, Pro^{B29}-insulin (designated DKP-HI) was designed. The substitution of B10 from His to Asp prevents the formation of trimmer, whereas the switch of Pro and Lys at position B28 and B29 destabilizes the dimer. DKP-insulin with three amino acid substitutions keeps receptor binding affinity as high as 160% ~ 200% relative to native insulin. At a protein concentration of 0.6 mmol/L and pH 7, DKP-HI is still monomeric according to analytical ultracentrifugation assay[19,20]. The NMR structures of DKP in Figure 1c (PDB entry: 1LNP) were calculated from NMR data in five different solution conditions: pH 6.8 and 25℃, pH 8.0 and 25℃, pH 8.0 and 32℃, pH 2 and 25℃, and 20% deuteric acetic acid and 25℃. These NMR data were complementary and no any contrary. The DKP substitutions were normally used as a monomeric template for additional substitution, so that these insulin analogues can be tested at physiological conditions. Kaarsholm *et al.* (Novo research group) used Glu^{B16} substitution to prevent the aggregation at pH 2.4[21]. Its NMR structure is shown in Figure 1d. Consequently they published the NMR structure of (B1, B10, B16, B27) Glu, des-B30 insulin (4E-insulin) at pH 6.5 and 34℃. This four Glu mutant retains only 47% biological potency of that of native insulin and it is used as monomeric template for their further mutagenesis[22]. The NOE restraints used for DKP-insulin is 749 including 555 inter-residue NOE, while that of Glu^{B16}, des-B30 insulin is 479 inter-residue NOE, which is the "best" structure from their group[21]. The average root mean square (rms) deviation of the 20 structures of DKP-insulin is 0.34 Å, whereas that for 25 calculated structures of Glu^{B16}, des-B30 insulin is 0.42 Å for backbone atoms (A2 ~ A19 and B4 ~ B26). Ignoring the conformational differences in detail, the NMR structure of native HI, DKP-insulin and Glu^{B16}-insulin shown in

Figure 1b ~ d is quite similar and also similar to the crystal structure in Figure 1a.

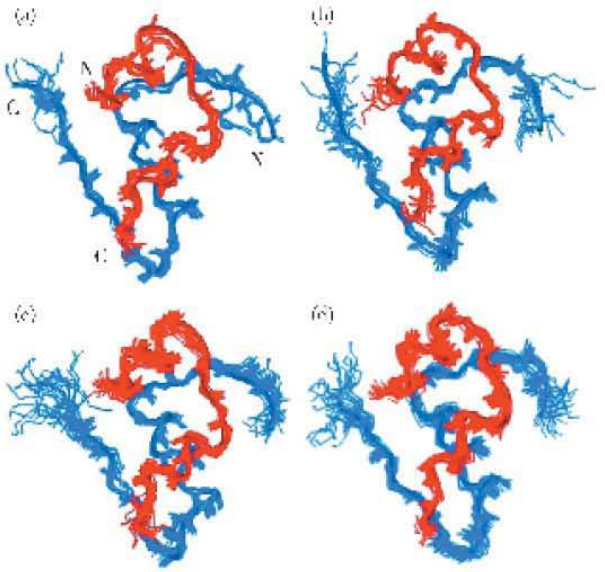

Fig. 1 14 crystal structures of T-state insulin from different crystal forms (a) and ensemble of 10 NMR structures of native insulin (b), and that of 20 Asp^{B10}, Lys^{B28}, Pro^{B29}-DKP-insulin (c) and that of 20 Glu^{B16}-insulin (d)

The A chain is shown in red, B chain in blue. Coordinates in (a) were obtained from the Protein Data Bank (identifiers 4INS, 1APH, 1BPH, 1CPH, 1DPH, 1TRZ, 1TYL, 1TYM, 1ZNI, 1LPH, 1G7A, 1EV6, and 1ZNJ). PDB ID in (b), (c) and (d) is 2HIU, 1LNP and 1HLS respectively.

2.3 Three insulin mutants with diabetes mellitus: Ser^{B24}-HI, Leu^{A3}-HI and Leu^{B25}-HI

There are three mutants human insulin associated with diabetes mellitus: Ser^{B24}-HI (insulin Los Angeles, where patient was found[23]), Leu^{A3}-HI (insulin Wakayama[24]) and Leu^{B25}-HI (insulin Chicago[25]). What are the conformations of these mutated insulins? The NMR study of Ser^{B24}-HI exhibits the paradoxical structure and function[26]. Ser^{B24}-HI retains the conformation of native insulin mainly, nevertheless its biological activity is reduced by ~ 15-fold. Among vertebrate insulin Phe^{B24} is invariant. Most substitutions of PheB24 lead to inactive species, but glycine or D-amino acids are well tolerated in this position. The B24 aromatic ring appears to anchor other putative receptor-binding residues either in crystal structure of insulin or in NMR structure of insulin and DKP-insulin. From conventional 2D NMR and also the edited-NOESY spectrum of selected ^{13}C labeled Ser^{B24}-HI, the NOE between Tyr^{B26} and Val^{B12} or Leu^{B15} are restored as in native insulin. It seems that the substitution from Phe to Ser at B24 position destroy the local binding environment between insulin and the receptor. NMR study of Leu^{A3}-HI and Leu^{B25}-HI also exhibit the similar structural characters: retaining the major conformation of native insulin (un-published data), but losing binding-affinities dramatically. In native insulin, A3 is Val, and B25 is Phe. Gly substitution at A3 position causes the binding affinity decreased to 0.15%, while Ser or Leu substitution at B25 causes the binding affinity dropping down to 0.13%. Presumably the corresponding site of A3, B24 and B25 of receptor require the critical conformational fitting. The side-chains of these mutants prevent or break such precise fitting.

2.4 Gly^{B24}-HI: a structural switch exposes key residues for receptor binding

What is the "active" conformation of insulin when it binds to receptor? The best map to get insight into the "active" insulin structure is to solve the complex structure of the receptor with its ligand, insulin. But to date neither the co-crystal structure, nor NMR structure of such complex has been solved due to the large size and complicity of the cross-membrane receptor. Derewenda *et al*, reported the crystal structure of B29-A1 covalent linked single chain insulin without receptor potency[8,27]. Surprisingly the structure of such mini-proinsulin is nearly identical to that of native T-state insulin as shown in Figure 2a, which imply that C-terminal of B chain "closed" against the central helix of B chain represents the "inactive" conformation. This observation allows us to propose the hypothesis that the conformational change at C-terminal of B chain has to occur on receptor binding[27,28]. The NMR structure of Gly^{B24}-insulin provided the first experimental evidence of this hypothesis[29,30]. The Phe to Gly substitution at position B24 is tolerant for receptor binding, it retains as high as 80% of that of native insulin. Its structure is "opened" at C-terminal of B chain as shown in Figure 2b, which indicates that the "closed" B chain in native state is "inactive", whereas the "opened" B chain may

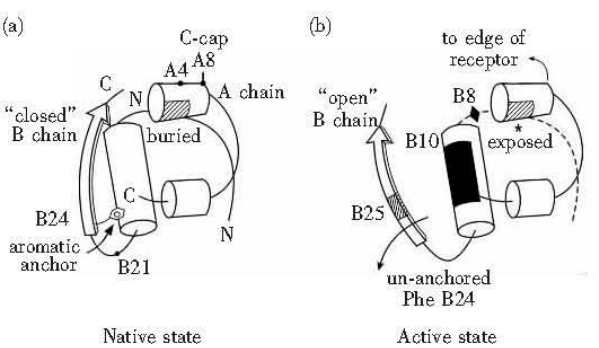

Fig. 2 Cylinder representations of insulin

(a) "Closed" model of the insulin in solution. (b) proposed model of active "open" conformation shows partial detachment of the C-terminal β-strand of B-chain exposes the side chains of Ile^{A2} and Val^{A3} (shaded; asterisk in B) in a recognition α-helix.

represent an active state. From ^{1}H NOESY spectrum and ^{13}C-edited NOESY of the selected ^{13}C labeled Gly^{B24}-insulin the characteristic long-range contacts for "closed" conformation, such as NOE between B26 aromatic to B12 and B15 methyl protons in native insulin, do not retain in the analogue. Or in other words, there is a structural switch at B24 whereby the C-terminal of B chain moves away on receptor binding to expose or partially expose underlying A2-A3 residues. Ludvigsen *et al.* [31] agreed above conclusion from their NMR study of Glu^{B16}, Gly^{B24}, des^{B30}-insulin mutant in 1998, but disagreed the extent of switching. Their structure shows rearrangement of C-terminal of B chain, which allows the Phe^{B25} side chain orient inward to occupy the position normally taken up by Phe^{B24} in native insulin. Their restraints for structural calculation listed more than ten long-range contacts between residue B25 and B26 to B12, B15, B17, and B28 etc., which were not found in Gly^{B24}-insulin. The arguments arise from the different solvent used in two groups from their point of view. NMR study of Gly^{B24}-insulin was only in 20% deuteric acetic acid in 1991, which may induce extra flexibility[31]. To address whether or not their conclusion is right, recently new mutant Glu^{B16}, Gly^{B24}-insulin was prepared and NMR data were acquired by our group in aqueous solution at pH 8.0 and 32℃ as Ludvigsen *et al* used and also in 20% acetic acid and 25℃. The NMR data were collected from Varian 600 MHz and Bruker 800 MHz instrument. Most long-range contacts between residue B25 and B26 to B12, B15, B17, and B28 etc. that had been found from Novo NMR group were not observed in Glu^{B16}, Gly^{B24}-insulin except the NOE between B26 aromatic proton and B12 methyl group at 200 ms mixing NOESY spectrum only (un-published data). We also found that there were major and minor conformers for B25 and B26 aromatic resonance. The long-range NOE between B26 and B12 was actually from minor conformer of the aromatic B26. Co-solvent 20% acetic acid may affect the flexibility of the segment out of the core of the molecule somehow, but not in very serious. Actually new crystal structure of DPI-insulin in 20% acetic acid was solved by G. Dodson recently (personal communication). The effect of 20% acetic acid on long-range NOE is just "more or less", instead of "present or absent" based on back-calculated NOESY spectra.

The calculated structure of Glu^{B16}, Gly^{B24}-insulin (Figure 3b) is more like that of Gly^{B24}-insulin[27], and unlike that of Glu^{B16}, Gly^{B24}, Des^{B30}-insulin (Figure 3c). There might be two possible reasons to form the transient NOE: from partial dimer or from segmental movement of B23-B30 (detached C-terminal of B chain represents the "open" state, while attached is "close" state). The C-terminal of B chain for Gly^{B24} substituted mutant is still quite flexible: the long-range NOE of B25 are recorded when it moved close to central segment of B chain, while it disappeared when it moved away in most time. The residual dimmer contacts also make the over-estimated long-range NOE. In Gly^{B24}-DKP at pH 8.0 and pH 2.0 only NOE between aromatic proton of B25 (minor conformer) and methyl of $B12^{Val}$ was found, implying a similar possibility.

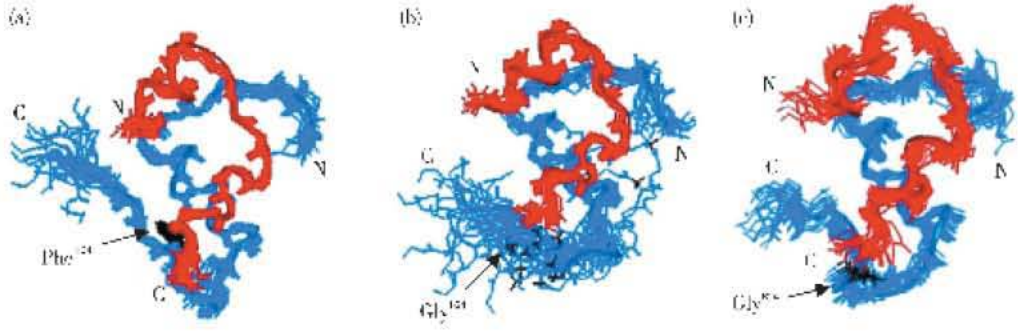

Fig. 3 NMR ensemble of DKP-insulin(a), Glu^{B16}, Gly^{B24}-HI (b) and Glu^{B16}, Gly^{B24}, DesB30-HI (c)
A chain is shown in red, B chain in blue and B24 residue in black. Structures are aligned according to A2-A7, A13-A19 and B9-B19. PDB ID of Glu^{B16}, Gly^{B24}, DesB30-HI is 1A7F.

2.5 "Super-active" insulin mutant: D-Ala^{B24}-DKP

As described above, D-amino acid replacement of B24 does not decrease the potency. Actually D-Ala^{B24}-DKP has binding affinity 255% relative to that of native insulin, it is also higher than that of its parent template DKP (200%), whereas L-Ala^{B24}-DKP shows only 4.3% relative to native insulin (Table 2). Circular dichroism (CD) is a simple technique to estimate the secondary structure and other thermo-dynamic properties in solution. The CD wavelength scan of D- and L-Ala substitution mutants at 4℃ and 25℃ indicate quite similar well ordered secondary structure as the parent

DKP-insulin does. Thermodynamic stabilities were measured by CD guanidine hydrochloride titration; unfolding was monitored as a function of guanidine concentration. The denaturant unfolding by 8mol/L guanidine of these samples shows very similar curves (see Figure 4a and 4b), indicating nearly same stability. The ΔG_u was calculated according to Sosnick *et al*[32]. But the NOESY spectra in region of aromatic protons and long-range methyl protons are very different as shown in Figure 4c and 4d.

Table 2 Properties of B24 mutants of insulin

Sample	ΔG_u	$\Delta\Delta G_u$	C_{mid}	m	Potencies
Human insulin	4.4 ± 0.1	—	5.3 ± 0.1	0.84 ± 0.01	100
DKP-insulin	4.9 ± 0.04	—	6.1 ± 0.1	0.84 ± 0.01	208
L-Ala^{B24}-DKP	4.3 ± 0.05	−0.6	6.1 ± 0.1	0.71 ± 0.01	4.3
D-Ala^{B24}-DKP	4.8 ± 0.04	−0.1	6.1 ± 0.1	0.78 ± 0.01	255
Gly^{B24}-DKP	4.3 ± 0.05	−0.6	6.6 ± 0.1	0.65 ± 0.01	50
Glu^{B16}, Gly^{B24}-HI	3.2 ± 0.04	−1.2	5.2 ± 0.1	0.61 ± 0.01	14.9

