ANTIOXIDANTS & REDOX SIGNALING
Volume 10, Number 1, 2008

DOI: 10.1089/ars.2007.1860

Forum Review

The *In Vitro* Oxidative Folding of the Insulin Superfamily

ZHAN-YUN GUO,[1] ZHI-SONG QIAO,[2] and YOU-MIN FENG[3,4]

ABSTRACT

Insulin and related proteins, which have been found not only in mammals, birds, reptiles, amphibians, fish, and cephalochordate, but also in mollusca, insects, and *Caenorhabditis elegans*, form a large protein family, the insulin superfamily. In comparing their amino acid sequences, a common sequence characteristic, the insulin structural motif, is found in all members of the superfamily. The structural motif is deduced to be the sequence basis of the identical disulfide linkages and similar three-dimensional structures of the superfamily. The insulin superfamily provides a series of disulfide-containing proteins for the studies of *in vitro* oxidative folding. The *in vitro* folding pathways of insulin-like growth factor-1 (IGF-1), porcine insulin precursor (PIP), human proinsulin, and *Amphioxus* insulin-like peptide (AILP) have been established by capture and analysis of the folding intermediates during their *in vitro* oxidative folding process. The family also provides an excellent system for study of the sequence structure relation: insulin and IGF-1 share high amino acid sequence homology, but they have evolved different folding behaviors. The sequence determinants of their different folding behaviors have been revealed by analyzing the folding behaviors of those global and local insulin/IGF-1 hybrids. *Antioxid. Redox Signal.* 10, 127–139.

INTRODUCTIONS

The studies of insulin almost crown all of the fields of protein science since insulin was found more than 80 years ago. Many of the accomplishments of insulin studies are the milestones of biochemistry, such as its amino acid sequence (4), its total chemical synthesis (44, 45, 76), its crystal structure (2), and its cell-membrane receptor (15, 20). However, insulin has not become a model protein for oxidative refolding studies until recently because of its double-chain nature. Protein folding can be simply considered as a process by which a biologically inactive polypeptide chain is converted to a unique three-dimensional structure with specific biological activity. Anfinsen and co-workers (1) first demonstrated that the three-dimensional structure of a globular protein is uniquely determined by its amino acid sequence in 1960s, which established the basis of *in vitro* protein-folding studies. Since then, significant advances have been made in the understanding of protein folding through experimental and theoretical approaches. With small disulfide-containing proteins, such as bovine pancreatic trypsin inhibitor (14), RNase A (60), and hirudin (8), the folding intermediates have been isolated and characterized, and the folding pathways have been revealed. The oxidative folding provides a unique method for protein-folding studies: capturing the folding intermediate during refolding process and reconstituting the folding pathways.

In the protein-folding studies, revealing the *in vitro* folding pathway is an important approach to understand the mechanism of protein folding. For protein folding pathway analysis, the disulfide-containing proteins are frequently chosen as models

[1]Institute of Protein Research, Tongji University, Shanghai, China.
[2]Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts.
[3]Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China.
[4]College of Life Science, Zhejiang Sci-Tech University, Hangzhou, China.
The first two authors contributed equally to this work.

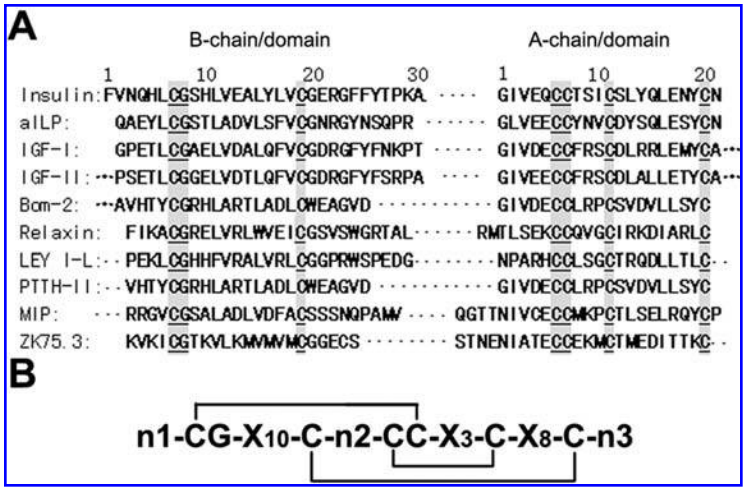

FIG. 1. Amino acid sequences of some insulin superfamily members (A) and the putative insulin structural motif (B). ILP, *Amphioxus* insulin-like peptide; IGF-I, human insulin-like growth factor-I; IGF-II, human insulin-like growth factor-II; Bom-2, bombyxin-2; relaxin, porcine relaxin; LEY I-L, Leydig insulin-like protein; PTTH-II, prothoracicotrophic hormone-II; MIP, mollucan insulin-related peptide; ZK75.3, *C. elegans* insulin-like protein.

for the following reasons: (a) the formation of disulfide(s) is coupled with the conformational folding process (6, 13, 71); (b) the mechanism of thiol/disulfide exchange reaction is clear and can be well controlled *in vitro* (21); and (c) the folding process can be stopped, and the folding intermediate can be trapped either by modification of the free thiols or by acidification of the refolding solution. Subsequently, the trapped intermediates can be purified and their properties can be analyzed (13). Therefore, the disulfide(s) can be used as a unique probe for the protein-folding studies.

The insulin superfamily members, such as recombinant porcine insulin precursor (PIP), human proinsulin (HPI), *Amphioxus* insulin-like peptide (AILP), and insulin-like growth factor-1 (IGF-1), provide a series of small disulfide-containing proteins that can be used for oxidative protein-folding studies. Additionally, the superfamily also provides an excellent model system to study the sequence structure relation: the insulin/IGF-1 system in which homologous amino acid sequences encode different folding behaviors (31, 32). In this review, we focus on the recent advances of the following parts: the insulin superfamily and its structural motif; the *in vitro* oxidative folding pathways of single-chain insulin and its related proteins; and the molecular basis and evolution of the different folding behaviors of insulin and IGF-1.

INSULIN SUPERFAMILY AND INSULIN STRUCTURAL MOTIF

Amino acid sequences and three-dimensional structures of the insulin superfamily

Insulin and insulin-like peptides have been found not only in vertebrates such as mammals, birds, reptiles, amphibians, and fish, but also in cephalochordates, such as *Amphioxus* (5). Additionally, they are also found in mollusks (62), insects (52), and *C. elegans* (19). Insulin and its related proteins form a large protein superfamily, the insulin superfamily (Fig. 1A). All of the superfamily members share a common amino acid sequence characteristic, termed the insulin structural motif (Fig. 1B) (9, 19, 51, 78). Whether the mature proteins are double-chain forms (such as insulin and relaxin) or single-chain forms (such as IGF-1 and IGF-2), they are synthesized *in vivo* as single-chain precursors. In the insulin structural motif, six cysteine residues and a glycine residue (B8 position in human insulin) are absolutely conserved, and the residue numbers between two intrachain/domain cysteines, X10, X3, and X8, are invariable. The six absolutely conserved cysteine residues form three disulfide bonds in the folded proteins: two interchain/domain bonds (A7–B7 and A20–B19 in insulins) and one intra-chain/do-

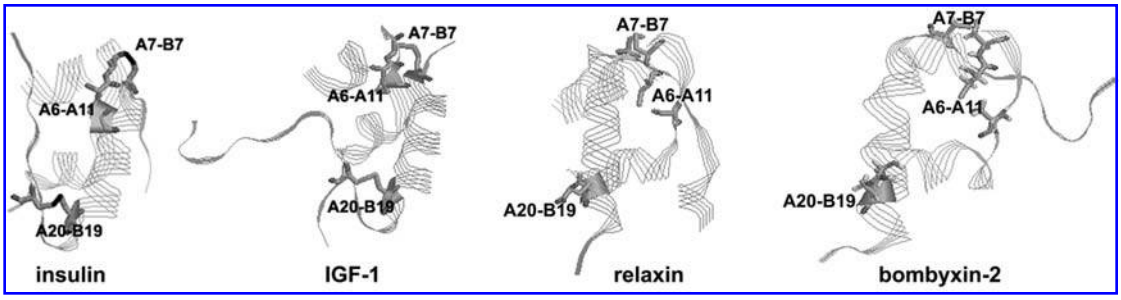

FIG. 2. Schematic representation of the three-dimensional structures of insulin, IGF-1 (native), relaxin, and bombyxin-2. For clarity, the disulfides in IGF-1, relaxin, and bombyxin were named as those of insulin.

main bond (A6–A11 in insulins). The three disulfide bonds are crucial for the native structure and biological function of the superfamily.

Besides sharing a common structural motif, the members of the insulin superfamily also have similar three-dimensional structures (Fig. 2). The structures of the superfamily are mainly encoded by the sequence of the A- and B-chain/domains and stabilized by three disulfide bonds. Their tertiary structures are composed mainly of 3 α-helical segments: one is located at the central part of B-chain/domain (H1); the other two segments are located at the N-terminal and the C-terminal, respectively, of the A-chain/domain (H2 and H3). These α-helical segments are hierarchically stabilized by the three disulfide bonds. These superfamily members can be further classified according to their subtle structural differences: the insulin group and the relaxin group. In the insulin group, including insulin, IGF-1, and IGF-2, a β-turn structure is found after the H1 α-helical segment in the B-chain/domain. The β-turn structure allows the C-terminal fragment of the B-chain/domain to interact with the major part of the molecule and be involved in biological functions. In the relaxin group, including relaxin and bombyxin, no β-turn structure is found after the H1 α-helical segment in the B-chain.

Insulin is the prototype of the insulin superfamily and has been extensively studied in past decades. Herein, insulin is used as a model to discuss the contribution of the invariable structural elements to its structure, function, and folding.

Contribution of the disulfides to the structure, function, and folding of insulin

Insulin is a small globular protein composed of A- and B-chains linked by three disulfide bonds, one intrachain bond, A6–A11, and two interchain bonds, A7–B7 and A20–B19. The amino acid sequence of human insulin is shown in Fig. 1A. The three-dimensional structure of insulin has been thoroughly studied with x-ray crystallography (Fig. 2) (2) and NMR spectroscopy (55, 59, 73). The insulin monomer is a compact globular protein with a hydrophobic core that is composed mainly of those invariable hydrophobic residues, such as B6Leu, B11Leu, B15Leu, and A16Leu. The A-chain contains two α-helixes (A2Ile–A8Thr and A13Leu–A19Tyr, designated as H2 and H3, respectively), and the B-chain contains one central α-helical segment (B9Ser–B19Cys, designated H1). After the H1 helical segment, a β-turn structure makes that C-terminal part of the B-chain form a contact with the major part of insulin and form the receptor-binding surface. In the native structure, the N-terminus of the A-chain is close to the C-terminus of the B-chain, which is the structural basis of the recombinant expression of single-chain insulins, in which the B- and A-chains are linked together by a short peptide linker.

Insulin contains six absolutely conserved cysteine residues that form three disulfide linkages. As shown in Fig. 2, the intra-chain disulfide bond A6–A11 is buried inside the insulin molecule and close to the hydrophobic core, whereas the interchain disulfide bond A20–B19 is only partially buried, and the disulfide A7–B7 is fully exposed to the solvent. Although the role of disulfides had been studied in 1970s by chemical modification (3), only recently was the role of each disulfide elucidated unambiguously by protein engineering (Table 1).

When the disulfide bond A6–A11 was removed by substitution of the two cysteines with serine or alanine, the native structure of insulin was disturbed: the α-helical segment at the N-terminus of the A-chain (H2) was unfolded, whereas other parts still retained the native conformation (36, 40, 72). The structural disturbance led to significant activity loss because the N-terminal of the A-chain is involved in receptor binding: the mutant with alanine substitution retained $\sim$1.6–4.6% receptor-binding activity, whereas the mutant with a serine replacement retained 0.1–0.3% receptor-binding activity (16, 36, 40, 72). Therefore, alanine substitution of the intra A-chain disulfide has less detriment than serine replacement, because this disulfide is located in the hydrophobic core. Although deletion of the intra-chain disulfide bond has a significant detriment on both structure and activity, it has little effect on the *in vitro* folding of PIP or proinsulin (7, 43, 48), implying that the formation of the intrachain disulfide A6–A11 is not a critical step during the insulin *in vitro* refolding (74).

Deletion of the interchain disulfide A7–B7 (substitution of the two cysteines with serine or glutamic acid) caused more serious structural disturbance, although the disulfide is completely exposed and was expected to be unimportant previously based on the results of chemical modifications (23, 24, 39): both the α-helix at the N-terminus (H2) and part of the α-helix at the C-terminus (H3) were unfolded within the A-chain, but the B-chain was not disturbed. The greater structural disturbance caused by removal of the disulfide A7–B7 resulted in greater activity loss: the receptor-binding activity of the mutant was decreased more than 1,000-fold compared with that of native insulin (23, 24, 39, 70). However, deletion of this interchain disulfide had no negative effect on *in vitro* refolding of PIP or proinsulin (7, 43, 48), suggesting that the formation of the disulfide A7–B7 is not a critical step during insulin refolding.

Deletion of the interchain disulfide A20–B19 had the greatest structural disturbance (7, 23, 70). More important, removal of the disulfide A20–B19 could significantly decrease the *in vitro* refolding efficiency (7, 70), whereas the mutant with the single disulfide A20–B19 could fold quickly and efficiently *in vitro* (74), suggesting that the disulfide A20–B19 plays a critical role during insulin refolding (7, 56, 70, 74). In summary, the contributions of the three disulfide bonds to the structure, activity, and refolding of insulin can be ranked as A20–B19 $>$ A7–B7 $>$ A6–A11.

The role of the disulfide bonds of human IGF-1 also was studied by pairwise disulfide deletion (54). The intact structure, function, and stability of IGF-1 also require the presence of three disulfide bonds. The analogue lacking all the three disulfides is actually inactive and unfolded. However, introduction of disulfide bond 18–61 (corresponding to disulfide B19–A20 in insulin) results in a compact partially folded state with 1% of IGF-1 receptor binding activity. After the introduction of either disulfide 6–48 (corresponding to disulfide A7–B7 in insulin) or disulfide 47–52 (corresponding to disulfide A6–A11 in insulin), additional but incomplete structural organization and biological function can be observed. In both (18–61) and (6–48, 18–61) analogues, a native-like microdomain in the neighborhood of disulfide bond 18–61 is observed with 2D-NMR, which defines a hydrophobic core that is critical for the initiation of the IGF-1 folding.

Contribution of the absolutely conserved B8Gly to the structure, function, and folding of insulin

In the insulin structural motif, an absolutely conserved glycine residue (B8Gly in human insulin) exists (Fig. 1B). The B8Gly is located at the conjunction of an invariant α-helical segment (H1) and a segment of variable conformation (B1Phe–B8Gly). The latter segment undergoes T (extended) $\rightarrow$ R (helical) transition in the insulin hexamer. In the R state of insulin, the main-chain dihedral angles of the B8 position are those of L-amino acids, whereas in the T-state, the main-chain dihedral angles are those of D-amino acids. L-Amino acid substitutions of B8Gly almost abolished the insulin chain recombination, significantly decreased the expression efficiencies of single-chain insulin (PIP) in yeast cells, the *in vitro* folding efficiencies, structural stabilities, and disulfide stabilities. However, the mutants with L-amino acid substitutions at the B8 position usually retained significant amounts of receptor-binding activity compared with their counterparts with D-amino acid substitutions. For those mutants with D-amino acid substitutions at B8 position, they retained native or even better foldabilities (high chain recombination efficiencies and more stable structures), but these mutants lost the biological activities completely (28, 29, 53, 78). These results suggest that Gly is likely the only applicable natural amino acid for the B8 position of insulin when both foldability and activity are concerned.

In summary, the data mentioned above suggest that the insulin structural motif is the basis of the identical disulfide linkage and similar secondary and tertiary structures of the members of the insulin superfamily.

OXIDATIVE FOLDING OF SINGLE-CHAIN INSULINS AND RELATED PROTEINS

Insulin is synthesized in vivo as a single-chain polypeptide, proinsulin

As mentioned earlier, the studies of insulin almost crown all of the fields of protein science, but its *in vitro* oxidative folding pathway has not been reported until recently. The major reason is that its mature form consists of two polypeptide chains. However, insulin is synthesized *in vivo* as a single-chain polypeptide (preproinsulin) with a signal sequence at the N-terminus of the B-chain and a connecting peptide (C-peptide) between the B- and A-chains (63, 66). After cleavage of the N-terminal signal sequence in the endoplasmic reticulum, the nascent polypeptide folds into a unique structure with correct disulfide bridges and is packaged into the secretory granules as proinsulin. Subsequently, the C-peptide is removed by a set of specific proteases (prohormone convertases), and the single-chain proinsulin is converted into double-chain insulin and C-peptide in the B-cell granule (17, 18). The C-peptide flanked by dibasic amino acids at two ends as processing sites has 31 residues in human proinsulin (64, 65).

Disulfide intermediates and folding pathways of PIP, HPI, AILP, and IGF-1

The insulin superfamily is a group of small globular proteins that share a common structural motif characterized by three disulfide bonds. Of these members, the disulfide-forming pathway of IGF-I has been extensively investigated (31–34, 40, 49, 58, 75). Although insulin is the best-characterized member structurally and functionally in the insulin superfamily, its *in vitro* disulfide-folding pathway was investigated just recently because of its double-chain nature. By using single-chain insulin precursors as models, our group and other groups have been able to investigate systematically the disulfide-coupled refolding and unfolding of insulin *in vitro*. In the next section, we review the *in vitro* folding pathways of PIP, HPI, AILP, and IGF-1.

Porcine insulin precursor (PIP). PIP is a recombinant single-chain insulin precursor in which B30Ala and A1Gly are linked by a dipeptide linkage, Ala-Lys (Fig. 3). The recombinant PIP can be expressed and secreted in a correctly folded and soluble form from *S. cerevisiae* (79), and can be converted into human insulin by means of transpeptidation *in vitro*. In the redox buffer containing L-arginine, the fully reduced/denatured PIP can refold efficiently in vitro (56). By quenching the refolding process at different time points, we

TABLE 1. THE CONSTRUCTED PEPTIDE MODELS OF THE DISULFIDE INTERMEDIATES OF INSULIN, PIP, AND PROINSULIN

Disulfide bond	*Peptide model*	*Disulfide linkage*	*Helical content*	*Reference*
1	[Ser^{A7}, Ser^{B7}, Ala^{A6}, Ala^{A11}]PIP	A20–B19	ND	74
	[Ser^{A6}, Ser^{A11}]DKP insulin	A20–B19, A7–B7	Helix 1, 3	36
	[Ala^{A6}, Ala^{A11}]DKP insulin	A20–B19, A7–B7	Helix 1, 3	72
	[Ser^{A7}, Ser^{B7}]DKP insulin	A20–B19, A6–A11	Helix 1	39
2	[Ser^{A7}, Ser^{B7}]PIP	A20–B19, A6–A11	ND	23, 43
	[Glu^{A7}, Glu^{B7}]PIP	A20–B19, A6–A11	ND	24
	[Ala^{A6}, Ala^{A11}]PIP	A20–B19, A7–B7	ND	43
	[Ser^{A7}, Ala^{A11}]PIP	A20–B19, A6–B7	ND	43
	[Ala^{A6}, Ser^{B7}]PIP	A20–B19, A7–A11	ND	43
3	[A6–B7, A7–A11, A20–B19]insulin	A6–B7, A7–A11, A20–B19	Helix 1, 3	35
	[A6–A7, A11–B7, A20–B19]insulin	A6–A7, A11–B7, A20–B19	Helix 1	37

The retained helix of these mutants is identified by NMR spectroscopy. ND, not determined.