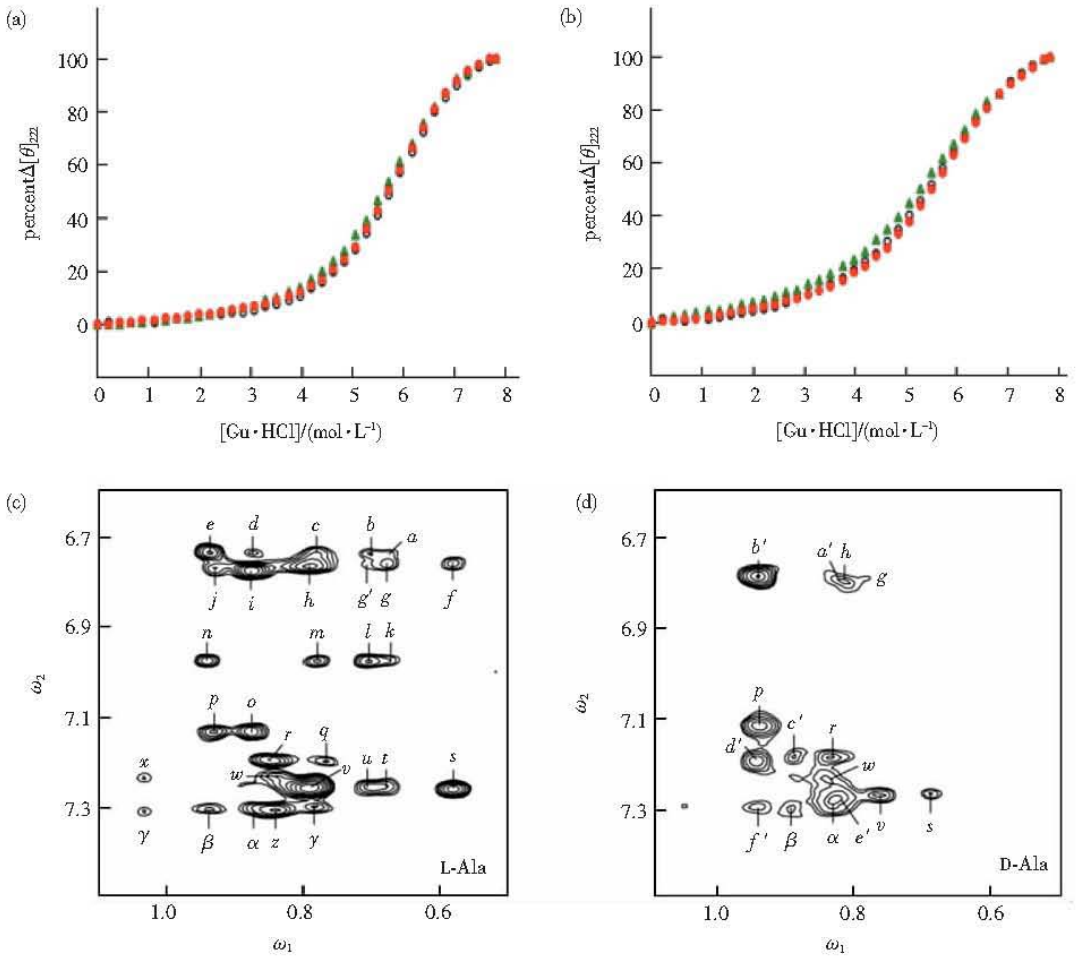

Fig. 4

(a), (b) Guanidine unfolding CD curves of D-Ala^{B24}-DKP-insulin (open circles), L-Ala^{B24}-DKP-insulin (green triangles), and DKP-insulin (red circles) at 4℃ (a) and 25℃ (b). (c), (d) 2D 600 MHz NOESY spectra of D-Ala^{B24}-DKP-insulin (c) and L-Ala^{B24}-DKP-insulin (d) at pH 7.6 and 32℃ showing inter-residue NOEs between aromatic protons (vertical axis) and aliphatic protons (horizontal axis). (c) NOE contacts for L-Ala^{B24}-DKP-insulin: (*a*) B26-H_εA2-$H_{\gamma 1}$; (*b*) B26-H_ε/B15-$H_{\delta 1,2}$; (*c*) B26-H_ε/B15-$H_{\delta 1,2}$; (*d*) B26-H/B12-$H_{\gamma 2}$; (*e*) B26-H_ε/A2-$H^{\gamma 1}$; (*f*) A19-H_ε/A2-H_δ; (*g*) A19-H_ε/A2-$H_{\gamma'}$; (*g'*) A19-H_ε/B15-$H_{\delta 2}$; (*h*) A19-H_ε/B15-$H_{\delta ?}$-A16-H_δ; (*i*) B16-H_ε/B18-$H_{\gamma 2}$; (*j*) B16-H_ε/B12-$H_{\gamma 1}$; (*k*) B26-H_δ/A2-$H_{\gamma'}$; (*l*) B26-H_δ/B15-$H_{\delta 2}$; (*m*) B26-H_δ/B15-$H_{\delta ?}$; (*n*) B26-H_δ/B12-$H_{\gamma 1}$; (*o*) B16-H_δ/B18-$H_{\gamma 2}$; (*p*) B16-H_δ/B17-H_δ; (*q*) B1-H_δ/B6-H_δ; (*r*) B1-H_δ/A13-$H_{\delta 2}$; (*s*) A19-H_δ/A2-H_δ; (*t*) A19-H_δ/A2-$H_{\gamma'}$; (*u*) A19-H_δ/B15-$H_{\delta 2}$; (*v*) A19-H_δ/B15-$H_{\delta ?}$-A16-H_δ; (*w*) B1-H_ξA13-$H_{\delta 2}$; (*x*) B1-H_δ/B18-H_γ; (*y*) B1-H_ε/B6-$H_{\delta 2}$; (*z*) B1-H_ε/B6-$H_{\delta ?}$; (*α*) B1-H_ε/A13-$H_{\delta 2}$; (*β*) B1-H_ε/A13-$H_{\delta ?}$; and (*γ*) B1-H_ε/B18-H_γ. (d) NOE contacts for D-Ala^{B24}-DKP-insulin: (*g*) A19-H_ε/A2-$H_{\gamma 1}$; (*h*) A19-H_ε/A19-H_δ; (*a'*) B16-H_ε/B11-H_δ; (*b'*) B16-H_ε/B17-H_δ; (*p*) B16-H_δ/B17-H_δ; (*r*) B1-H_δ/A13-$H_{\delta 2}$; (*c'*) B1-H_δ/A13-$H_{\delta ?}$; (*d'*) B25-H_ε/B12-$H_{\gamma 2}$; (*w*) B1-H_ε/A13-$H_{\delta 2}$; (*s*) A19-H_δ/A2-H_δ; (*v*) A19-H_δ/B15-$H_{\delta ?}$; (*e'*) A19-H_δ/A16-H_δ; (*α*) B1-H_ε/A13-$H_{\delta 2}$; (*β*) B1-H_ε/A13-$H_{\delta ?}$; and (*f'*) B1-H_ε/B6-$H_{\beta 2}$.

CD data were obtained at 4℃. Guanidine denaturation data were fitted by nonlinear least squares to a two-state model. In brief, CD data $\theta(x)$, where x indicates the concentration of denaturant, were fitted according to Equation 1:

$$\theta(x) = \frac{\theta_A + \theta_B \exp[(-\Delta G^\circ_{H_2O} - mx)/RT]}{1 + \exp[(-\Delta G^\circ_{H_2O} - mx)/RT]} \quad (1)$$

where x is the concentration of guanidine hydrochloride and where θ_A and θ_B are baseline values in the native and unfolded states. These baselines were approximated by pre- and post-transition lines $\theta_A(x) = \theta_A^{H_2O} + m_A x$ and $\theta_b(x) = \theta_B^{H_2O} + m_B x$. Fitting the original CD data and baselines simultaneously circumvents artifacts associated with linear plots of ΔG as a function of denaturant according to $\Delta G^\circ - (x) = \Delta G^\circ_{H_2O} + m^\circ x$. C_{mid} is defined as that molar concentration of guanidine HCl associated with 50% protein unfolding. The m value (kcal/mol/M) is defined as the slope in plotting the unfolding free energy versus molar concentration of denaturant; this slope is often found to be proportional to the protein surface area exposed on unfolding.

While the long-range contacts in L-Ala mutant are nearly identical to those of native insulin ("closed" conformation, Figure 5b), the number of NOEs in D-Ala mutant are much less, which is similar to Gly B24 substitution. Only NOEs from B25 to B12 and B15 exist in D-Ala^{B24}-DKP implying the "opened" conformation (Figure 5c). The NMR data were collected from three conditions: aqueous solution at pH 8.0 and 32℃ as S. Ludvigsen *et al.* used, at pH 2.0 and 25℃ and also in 20% acetic acid and 25℃. Three NOESY spectra of D-Ala^{B24}-DKP are nearly identical. Comparing Figure 3 with Figure 5 the structure of D-Ala^{B24}-DKP in Figure 5c is really similar to that of Glu^{B16}, Gly^{B24}-insulin in Figure 3b, both of which adopt the "open" state.

2.6 Receptor binding related insulin residues

The putative residues for receptor binding has been extensively studied since the report of first "classical" binding surface of insulin[33]. Alanine scanning mutagenesis and organic chemistry synthesis are both powerful tools to explore the insulin residue for its receptor binding[34~36]. In general, the following invariant and conserved residues of insulin are involved in receptor binding: four of these invariant residues (Ile^{A2}, Val^{A3}, Tyr^{A19} and Phe^{B24}) interact directly with the receptor and four additionally conserved residues (Asn^{A21}, Leu^{B6}, Leu^{B11} and Val^{B12}) and one less conservative residue (Phe^{B25}) are important in maintaining the receptor-binding conformation. There are additional residues considered to be important in receptor binding based upon the crystal structure of insulin (Glu^{A4}, Gln^{A5}, Thr^{A8}, Leu^{B15}, Tyr^{B16} and Tyr^{B26}), most of which are less well conserved.

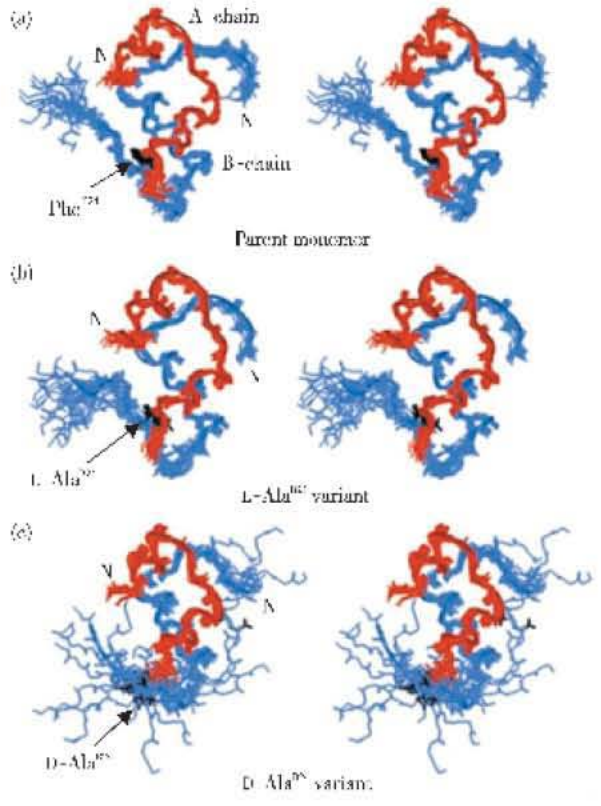

Fig. 5 Stereo views of NMR ensemble of 20 structures of DKP-insulin (a), L-Ala^{B24}-DKP-insulin (b), and D-Ala^{B24}-DKP-insulin (c)

The A chain is shown in red and B chain in blue. Phe^{B24} and its stereospecific substitutions (L-Ala^{B24}, panel b and D-Ala^{B24}, panel c) are shown in black. Structures in panels (a) and (b) are aligned according to the backbone of A2-A7, A13-A19, and B9-B24. Structures in panel (c) are aligned according to the backbone of A2-A7, A13-A19, and B9-B19.

Recently the CD and NMR studies of two mutants at A16 and A2/A3 position[37,38] are quite interesting: A16 analogues exhibits good receptor binding potency, perturbed structure, stability but bad yield, while A2/A3 Gly analogues exhibits good yield, native structure with bad potency. The hydrophobic substitutions (Ile, Val and Phe) of a key internal side chain Leu^{A16} retained moderate receptor-binding activities from 30% ~60% (Table 3). But the recombination yields of these mutants were too low to obtain enough material for NMR 2D experiments. The denaturant stabilities of these A16 mutants were lower than that of the parent DKP-insulin from 1.4 to 3.4 kcal/mol (Table 3). These observations suggest that the A16 side chain provides an essential structural buttress but unlike neighboring core side chains, does not itself contact the receptor[37]. Whereas NMR spectra of Gly^{A2,A3}-DKP exhibit very nice dispersion peaks and demonstrates

local unfolding of the A1-A5 segment in an otherwise native-like structure[38]. The RMSD of calculated structures were only 0.52, 0.97 and 0.60 Å for backbone atoms of α-helix, A chain and B chain respectively[38]. Its receptor binding activity is only 0.04% relative to that of insulin. Table 4 listed the yields and CD results of Gly substitution at N-terminal of A chain. The yield of $Gly^{A2, A3}$-DKP (3G-DKP in Table 4, due to Gly from A1 to A3) is 86% and 78% relative to that of native insulin at neutral and basic pH respectively. Basic pH was used due to insulin chain combination normally carried out at basic pH. More Gly substitution at A4 and A5 shows similar results: good yields, less stable than native insulin and essentially without biological activity (Table 4 and Figure 6).

Table 3 Properties of insulin mutants at position A16

Analogue	Activity	ΔG_u	$\Delta\Delta G_u$	C_{mid}/M	m/(kcal·mol^{-1}·M^{-1})
(a) Engineered monomer					
human insulin	100	4.4 ± 0.1		5.3 ± 0.1	0.84 ± 0.01
[Lys^{B20}, Pro^{B29}] HI	100	3.5 ± 0.1	−0.9 ± 0.2	4.9 ± 0.1	0.71 ± 0.01
Asp^{B10}-HI	ND	5.0 ± 0.1	0.6 ± 0.2	6.2 ± 0.1	0.80 ± 0.01
DKP-insulin	161 ± 19	4.9 ± 0.1		5.8 ± 0.1	0.84 ± 0.01
(b) A16 analogues					
Ile^{A16}-DKP-insulin	64 ± 19	3.5 ± 0.1	−1.4 ± 0.2	5.4 ± 0.1	0.66 ± 0.01
Val^{A16}-DKP-insulin	33 ± 8	2.1 ± 0.1	−2.8 ± 0.2	3.7 ± 0.1	0.57 ± 0.01
Phe^{A16}-DKP-insulin	34 ± 11	1.5 ± 0.1	−3.4 ± 0.2	3.1 ± 0.1	0.50 ± 0.01
(c) control analogues					
Asp^{B10}-DOI	ND	4.1 ± 0.1	−0.3 ± 0.2	6.4 ± 0.1	0.65 ± 0.01
allo-Ile^{A1}-DKP-insulin	4 ± 0.01	4.9 ± 0.1	0.0 ± 0.2	6.5 ± 0.2	0.75 ± 0.02

DOI: des-octapeptide (B23-B30) insulin.

Table 4 Properties of insulin with substituted Gly at N-terminal of A chain

Protein	Yield/%[1]	ΔG_u[2]	$\Delta\Delta G_u$[3]	C_{mid}/M[4]	m/kcal·mol^{-1}·M^{-1}[5]
(a) Neutral pH					
human insulin	100	4.4 ± 0.1	—	5.3 ± 0.1	0.84 ± 0.01
DKP-insulin	110	4.9 ± 0.1	—	5.8 ± 0.1	0.84 ± 0.01
Gly^{A2}-DKP-insulin	106				
Gly^{A3}-DKP-insulin (2G)	88	2.9 ± 0.2	−2.0 ± 0.2	5.0 ± 0.1	0.57 ± 0.07
3G-DKP-insulin	86	2.8 ± 0.1	−2.1 ± 0.2	4.4 ± 0.1	0.64 ± 0.01
4G-DKP-insulin	107	2.4 ± 0.1	−2.5 ± 0.2	4.0 ± 0.1	0.61 ± 0.01
5G-DKP-insulin	101	2.7 ± 0.1	−2.2 ± 0.2	4.2 ± 0.1	0.63 ± 0.01
5G-insulin	50	2.0 ± 0.1	−2.4 ± 0.2	4.2 ± 0.1	0.60 ± 0.01
DKP-des-[A6-A11]^{Ser} [6]	90	< 2.2	> −3.0 ± 0.4	< 3.2	0.65 ± 0.07
DKP-des-[A7-B7]^{Ser}	33	< 1.0	> −4.4 ± 0.2	< 1.5	0.46 ± 0.03
(b) Basic pH					
human insulin	100	3.2 ± 0.1	—	4.3 ± 0.1	0.75 ± 0.02
DKP-insulin	110	3.9 ± 0.1	—	5.3 ± 0.1	0.73 ± 0.02
3G-DKP-insulin	78	3.4 ± 0.1	−0.5 ± 0.1	4.6 ± 0.1	0.74 ± 0.02
4G-DKP-insulin	96	3.5 ± 0.1	−0.4 ± 0.1	4.6 ± 0.1	0.76 ± 0.02
5G-DKP-insulin	101	3.2 ± 0.1	−0.7 ± 0.1	3.5 ± 0.1	0.70 ± 0.02
5G-insulin	103	2.5 ± 0.1	−0.7 ± 0.2	3.6 ± 0.1	0.58 ± 0.02

[1] Yield is defined following purification of analogs by CM-cellulose chromatography. In control synthesis of DKP-insulin a mixture of 60 mg A-chain-S-sulfonate and 30 mg DKP B-chain-S-sulfonate yields 12 mg of monocomponent DKP-insulin. ΔG_u[2], C_{mid}[4] and m[5] value are defined in Table 2. $\Delta\Delta G_u$[3] (kcal/mol) indicates difference in ΔG_u values relative to parent structure (DKP-insulin or human insulin). Uncertainties in two-state fitting parameters do not include possible systematic error due to non-two-state behavior. [6] Lack of sigmoidicity in transitions makes uncertain fitting of pre-transition base line; only upper bounds to stability have been estimated.