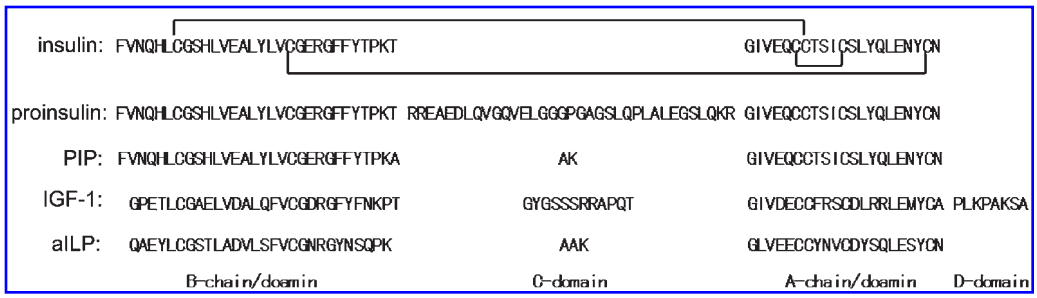

FIG. 3. The amino acid sequence of insulin, porcine insulin precursor (PIP), human proinsulin (HPI), IGF-1, and AILP. The disulfide linkages of insulin are shown; the native disulfide linkages of other members are identical to those of insulin.

captured and isolated three distinct folding intermediates. One intermediate (1SSPIP) contains one disulfide bond (1SS), possibly A6–A11; the other two intermediates (2SSPIPa and 2SSPIPb) contain two disulfide bridges (2SS): the disulfide A20–B19 as well as another interchain bond (A7–B7 or A6–B7). The one-disulfide intermediate contains few ordered secondary structures, whereas the two two-disulfide intermediates retain significant amount of helical structures. Based on the time-dependent formation and distribution of the trapped intermediates, two different disulfide-forming pathways were proposed, as shown in Fig. 4. The first folding pathway involves the rapid formation of the intra-chain disulfide bond, followed by the slower formation of one interchain disulfide bond. The second folding pathway starts with the formation of the disulfide bridge A20–B19, followed immediately by another interchain bond.

Human proinsulin (HPI). Insulin is synthesized *in vivo* as single-chain proinsulin (HPI) (Fig. 3). The difference between HPI and PIP is the length of their connecting peptide between B- and A-chains: the HPI contains a much longer (34 residues) connecting peptide. Although the connecting peptide is deduced to be without ordered structure, it has a significant effect on the folding kinetics. During the oxidative refolding of HPI, four disulfide intermediates (P1, P2, P3, and P4) were captured and characterized (57). Surprisingly, all of the four intermediates contain three disulfide bonds (3SS). The intermediates with two interchain disulfide bonds, such as P1, P2, and P3, have partially structured conformation; whereas the intermediate without interchain disulfide bridge (P4) has little ordered structure. All of the kinetic intermediates can refold into native HPI through disulfide rearrangement in the refolding buffer. The P2 intermediate that contains the disulfide A20–B19 is a crucial intermediate during HPI refolding because the other three intermediates need to rearrange their disulfides to form P2 during their refolding to HPI. The putative disulfide folding pathway of HPI *in vitro* is shown in Fig. 5. Under the *in vitro* refolding condition, the fully reduced PHI is quickly converted to the intermediates with three disulfides. Those one-disulfide and two-disulfide intermediates are too short-live to be captured. The quick formation of the three disulfides can mask all

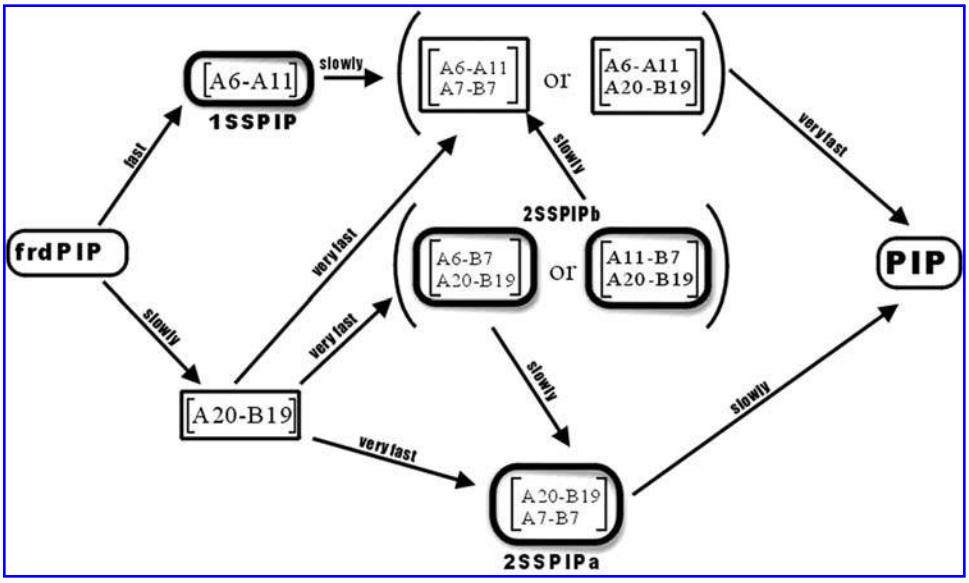

FIG. 4. The putative disulfide formation pathway of PIP. Intermediates are represented by their disulfide linkage pattern. Intermediates with an overstriking rectangle represent the intermediates captured in the study.

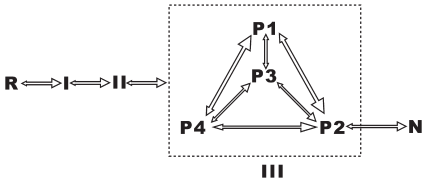

FIG. 5. The putative disulfide folding pathway of HPI *in vitro*. I, II, and III represent the intermediates population with one, two, and three disulfides, respectively. R and N are fully reduced and native HPI, respectively. The *double arrow* indicates the equilibrium between each intermediates species, and the equilibrium has a greater tendency for the end with the *bigger arrow*.

of the free thiols in the polypeptide chain and prevent the aggregation caused by disulfide crosslinking among polypeptides. The native disulfide bonds are subsequently regenerated by disulfide reshuffling within the polypeptide chain. This is a common mechanism used by many of proteins (such as RNase A) to prevent disulfide-crosslinked aggregations.

Although different intermediates have been identified during the oxidative refolding of PIP and HPI, one intrinsic refolding mechanism is the same: formation of disulfide A20–B19 is a crucial step in the refolding process of both PIP and HPI. Disulfide bond A20–B19 stabilizes a molten subdomain that is important for the folding initiation of PIP, HPI, and IGF-1 (38).

Amphioxus insulin-like peptide (AILP). Amphioxus ILP belongs to the insulin superfamily and is proposed as a common ancestor of insulin and IGF-1 (5). The AILP deduced from the cDNA sequence displays structural characteristics of both mammalian insulin and IGF-1. A recombinant single-chain AILP (rAILP) was constructed previously (61), in which the deduced B- and A-domains were linked together by a tripeptide, Ala-Ala-Lys (Fig. 3). The AILP could be successfully expressed in yeast cells and adopted an insulin-like fold. Furthermore, the AILP acquired moderate mammalian insulin activity by grafting some insulin residues that are involved in insulin-receptor interactions (27).

During oxidative refolding, the fully reduced AILP acquired its three disulfides through the 1SS, 2SS, and 3SS stages (10). During reductive unfolding, its three disulfides were reduced gradually through the 2SS and 1SS stages. Interestingly, the intermediates captured during the unfolding process, U1, U2, U3, and U4, are almost identical to those intermediates (P1, P2, P3, and P4) captured in the refolding process; therefore, the refolding and unfolding of AILP may share the same intermediates and follow a similar but reverse pathway. The schematic flow chart of major oxidative refolding and reductive unfolding pathways of AILP is shown in Fig. 6. In the oxidative re-

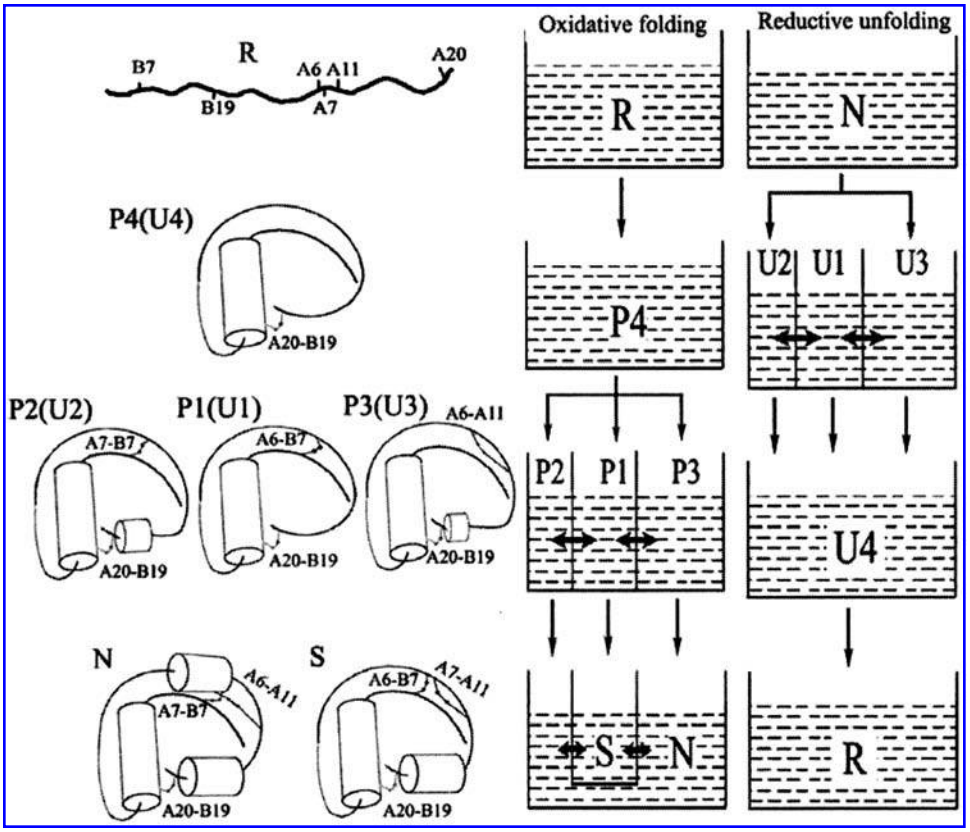

FIG. 6. Schematic flow chart of major oxidative refolding and reductive unfolding pathway of rAILP. R, fully reduced AILP; N, native AILP; S, swap AILP; intermediates P1-4 were identified during the oxidative refolding process, whereas U1-4 were identified during the reductive unfolding process.

folding of AILP, the disulfide bond A20–B19 is formed first (intermediate P4), followed by the formation of the second disulfide bonds, whose pairings have much more flexibility (intermediates P2, P1, or P3). The nonnative disulfides can be subsequently converted to native ones through disulfide reshuffling.

Insulin-like growth factor 1 (IGF-1). IGF-1 is a 70-residue single-chain globular protein composed of B-, C-, A-, and D-domains from the N-terminus to the C-terminus (42). IGF-1 adopts an insulin-like structure mainly including three α-helical segments (8–18, helix I; 42–49, helix II; and 54–61, helix III) in its A- and B-domains. The conformation of its C- and D-domains is highly flexible, so the insulin-like structure of IGF-1 is mainly encoded by its A- and B-domains.

During the oxidized refolding of IGF-1 and its analogues, one 1SS and at least three 2SS intermediates were identified, isolated, and characterized, and the oxidative refolding pathway of IGF-1 was proposed as shown in Fig. 7. Because the disulfide 18–61 (corresponding to disulfide A20–B19 in insulin) is present in almost all of the captured disulfide intermediates, it should be the first disulfide bond formed during the refolding process. However, formation of the second disulfide bond is random. Among all the 2SS intermediates, (18–61, 6–48) and (18–61, 6–47) are the immediate precursor of the native and swap IGF-1s, respectively. Other 2SS intermediates need to undergo disulfide rearrangement to form these two intermediates before they can refold to native and swap IGF-1s.

The common folding behavior of the insulin superfamily

The members of insulin superfamily share high sequence homology and similar tertiary structures. Their *in vitro* folding pathways also share common characteristics.

I. The disulfide A20–B19 is the most important and is formed at the initial refolding stage. In the refolding pathway of PIP, both 2SS intermediates contain the native disulfide bond A20–B19. The most important intermediate P2 of HPI also contains the disulfide bond A20–B19. During refolding of AILP and IGF-1, an essential 1SS intermediate containing disulfide 18–61/A20–B19 has been identified. For both PIP and IGF-1, the peptide model with only disulfide A20–B19 still retained partially folded native conformation and could refold efficiently *in vitro* (54, 74). Furthermore, a short IGF-1 peptide model containing the fragment 7–25 and fragment 53–70 of IGF-1 that are connected by disulfide bond 18–61 presents a transient native-like partial core, implying that the early folding events of the insulin-related polypeptides are nucleated by a native-like molten subdomain containing disulfide A20–B19 (38). All these observations strongly support the critical and conservative role of disulfide A20–B19 in the folding initiation of the insulin superfamily.

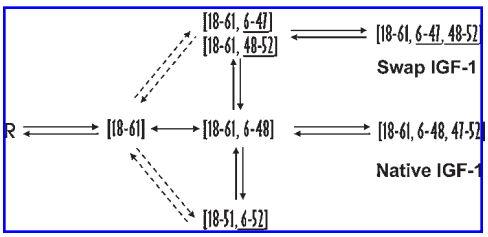

FIG. 7. The putative *in vitro* refolding pathway of IGF-1. The intermediates are represented by the disulfide bonds linkage, and those underlined are nonnative disulfide linkage.

II. The second disulfide pairing is flexible and can be rearranged from nonnative to native pairing. In all of the refolding pathways discussed, formation of the second disulfide bond in 1SS intermediate (A20–B19) seems random. All of the identified intermediates during the oxidative refolding process of PIP, HPI, AILP, and IGF-1 are listed in Table 2. The nonnative disulfide bond in the 2SS species can be finally converted to the native one (or swap disulfide in AILP and IGF-1) by disulfide rearrangement, which is an intramolecular reaction and can minimize the chance of aggregation. The constructed peptide models that contain two disulfides also showed the flexibility of the second disulfide pairing: most of the model peptides with native or nonnative disulfides could fold quickly and efficiently *in vitro* (43).

III. The α-helical segment I (H1, B9Ser–B19Cys) that is close to the disulfide bond A20–B19 is probably the folding initiation site. The α-helical segment I is the longest helical structure in the superfamily, and it is also the most robust ordered structure in the superfamily. When the other two α-helical segments located at the A-chain/domain were disturbed by disulfide deletion or disulfide mispairing, this segment still remained intact. During the initiation of refolding, the α-helix I can form transiently because of its high helical propensity, and subsequently stabilized by formation of the disulfide bond A20–B19 (23, 24, 43, 74).

IV. The *in vitro* refolding and unfolding pathways may share some common folding intermediates but flow in the opposite directions. Many disulfide intermediates are present both in the refolding and unfolding pathways of HPI, IGF-1, and AILP (10). The shared intermediates suggest that refolding and unfolding may adopt some similar but reverse pathways.

THE MOLECULAR BASIS AND EVOLUTION OF THE DIFFERENT FOLDING BEHAVIORS OF INSULIN AND IGF-1

Insulin and IGF-1: homologous amino acid sequences encode different folding behaviors

As mentioned earlier, the insulin superfamily is a group of small globular proteins that share a common structural motif and an identical cysteine-distribution pattern. Among these members, insulin and insulin-like growth factor-1 (IGF-1) share most similarities structurally and functionally and have been extensively studied in the past (42) (Fig. 3).

Insulin is composed of A- and B-chains in its mature form and contains three disulfide bonds that stabilize its native structure (Fig. 8A). Although insulin is synthesized *in vivo* as a single-chain form (proinsulin), its two separate polypeptide chains

Table 2. The Disulfide Linkage of the Captured Intermediates During the Refolding of the Members of Insulin Superfamily

Insulin superfamily	*Intermediates*	*Disulfide number*	*Disulfide linkage*
PIP	1SSPIP	1	A6–A11
	2SSPIPa	2	A20–B19, A7–B7
	2SSPIPb	2	A20–B19, A6–B7
			A20–B19, A11–B7
HPI	P2	3	A6–B7, A7–A11, A20–B19
			A6–B7, A11–B7, A20–B19
	P3	3	A20–B7, A6–A7, A11–B19
			A20–B7, A6–A11, A7–B19
			A20–B7, A7–A11, A6–B19
	P4	3	B7–B19, A6–A7, A11–A20
			B7–B19, A6–A11, A7–A20
			B7–B19, A6–A20, A7–A11
AILP	P1	2	A20–B19, A6–B7
	P2	2	A20–B19, A7–B7
	P3	2	A20–B19, A6–A11
	P4	1	A20–B19
IGF-1		1	18–61 (A20–B19)
		2	18–61, 6–47
			(A20–B19, A6–B7)
		2	18–61, 6–48
			(A20–B19, A7–B7)
		2	18–61, 48–52
			(A20–B19, A7–A11)
		2	18–61, 6–52
			(A20–B19, A11–B7)

Because of technical limitations, some disulfide linkages are not experimentally analyzed. For the intermediates of IGF-1, the corresponding disulfide linkage of insulin is listed in parentheses.

contain enough folding information (68) that is the basis of insulin chain recombination, a prerequisite for the chemical synthesis of insulin at 1960s (44, 45, 76). As discussed earlier, whether the folding is *in vitro* and *in vivo*, proinsulin and PIP always fold into a unique structure with native disulfide linkages (A7–B7, A6–A11, and A20–B19); other disulfide isomers are undetectable at the final folding stage, although some of them are really accumulated to some extent as folding intermediates at the distinct folding stages both *in vivo* and *in vitro* (46, 47, 77). So the amino acid sequence of insulin encodes a unique stable structure that is stabilized by three disulfides (Fig. 8A). Some nonnative disulfide isomers that are thermodynamically unstable under nondenatured conditions can be trapped either at distinct folding stages as folding intermediates or at denatured conditions in which these isomers share similar energetic states with that of the native form. For example, the swap insulin that has alternative disulfide bridges (A6–B7, A7–A11, and A20–B19) and is thermodynamically unstable can be trapped at a denatured condition (35). As shown in Fig. 8B, the free-energy difference (ΔG) between the native and swap insulin is very large (the native form is much more stable than the swap form), so formation of the swap insulin is thermodynamically forbidden under nondenatured condition (the swap form is undetectable). Three-dimensional structure analysis showed that the alternative disulfide linkages caused a corresponding conformational change in the structure: the α-helix segment (H2) at the N-terminal of the A-chain presented in the native from was unfolded in the swap form, as shown in Fig. 8A.