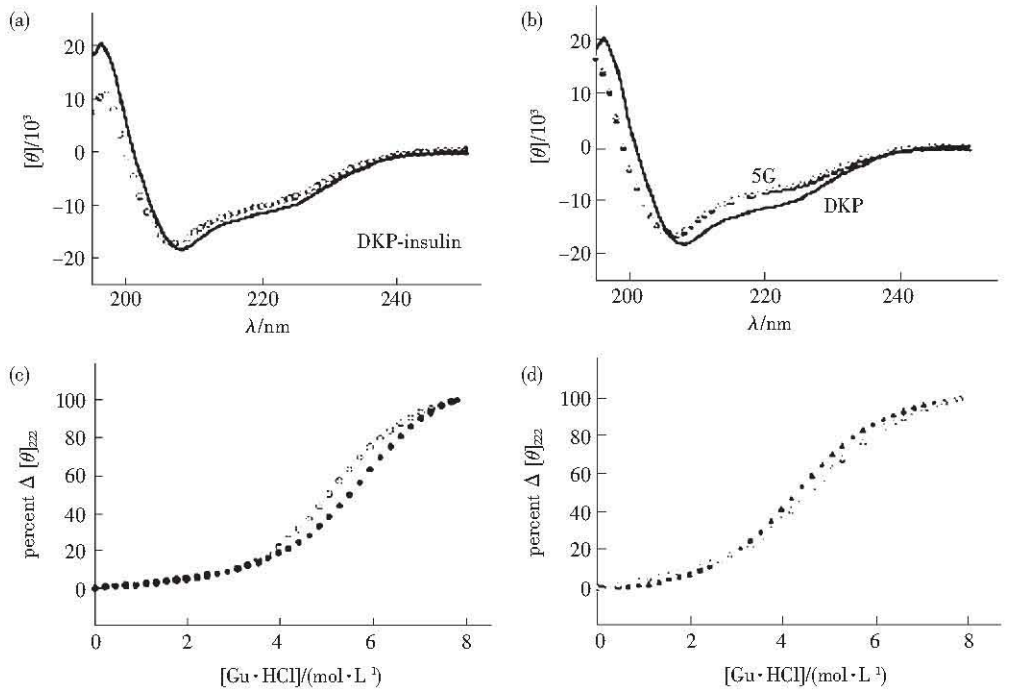

Fig. 6 Effect of pH on CD-detected structure and stability

(a) Far-UV CD spectra of DKP-insulin at pH 7.4 (line) and 10.5 (open circle) at 4℃. (b) Spectra of 5G-DKP-insulin at pH 7.4 (filled triangle) and 10.5 (open triangle) relative to DKP-insulin at pH 7.4 (line) at 4℃. (c) Guanidine denaturation studies of DKP-insulin at pH 7.4 (filled circle) and 10.5 (open circle) at 4℃. (d) Guanidine denaturation studies of 5G-DKP-insulin at pH 7.4 (filled triangle) and 10.5 (open triangle) at 4℃. Inferred thermodynamic parameters are given in **Table 4.**

Above mutants retain partial folds with efficient disulfide pairing, which gave rise to the relatively high yields, suggesting that the native N-terminal α-helix does not participate in the transition state of the reaction of chain recombination, or in other words, the role of A chain α-helix in disulfide pairing is asymmetric. As we know, lots insulin mutants were synthesized by combination of isolated mutated A and B chains. This reaction, designated insulin chain combination, yields native disulfide pairing[39]. So this study also provides the information for mechanism of insulin chain combination, which is less understood during the long period since the method applied in 60′s.

2.7 Does mutant with enhanced stability contain enhanced potency?

Kaarsholm *et al.*[40,41] reported the potency and CD results of A8 substituted analogues that the enhanced potency has been found to couple with enhanced stability. The potency of Ala, Arg and His substitution at A8 is 112%, 302% and 309% respectively, whereas the stability of these mutants is 3.92, 5.01 and 5.52 kcal/mol comparing with 3.76 kcal/mol of native insulin (Table 5).

Table 5 CD results of A8 analogues of insulin in 10 mmol/L Tris at 25℃

Analog	Potency[1)]	G_u/(kcal·mol^{-1})	C_{mid}/M	m/(kcal·mol^{-1}·M^{-1})
Human insulin[2)]	100	3.76	4.37	0.86
HisA8 human	309	5.52	5.41	1.02
AlaA8 human	112	3.92	4.51	0.87
ArgA8 human	302	5.01	4.77	1.07
DKP-insulin	161[3)]	4.80	5.80	0.84
GluA8-DKP-insulin	61[3)]	5.90	6.60	0.89
GlnA8-DKP-insulin	122[3)]	5.00	6.40	0.78

[1)] Potency was measured by insulin-dependent stimulation of glucose uptake in isolated adipocytes. [2)] CD studies of the unfolding of human insulin represent measurements at a protein concentration of 3 μmol/L. [3)] Placement receptor binding was determined by relative affinity for the placental insulin receptor; number of assays is given in parenthesis with standard deviations provided. Under these conditions the K_d for native insulin is (0.48 ± 0.06) nmol/L. Control studies were also performed using His^{A8}-human insulin (positive control) and Ala^{A2}-DKP-insulin (negative control) as described in text.

In comparison with the template structure (4E-insulin), the His^{A8}-4E-insulin mutation enhances the helical character of residues A2-A8. The A8 site defines one edge of the receptor binding surface and delimits the A1-A8 α-helix as shown in Figure 2a. The enhanced activities of the His^{A8} and Arg^{A8} analogs (relative to Thr^{A8}) were ascribed to thermodynamic stabilization of this critical helix because of their higher helical propensities, including more favorable C-cap properties. Additional stability may derive from an electrostatic interaction between the A8 side chain and the negative charge side chain of Glu^{A4} (a potential C-capping box[41]). Whether a correlation between stability and activity is general or limited to these particular analogs is not known. To test their proposal, more mutations at A8 (Glu^{A8}-DKP-insulin and Gln^{A8}-DKP-insulin) were prepared and studied[42]. CD and NMR spectra of the DKP analogs were essentially identical to those of DKP-insulin, indicating a correspondence of structures. Receptor binding affinities were determined by competitive displacement of ^{125}I-insulin from human placental membranes. The results in Table 5 show that the analogs receptor binding affinities are uncorrelated with stability, which suggest that the receptor binding affinities of A8 analogs reflect local features of the hormone-receptor interface rather than the stability of the free hormone or the intrinsic C-capping propensity of the A8 side chain. Contrary to a previous proposal[41], the activity and stability of insulin are uncorrelated among a series of A chain analogs. The substitutions affect the C-terminal residue of A chain and putative electrostatic capping box of a conserved recognition α-helix.

2.8 IR structure and activation mechanism

As discussed above, despite of the central role of post-receptor signaling pathway in the hormone control of the metabolism, how insulin binds to and triggers its receptor are poorly understood. How does insulin binds to its receptor? What are the contact points between insulin and the receptor, and how is specific recognition achieved? How are these structural features coupled to signal transduction within the hormone-holoreceptor complex[43,44]? These questions have long been the subject of speculation.

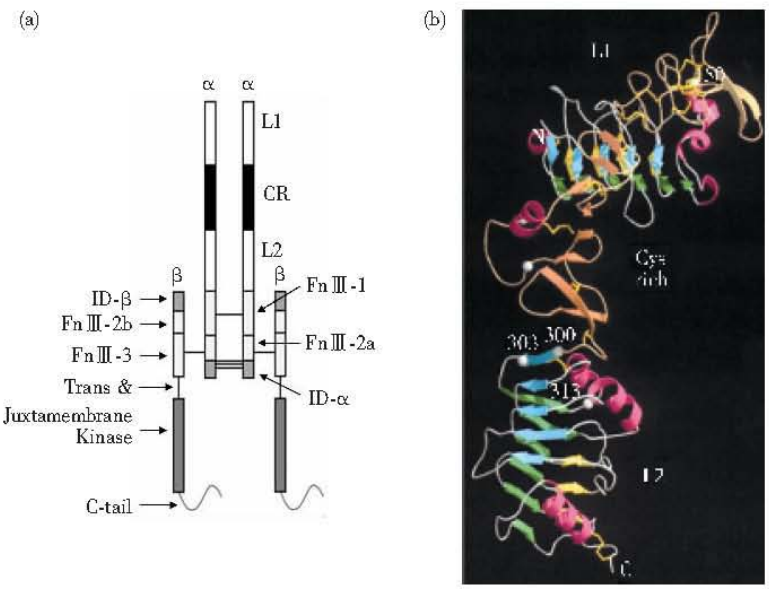

Fig. 7

(a) Schematic outline of α2β2 insulin receptor. L1 and L2: ligand binding domains, CR: Cystein-rich domain, and FnⅢ: fibronection domains. (b) Crystal structure of IGFR L1-CR-L2 domains. Polypeptide fold of hIGF-Ⅰ R (residues 1 ~459) adapted from Ref[45]. Helices are represented by curled ribbons and β strands by broad arrows.

Figure 7a shows the schematic outline of α2β2 insulin receptor[45~48], which contains large domain L1 and L2 involving in binding of insulin, cysteine-rich domain (CR) and Fn Ⅲ fibronection domains[48]. The insulin receptor (IR) is a member of a family of α2β2 receptor tyrosine kinases (TK). The α chain is extracellular and it contains the hormone-binding site(s). The β chain passes through the membrane and contains the intracellular TK domain. The crystal structure of insulin receptor is not solved yet; however, that of the first three domains of its homologue, insulin-like growth factor receptor, was solved (PDB ID entry: 1IGR)[45] and shown in Figure 7b. The L domains each contain a novel single-stranded right-handed β-helix capped at each end by α-helices and disulfide bridges. IR has an additional pair of cysteines relative

to IGFR. An excellent review paper of insulin and IGF-Ⅰ receptors summarized the detailed progreses[49]. According to the L1-CR-L2 structure of homologue receptor, IGF-Ⅰ receptor, and electron microscopic reconstructions of the IR undertake the isolated IR ectodomains with and without bound insulin[50]. In addition of the major (or classical) receptor-binding surface as mentioned above, the minor accessory surface is proposed to contain Leu^{A13} and Leu^{B17}[51,52]. But what is the binding surface of IR? The photo-cross-linking and alanine scanning had been used to explore the surface. Mapping of the *para*-azido-Phe-insulin with IR by Steiner[53] identified a novel B25 binding site near the C-terminus of the α subunit of IR. The alanine scanning mutagenesis[54,55] also confirmed the importance of this region (residues 704 ~ 719). The IR binding surface and the structure of bound insulin will be explored by the possible structural determination of the complex of minimum receptor binding domain (L1-CR-L2) with insulin by crystal diffraction or TROSY-based NMR.

3 Disulfide pairing coupled insulin folding intermediates

3.1 Disulfide-coupled protein folding pathway

Protein folding study has long been studied in last three decades since Anfinsen′s famous assumption proposed that native structure of protein is uniquely determined by its primary amino acid sequence[56]. Now the folding study is one of the most fascinating topics not only in theoretical research, but also in medicine and in pharmaceutical industry. The central task to study protein folding is to get insight into the folding pathway and the kinetically accumulated folding intermediate, and lead to understand the folding mechanism. Although the populations of possible folding conformations are numerous, the kinetics of folding are relatively simple indicating that many molecules adopt the similar conformation of intermediates, following the same folding pathway, and that relatively few conformations are sampled during the timescale of folding. The folding pathway of disulfide bond containing protein can be determined in detail because the disulfide bonds can be used as probes of protein folding[57]. In particular, the kinetic roles and importance of most disulfide intermediates can be determined. The folding pathway of the following proteins had been studied and compared extensively. In BPTI a crucial intermediate is partly folded, in α-lactalbumin the intermediate tend to adopt to the molten globule conformation, while in ribonuclease the early disulfide intermediates are largely unfolded, and non predominate. However, the slowest step in each case is the formation of a disulfide bond to be buried in a stable folded conformation, whereas the most rapid step is the formation of an accessible disulfide bond on the surface of the folded state. The most extensively characterized disulfide folding transition is that of BPTI, a small protein with three disulfides (30-51, 38-55 and 5-14), which is similar to insulin. Its quasi-native species with native conformation is (30-51, 38-55) intermediate, but no 5-14 disulfide bond. Residue Cys5 and Cys14 are completely exposed in solvent, and quickly to form a disulfide bond without any conformational rearrangement[57].

What is the folding pathway of insulin? Does the "quasi-native" intermediate of insulin adopt the similar conformation just like BPTI does? What is the structure of insulin disulfide pairing-coupled folding intermediate? Because insulin structure was solved well in solution and three disulfide bridges of insulin represent a typical protein with disulfide bonds for folding process, a series of studies of disulfide-pairing coupled folding intermediates of insulin has been investigated in our group. Table 6 listed our NMR structural studied folding intermediates of insulin and IGF.

Table 6 List of disulfide intermediate models

	Insulin	IGF-Ⅰ
One disulfide bond isomer		$Ala^{47,52}$, $Ser^{6,48}$-IGF-Ⅰ IGF-Ⅰ-peptide
Two disulfide bond isomer	$Ser^{A6,A11}$-DKP $Ala^{A6,A11}$-DKP $Ser^{A7,B7}$-DKP	$Ala^{47,52}$-IGF-Ⅰ
Three disulfide bond isomer	[A6-B7, A7-A11, A20-B19]-HI [A6-A7, A11-B7, A20-B19]-HI	[48-52, 6-47, 18-61]-IGF-Ⅰ [47-48, 6-48, 18-61]-IGF-Ⅰ

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

Folding of insulin *in vivo* is accomplished by the single-chain proinsulin (HPI), which contains a connecting peptide (C peptide) between the C-terminal of the B chain and N-terminal of the A chain[58]. After cleavage the C-peptide, insulin was released as functional molecule to control the blood sugar.

Oxidative refolding of reduced proinsulin or oxidation of isolated A- and B-chains (i. e. insulin chain recombination) preferentially yields native proinsulin or insulin respectively with native disulfide pairing in each case implying that insulin A- and B-chain contains sufficient information to interact and form the native molecule[59~61].