IGF-1 is a single-chain polypeptide that is composed of B-, C-, A-, and D-domains (Fig. 3) (42). Its B- and A-domains are homologous to the B- and A-chains of insulin, respectively; its C-domain is analogous to the C-peptide of proinsulin, but they share no sequence homology; its D-domain has no counterpart in insulins. The IGF-1 isolated from serum, where it is bound with IGF-binding proteins, shares identical disulfide bridges with insulin: two disulfides (6–48 and 18–61, corresponding to A7–B7 and A20–B19 in insulins, respectively) are formed between A- and B-domains; and one disulfide (47–52, corresponding to A6–A11 in insulins) is formed within the A-domain. The native IGF-1 also shares similar structure with insulin (Fig. 8A): three α-helical segments within the A- and B-domains form the rigid frame of the IGF-1 fold that is stabilized by its three native disulfide bonds (12, 67). However, *in vitro* IGF-1 always refolds into two disulfide isomers, native IGF-1 ($\sim$60%) and swap IGF-1 ($\sim$40%) (31–34, 49, 50, 58). The swap IGF-1 has alternative disulfide linkages (6–47, 48–52, 18–61) that are identical to those of the swap insulin (A6–B7, A7–A11, A20–B19). The alternative disulfide linkages disturb the native structure: the α-helical segment II present in the native IGF-1 is unfolded in the swap form (Fig. 8A). Although the native and swap IGF-1s have different disulfide linkages and differ-

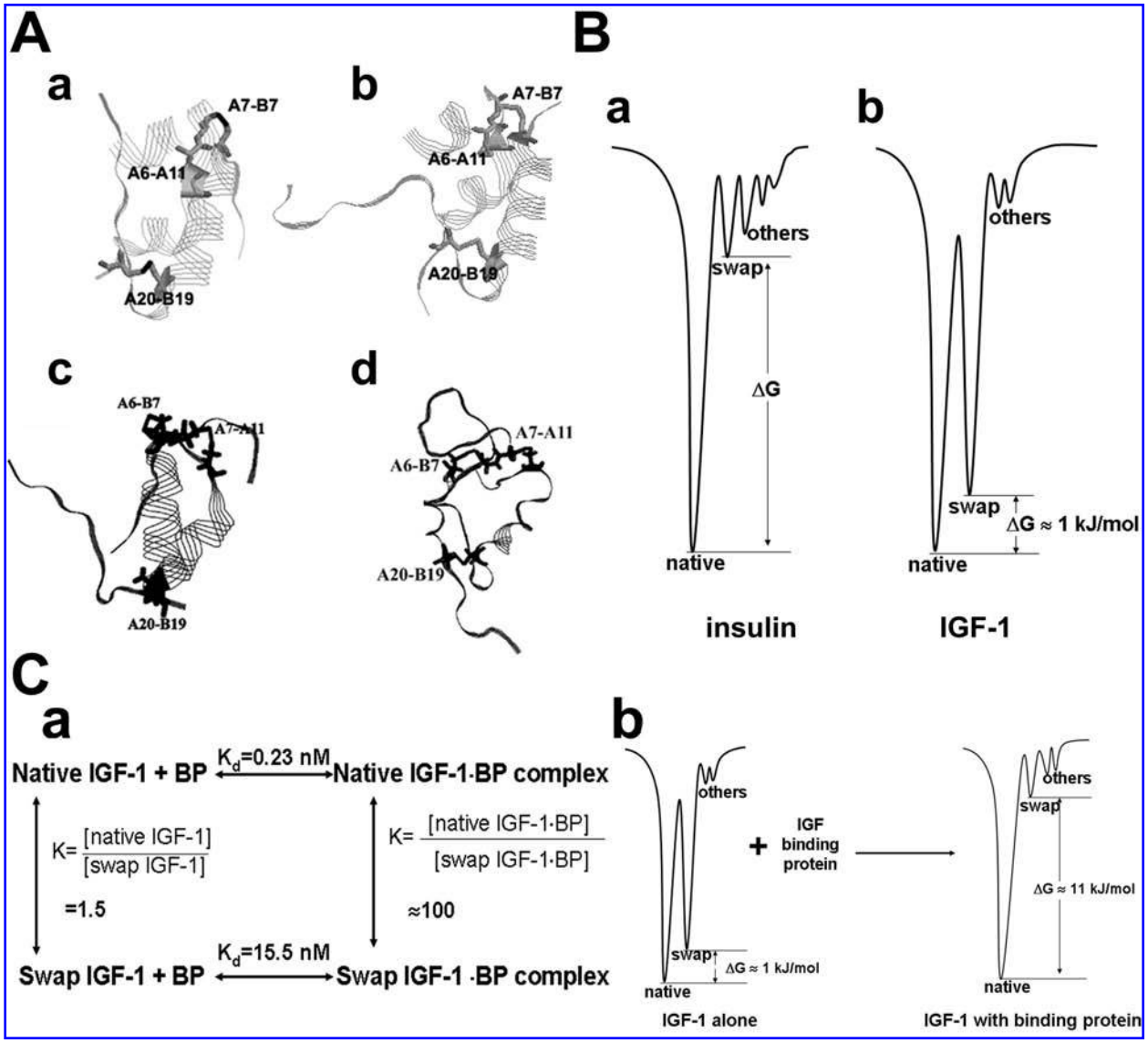

FIG. 8. The different folding behaviors of insulin and IGF-1. (**A**) The three-dimensional structures of native and swap forms of insulin and IGF-1. For clarity, the disulfides of IGF-1s were labeled as those of insulin. (**B**) The energetic states of disulfide isomers of insulin and IGF-1. (**C**) The role of IGFBP during IGF-1 folding. The calculated equilibrium constant (K) and free-energy difference (ΔG) between native and swap IGF-1 in the presence and absence of IGFBP-3 is shown in panels **a** and **b**, respectively.

ent three-dimensional structures, they share similar thermodynamic stabilities (Fig. 8B): the calculated free energy difference (ΔG) between the native and swap forms is only 1.0 kJ/mol; it means that IGF-1 always folds into a mixture of native and swap forms under thermodynamic control because the two isomers have similar energetic states. So the amino acid sequence of IGF-1 actually encodes two stable disulfide isomers.

As mentioned earlier, the swap IGF-1 and swap insulin have identical disulfides and similar structures, but they have quite different energetic states compared with their corresponding native forms: the energetic state of the swap insulin is much higher than that of the native form; in contrast, the swap and native IGF-1s share similar thermodynamic stability (Fig. 8B). It is the different energetic states between the native and the swap forms that result in the different folding consequences of insulin and IGF-1 under thermodynamic control: insulin folds into a unique native form. IGF-1, in contrast, folds into two stable disulfide isomers. Because IGF-1 folds into two disulfide isomers, why has only the native form been isolated from serum? It is known that IGF-1 is bound with IGF-binding proteins *in vivo*, and the binding proteins preferably bind the native form (30). The IGF-binding proteins shift the equilibrium between the native and swap IGF-1s established during IGF-1 folding. As shown in Fig. 8C, the equilibrium constant (K) between native and swap IGF-1s is only 1.5 (60% of native IGF-1 *vs.* 40% of swap IGF-1) when IGF-1 folds alone; however, the IGF-1 almost can quantitatively fold into native form (99% of native IGF-1 *vs.* 1% swap IGF-1) in the presence of IGF-binding pro-

teins [the calculation is based on the measured equilibrium constants (K_d) between native/swap IGF-1s and the IGF binding protein-3; other IGF binding proteins probably also bind the native IGF-1 with higher affinities]. The IGF-1 inding proteins actually thermodynamically control the *in vivo* IGF-1 folding: IGF-1 folds into a unique form, native IGF-1, in the presence of IGF-binding proteins. *In vitro,* the folding of IGF-1 is also thermodynamically controlled by IGF-binding protein (IGFBP-1), which has been experimentally demonstrated (32).

Sequence determinant of the different folding behaviors of insulin and IGF-1

Insulin and IGF-1 share most similarities structurally and functionally, but they have quite different folding behaviors, as discussed earlier. Considering their high sequence homology and small size, their different folding behaviors are quite intriguing: it provides a good opportunity to investigate the sequence structure relation, a central problem of protein folding studies.

IGF-1 is composed of four domains (A, B, C, and D). After dissection into different parts, the B- and A-domains, separately or linked by a short peptide, also fold into two disulfide isomers (22). So it is the B- and A-domains that dictate the unusual folding behavior of IGF-1. For insulin, it is also the B- and A-chains that encode the folding information. To find out the sequence determinant within the A- and B-chain/domains that dictates the different folding behavior of insulin and IGF-1, a series of global and local insulin/IGF-1 hybrids were constructed in our laboratory (Fig. 9).

First, two global hybrids between insulin and IGF-1, Ins(A)/IGF-1(B) and Ins(B)/IGF-1(A), were constructed (25). The hybrid Ins(A)/IGF-1(B) folded into two thermodynamically stable disulfide isomers both *in vitro* and *in vivo*; whereas the hybrid Ins(B)/IGF-1(A) folded into a unique form that has very similar structure to that of native insulin. Therefore, once the molecules contain the B-domain of IGF-1, such as IGF-1, mini-IGF-1, and Ins(A)/IGF-1(B), they fold into two stable disulfide isomers (IGF-1 folding); once the molecules contain the B-chain of insulin, such as PIP, proinsulin, and Ins(B)/IGF-1(A), they fold into a unique stable structure (insulin folding). Thus, the different folding behaviors of insulin and IGF-1 are mainly controlled by their B-chain/domain. Besides different folding behaviors, the intra-chain/domain disulfide bonds of insulin and IGF-1 have quite different energetic states: the intra-chain bond, A6–A11, is a stable bond in insulin; in contrast, the corresponding disulfide, 47–51, in IGF-1, is a "strained" bond with high energy (31). The different energetic states of the intra-chain/domain disulfide were deduced to be responsible for the different folding behaviors of insulin and IGF-1. Reshuffling the native disulfides to swap form can release the high energy stored in the strained disulfide 47–51 in IGF-1, although it increases the energetic state by unfolding of the α-helical segment H2. As a net result, the swap IGF-1 has similar energetic state as native IGF-1 because the energy increase caused by unfolding of the α-helical segment H2 is counteracted fully or partially by the energy decrease caused by rearrangement of the high-energetic intra-domain disulfide. For insulin, no energy decrease is caused by disulfide rearrangement to counteract the energy increase caused by unfolding of the α-helical segment H2 because the intra-domain disulfide is already a stable bond in insulin. As a net result, the native isomer of insulin is much more stable than the swap isomer. By analyzing the energetic states of the intra-chain/domain disulfide of the global insulin/IGF-1 hybrids, it is suggested that the energetic state of the intra-chain/domain disulfide of insulin and IGF-1 is also controlled by the B-chain/domain (26).

Second, several local insulin/IGF-1 hybrids have been constructed to find a detailed sequence determinant in the B-chain/domain (11, 41). The local hybrid, (1-9)PIP, in which the N-terminal 10 residues of insulin B-chain is replaced by the corresponding nine residues of IGF-1, folds into two disulfide isomers *in vivo* and *in vitro*. However, swapping the N-terminal five residues of insulin B-chain with the corresponding part of IGF-1 has no effect on the folding behavior of PIP. Meanwhile, replacement of the N-terminal residues (four residues or nine residues) of IGF-1 with its counterpart of insulin alters the folding behavior of mini-IGF-1: it folds into a unique structure. Therefore, the N-terminal sequence of the B-chain/domain determines the different folding behaviors of insulin and IGF-1. For the (1–4)PIP that contains the four residues of IGF-1, swapping an additional residue (its B10His was replaced by Glu, the corresponding residue in IGF-1) can change its folding behavior from one stable molecule to two disulfide isomers.

The evolution of the different folding behaviors of insulin and IGF-1

As mentioned earlier, *Amphioxus* insulin-like peptide (AILP) has both characteristics of insulin and IGF-1 and is proposed as their common ancestor (5). To investigate origin of the dif-

name	amino acid sequence			number of stable isomers
	B-chain/domain		A-chain/domain	
PIP:	FVNQHLCGSHLVEALYLVCGERGFFYTPKA	*AK*	GIVEQCCTSICSLYQLENYCN	1
mini-IGF-1:	GPETLCGAELVDALQFVCGDRGFYFNKPK	*AK*	GIVDECCFRSCDLRRLEMYCA	2
Ins(B)/IGF-1(A):	FVNQHLCGSHLVEALYLVCGERGFFYTPKA	*AK*	GIVDECCFRSCDLRRLEMYCA	1
Ins(A)/IGF-1(B):	GPETLCGAELVDALQFVCGDRGFYFNKPK	*AK*	GIVEQCCTSICSLYQLENYCN	2
[1-9]PIP:	GPETLCGAELVEALYLVCGERGFFYTPKA	*AK*	GIVEQCCTSICSLYQLENYCN	2
[1-4]PIP:	GPETLCGSHLVEALYLVCGERGFFYTPKA	*AK*	GIVEQCCTSICSLYQLENYCN	1
[1-4][B10E]PIP:	GPETLCGSELVEALYLVCGERGFFYTPKA	*AK*	GIVEQCCTSICSLYQLENYCN	2
[1-10]mini-IGF-1:	FVNQHLCGSHLVDALQFVCGDRGFYFNKPK	*AK*	GIVDECCFRSCDLRRLEMYCA	1
[1-5]mini-IGF-1:	FVNQHLCGAELVDALQFVCGDRGFYFNKPK	*AK*	GIVDECCFRSCDLRRLEMYCA	1

FIG. 9. Relation between amino acid sequence and folding behavior of insulin/IGF-1 hybrids. The sequence of IGF-1 fragment is underlined.

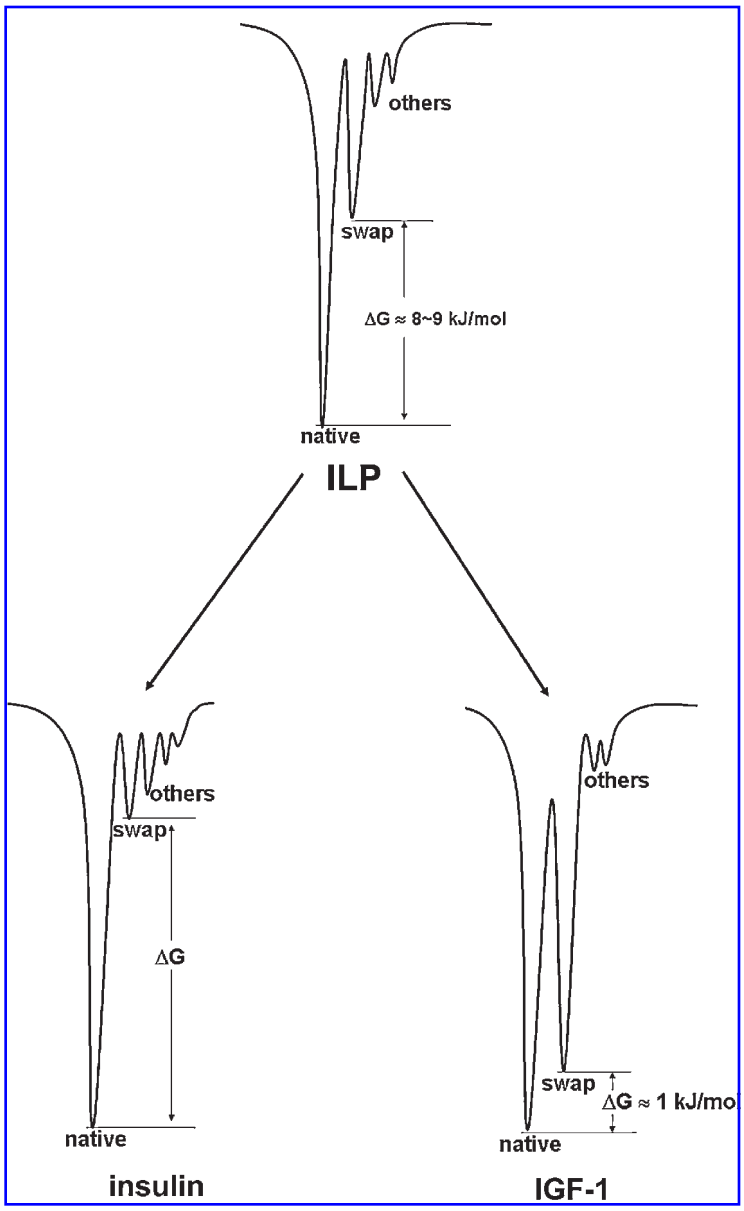

FIG. 10. The proposed bifurcating evolution pathway of the different folding behaviors of insulin and IGF-1.

ferent folding behaviors of insulin and IGF-1, a single-chain AILP was constructed and recombinantly expressed in yeast cells (61). The *in vitro* refolding analysis revealed that the refolding of AILP has the characteristics of both insulin and IGF-1 (69). On the one hand, AILP folds into two disulfide isomers under thermodynamic control, which is similar to that of IGF-1. On the other hand, the native isomer is the major folding product: more than 96% of AILP folds into the native form, which is similar to that of insulin. Based on these results, a bifurcating evolution pathway of the different folding behaviors of insulin and IGF-1 has been proposed (Fig. 10).

CONCLUDING REMARKS

The insulin superfamily includes many members from different species. All the members share highly homologous primary and tertiary structures. Here we reviewed the *in vitro* folding pathways coupled with disulfide formation and their underlying molecular mechanism of four representative members of this family: insulin, proinsulin, IGF-I, and AILP. Insulin (or proinsulin) and IGF-1 have very different folding behaviors, with the former folding into a unique native molecule and the latter forming two thermodynamically stable isomers when folded *in vitro*. This difference is determined mainly by the local primary sequence, particularly the N-terminus of B-chain/domain. However, the *in vitro* folding pathway of insulin, IGF-1, and their ancestor AILP also shares many similarities that are possibly determined by the homologous structural motif in the insulin superfamily. The *in vitro* folding pathway and the molecular-mechanism studies of the insulin superfamily members can help us to further understand the general principles and molecular determinants of protein folding.

ABBREVIATIONS

SS, intermediate contains one disulfide bond; 2SS, intermediate contains two disulfide bonds; 3SS, intermediate contains three disulfide bonds; AILP, *Amphioxus* insulin-like peptide; HPI, human proinsulin; IGF-1, insulin-like growth factor-1; IGF-2, insulin-like growth factor-2; IGFBP, IGF-binding protein; PIP, porcine insulin precursor; rAILP, recombinant single-chain AILP; RNase A, ribonuclease A.