The insulin disulfide isomers with three disulfides can be induced by disulfide exchange in 10 mmol/L dithiothreitol (DTT), 6 mol/L guanidine-HCl and 0.1 mol/L Tris-HCl (pH 8.3) in room air at 37℃ for 23 h[62, 63]. Figure 8c shows HPLC chromatograph profiles of the products of insulin disulfide reassortment at denatured condition. Isomer-Ⅰ (insulin-swap) and isomer-Ⅱ (insulin-swap2) retain native A20-B19 disulfide and contain mis-matched (A6-B7 and A7-A11) and (A6-A7 and A11-B7) disulfides respectively. In all the products *ca*. 70% of the

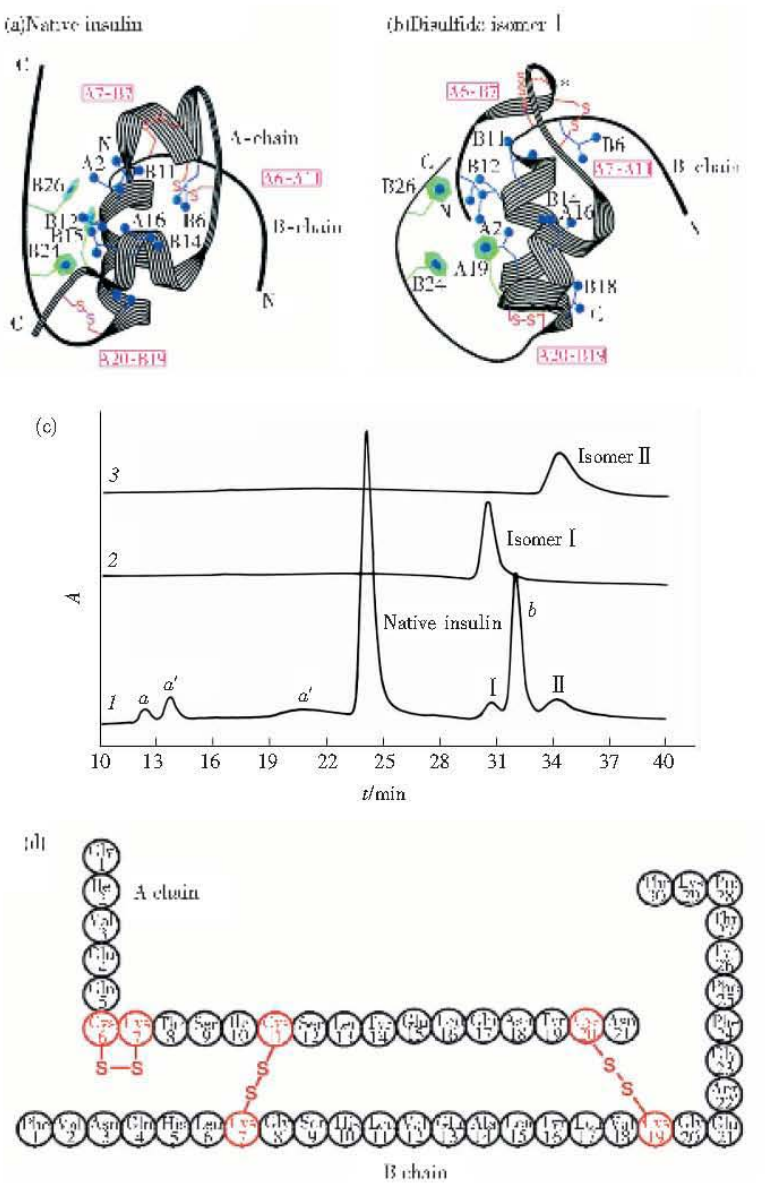

Fig. 8

Structures of native insulin (a) and insulin-disulfide isomer-Ⅰ (b). Cylinder models are derived from NMR studies as described in text. In each panel disulfide bridges (boxes) are shown in red. Selected side chains (blue) in the hydrophobic core are as labeled. The structure of insulin-swap is remarkable for segmental unfolding of the A1-A8 α-helix and tilting of the A12-A18 helix in an otherwise native-like globular domain. (c) Reverse-phase HPLC chromatograms showing disulfide reassortment of native insulin(*1*), purified isomer Ⅰ (insulin-swap, *2*), and purified isomer Ⅱ (insulin-swap2, *3*). In chromatogram (*1*) species *a*, *a′*, and *a″* are isolated A chains containing internal disulfide bridges; species *b* is the isolated [B7-B19] cyclic-B chain. Isomers Ⅰ and Ⅱ are herein designated insulin-swap and insulin-swap2. Their differences in elution times and breadth of peaks relative to native insulin correlate with trends in CD-defined helix contents (Table 7). (d) Primary structures and disulfide pairing schemes of insulin-swap2. Insulin-swap2 contains an unusual A6-A7 vicinal disulfide bridge as well the A7-A11 pairing also found in insulin-swap and the native A20-B19 disulfide bridge.

starting material remained as native insulin; other products included insulin-swap (3% ~5% of starting material), insulin-swap2 (3% ~5%), and isolated A- and B-chains containing internal disulfide bonds (20% ~ 25%). The absence of other disulfide isomers suggests that the observed species exhibit relatively high stability and/or kinetic accessibility. The NMR structures of native and isomer Ⅰ are showed in Figure 8a and b respectively.

The metastable isomer-Ⅰ retains ordered secondary structure and a compact hydrophobic core as shown in Figure 8b. Comparing with the native insulin structure in Figure 8a, although the orientation of A and B chain of isomer-Ⅰ is globally rearranged, one face of the isomer is similar between native and isomer-Ⅰ structure, which contains receptor binding determinants as shown in Figure 9a. Both isomers retain 13% ~ 27% of the receptor binding potency relative that of insulin (Table 7). The isomer-Ⅱ is less stable than the isomer-Ⅰ. Its native-like elements of structure are maintained in the B chain whereas the A chain is largely disordered. Thermodynamic studies with guanidine denaturation demonstrate the instability of the isomers-Ⅱ relative to native insulin (Table 7, $\Delta\Delta G_u > 3.5$ kcal/mol). But the key long-range NOEs that defined native-like B chain folding existed in NOESY spectrum: Val^{B12} to Phe^{B24} and/or Tyr^{B26}, Leu^{B15} to Tyr^{B26}, and Leu^{B15} to Tyr^{A19} *etc.* shown in Figure 9b. The structural feature like the orientation of C-terminal of A chain and central α-helix from B9-B18 followed with β-turn of B chain is also retained in isomer-Ⅱ as that in native state. Both insulin isomers are caught in kinetic trapping. Insulin isomer-Ⅱ is less well-ordered due to its stereo-chemically unfavourite adjacent Cys, which form a disulfide bond (Figure 8d); and A1 ~ A11 peptide segment is more flexible in insulin-swap2.

Insulin-like growth factor-Ⅰ (IGF-Ⅰ) and insulin share high sequence similarity in both the A and B chains of insulin, but IGF-Ⅰ contains an additional connecting C-domain and an extension at the C-terminus known as the D-domain. Both NMR structure and crystal structure of IGF-Ⅰ reveal its structural homology with insulin[64, 65]. While oxidative refolding of proinsulin or chain combination of insulin yields only the native disulfide pairing scheme, oxidative refolding of insulin-like growth factor-Ⅰ (IGF-Ⅰ) and chain combination of IGF-Ⅰ A- and B- domains in each case yields two products under thermodynamic control: native IGF-Ⅰ and an alternative structure with "swapped" disulfide bonds (IGF-swap). IGF-swap contains two non-native disulfides 6-47 and 48-52 corresponding to B7-A6 and A7-A11 disulfide bonds and a native disulfide bond 18-61 (similar to B19-A20 in insulin)[66 ~ 68]. The products of native IGF-Ⅰ and IGF-swap are 60% and 40% respectively, which indicate that two products are in thermodynamic equilibrium. Whereas the swapped disulfide pairing scheme in HPI or insulin corresponds to kinetic trap, native IGF-Ⅰ and IGF-swap exhibit similar stabilities. NMR study of IGF-swap suggests that the tertiary structures of IGF-Ⅰ and swap are similar in the neighborhood of their shared disulfide bridge (18-61). IGF-Ⅰ and IGF-swap adopt similar secondary structures but distinct tertiary folds[67, 68]. Why does the folding information ambiguous in IGF-Ⅰ, whereas it is uniquely in insulin and proinsulin? Is it possible that sequence of IGF-Ⅰ missed some folding information? These questions leave the broad interesting in further study.

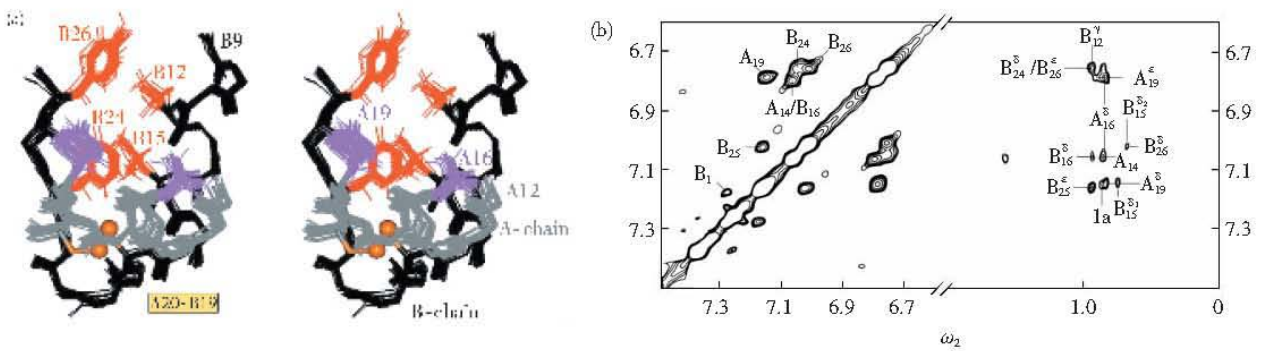

Fig. 9 Long-range interactions of B-chain super-secondary structure

(a) Superposition of crystal structures (residues A12-A20 and B9-B26) showing key side-chain contacts among Leu^{A16}, Tyr^{A19}, Val^{B12}, Leu^{B15}, Phe^{B24}, and Tyr^{B26}. Main-chain atoms of the A- and B- chains are shown in gray and black, respectively; side chains of the A- and B- chains are shown in lilac and red, respectively. A20-B19 disulfide bridge is shown as yellow-orange balls (sulfur atoms) and sticks. Spread of conformations reflects structural variation among crystal forms. Coordinates were obtained from the Protein Data Bank (identifiers 4INS, 1APH, 1BPH, 1CPH, 1DPH, 1TRZ, 1TYL, 1TYM, 1ZNI, 1LPH, 1G7A, 1EV6, and 1ZNJ). (b) NOESY spectrum of insulin-swap2 in 80% D_2O-20% deuterioacetic acid showing long-range contacts between aromatic and aliphatic sidechains. Of particular interest are contacts characteristic of native-like B-chain super-secondary structure (from Val^{B12} to Phe^{B24} and/or Tyr^{B26}; and from Leu^{B15} to Tyr^{B26}) and native-like A16-A19 and B15-A19 contacts. Aromatic resonances of Phe^{B24}-H_γ and Tyr^{B26}-H_ε overlap at 6.75 ppm. Assignment of possible Ile^{A2}-Tyr^{A19} interaction (Ia) is provisional due to resonance overlap.

Table 7 Guanidine denaturation

Analogue	ΔG_u	$\Delta\Delta G_u$	C_{mid}/M	m/kcal· mol^{-1}· M^{-1}	Potencies/%
(a) Insulin analogues					
Native insulin	4.4 ± 0.1	—	5.2 ± 0.1	0.84 ± 0.01	100
Insulin-swap	1.3 ± 0.2	−3.1 ± 0.2	2.2 ± 0.2	0.58 ± 0.04	13 ~ 27
Insulin-swap2	0.5 ± 0.2	−3.9 ± 0.3	1.2 ± 0.1	0.44 ± 0.03	13 ~ 27
(b) IGF analogues					
Native IGF-Ⅰ	2.5 ± 0.1	—	4.2 ± 0.1	0.58 ± 0.01	100
IGF-swap	3.3 ± 0.1	0.8 ± 0.2	5.1 ± 0.1	0.64 ± 0.01	10

ΔG_u indicates apparent change in free energy on denaturation in guanidine-HCl as extrapolated to zero denaturant concentration by two-state model. $\Delta\Delta G_u$ indicates difference in ΔG_u values relative to native insulin (a) or IGF-Ⅰ (b). Uncertainties in two-state fitting parameters do not include possible systematic error due to non-two-state behavior. Insulin potencies (a) are expressed as percent activity in isolated fat-cell assays and *in vivo* rodent models; IGF-Ⅰ potencies (b) are defined by percent affinity for the type Ⅰ IGF receptor defined relative to native ligand. For sigmoidal transitions (native insulin, IGF-Ⅰ, and IGF-swap), model employs simultaneous fitting of slops of pre- and post-transition baseline. See text in detail.

3.3 Two disulfide insulin analogue: Ser$^{A6, A11}$-DKP, Ala$^{A6, A11}$-DKP, Ala$^{47, 52}$-IGF-Ⅰ and Ser$^{A7, B7}$-DKP

Two disulfide insulin analogues were prepared by novel synthetic S-sulphonate A and B chain, followed by chain combination. The DKP-template makes these analogues well dissolved at neutral pH as well as in low pH. To mimic the folding intermediate, an analogue lacks the A6-A11 or A7-B7 disulphide bridge; the Cys is replaced by Ser or Ala. From NMR results of Ser$^{A6, A11}$-DKP and Ala$^{A6, A11}$-DKP, the solution structure of Ser substitution at A6 and A11 position is remarkable for the segmental unfolding of the N-terminal A-chain α-helix (A2 to A8) in an otherwise native sub-domain[69]. The A6-A11 disulfide bridge is buried and surrounded by hydrophobic residues in native insulin. The Ala substituted analogue retains same structure with that of Ser analogue[70]. The structure suggests that the overall orientation of the A and B chains is consistent with reorganization of the A-chain′s N-terminal segment. Nevertheless, the analogue′s low biological activities (Table 8) imply that this segment functions as a preformed recognition α-helix. The corresponding two disulfide intermediate of IGF-Ⅰ (Ala$^{47, 52}$-IGF-Ⅰ) was trapped by acid quenching[71].

Table 8 Properties of two disulfide insulin isomers

Analogue	ΔG_u	$\Delta\Delta G_u$	C_{mid}	m/(kcal· mol^{-1}· M^{-1})	Potencies/%
DKP-insulin	4.9 ± 0.04	—	5.8 ± 0.1	0.84 ± 0.01	150 ± 24
Ala$^{A6, A11}$-DKP	1.4 ± 0.4	−3.5 ± 0.5	2.7 ± 0.4	0.52 ± 0.08	5
Ser$^{A6, A11}$-DKP	1.9 ± 0.3	−3.0 ± 0.3	2.9 ± 0.3	0.65 ± 0.07	0.1
Ser$^{A7, B7}$-DKP	0.3 ± 0.6	−4.6 ± 0.7	0.7 ± 1.1	0.48 ± 0.05	0.002

$\Delta\Delta G_u$ indicates the difference in ΔG_u values relative to that of DKP-insulin. Thermodynamic differences between DKP-des-[A6-A11] Ala and DKP-des-[A6-A11] Ser are not significant. Uncertainties in two-state fitting parameters do not include possible systematic error due to non-two-state behavior. Potencies are expressed as percent affinity for the human placental insulin receptor, defined relative to native human insulin (100%).

Removal of the 47-52 disulfide bond (corresponding to A6-A11 in insulin) was associated with local unfolding of the adjoining α-helix (helix 2; corresponding to A2 ~ A8 helix in insulin). The IGF-Ⅰ intermediate is partially folded and contains elements with native secondary and tertiary structure. Above three intermediates of either insulin or IGF-Ⅰ lack A6-A11 bridge (or 47-52 in IGF-Ⅰ), but keep similar structural features: native secondary structure (including helices 1 and 3) is otherwise retained and defines a folded sub-domain, which likes an "internal template" in contact with the N-terminal A chain helix in native insulin. The preference to form a helix in the N-terminal segment makes it possible to dock into the "internal template" by the hydrophobic interaction, consequently the helix formation may help to link the disulfide bridge of A6-A11. These results support the hypothesis that folding of the insulin motif is directed by a subset of native structural elements and that these elements form in an early step in the pathway. Formation of helix 2 (A1-A8) is likely to represent a late step[69].

The other two disulfides insulin analogue is Ser$^{A7, B7}$-DKP[72] that lacks the most exposed disulfide A7-B7 in insulin. CD and NMR results indicate that this intermediate model has much less well-defined

secondary structure comparing with those of above models (Table 8). The stability of the intermediate is only 0.3 kcal/mol, which is 4.6 kcal/mol lower than that of native insulin and also 1.1 ~ 1.6 kcal/mol lower than that of pairwise substitution of internal cystine A6-A11. Although substantially disordered and without significant biological activity, this un-tethered analogue contains a molten globule sub-domain comprising cystine A20-B19 and a native-like cluster of hydrophobic side chains. The long-range NOEs such as Phe^{B24} to Leu^{B15} and Leu^{B15} to Tyr^{A19} as well as B9-B19 helix-1 remain in this molecule.

The A7-B7 bridge provides an exposed tether between extended strand (B1-B8) to an α-helix (A2-A8). Local disorder is observed within un-tethered A and B chain segment. Remarkably, A and B chains contribute un-equal to this folded moiety: the B chain retains native-like super-secondary structure, whereas the A chain is largely disordered.

These results suggest that the B sub-domain provides a template to guide folding of the A chain. Stepwise organization of insulin-like molecules supports a hierarchic protein folding. Un-tethering the exposed A7-B7 bridge leads to a 10^5-folds decrease in the level of receptor binding, which may in part reflect destabilization of the receptor-binding surface, and in part a change in the environment of cystine A7-B7 on receptor binding. Does receptor require the environment of A7-B7 change from hydrophilic to hydrophobic? Is the hydrophilic environment of A7-B7 cystine in T-state not favor for receptor binding?