REFERENCES

1. Anfinsen CB. Principles that govern the folding of protein chains. *Science* 181: 223–230, 1973.
2. Baker EN, Blundell TL, Cutfield JF, Cutfield SM, Dodson EJ, Dodson GG, Hodgkin DM, Hubbard RE, Isaacs NW, Reynolds CD, Sakabe K, Sakabe N, and Vijayan NM. The structure of 2Zn pig insulin crystals at 1.5 A resolution. *Philos Trans R Soc Lond B Biol Sci* 319: 369–456, 1988.
3. Brandenburg D, Gattner HG, and Wollmer A. Preparation and properties of acetyl derivatives of beef insulin. I. *Hoppe Seylers Z Physiol Chem* 353: 599–617, 1972.
4. Brown H, Sanger F, and Kitai R. The structure of pig and sheep insulins. *Biochem J* 60: 556–565, 1955.
5. Chan SJ, Cao QP, and Steiner DF. Evolution of the insulin superfamily: cloning of a hybrid insulin/insulin-like growth factor cDNA from *Amphioxus*. *Proc Natl Acad Sci U S A* 87: 9319–9323, 1990.
6. Chang JY, Li L, and Bulychev A. The underlying mechanism for the diversity of disulfide folding pathways. *J Biol Chem* 275: 8287–8289, 2000.
7. Chang SG, Choi KD, Jang SH, and Shin HC. Role of disulfide bonds in the structure and activity of human insulin. *Mol Cells* 16: 323–330, 2003.
8. Chatrenet B and Chang JY. The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. *J Biol Chem* 268: 20988–20996, 1993.
9. Chen H and Feng YM. Contribution of the residue Glu9, Glu46, and Phe49 to the biological activity of insulin-like growth factor-1. *IUBMB Life* 51: 33–37, 2001.
10. Chen Y, Jin R, Dong HY, and Feng YM. In vitro refolding/unfolding pathways of *Amphioxus* insulin-like peptide: implications for folding behavior of insulin family proteins. *J Biol Chem* 279: 55224–55233, 2004.
11. Chen Y, You Y, Jin R, Guo ZY, and Feng YM. Sequences of B-chain/domain 1-10/1-9 of insulin and insulin-like growth factor 1 determine their different folding behavior. *Biochemistry* 43: 9225–9233, 2004.
12. Cooke RM, Harvey TS, and Campbell ID. Solution structure of human insulin-like growth factor 1: a nuclear magnetic resonance and restrained molecular dynamics study. *Biochemistry* 30: 5484–5491, 1991.
13. Creighton TE. Disulfide bonds as probes of protein folding pathways. *Methods Enzymol* 131: 83–106, 1986.
14. Creighton TE. The disulfide folding pathway of BPTI. *Science* 256: 111–114, 1992.
15. Cuatrecasas P. Interaction of insulin with the cell membrane: the primary action of insulin. *Proc Natl Acad Sci U S A* 63: 450–457, 1969.
16. Dai Y and Tang JG. Intra-A chain disulfide bond (A6-11) of insulin is essential for displaying its activity. *Biochem Mol Biol Int* 33: 1049–1053, 1994.
17. Davidson HW and Hutton JC. The insulin-secretory-granule carboxypeptidase H: purification and demonstration of involvement in proinsulin processing. *Biochem J* 245: 575–582, 1987.
18. Davidson HW, Rhodes CJ, and Hutton JC. Intraorganellar calcium and pH control proinsulin cleavage in the pancreatic beta cell via two distinct site-specific endopeptidases. *Nature* 333: 93–96, 1988.
19. Duret L, Guex N, Peitsch MC, and Bairoch A. New insulin-like proteins with atypical disulfide bond pattern characterized in *Caenorhabditis elegans* by comparative sequence analysis and homology modeling. *Genome Res* 8: 348–353, 1998.
20. Freychet P, Roth J, and Neville DM Jr. Insulin receptors in the liver: specific binding of (^{125}I)insulin to the plasma membrane and its relation to insulin bioactivity. *Proc Natl Acad Sci U S A* 68: 1833–1837, 1971.
21. Gilbert HF. Molecular and cellular aspects of thiol-disulfide exchange. *Adv Enzymol Relat Areas Mol Biol* 63: 69–172, 1990.
22. Gill R, Wallach B, Verma C, Urso B, De Wolf E, Grotzinger J, Murray-Rust J, Pitts J, Wollmer A, De Meyts P, and Wood S. Engineering the C-region of human insulin-like growth factor-1: implications for receptor binding. *Protein Eng* 9: 1011–1019, 1996.
23. Guo ZY and Feng YM. Effects of cysteine to serine substitutions in the two inter-chain disulfide bonds of insulin. *Biol Chem* 382: 443–448, 2001.
24. Guo ZY, Jia XY, and Feng YM. Replacement of the interchain disulfide bridge-forming amino acids A7 and B7 by glutamate impairs the structure and activity of insulin. *Biol Chem* 385: 1171–1175, 2004.
25. Guo ZY, Shen L, and Feng YM. The different folding behavior of insulin and insulin-like growth factor 1 is mainly controlled by their B-chain/domain. *Biochemistry* 41: 1556–1567, 2002.
26. Guo ZY, Shen L, and Feng YM. The different energetic state of the intra A-chain/domain disulfide of insulin and insulin-like growth factor 1 is mainly controlled by their B-chain/domain. *Biochemistry* 41: 10585–10592, 2002.
27. Guo ZY, Shen L, Gu W, Wu AZ, Ma JG, and Feng YM. In vitro evolution of *Amphioxus* insulin-like peptide to mammalian insulin. *Biochemistry* 41: 10603–10607, 2002.
28. Guo ZY, Tang YH, Wang S, and Feng YM. Contribution of the absolutely conserved B8Gly to the foldability of insulin. *Biol Chem* 384: 805–809, 2003.
29. Guo ZY, Zhang Z, Jia XY, Tang YH, and Feng YM. Mutational analysis of the absolutely conserved B8Gly: consequence on foldability and activity of insulin. *Acta Biochim Biophys Sin (Shanghai)* 37: 673–679, 2005.
30. Heding A, Gill R, Ogawa Y, De Meyts P, and Shymko RM. Biosensor measurement of the binding of insulin-like growth factors I and II and their analogues to the insulin-like growth factor-binding protein-3. *J Biol Chem* 271: 13948–13952, 1996.
31. Hober S, Forsberg G, Palm G, Hartmanis M, and Nilsson B. Disulfide exchange folding of insulin-like growth factor I. *Biochemistry* 31: 1749–1756, 1992.
32. Hober S, Hansson A, Uhlen M, and Nilsson B. Folding of insulin-like growth factor I is thermodynamically controlled by insulin-like growth factor binding protein. *Biochemistry* 33: 6758–6761, 1994.
33. Hober S, Lundstrom Ljung J, Uhlen M, and Nilsson B. Insulin-like growth factors I and II are unable to form and maintain their native disulfides under in vivo redox conditions. *FEBS Lett* 443: 271–276, 1999.
34. Hober S, Uhlen M, and Nilsson B. Disulfide exchange folding of disulfide mutants of insulin-like growth factor I in vitro. *Biochemistry* 36: 4616–4622, 1997.
35. Hua QX, Gozani SN, Chance RE, Hoffmann JA, Frank BH, and Weiss MA. Structure of a protein in a kinetic trap. *Nat Struct Biol* 2: 129–138, 1995.
36. Hua QX, Hu SQ, Frank BH, Jia W, Chu YC, Wang SH, Burke GT, Katsoyannis PG, and Weiss MA. Mapping the functional surface of insulin by design: structure and function of a novel A-chain analogue. *J Mol Biol* 264: 390–403, 1996.
37. Hua QX, Jia W, Frank BH, Phillips NF, and Weiss MA. A protein caught in a kinetic trap: structures and stabilities of insulin disulfide isomers. *Biochemistry* 41: 14700–14715, 2002.
38. Hua QX, Mayer JP, Jia W, Zhang J, and Weiss MA. The folding nucleus of the insulin superfamily: a flexible peptide model foreshadows the native state. *J Biol Chem* 281: 28131–28142, 2006.
39. Hua QX, Nakagawa SH, Jia W, Hu SQ, Chu YC, Katsoyannis PG, and Weiss MA. Hierarchical protein folding: asymmetric unfolding of an insulin analogue lacking the A7-B7 interchain disulfide bridge. *Biochemistry* 40: 12299–12311, 2001.
40. Hua QX, Narhi L, Jia W, Arakawa T, Rosenfeld R, Hawkins N,