Among these intermediates with two disulfides the conformation of $Ser^{A6, A11}$-DKP is more similar to that of native insulin, which is somehow similar to quasi-native species of BPTI intermediate as mentioned above in the beginning of the section. But A6-A11 disulfide bridge is buried, differing from the exposed Cys5-Cys14 bridge in native BPTI. However the structure of $Ser^{A7, B7}$-DKP is much less well folded, though the disulfide bond of A7-B7 is the most exposed one in native insulin that is similar to cystine 5-14 in BPTI.

3.4 One disulfide insulin analogue: IGF-I-peptide

IGF-Ⅰ analogue model with only one disulfide bond, IGF-Ⅰ-peptide, contains total 37 residues and one disulfide bridge between residues 18-61. The segment of Gly^7 to Phe^{25} of the peptide corresponds to the segment from B8 to B26 segment in insulin sequence, while that of Asp^{53} to Ala^{70} corresponds to A12 to A21 of insulin and additional D-domain of IGF-Ⅰ. NMR study of the IGF-Ⅰ-peptide reveals that the peptide maintains helix 10-18 (corresponding to B11-B19 helix in insulin) and short helix 56-61 (corresponding to A16 to A21 in insulin). In addition, NOESY spectrum showed the NOE corresponding to B26 - B15 in insulin is still retained (unpublished data).

Recently Feng *et al.* prepared a model peptide [A20-B19]-porcine insulin precursor (PIP) containing the single disulfide A20-B19[73]. Although the yield of the model peptide from yeast was 2 ~ 3 magnitudes lower than that of the wild-type PIP, the physicochemical property analysis suggested that the model peptide adopts a partially folded conformation. *In vitro*, the fully reduced model peptide can efficiently and easily form the disulfide A20-B19, which suggest that formation of the disulfide A20-B19 is kinetically preferred.

3.5 The folding pathway of PIP and HPI

Figure 10a shows the proposed folding pathway of single chain porcine insulin precursor (PIP) *in vitro*. The precursor contains B1-B30 segment and di-peptide linker (Ala-Lys), followed by A1-A21 A chain with total 53 residues. The recombinant insulin precursor was fully reduced and diluted to buffer for refolding. Reduction was trapped at different refolding time by adding iodoacetic acid. Then refolding products were digested by endoproteinase V8, followed by ESI-MS analysis. In redox buffer, the yield of native PIP from fully reduced/denatured PIP can reach as high as 85%. Three predominant intermediates have been captured and characterized: designated 1SSPIP and 2SSPIPa and 2SSPIPb[74]. One disulfide intermediate in up-left corner of Figure 10a is [A6-A11] intermediate, whereas two intermediates contain [A7-B7, A20-B19] and [A6-B7, A20-B19] or [A11-B7, A20-B19].

Although A6-A11 and A20-B19 disulfide bridges are each buried, above NMR structural studies of disulfide pairing intermediate models suggest a fundamental difference between their environments in a populated intermediates. The A6-A11 bridge is very flexible and poorly ordered as shown in Figure 10b, whereas the A20-B19 bridge participates in and presumably stabilizes a molten sub-domain in Figure 10b and c. A6-A11 is favored to form in refolding of reduced PIP because of the sequence proximity of these sites and instability of alternative local pairing schemes (a vicinal A6-A7 disulfide bridge and "swapped" A7-A11). Even if the A6 - A11 disulfide bond formed, it can not stabilize the whole molecule, the rearrangement of disulfide may break it and form the other disulfide bridge. In contrast, although the A20-B19 is site-distant in the sequence, it is still favored in refolding because of intrinsic long-range structural propensities to stabilize the "folding nucleation". It is reasonable to deduce that the [A20-B19] species represents the first "structural" intermediate, which contains a stable partial fold with native-like long-range contacts.

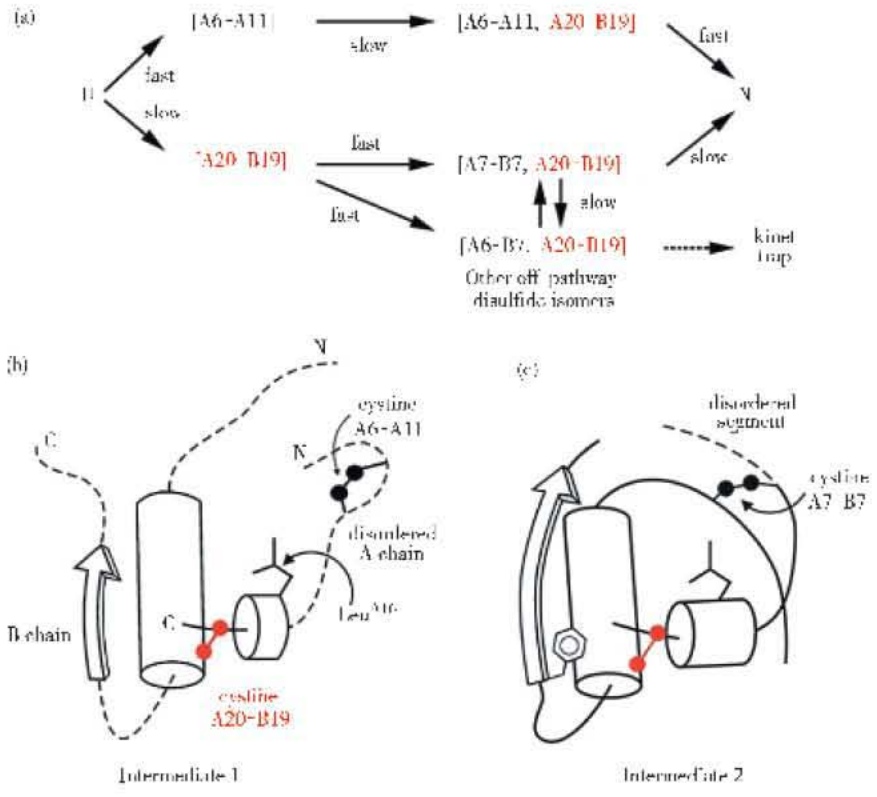

Fig. 10　Proposed disulfide pathway of porcine insulin precursor (PIP)

(a) Disulfide intermediates based on biochemical mapping studies of a proinsulin-like precursor. "U" designates unfolded and reduced polypeptide; "N" designates native state with three disulfide bridges (A6-A11, A7-B7 and A20-B19). Intermediate species are designated by cystine(s); e. g., [A20-B19] indicates a species containing one disulfide bridge. Cystine A20-B19 is highlighted in red in each panel. Designations "fast" or "slow" refer to kinetic study of Feng and coworkers. Because technical limitations in peptide mapping precluded unambiguous assignment of disulfide pairing schemes, pathway is in part hypothetical. (b) and (c) Cylinder models of two-disulfide insulin analogs: (b) DKP-des-[A7-B7]^{Ser} and (c) DKP-des-[A6-A11]$^{Ser/Ala}$ based on NMR data Dashed lines indicate disordered regions. Positions of cystines are indicated by filled circles and bar.

To summarize the NMR results of the disulfide-pairing coupled folding intermediates Figure 11 shows two intermediates model containing three disulfides in (a) and (b), three models containing two disulfides in (g) ~ (i) as described above. Ensembles of NMR structures of native insulin and DKP-insulin are shown in (d) and (f). An unpublished NMR structural model of thermo-unfolding native insulin (with three native disulfides) is shown in (c). The NMR data were collected from Bruker 800 MHz cryo-probe for native insulin at pH 2 and 60℃, the 2D NMR spectra were acquired within one hour to avoid the possible aggregation in longer time. This solution condition is normally used for insulin fibrillation study. The structure represents partly thermo-unfolding state of insulin, or pre-fibril state prior to forming amyloid fibril. The structure maintains central α-helix of B9 - B18 with adjoining β-turns (B20-B23) and short helical segment of A17-A20 with flexible N-terminal of both chains. A series of ^{15}N-HSQC spectra of ^{15}N-insulin were collected to monitor the time course of the conformational change in same conditions. The ^{15}N-TOCSY-HSQC and NOESY-HSQC spectra were acquired within three hours each, which allow the assignment of 60% residues in ^{15}N-HSQC spectra. Data indicated that the intensities and chemical shifts of these peaks were similar before 20 h, after which fibril started to form and NMR signal could not be acquired due to line-broadening. In spite of the conformational differences between a variety of the intermediates in Figure 11, one structural feature is always remained: a super-secondary structure as shown in (e), which consist of B9-B19 helix, adjoining B23-B26 segment folded back against the central B chain and A20-B19 disulfide bond linked B chain with A13-A19 helix (or more short helix of A17-A20 in thermo-unfolded insulin). The super-secondary structure might be the "folding nucleus" in folding mechanism. Among three disulfides, A20-B19 is the most important one in stabilizing the nascent polypeptide in earlier stage of the folding.

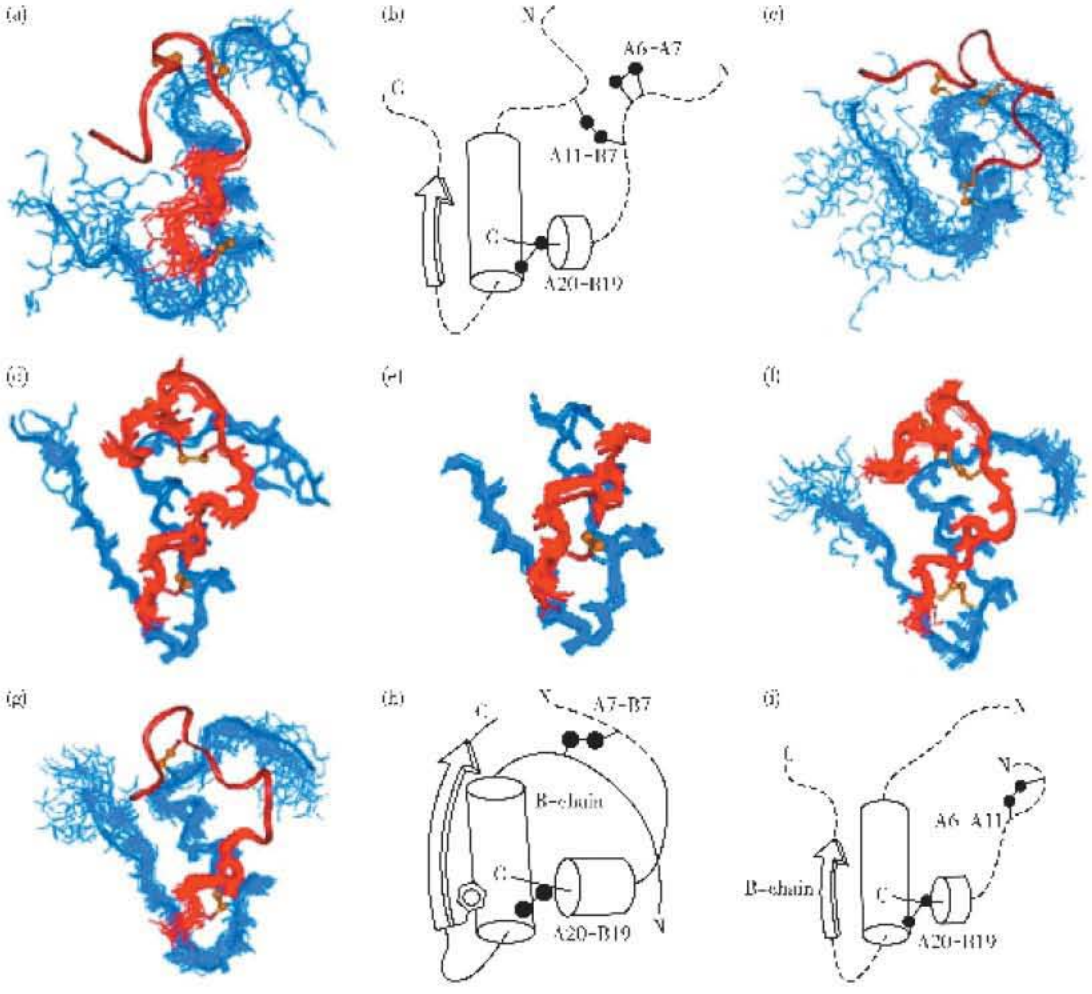

Fig. 11 Summary of the structures of insulin disulfide-pairing coupled insulin folding intermediates

(a) An ensemble of 20 NMR structures of insulin disulfide isomer-1 intermediates (PDB ID: 1SGL). (b) Proposed model of insulin isomer-2. (c) Thermo-unfolded insulin at 60℃ (unpublished data). (d) An ensemble of 10 NMR structures of native insulin (PDB ID: 2HIU). (e) 14 T-state crystal structures of native insulin (PDB ID: 4INS, 1APH, 1BPH, 1CPH, 1DPH, 1TRZ, 1TYL, 1TYM, 1ZNI, 1LPH, 1G7A, 1EV6, and 1ZNJ). Only backbone of A12-A21 and B9-B28 were shown. (f) An ensemble of 20 NMR structures of DKP-insulin (PDB ID: 1LNP). (g) That of $Ser^{A6, A11}$-DKP-insulin (PDB ID: 1VKT). (h) and (i) proposed models of $Ala^{A6, A11}$-DKP-insulin and $Ser^{A7, B7}$-DKP-insulin respectively. A chain backbone or ribbon are shown in red, that of B chain in blue. The structures in (a), (c), (d) ~(g) were aligned respect to the main chain atoms of B9-B19. All the structures of intermediates were remained the super-secondary structure shown in (e).

Acid-trapped folding intermediates and products of IGF-I reveal that in addition to non-native intermediates, three native-like intermediates were identified: a single-disulfide [18-61] intermediate, an intermediate with 18-61 and 6-48 disulfide bonds and another with 18-61 and 47-52 disulfide[75]. The first populated intermediate [18-61] corresponds to A20-B19 disulfide in insulin. Recently Qiao *et al.* reported the folding pathway of human proinsulin[76]. Differing from the folding pathway of porcine insulin precursor, no intermediates with one- or two-disulfide bonds could be captured under a variety of refolding conditions. Four intermediates with three disulfide bridges were isolated and characterized. The different pathway of HPI reveals that the C-peptide not only facilitates the folding of HPI, but also governs its kinetic folding pathway of HPI. The formation of disulfide A20-B19 may guide the transfer of the folding information from the B-chain template to the unstructured A-chain. The insulin folding pathway study and the NMR structural models of insulin folding intermediates supports the proposal that protein folding is both co-translationally and post-translationally[77]. The folding of the nascent chain begins early during biosynthesis and followed by further conformational readjustment, especially during the disulfide pairing and rearrangement.

4 Other members of insulin super-family, *C. elegans* insulin-like peptide and abnormal insulin analogues

4.1 *C. elegans* insulin-like peptide structurally resembles HI and interacts with human IR

Recently the first NMR structure of *C. elegans* INS-6 (insulin-like peptide, ILP) of invertebrate worm was reported[78]. In addition to the role of insulin in control glucose level in blood, the central role for growth and aging from insulin was also found. But what is the function of insulin-like protein (or peptide) in invertebrate? It is surprisingly reported that there are insulin-like peptides from *C. elegans* and its signaling controls the life span and aging of the worm[79, 80]. After total genomic sequencing of *C. elegans* was discovered, Ruvkun *et al.* found 37 candidate insulin-like genes by bioinformatic analysis of the *C. elegans* genome. These genes are expressed primarily in neurons, which can trigger dauer arrest and reproduction. Like insulin receptor for insulin, the daf-2 receptor is the receptor for *C. elegans* insulin-like peptide. The sequence of daf-2 is 35% identical in sequence of the human insulin receptor and 34% identical in sequence of the human IGF receptor[81]. Of course, studies of INS proteins and the daf-2 receptor are broadly interested to human biology for the structure, function and evolution of insulin-like signaling systems and also for relationship between life span and metabolism.