Miller JA, and Weiss MA. Native and non-native structure in a protein-folding intermediate: spectroscopic studies of partially reduced IGF-I and an engineered alanine model. *J Mol Biol* 259: 297–313, 1996.
41. Huang QL, Zhao J, Tang YH, Shao SQ, Xu GJ, and Feng YM. The sequence determinant causing different folding behaviors of insulin and insulin-like growth factor-1. *Biochemistry* 46: 218–224, 2007.
42. Humbel RE. Insulin-like growth factors I and II. *Eur J Biochem* 190: 445–462, 1990.
43. Jia XY, Guo ZY, Wang Y, Xu Y, Duan SS, and Feng YM. Peptide models of four possible insulin folding intermediates with two disulfides. *Protein Sci* 12: 2412–2419, 2003.
44. Katsoyannis PG, Tometsko A, and Zalut C. Insulin peptides. XII. Human insulin generation by combination of synthetic A and B chains. *J Am Chem Soc* 88: 166–167, 1966.
45. Kung YT, Du YC, Huang WT, Chen CC, and Ke LT. Total synthesis of crystalline insulin. *Sci Sin* 15: 544–561, 1966.
46. Liu M, Li Y, Cavener D, and Arvan P. Proinsulin disulfide maturation and misfolding in the endoplasmic reticulum. *J Biol Chem* 280: 13209–13212, 2005.
47. Liu M, Ramos-Castaneda J, and Arvan P. Role of the connecting peptide in insulin biosynthesis. *J Biol Chem* 278: 14798–14805, 2003.
48. Liu Y, Wang ZH, and Tang JG. Flexibility exists in the region of (A6-A11, A7-B7) disulfide bonds during insulin precursor folding. *J Biochem (Tokyo)* 135: 1–6, 2004.
49. Miller JA, Narhi LO, Hua QX, Rosenfeld R, Arakawa T, Rohde M, Prestrelski S, Lauren S, Stoney KS, Tsai L, and Weiss MA. Oxidative refolding of insulin-like growth factor 1 yields two products of similar thermodynamic stability: a bifurcating protein-folding pathway. *Biochemistry* 32: 5203–5213, 1993.
50. Milner SJ, Carver JA, Ballard FJ, and Francis GL. Probing the disulfide folding pathway of insulin-like growth factor-I. *Biotechnol Bioeng* 62: 693–703, 1999.
51. Murray-Rust J, McLeod AN, Blundell TL, and Wood SP. Structure and evolution of insulins: implications for receptor binding. *Bioessays* 14: 325–331, 1992.
52. Nagasawa H, Kataoka H, Isogai A, Tamura S, Suzuki A, Mizoguchi A, Fujiwara Y, Takahashi SY, and Ishizaki H. Amino acid sequence of a prothoracicotropic hormone of the silkworm *Bombyx mori*. *Proc Natl Acad Sci U S A* 83: 5840–5843, 1986.
53. Nakagawa SH, Zhao M, Hua QX, Hu SQ, Wan ZL, Jia W, and Weiss MA. Chiral mutagenesis of insulin: foldability and function are inversely regulated by a stereospecific switch in the B chain. *Biochemistry* 44: 4984–4999, 2005.
54. Narhi LO, Hua QX, Arakawa T, Fox GM, Tsai L, Rosenfeld R, Holst P, Miller JA, and Weiss MA. Role of native disulfide bonds in the structure and activity of insulin-like growth factor 1: genetic models of protein-folding intermediates. *Biochemistry* 32: 5214–5221, 1993.
55. Olsen HB, Ludvigsen S, and Kaarsholm NC. Solution structure of an engineered insulin monomer at neutral pH. *Biochemistry* 35: 8836–8845, 1996.
56. Qiao ZS, Guo ZY, and Feng YM. Putative disulfide-forming pathway of porcine insulin precursor during its refolding in vitro. *Biochemistry* 40: 2662–2668, 2001.
57. Qiao ZS, Min CY, Hua QX, Weiss MA, and Feng YM. In vitro refolding of human proinsulin: kinetic intermediates, putative disulfide-forming pathway folding initiation site, and potential role of C-peptide in folding process. *J Biol Chem* 278: 17800–17809, 2003.
58. Rosenfeld RD, Miller JA, Narhi LO, Hawkins N, Katta V, Lauren S, Weiss MA, and Arakawa T. Putative folding pathway of insulin-like growth factor-I. *Arch Biochem Biophys* 342: 298–305, 1997.
59. Roy M, Lee RW, Brange J, and Dunn MF. 1H NMR spectrum of the native human insulin monomer: evidence for conformational differences between the monomer and aggregated forms. *J Biol Chem* 265: 5448–5452, 1990.
60. Scheraga HA, Wedemeyer WJ, and Welker E. Bovine pancreatic ribonuclease A: oxidative and conformational folding studies. *Methods Enzymol* 341: 189–221, 2001.
61. Shen L, Guo ZY, Chen Y, Liu LY, and Feng YM. Expression, purification, characterization of *Amphioxus* insulin-like peptide and preparation of polyclonal antibody to it. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* 33: 629–633, 2001.
62. Smit AB, Vreugdenhil E, Ebberink RH, Geraerts WP, Klootwijk J, and Joosse J. Growth-controlling molluscan neurons produce the precursor of an insulin-related peptide. *Nature* 331: 535–538, 1988.
63. Steiner DF. Evidence for a precursor in the biosynthesis of insulin. *Trans N Y Acad Sci* 30: 60–68, 1967.
64. Steiner DF. On the role of the proinsulin C-peptide. *Diabetes* 27(suppl 1): 145–148, 1978.
65. Steiner DF, Cho S, Oyer PE, Terris S, Peterson JD, and Rubenstein AH. Isolation and characterization of proinsulin C-peptide from bovine pancreas. *J Biol Chem* 246: 1365–1374, 1971.
66. Steiner DF, Cunningham D, Spigelman L, and Aten B. Insulin biosynthesis: evidence for a precursor. *Science* 157: 697–700, 1967.
67. Vajdos FF, Ultsch M, Schaffer ML, Deshayes KD, Liu J, Skelton NJ, and de Vos AM. Crystal structure of human insulin-like growth factor-1: detergent binding inhibits binding protein interactions. *Biochemistry* 40: 11022–11029, 2001.
68. Wang CC and Tsou CL. The insulin A and B chains contain sufficient structural information to form the native molecule. *Trends Biochem Sci* 16: 279–281, 1991.
69. Wang S, Guo ZY, Shen L, Zhang YJ, Feng YM. Refolding of *Amphioxus* insulin-like peptide: implications of a bifurcating evolution of the different folding behavior of insulin and insulin-like growth factor 1. *Biochemistry* 42: 9687–9693, 2003.
70. Wang ZH, Liu Y, Ji JG, and Tang JG. Effects of deletion and shift of the A2B19 disulfide bond on the structure, activity, and refolding of proinsulin. *J Biochem (Tokyo)* 135: 25–31, 2004.
71. Wedemeyer WJ, Welker E, Narayan M, and Scheraga HA. Disulfide bonds and protein folding. *Biochemistry* 39: 4207–4216, 2000.
72. Weiss MA, Hua QX, Jia W, Chu YC, Wang RY, and Katsoyannis PG. Hierarchical protein "un-design": insulin's intrachain disulfide bridge tethers a recognition alpha-helix. *Biochemistry* 39: 15429–15440, 2000.
73. Weiss MA, Hua QX, Lynch CS, Frank BH, and Shoelson SE. Heteronuclear 2D NMR studies of an engineered insulin monomer: assignment and characterization of the receptor-binding surface by selective 2H and 13C labeling with application to protein design. *Biochemistry* 30: 7373–7389, 1991.
74. Yan H, Guo ZY, Gong XW, Xi D, and Feng YM. A peptide model of insulin folding intermediate with one disulfide. *Protein Sci* 12: 768–775, 2003.
75. Yang Y, Wu J, and Watson JT. Probing the folding pathways of long R(3) insulin-like growth factor-I (LR(3)IGF-I) and IGF-I via capture and identification of disulfide intermediates by cyanylation methodology and mass spectrometry. *J Biol Chem* 274: 37598–37604, 1999.
76. Zahn H, Brinkhoff O, Meienhofer J, Pfeiffer EF, Ditschuneit H, and Gloxhuber C. Combining of synthetic insulin chains into biologically active preparations. *Z Naturforsch B* 20: 666–670, 1965.
77. Zhang BY, Liu M, and Arvan P. Behavior in the eukaryotic secretory pathway of insulin-containing fusion proteins and single-chain insulins bearing various B-chain mutations. *J Biol Chem* 278: 3687–3693, 2003.
78. Zhang H, Tang YH, and Feng YM. Possible role of b8gly in insulin structural motif. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* 32: 480–484, 2000.
79. Zhang Y, Hu H, Cai R, Feng Y, Zhu S, He Q, Tang Y, Xu M, Xu Y, Zhang X, Liu B, and Liang Z. Secretory expression of a single-chain insulin precursor in yeast and its conversion into human insulin. *Sci China C Life Sci* 39: 225–233, 1996.

Address reprint requests to:
You-Min Feng
Institute of Biochemistry and Cell Biology
Chinese Academy of Sciences
320 Yue-Yang Rd.
Shanghai 200031, China

E-mail: fengym@sunm.shcnc.ac.cn

Date of first submission to ARS Central, July 30, 2007; date of final revised submission, July 30, 2007; date of acceptance, August 1, 2007.

This article has been cited by:

1. Ming Liu, Israel Hodish, Leena Haataja, Roberto Lara-Lemus, Gautam Rajpal, Jordan Wright, Peter Arvan. 2010. Proinsulin misfolding and diabetes: mutant INS gene-induced diabetes of youth. *Trends in Endocrinology & Metabolism* **21**:11, 652-659. [CrossRef]
2. Qingxin Hua. 2010. Insulin: a small protein with a long journey. *Protein & Cell* **1**:6, 537-551. [CrossRef]
3. Tse Siang Kang, R. Manjunatha Kini. 2009. Structural determinants of protein folding. *Cellular and Molecular Life Sciences* **66**:14, 2341-2361. [CrossRef]
4. Zhao-Jun Zhang, Lan Wu, Zhi-Song Qiao, Ming-Qiang Qiao, You-Min Feng, Zhan-Yun Guo. 2008. Contribution of the Conserved A16Leu to Insulin Foldability. *The Protein Journal* **27**:3, 192-196. [CrossRef]
5. Salvador Ventura . 2008. Oxidative Protein Folding: From the Test Tube to In Vivo InsightsOxidative Protein Folding: From the Test Tube to In Vivo Insights. *Antioxidants & Redox Signaling* **10**:1, 51-54. [Citation] [PDF] [PDF Plus]