The *C. elegans* insulin-like peptide sequences are classified into three classes: α, β and γ, based on the pattern of cysteines[80]. α-class proteins contain three SS bridges, one of which is a novel bridge. But it lacks an A6-A11 disulfide bridge, an invariant feature of vertebrate insulins and insulin-like growth factors[82]. β-class proteins appear to retain insulin's three canonical disulfide bridges plus a fourth, whose location is variable. γ-class proteins contain three canonical (or pseudo-canonical) disulfide bridges. The sequence of INS-6 is shown in Figure 12, which is a 50 residues single chain peptide and contains four disulfide bridges: three are homologue with those of human insulin or IGF-Ⅰ, while the fourth one is at position of A21 and B23 in sequence of insulin with underline. The sequence identity between human insulin and INS-6 is ~26%.

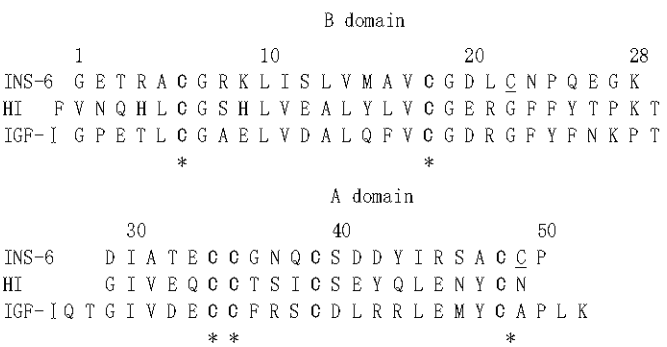

Fig. 12 Sequence of *C. elegans* insulin-like peptide (INS-6), human insulin and IGF-Ⅰ

In Figure 13 the ribbon models of NMR structure of *C. elegans* INS-6 and crystal structure of human mini-proinsulin[8] are shown. The additional disulfide 26-53 locates at the C-terminal of INS-6. The global conformation of *C. elegans* INS-6 is similar to that of human insulin in general. The structure retains three helices with some characteristic structural features of human insulin. Despite of the fact that human insulin and INS-6 only share 13 residues identity (with six Cys), INS-6 binds to human insulin receptor with weak, but detectable relative affinity[78]. As shown in Figure 12, the single chain *C. elegans* INS-6 has B29-A1 (corresponding to the sequence of human insulin) amide bond, which causes 1 500 fold reduction of binding in a mini-proinsulin with B29-A1 linker[83]. The substitution of a single residue at A3, A19, B24 and B25 impair receptor binding dramatically as described above. Considering the factors described, the interaction between INS-6 and human IR is still significant. The hydrophobic side chains of INS-6 shown in Figure 14 actually compose a new hydrophobic surface, which include Ile^{34}, Ala^{35}, Tyr^{47}, Ile^{48}, Ala^{51}, Leu^{14}, Ile^{15}, Leu^{17}, Val^{18} and Val^{21} (corresponding to A2, A3, A15, A16, A19, B11, B12, B14, B15 and B18 respectively). Shown in Figure 14c Tyr^{47} in INS-6 (corresponding to A15 in sequence of HI) seems to occupy the position of A19 in human insulin, which may retain the receptor binding characteristics. Such conservation of an insulin-like fold as well as binding in a *C. elegans* INS-6 suggests that the multiple compensating interactions within INS-6 permit maintenance of an insulin-like fold. Perhaps

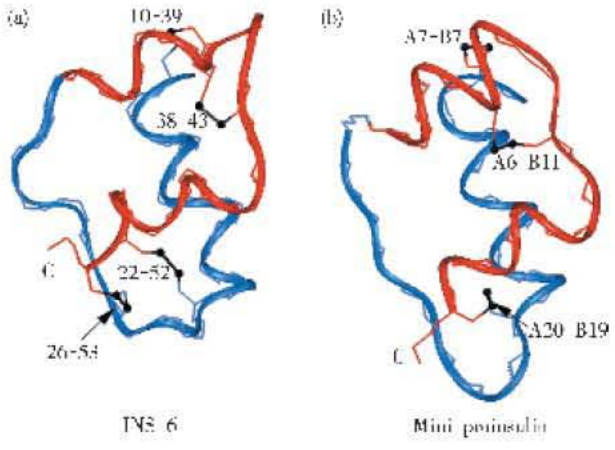

Fig. 13 Ribbon models of (a) NMR structure of *C. elegans* insulin-like peptide INS-6 and (b) crystal structure of mini-proinsulin(des-B30, B29-A1 linked single chain insulin, PDB ID: 6INS

A-domain is shown in red, B-domain in blue, and disulfide in black. Four disulfide bridges 10-39, 38-43, 22-52 and 26-53 (corresponding to B7-A7, A6-A11, B19-A20 and B23-A21 in human insulin; B23-A22 does not exist in human insulin) are shown in black.

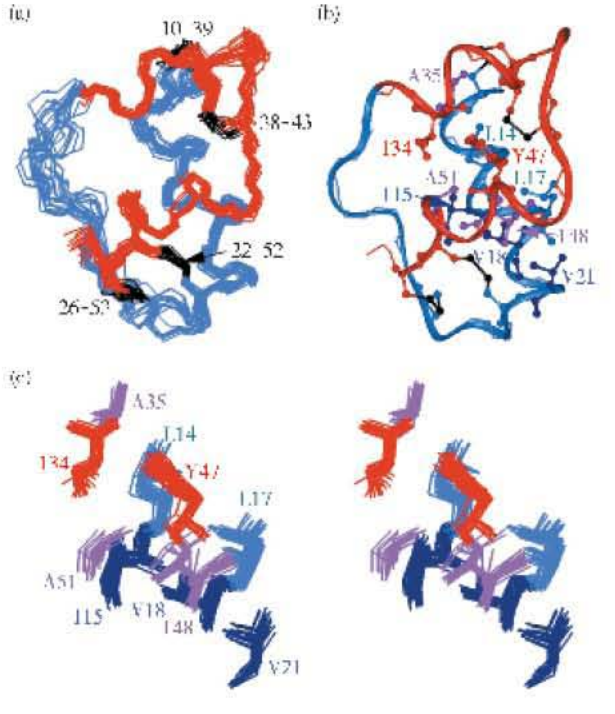

Fig. 14 Structure of INS-6

(a) NMR-ensemble of 20 structures of INS-6. Residue 9-32 (corresponding to B6-B29 in human insulin) is shown in blue, and residue 33-54 (A1-A22 in human insulin) is shown in red. (b) Ribbon model of one solution structure of INS-6 and side-chain of 10 buried residues. Color codes are same as in (a). Side chains of Ile^{34} and Tyr^{47} are shown in red; Ala^{35}, Ile^{48} and Ala^{51} in magenta; Leu^{14} and Leu^{17} in blue and Ile^{15}, Val^{18} and Val^{21} in purple. Disulfide bonds are shown in black. (c) Stereo view of the side-chain of these residues of the ensemble.

the extensive differences between the surfaces of insulin and INS-6 permit recapitulation of an analogous signaling conformation. Understanding the molecular basis of INS-6 activity will require co-crystal structures of INS-6/def-2 and INS-6/IR complexes.

The study of insulin-like peptide and signaling system in *C. elegans* will reveal the structure and evolution of mammalian signaling systems. According to the comparison of 80 sequences of members of the insulin family[82], the most conservative residues are: six Cys and B8 Gly in insulin sequence. These seven residues also retain in *C. elegans* INS-6. The fourth disulfide bridge between A21 and B23 (corresponding sequence number in HI) seems not important for keeping the same folding of *C. elegans* INS-6. The distance between these two residues is always close in NMR or crystal structure of insulin and its analogues. A fake model calculation of INS-6, which maintain all distance and dihedral angle restraints except the A21-B23 disulfide, showed nearly identical structure of INS-6 implying that the fourth disulfide is un-necessary for the request of structure and disappeared in evolution.

4.2 Other member of insulin super-family adopt the "insulin fold"

As described above, insulin and related peptides have been identified in several phyla of the animal kingdom. Most have been found in the vertebrates, but examples have also been identified in proto-chordates, insects and mollusks. Among the members of insulin super-family (insulin, IGF, relaxin, bombyxin and molluscan insulin-like peptide), IGFs are closest to insulin in sequence and structure. The designation of IGF is due to the fact that they showed mitogenic effects and exhibited insulin-like effect in adipose and muscle tissue and their structures were homologous to that of HPI or insulin[84]. IGF-Ⅰ and -Ⅱ are single chain proteins of about 7.5 ku, and consist of B, C, A and D domains, in which B and A domains are sequence and structural homologues of insulin B and A chain. IGF-Ⅰ and -Ⅱ share 67% sequence identity. The NMR structure of IGF-Ⅰ and IGF-Ⅱ both adopt the "insulin fold", which contain three helices with β-turn in B domain[85~87]. The modeling and NMR structure of IGF-Ⅰ disulfide isomer (corresponding to insulin disulfide isomer-Ⅰ) is relatively flexible, but still retains two helices (relative to helix-1 and -3 in human insulin) and internal packing[88, 89]. A similar relationship has been described between the structure of native insulin and a homologous disulfide isomer, suggesting that these alternative folds represent general features of insulin-like sequences. There are two IGF receptors: IGFR-1 and IGFR-2. Whereas IGF-Ⅰ binds to IGFR-2 is a 1 000 fold weaker relative to that of IGF-Ⅰ and IGFR-1, IGF-Ⅱ binds to IGFR-1

and -2 with same affinity[87]. IGF-Ⅰ and -Ⅱ bind to insulin receptor with lower and tighter affinity respectively. The comparison study of insulin and IGF signaling systems are broad relevant in hormone control and also in growth and metabolism. The C domain in IGF is shorter than C-peptide in proinsulin. NMR structure of min-IGF-Ⅰ[90] exhibits that the inactive IGF-Ⅰ mutant lacking the C domain with only 57 residues retains three helical segments with conformational change at C-terminus of B domain, which are important for receptor binding. Thus, deletion of the C region of IGF-Ⅰ results in a substantial tertiary structural rearrangement that accounts for the loss of receptor affinity. As described in above section, mini-proinsulin, a single-chain insulin analogue with B29-A1 peptide bond, exhibits the affinity of the insulin receptor 1 500-fold lower than that of the two-chain analogue[83], however these two structures are essentially identical. Differences are observed near the site of cross-linking: the adjoining A1-A8 alpha-helix (variable among crystal structures) is less well-ordered in mini-proinsulin than in the two-chain variant. Despite strict conservation of non-polar residues in their hydrophobic cores, the structural differences between mini-proinsulin and mini-IGF imply that two homologous proteins provide examples of homologous protein sequences encoding non-homologous structures[83].

The studies of hybrid insulin-IGF molecules are also interesting for hormone signaling and protein folding[91~93]. One hybrid consisting of the B-chain of bovine insulin and the A-domain of IGF-Ⅰ indicated that its insulin-like activity (insulin receptor-binding and stimulation of lipogenesis) was 0.2%, and its growth-factor activity was less than 1%, both relative to natural insulin respectively[91]. Another hybrid, consisting of the insulin B and A chains linked with the C domain of IGF-Ⅰ, binds with high affinity to both truncated soluble insulin receptors and membrane-bound holo-receptors. The affinity for interacting with the soluble truncated insulin receptors was 55% ~94% relative to insulin, and affinity for membrane-bound insulin receptors was 113% of that of insulin[92].

Recently Feng *et al*, reported two hybrids of insulin and IGF-Ⅰ, Ins (A)/IGF-Ⅰ(B) and Ins (B)/IGF-Ⅰ(A), as well as a mini-IGF-Ⅰ[93]. Both mini-IGF-Ⅰ and Ins (A)/IGF-Ⅰ(B) fold into two thermodynamically stable disulfide isomers, while Ins (B)/IGF-Ⅰ(A) folds into one unique thermodynamically stable tertiary structure. The results again imply that the different folding behavior of insulin and IGF-Ⅰ is mainly controlled by their B-chain/domain. The preliminary NMR structure of the hybrid Ins (B)/IGF-Ⅰ(A) exhibits the insulin-like fold in general (author's unpublished data). Further mutation of hybrid insulin/IGF is expected to explore more detail in signaling and folding of insulin and IGF.

The crystal structure of relaxin exhibits insulin-like folding with three helices despite of the biologically diverse activities[94]. Relaxin is a two chain molecule with 24 residues A chain and 28 residues B chain. Bombyxin-Ⅱ from the brain of the silkworm contains only 20 residues A chain and 28 residues B chain. The overall main-chain NMR structure of bombyxin-Ⅱ is similar to that of insulin[95, 96]. However, there are significant conformational and functional differences in their C-terminal parts of the B-chain. The B-chain C-terminal part of bombyxin-Ⅱ adopts an extension of the B-chain central helix similar to that of relaxin. It suggests that insulin might have evolved the additional receptor-recognition site in the B chain C-terminal β-strand to distinguish itself from bombyxin and relaxin.

The molluscan neuron insulin-related peptide precursor (MIP-1) has the same overall structure as its vertebrate counterparts[97]. The common "insulin fold" is probably adopted in all insulin family members. The different biological functions and binding affinities for the various receptors may arise from differences of the key side chains rather than their structural differences between the members of insulin super-family. Although the specificities of receptor binding are different, there is a possibility of co-evolution of polypeptides and their receptors[82].

Although the structures of IGF-Ⅰ and -Ⅱ have been determined, structure of proinsulin has never been reported either by NMR or by X-diffraction. One technical difficulty is the flexible conformation of C-peptide with 35 residues in length, which is Gly- and Pro-rich sequence. The intracellular conversion of proinsulin to insulin occurs *via* cleavage at the two dibasic sites: Arg31-Arg32, B chain-C-peptide (BC) junction; and Lys64-Arg65, A chain-C-peptide (CA) junction, catalyzed by the subtilisin-like prohormone convertases SPC3 and SPC2 respectively[98, 99]. Does the structure of insulin motif in proinsulin retain same structure as in separated insulin? Is the conformation of C-peptide completely random? Is there any special function or sequence request of C-peptide in the proteolytic processing from proinsulin to insulin by SPC2/SPC3? Steiner[100] proposed a conformational change at N-terminal of A chain that the productive convertase interaction requires extended peptide conformations in both the CA junction (residues 62 ~ 67, LQKRGI) and the BC junction (residues 29 ~34, KTRREA) and leads to significant perturbations in the normally helical N-terminal region of the A chain and the extended C-terminal region of the B chain of the insulin moiety of proinsulin. Recently the heteronuclear 3D NMR data of ^{15}N-DKP-HPI and ^{13}C/

^{15}N-DKP-HPI were collected and analyzed in our group. The determination of three-dimensional structure of HPI and further point-mutagenesis at C-peptide, (especially in CA or BC function) will explore the proinsulin from structure to processing in detail. Actually some studies of single-chain insulin are related to HPI study by use of C-peptide (or linker between C-terminal of B chain and N-terminal of A chain) with different length. The linker of PIP is dipeptide (AK)[74,93], that of M2PI is nanopeptide (RRYPGDVKR)[101] and that of hybrid-insulin is C-domain of IGF-Ⅰ[92]. Taken together, these studies will facilitate to understand the structure, function, evolution, processing and storage for insulin and IGF-Ⅰ.

4.3 D-Ala substitutions

D-Ala or other D-isomer amino acids are not existed in normal cell except in some cell wall of bacteria. But it is good to be used in conformational study of protein folding. The D-Ala or D-Trp substitution at A1 position rescues the L-amino acid replacement at same position in receptor binding[102, 103]. The L-substitution perturbed the binding surface of receptor, whereas the orientation of the side chain of D amino acids turns to the other direction. As described above D-Ala^{B24} substitution increased the binding affinity from 4.3 % of L-Ala^{B24} to 255%. The NMR structure of D-Ala^{B26}-Des-(B27-B30)-insulin-amide was reported with 12.5-fold increase in binding affinity by Wollmer[104], which retains the folding of native insulin with truncation of tetra-peptide at C-terminal of B chain. Interestingly the potency of D-Ala^{B26}- insulin is only 18% related to that of insulin, implying that the tetra-peptide deletion exposed the major part of receptor binding surface and facilitated the binding. The amide C-terminal of B chain also contributed to the higher potency similar to DPI-amide (100% affinity) and DPI (20% potency). By the way, from their conformational analysis Wollmer[104] also criticized the major conclusion of Ludvigsen[31] that the side-chain of B25 Phe is inward in Glu^{B16}, Gly^{B24}-desB30 insulin.

The NMR structure of D-Ala^{B8}-DKP-insulin retains identical native insulin folding with only 0.2% potency (unpublished data). The slowly exchanged amide NMR experiments exhibited that the B8 D-Ala mutant is more rigid protein with slower rate of exchange, which consists with the higher thermodynamic stability of 6.4 kcal/mol (1.5 kcal/mol higher than that of DKP-insulin) from guanidine-hydrochloride titration monitored by CD 222 nm. The receptor binding affinity, stability of NMR structure of D-Ala^{B8}-DKP-insulin contravenes to the proposal that the potency and stability of insulin are directly correlated[40, 41].

As described above there are two structural switch of insulin: B8 and B23. B8 Gly is a key residue for transition between T- and R-state of insulin. To date it is still not clear that what state of insulin adopts while it binds to receptor. The NMR structure and stability D-Ala^{B8}-DKP-insulin imply that at least insulin is not a T-state in binding.

4.4 B24-B25 ester bond causes inactive

Wollmer also reported NMR structure of insulin with wild type sequence and B24- B25 ester bond[105]. The structure looks the same as that of native insulin, but the depsi-insulin ([B24-B25 CO-O] insulin) has only 3% ~4% potency. The modification broke the H-bond between the amide proton of B25 and the carbonyl oxygen of A19, however the C-terminal of B chain looked no further detachment. The crystal structural studies of Asn^{A21} substituted by Gly, Ala, Asp and Gln respectively exhibited the conformational correlation and coupled motion between residue A21 and B25 side chain[106]. Invariant residue A21 may affect the receptor binding by interacting with the B25 side chain and the B chain C-terminal segment to assist the B25 side chain rearranging into the "binding" conformation. The lack of H-bond between B25 and A19 or the replacement of A21 both affect the position of B25 Phe side chain, and consequently affect the receptor binding. Considering the above observations together, the local environment of B25 and its orientation of side chain are very sensitive for receptor binding.

4.5 Chiral mutagenesis exhibits the binding interface of receptor binding in detail

The replacement of residue with β-chiral center, such as Allo-Ile and Allo-Thr, allow us to understand the local environment of the side-chain and the contacts of the residue in detail. A2 Ile is an invariant residue. Ala substitution at A2 impairs insulin's activity by 200 fold[107]. Allo-Ile residue exchanges the position of $C^{\gamma'}H_3$ with the $C^{\gamma}H_2$ and $C^{\delta}H_3$ methyl group. The Allo-Ile substitution at A2 causes reduction of binding 50 fold lower[108]. The NMR structure exhibits that the Ile^{A2} side-chain inserts within a chiral pocket of the receptor as part of insulin's hidden functional surface. The low activity of the chiral mutant reflects steric incompatibility between the inverted side-chain and a chiral pocket in the insulin receptor. From the studies of AlaA2 and Allo-Ile^{A2} mutants, one can expect that the chiral pocket of receptor for insulin A2 position need at least one γ- or δ-CH_3 from "active-insulin", and binding will completely lose if only one β-CH_3 exists in A2 position of insulin. The NMR study of Allo-Thr mutagenesis at A3 and B12 position are underway.

5 Perspectives

As described above, insulin has been extensively studied by variety methods, but there remains

uncertainty regarding the biologically active conformation and the structural features that constitute the receptor-binding domain and other related unclear issues.

5.1 Insulin-insulin receptor binding

What is the structure of insulin receptor and its complex with "active" insulin? Is there any conformational change of IR in binding? What is the binding surface of IR? The NMR study of "minimum" binding domain of IR (L1-CR-L2 fragment) with and without insulin will explore the final structure of the complex. To initiate the TROSY-based NMR study of the IR complex with insulin, the complex of $^{13}C/^{15}N$-labeled insulin with unlabelled 70 ku minimizing binding domain of IR will make it possible to address the structure of "active-binding" insulin. Then NMR study of $^{13}C/^{15}N/^{2}H$ labeled "minimum" IR domains allows us to determine its structure, and followed by the complex study. The 900 MHz cryo-probe may facilitate the measurements. The additional mutagenesis of IR or insulin will get insight into the interaction surface in detail. Using residual dipolar couplings in liquid crystalline solutes will dramatically expand the repertoire of NMR experimental constraints[109]. The partial alignment of the nuclei allows measurement of inter-nuclear orientations by use of residual dipolar couplings for nuclei of known distance, and consequently allows refinement of the structure of the complex. The new technical progress for NMR study is segmental isotopic labeling using expressed protein ligation[110], which allows us labeling N-terminal domain of the multi-domains protein and explore its interface or monitor the interaction with other part of the molecule or other proteins (ligand, inhibitor etc.). By use of such technique L1-L2 domain might be labeled in intact IR and determine the binding surface of L1-L2 with insulin, which will simplify the analysis of intact IR. The second binding surface of insulin was reported recently that the hexamer-forming related residue A13 and B17 also participate in receptor binding, which is in opposite side of the "classical" binding surface[49, 52]. The structural determination of IR/insulin complex will get insight into these two opposite binding surfaces of insulin finally. The crystal or NMR structure of the complex of IR/insulin will allow us to verify the De Meyts-Schuffer model of receptor binding mechanism[49], and further understand the signaling of insulin and design of insulin-receptor agonists, as well as antagonists.

5.2 *C. elegans* insulin and other invertebrate insulin-like peptide

The reported NMR structure of *C. elegans* INS-6 (insulin-like peptide) of invertebrate worm is the first step in study of invertebrate insulin-like peptide[78]. There are other classes of insulin-like peptides with different disulfide bonds. The structural study of the peptides and comparison of the vertebrate insulin make it possible to know more about the insulin evolution. The over-expression of def-2 and other receptors, and following NMR study of the complex allow compare the binding and signaling of human insulin and its receptor. The functional study of *C. elegans* INS proteins will test the hypothesis of the signaling mechanism of the nematode family.

Insulin receptors are expressed in most tissues of the body, including classic insulin-sensitive tissues (liver, muscle, and fat), as well as "insulin-insensitive" tissue, such as red blood cells and neuronal tissue of the central nervous system (CNS). The *C. elegans* INS-6 are expressed in neurons, which provides a model for the action of insulin in brain. This represents another frontier of diabetes research. Insulin is an important neuro-modulator, contributing to neurobiological processes, in particular energy homeostasis and cognition[111, 112]. Further study in brain (or CNS) insulin will allow us understand the multi-functional insulin in detail.

5.3 Insulin fibrillation

Fibrillation study of insulin is an old project, which was first reported in 1940's[113, 114], however now insulin fibrillation study is also a frontier one. When insulin solutions are subjected to acid, heat or agitation, the normal way of insulin assembly (dimmer to tetramer or to hexamer) is disrupted. The conformational changes lead to an alternative aggregation pathway and formation of insoluble amyloid fibril. The factors effected the fibril formation are: sequence of the molecule, disulfide bond breaking or mis-matched, ionic strength, pH, temperature, concentration of protein, stirring speed, organic solvent and seed. Two solution conditions were normally used for fibril study of insulin: pH 2 and 55℃ ~ 70℃, or pH 7, 45℃ with 60% ethanol. The fibrillation study of insulin is an important issue for three reasons: (1) the fibril formation is one of the major problems that limit high yield in pharmaceutical industry, (2) the fibril related amyloid disease is type-Ⅱ diabetes related, (3) the conformational conversion from helix-rich structure to well-organized β-sheet-rich structure is an interesting project in protein folding. The inherent toxicity of aberrant aggregates of insulin or proinsulin is a general feature of protein misfolding disease. The yield of chain combination is a key factor for pharmaceutical chemistry. The correct disulfide pairing is always coupled with folding of reduced proinsulin. The kinetic competition between productive folding and off-pathway aggregation (particularly irreversible B-chain aggregation leading to fibrils) determines the yield of insulin. Mutation at some key position can

block the productive pathway, as described example: Leu^{A16} mutants. Also the mismatching of disulfide bridges will lead to off-pathway. Mapping thermodynamic determinants of disulfide pairing and mapping kinetic determinants of foldability both will enable us to understand how to increase the productive folding and avoid the off-pathway fibrillation.

The Akita mouse insulin is natural mutant of Cys^{A7} to Phe, which induced from mutation of insulin-2 gene[115, 116]. The Akita diabetic mouse is a good model for diabetes research. Another model for diabetes mouse is *Octan-degu*, a small mouse-like rodent in South American, in which adult onset diabetes mellitus is associated with insulin-specific islet amyloidosis[117]. The His^{B10} in native human insulin is replaced to Asn. The His^{B10} is a key residue for zinc binding hexamer formation. The disruption of the formation of stable storage state is a possible reason to form the fibril.

5.4 PDI assistant HPI folding

Protein folding requires the assistant of chaperone and foldase. Protein disulfide isomerase (PDI) is a very important foldase for correct folding of many disulfide-bonded proteins. Although extensive studies had been carried out to understand the catalytic functions of PDI, little is known about its substrate binding. PDI is a multi-domains large protein. The b′ domain of human PDI is essential and sufficient for the binding of small peptides[118]. While the principal peptide-binding site of PDI is provided by b′ domain, all domains contribute to binding of partially folded proteins. The yields of insulin chain combination and oxidative folding of proinsulin are ordinarily limited *in vitro* due to aggregation of the unfolded polypeptide (or B) chains. PDI enhances the yield of proinsulin folding in part through its catalytic activity and in part through its chaperone activity[119]. PDI is both an oxidoreductase and a chaperone[120]. It would be interesting to study the interaction between the b′ domain of PDI and its substrate protein (HPI or its segment) as the initial site and template for the folding of human proinsulin. The structural study of the interaction between PDI (or its isolated b′ domain) and its substrates will further get insight into the initiate state of protein folding *in vitro* and *in vivo*. These studies and characterization the b′ domain and its mechanism of binding to peptides will not only shed light on how PDI enhances the folding of proinsulin, but also provide general insight into structure-activity relationships in PDI. The results thus promise to fill a critical gap in the present state of knowledge in this field.

Last decade the NMR technical developments and other methods brought us more extensive insight into the folding, binding and stability of insulin and the related proteins. This review just summarized the major points of the subjects, which will be useful for the further study, so that the remaining questions can be answered and the ultimate solution to the problem of insulin folding and binding can be at least envisaged.

Acknowledgments The author thanks Michael A. Weiss, M. D., Ph. D, for the support of the long-term insulin study.

Refrences

1 American Diabetes Association. Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care, 1997, **20**: 1183 ~ 1195

2 Amos A, McCarty D, Zimmet P. The rising global burden of diabetes and its complications: estimates and projections to the year 2010. Diabetic Med, 1997, **14**: S1 ~ S85

3 Pan X R, Yang W Y, Li G W, *et al*. Prevalence of diabetes and its risk factors in China. Diabetes Care, 1997, **20**: 1664 ~ 1669

4 Zimmet P, Alberti K, Shaw J. Global and societal implications of the diabetes epidemic. Nature, **414**: 782 ~ 787

5 Dodson G, Steiner D. The role of assembly in insulin′s biosynthesis. Curr Opin Struct Biol, 1998, **8**: 189 ~ 194

6 Baker E N, Blundell T L, Cutfield J F, *et al*. The structure of 2Zn pig insulin crystals at 1.5 Å resolution. Phil Trans Royal Soc London Ser, 1998, **319**: 369 ~ 456

7 Peking Insulin Structure Research Group. Insulin′s crystal structure at 2.5Å resolution. Peking Rev, 1971, **40**: 11 ~ 16

8 Derewenda U, Derewenda Z, Dodson E J. *et al*. Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer. Nature, 1989, **338**: 594 ~ 596

9 Blundell T L, Cutfield J F, Cutfield S M, *et al*. Atomic positions in rhombohedral 2-zinc insulin crystals. Nature, 1971, **231**: 506 ~ 511

10 Bentley G, Dodson E, Dodson G, *et al*. Structure of insulin in 4-zinc insulin. Nature, 1976, **261**: 166 ~ 168

11 Badger J, Harris M R, Reynolds C D, *et al*. Structure of the pig insulin dimer in the cubic crystal. Acta Crystallographica Sec. B, Struct Sci, 1991, **47**: 127 ~ 136

12 Ciszak E, Smith G D. Crystallographic evidence for dual coordination around zinc in the T3R3 human insulin hexamer. Biochemistry, 1994, **33**: 1512 ~ 1517

13 Petruzzelli L, Herrera R, Rosen O M. Insulin receptor is an insulin-dependent tyrosine protein kinase: copurification of insulin-binding activity and protein kinase activity to homogeneity from human placenta. Proc Natl Acad Sci USA, 1984, **81**: 3327 ~ 3331

14 Fiaux J, Bertelsen E B, Horwich A L, *et al*. NMR analysis of a 900K GroEL GroES complex. Nature, 2002, **418**: 207 ~ 211

15 Riek R, Pervushin K, Wuthrich K. TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. Trends Biochem Sci, 2000, **25**: 462 ~ 468

16 Hua Q X, Weiss M A. Toward the solution structure of human insulin: Sequential 2D ^{1}H NMR assignment of a Des-pentapeptide analogue and comparison with crystal structure. Biochemistry, 1990, **29**: 10545 ~ 10555

17 Hua Q X, Weiss M A. 2D-NMR studies of Des-(B26-B30)-insulin: Sequence-specific resonance assignments and effects of solvent composition. Biochim Biophys Acta, 1991, **1078**: 101 ~ 110

18 Hua Q X, Weiss M A. Comparative 2D-NMR studies of human insulin and Des-pentapeptide insulin: Sequential resonance assignment and implication for protein dynamics and receptor

recognition. Biochemistry, 1991, **30**: 5505 ~ 5515
19 Weiss M A, Hua Q X, Lynch C S, *et al.* Hetero-nuclear 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, 1991, **30**: 7373 ~ 7389
20 Hua Q X, Hu S Q, Frank B H, *et al.* Mapping the functional surface of insulin by design: structure and function of a novel a-chain analogue. J Mol Biol, 1996, **264**: 390 ~ 403
21 Ludvigsen S, Roy M, Thgersen H, *et al.* High-resolution structure of an engineered biologically potent insulin monomer, B16 Tyr→His, as determined by nuclear magnetic resonance spectroscopy. Biochemistry, 1994, **33**: 7998 ~ 8006
22 Olsen H B, Ludvigsen S, Kaarsholm N C. Solution structure of an engineered insulin monomer at neutral pH. Biochemistry, 1996, **35**: 8836 ~ 8845
23 Haneda M, Chan S J, Kwok S C M, *et al.* Studies on mutant human insulin genes: identification and sequence analysis of a gene encoding [SerB24] insulin. Proc Natl Acad Sci USA, 1983, **80**: 6366 ~ 6370
24 Sakura H, Iwamoto Y, Sakamoto Y, *et al.* Structurally abnormal insulin in a diabetic patient. Characterization of the mutant insulin A3 (Val to Leu) isolated from the pancreas. J Clin Invest, 1986, **78**: 1666 ~ 1672
25 Shoelson S, Fickova M, Haneda M, *et al.* Identification of a mutant human insulin predicted to contain a serine-for-phenylalanine substitution. Proc Natl Acad Sci USA, 1983, **80**: 7390 ~ 7394
26 Hua Q X, Shoelson S E, Inouye K, *et al.* Paradoxical structure and function in a mutant insulin associated with diabetes mellitus in man. Proc Natl Acad Sci USA, 1993, **90**: 582 ~ 586
27 Derewenda U, Derewenda Z, Dodson E J, *et al.* X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue. J Mol Biol, 1991, **220**: 425 ~ 433
28 Liang D C, Chang W R, Zhang Z P, *et al.* The possible mechanism of binding interaction of insulin molecule with its receptor. Sci China (ser. B), 1992, **35**: 547 ~ 557
29 Hua Q X, Shoelson S E, Kochoyan M, *et al.* Receptor binding redefined by a structural switch in a mutant human insulin. Nature, 1991, **354**: 238 ~ 241
30 Hua Q X, Shoelson S E, Weiss M A. Nonlocal structural perturbations in a mutant human insulin : sequential resonance assignment and ^{13}C-isotope-aided 2D-NMR Studies of [PheB24-Gly] insulin with Implications for receptor recognition. Biochemistry, 1992, **31**: 11940 ~ 11951
31 Ludvigsen S, Olsen H B, Kaarsholm N C. A structural switch in a mutant insulin exposes key residues for receptor binding. J Mol Biol, 1998, **279**: 1 ~ 7
32 Sosnick T R, Fang X, Shelton V M. Application of circular dichroism to study RNA folding transitions. Methods Enzymol, 2000, **317**: 393 ~ 409
33 Pullen R A, Lindsay D G, Wood SP, *et al.* Receptor-binding region of insulin. Nature, 1976, **259**: 369 ~ 373
34 Kristensen C, Kjeldsen T, Wiberg F C, *et al.* Alanine scanning mutagenesis of insulin. J Biol Chem, 1997, **272** (20): 12978 ~ 12983
35 Nakagawa S H, Tager H S. Importance of aliphatic side-chain structure at positions 2 and 3 of the insulin A chain in insulin-receptor interactions. Biochemistry, 1992, **31**: 3204 ~ 3214
36 Nakagawa S H, Tager H S, Steiner D F. Mutational analysis of invariant valine B12 in insulin: implications for receptor binding. Biochemistry, 2000, **34**: 15826 ~ 15835
37 Weiss M A, Nakagawa S H, Jia W, *et al.* Protein structure and the spandrels of San Marco: insulin's receptor-binding surface is buttressed by an invariant leucine essential for its stability. Biochemistry, 2002, **41**: 809 ~ 819
38 Hua Q X, Chu Y C, Jia W, *et al.* Mechanism of insulin chain combination. Asymmetric roles of A-chain α-helices in disulfide pairing. J Biol Chem, 2002, **277**: 43443 ~ 43453
39 Katsoyannis P G, Tometsko A. Insulin synthesis by recombination of A and B chains: a highly efficient method. Proc Natl Acad Sci USA, 1966, **55**: 1554 ~ 1561
40 Kaarsholm N C, Norris K, Jorgensen R J, *et al.* Engineering stability of the insulin monomer fold with application to structure-activity relationships. Biochemistry, 1993, **32**: 10773 ~ 10778
41 Olsen H B, Ludvigsen S, Kaarsholm N C. The relationship between insulin bioactivity and structure in the NH_2-terminal A-chain helix. J Mol Biol, 1998, **284**: 477 ~ 488
42 Weiss M A, Hua Q X, Jia W, *et al.* Activities of monomeric insulin analogs at position A8 are uncorrelated with their thermodynamic stabilities. J Biol Chem, 2001, **276**: 40018 ~ 40024
43 Adams T E, Epa V C, Garrett T P J, *et al.* Structure and function of the type 1 insulin-like growth factor receptor. Cell Mol Life Sci, 2000, **57**: 1050 ~ 1093
44 Petruzzelli L, Herrera R, Rosen O M. Insulin receptor is an insulin-dependent tyrosine protein kinase: co-purification of insulin-binding activity and protein kinase activity to homogeneity from human placenta. Proc Natl Acad Sci USA, 1984, **81**: 3327 ~ 3331
45 Garrett T P, McKern N M, Lou M, *et al.* Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. Nature, 1998, **394**: 395 ~ 399
46 Ullrich A, Bell J R, Chen E Y, *et al.* Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature, 1985, **313**: 756 ~ 761
47 Siddle K. The insulin receptor and type Ⅰ IGF receptor: comparison of structure and function. Prog Growth Factor Res, 1992, **4**: 301 ~ 320
48 Marino-Buslje C, Martin-Martinez M, Mizuguchi K, *et al.* The insulin receptor: from protein sequence to structure. Biochem Soc Trans, 1999, **27**: 715 ~ 726
49 de Meyts P, Whittaker J. Structural biology of insulin and IGF1 receptors: implications for drug design. Nat Rev Drug Discovery, 2002, **1**: 769 ~ 783
50 Luo R Z, Beniac D R, Fernandes A, *et al.* Quaternary structure of the insulin-insulin receptor complex. Science, 1999, **285**: 1077 ~ 1080
51 de Meyts P. 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, 1994, **37** (Suppl 2), S135 ~ S148
52 Schaffer L. A model for insulin binding to the insulin receptor. Eur J Biochem, 1994, **221**: 1127 ~ 1132
53 Kurose T, Pashmforoush M, Yoshimasa Y, *et al.* 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, 1994, **269**: 29190 ~ 29197
54 Williams P F, Mynarcik D C, Yu G Q, *et al.* Mapping of an NH_2-terminal ligand binding site of the insulin receptor by alanine scanning mutagenesis. J Biol Chem, 1995, **270**: 3012 ~ 3016
55 Mynarcik D C, Yu G Q, Whittaker J. Alanine-scanning mutagenesis of a C-terminal ligand binding domain of the insulin receptor alpha subunit. J Biol Chem, 1996, **271**: 2439 ~ 2442
56 Anfinsen C B. Principles that govern the folding of protein chains. Science, 1973, **181**: 223 ~ 230

57 Creighton T E. Folding pathways determined using disulfide bonds. In: Creighton T E ed. Protein Folding. New York: W. H. Freeman, 1992. 301 ~ 351

58 Steiner D F. Evidence for a precursor in the biosynthesis of insulin. TransN Y Acad Sci, 1967, **30**: 60 ~ 68

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

60 Wang C C, Tsou C L. The insulin A and B chains contain sufficient structural information to form the native molecule. Trends Biochem Sci, 1991, **16**: 279 ~ 281

61 Hua Q X, Qian Y Q, Tsou C L. The interaction of the S-thiomethyl insulin A and B chains in solution. Biochim Biophys Acta, 1984, **789**: 234 ~ 240

62 Hua Q X, Gozani S N, Chance R E, *et al*. Structure of a protein in a kinetic trap. Nature Struct Biol, 1995, **2**: 129 ~ 138

63 Hua Q X, Jia W, Frank B H, *et al*. A protein caught in A kinetic trap: Structures and stabilities of insulin disulfide isomers. Biochemistry, 2002, **41**: 14700 ~ 14715

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

65 Brzozowski A M, Dodson E J, Dodson G G, *et al*. Structural origins of the functional divergence of human insulin-like growth factor-Ⅰ and insulin. Biochemistry, 2002, **41**: 9389 ~ 9397

66 Hober S, Forsberg G, Palm G, *et al*. Disulfide exchange folding of insulin-like growth factor Ⅰ. Biochemistry, 1992, **31**: 1749 ~ 1756

67 Miller J A, Narhi L O, Hua Q X, *et al*. Oxidative refolding of insulin-like growth factor 1 yields two products of similar thermodynamic stability: a bifurcating protein-folding pathway. Biochemistry, 1993, **32**: 5203 ~ 5213

68 Narhi L O, Hua Q X, Arakawa T, *et al*. Role of native disulfide bonds in the structure and activity of insulin-like growth factor 1: genetic models of protein-folding intermediates. Biochemistry, 1993, **32**: 5214 ~ 521

69 Hua Q X, Hu S Q, Frank H, *et al*. Mapping the functional surface of insulin by design: Structure and function of a novel A-chain analogue. J Mol Biol, 1996, **264**: 390 ~ 403

70 Weiss M A, Hua Q X, Jia W, *et al*. Hierarchical protein "undesign": insulin's intrachain disulfide bridge tethers a recognition alpha-helix. Biochemistry, 2000, **39**: 15429 ~ 15440

71 Hua Q X, Nahri L, Jia W, *et al*. 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, 1996, **259**: 297 ~ 313

72 Hua Q X, Nakagawa S H, Jia W, *et al*. Hierarchical protein folding: asymmetric unfolding of an insulin analogue lacking the A7 ~ B7 interchain disulfide bridge. Biochemistry, 2001, **40**: 12299 ~ 12311

73 Yan H, Guo Z Y, Gong X W, *et al*. A peptide model of insulin folding intermediate with one disulfide. Protein Sci, 2003, **12**: 768 ~ 775

74 Qiao Z S, Guo Z Y, Feng Y M. Putative disulfide-forming pathway of porcine insulin precursor during its refolding *in vitro*. Biochemistry, 2001, **40**: 2662 ~ 2668

75 Milner S J, Carver J A, Ballard F J, *et al*. Probing the disulfide folding pathway of insulin-like growth factor-Ⅰ Biotechnol. Bioeng, 1999, **62**: 693 ~ 703

76 Qiao Z S, Min C Y, Hua Q X, *et al*. *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, 2003, **278**: 17800 ~ 17809

77 Tsou C L. Folding of the nascent peptide chain into a biologically active protein. Biochemistry, 1988, **27**: 1809 ~ 1812

78 Hua Q X, Nakagawa Wilken J, Ramos R R, *et al*. A divergent INS protein in Caenorhabditis elegans structurally resembles human insulin and activates the human insulin receptor. Genes & Develop, 2003, **17**: 826 ~ 831

79 Weiqing L, Kennedy S G, Ruvkun G. daf-28 encodes a *C. elegans* insulin superfamily member that is regulated by environmental cues and acts in the DAF-2 signaling pathway. Genes & Develop, 2003, **17**: 844 ~ 858

80 Pierce S B, Costa M, Wisotzkey R, *et al*. 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 & Develop, 2001, **15**: 672 ~ 686

81 Kimura K D, Tissenbaum H A, Liu Y, *et al*. daf-2, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. Science, 1997, **277**: 942 ~ 946

82 Murray-Rust J, McLeod A N, Blundell T L, *et al*. Structure and evolution of insulins: implications for receptor binding. Bioessays, 1992, **14**: 325 ~ 331

83 Hua Q X, Hu S Q, Jia W H, *et al*. Mini-proinsulin and mini-IGF-Ⅰ: Homologous protein sequences encoding Non-homologous structures. J Mol Biol, 1998, **277**: 103 ~ 118

84 Humbel R E. Insulin-like growth factors Ⅰ and Ⅱ. Eur J Biochem, 1990, **190**: 445 ~ 462

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

86 Sato A, Nishimura S, Ohkubo T, *et al*. Three-dimensional structure of human insulin-like growth factor-Ⅰ determined by ^{1}H-NMR and distance geometry. Int J Pept Protein Res, 1993, **41**: 433 ~ 440

87 Torres A M, Forbes B E, Aplin S E, *et al*. Solution structure of human insulin-like growth factor Ⅱ. Relationship to receptor and binding protein interactions. J Mol Biol, 1995, **248**: 385 ~ 401

88 Gill R, Verma C, Wallach B, *et al*. Modelling of the disulphide-swapped isomer of human insulin-like growth factor-1: implications for receptor binding. Protein Eng, 1999, **12**: 297 ~ 303

89 Sato A, Koyama S, Yamada H, *et al*. Three-dimensional solution structure of a disulfide bond isomer of the human insulin-like growth factor-Ⅰ. J Pept Res, 2000, **56**: 218 ~ 230

90 de Wolf E, Gill R, Geddes S, *et al*. Solution structure of a mini IGF-Ⅰ. Protein Sci, 1996, **5**: 2193 ~ 2202

91 Zong L, Burke G T, Katsoyannis P G. An insulin-like hybrid consisting of a modified A-domain of human insulin-like growth factor Ⅰ and the B-chain of insulin. J Protein Chem, 1990, **9**: 389 ~ 395

92 Kristensen C, Andersen A S, Hach M, *et al*. A single-chain insulin-like growth factor Ⅰ/insulin hybrid binds with high affinity to the insulin receptor. Biochem J, 1995, **305**: 981 ~ 986

93 Guo Z Y, Shen L, Feng Y M. The different folding behavior of insulin and insulin-like growth factor 1 is mainly controlled by their B-chain/domain. Biochemistry, 2002, **41**: 1556 ~ 1567

94 Eigenbrot C, Randal M, Quan C, *et al*. X-ray structure of human relaxin at 1.5Å. Comparison to insulin and implications for receptor binding determinants. J Mol Biol, 1991, **221**: 15 ~ 21

95 Nagata K, Hatanaka H, Kohda D, *et al*. 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, 1995, **253**: 749 ~ 758

96 Nagata K, Hatanaka H, Kohda D, *et al.* 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, 1995, **253**: 759 ~770

97 Smit A B, Vreugdenhil E, Ebberink R H, *et al.* Growth-controlling molluscan neurons produce the precursor of an insulin-related peptide. Nature, 1988, **331**: 535 ~538

98 Oliva A A, Steiner D F, Chan S J. Proprotein convertases in amphioxus: predicted structure and expression of proteases SPC2 and SPC3. Proc Natl Acad Sci USA, 1995, **92**: 3591 ~3595

99 Steiner D F. The proprotein convertases. Curr Opin Chem Biol, 1998, **2**: 31 ~39

100 Steiner D F, Lipkind G. Predicted structural alterations in proinsulin during its interactions with prohormone convertases. Biochemistry, 1999, **38**: 890 ~896

101 Chang S G, Kim D Y, Choi K D, *et al.* Human insulin production from a novel mini-proinsulin which has high receptor-binding activity. Biochem J, 1998, **329**: 631 ~635

102 Cosmatos A, Cheng K, Okada Y, *et al.* The chemical synthesis and biological evaluation of [1-L-alanine-A] -and [1-D-alanine-A] -insulins. J Biol Chem, 1978, **253**: 6586 ~6590

103 Wan Z L, Liang D C. Studies on the crystal structure of A1-(D-Tryptophan) insulin. Sci China B, 1990, **33**: 810 ~820

104 Kurapkat G, Siedentop M, Gattner H G, *et al.* The solution structure of a superpotent B-chain-shortened single-replacement insulin analogue. Protein Sci, 1999, **8**: 499 ~508

105 Kurapkat G, de Wolf E, Grotzinger J, *et al.* Inactive conformation of an insulin despite its wild-type sequence. Protein Sci, 1997, **6**: 580 ~587

106 Zeng Z H, Liu Y S, Jin L, *et al.* 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, 2000, **1479**: 225 ~236

107 Xu B, Hua Q X, Nakagawa S H, *et al.* A cavity-forming mutation in insulin induces segmental unfolding of a surrounding α-helix. Protein Science, 2002, **11**: 104 ~116

108 Xu B, Hua Q X, Nakagawa S H, *et al.* Chiral mutagenesis of insulin's hidden receptor-binding surface: structure of an Allo-isoleucine^{A2} analogue. J Mol Biol, 2002, **316**: 435 ~441

109 Tjandra N, Bax, A. Direct measurement of distances and angles in biomolecules by NMR in dilute liquid crystalline medium. Science, 1997, **278**: 1111 ~1114

110 Cowburn D, Muir T W. Segmental isotopic labeling using expressed protein ligation. Methods Enzymol, 2001, **338**: 41 ~54

111 Bruning J C, Gautam D, Burks D J, *et al.* Role of brain insulin receptor in control of body weight and reproduction. Science, 2000, **289**: 2122 ~2125

112 Kyriaki G. Brain insulin: regulation, mechanisms of action and functions. Cell Mol Neurobiol, 2003, **23**: 1 ~25

113 Waugh D F. The properties of protein fibers produced reversibly from soluble protein molecules. Am J Physiol, 1941, **133**: 484 ~485

114 Waugh D F. A fibrous modification of insulin. Ⅰ. The heat precipitation of insulin. J Am Chem Soc, 1946, **68**: 247 ~250

115 Yaguchi M, Nagashima K, Izumi T, *et al.* Neuropathological study of C57BL/6Akita mouse, type 2 diabetic model: enhanced expression of alpha B-crystallin in oligodendrocytes. Neuropathology, 2003, **23**: 44 ~50

116 Izumi T, Yokota-Hashimoto H, Zhao S, *et al.* Dominant negative pathogenesis by mutant proinsulin in the Akita diabetic mouse. Diabetes, 2003, **52**: 409 ~416

117 Nishi M, Steiner D F. Cloning of complementary DNAs encoding islet amyloid polypeptide, insulin, and glucagon precursors from a New World rodent, the degu, Octodon degus. Mol Endocri, 1990, **4**: 1192 ~1198

118 Klappa P, Ruddock L W, Darby N J, *et al.* The b′ domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins. EMBO J, 1998, **17**: 927 ~935

119 Winter J, Klappa P, Freedman R, *et al.* Catalytic activity and chaperone function of human protein-disulfide isomerase are required for theefficient refolding of proinsulin. J Biol Chem, 2002, **277**: 310 ~317

120 Wang C C, Tsou C L. Protein disulfide isomerase is both an enzyme and a chaperone. FASEB J, 1993, **7** (15): 1515 ~1517