Refolding of Amphioxus Insulin-like Peptide: Implications of a Bifurcating Evolution of the Different Folding Behavior of Insulin and Insulin-like Growth Factor 1†

Shuai Wang,‡,§ Zhan-Yun Guo,‡,|| Lu Shen,|| Ying-Jiu Zhang,§ and You-Min Feng*,||

Key Laboratory of Proteomics, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China, and Life Science College, Jilin University, Changchun 130061, China

Received April 21, 2003; Revised Manuscript Received June 12, 2003

ABSTRACT: Insulin and insulin-like growth factor 1 (IGF-1) share high sequence homology, but their folding behaviors are significantly different: insulin folds into one unique thermodynamically controlled structure, while IGF-1 folds into two thermodynamically controlled disulfide isomers. However, the origin of their different folding behaviors is still elusive. The amphioxus insulin-like peptide (ILP) is thought to be the common ancestor of insulin and IGF-1. A recombinant single-chain ILP has been expressed previously, and now its folding behavior is investigated. The folding behavior of ILP shows the characteristics of both insulin and IGF-1. On one hand, two thermodynamically controlled disulfide isomers of ILP have been identified; on the other hand, the content of isomer 1 (its disulfides are deduced identical to those of swap IGF-1) is much less than that of isomer 2 (its disulfides are deduced identical to those of native IGF-1); that is, more than 96% of ILP folds into the native structure. The present results suggest that the different folding behaviors of insulin and IGF-1 are acquired through a bifurcating evolution: the tendency of forming the thermodynamically controlled non-native disulfide isomer is diminished during evolution from ILP to insulin, while this tendency is amplified during evolution from ILP to IGF-1. Moreover, the N-terminal Gln residue of ILP can spontaneously form a pyroglutamate residue, and its cyclization has a significant effect on the folding behavior of ILP: the percentage of isomer 1 is approximately 2-fold that of isomer 1 of the noncyclized ILP; that is, isomer 1 becomes more favored when the N-terminal residue of ILP is cyclized. So, we deduce that the N-terminal residues have a significant effect on the folding properties of insulin, IGF-1, and ILP.

In the 1960s Anfinsen and co-workers first demonstrated the three-dimensional structure of a globular protein is uniquely determined by its amino acid sequence (*1*). Since then, significant advances have been made in the understanding of protein folding through experimental and theoretical approaches. For small proteins with two-state folding, topology is a major determinant of the folding rate and greatly influences the structure of the transition-state ensemble (*2*−*4*). Studies on the disulfide-coupled folding of some small globular proteins, such as BPTI,[1] RNase A, and EGF, have revealed a sequence of preferred kinetic intermediates which define a folding pathway (*5*−*12*). In vivo the protein folding is assisted by the molecular chaperones, especially for large proteins (*13*−*15*); some chaperones even can provide the missing steric information for protein folding (*16*).

Insulin is an extensively studied small globular protein with A- and B-chains linked by three disulfides (one intrachain bond, A6−A11; two interchain bonds, A7−B7 and A20−B19). Its three-dimensional structure has been well solved by X-ray crystallography (*17*, *18*) and NMR (*19*−*21*) since the 1970s. Although the separate A- and B-chains of insulin can be recombined successfully in vitro (*22*), a single-chain polypeptide (pre-proinsulin) is synthesized in vivo. When B29Lys and A1Gly were linked together by a peptide bond directly, the mini-proinsulin still retained the three-dimensional structure identical to that of insulin (*23*, *24*). A single-chain insulin (PIP) can fold correctly and can be secreted efficiently from transformed yeast cells (*25*). In vitro PIP can efficiently fold into the native state with correct pairing of its three disulfides through a defined folding pathway (*26*). It can reasonably be presumed that the three-dimensional structure of PIP is identical or very similar to that of insulin/mini-proinsulin.

† This work was supported by grants from the Chinese Academy of Sciences (KJ951-B1-606) and the National Foundation of Nature Science (30170209).

* Corresponding author. Tel: (86) 021-64374430. Fax: (86) 021-64338357. E-mail: fengym@sunm.shcnc.ac.cn.

‡ These authors contributed equally to this work.

§ Jilin University.

|| Chinese Academy of Sciences.

[1] Abbreviations: PIP, recombinant porcine insulin precursor in which the C-terminus of porcine insulin B-chain and the N-terminus of porcine insulin A-chain were linked together by a dipeptide, Ala-Lys; ILP, recombinant single-chain amphioxus insulin-like peptide in which the C-terminus of the B-chain and the N-terminus of the A-chain of the amphioxus insulin-like peptide were linked together by a tripeptide, Ala-Ala-Lys, and the B29Arg was replaced by Lys residue; IGF-1, insulin-like growth factor 1; IGF-2, insulin-like growth factor 2; BPTI, bovine pancreatic trypsin inhibitor; RNase A, ribonuclease A; EGF, epidermal growth factor; GSH, reduced glutathione; GSSG, oxidized glutathione; EDTA, ethylenediaminetetraacetic acid; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid; PAGE, polyacrylamide gel electrophoresis; NMR, nuclear magnetic resonance; CD, circular dichroism.

10.1021/bi0346289 CCC: $25.00
Published on Web 07/18/2003

IGF-1 is an insulin-like 70-residue single-chain protein composed of B-, C-, A-, and D-domains linked by three disulfides (47−52, 6−48, 18−61) corresponding to those of insulin, (A6−A11, A7−B7, A20−B19) (*27*). The B- and A-domains of IGF-1 are homologous to the B- and A-chains of insulin, respectively; the C-domain is analogous to the C-peptide of proinsulin, but they share no sequence homology; the D-domain has no counterparts in insulins. IGF-1 adopts an insulin-like structure also stabilized by its three disulfides (*28*). The ordered structure of IGF-1 is mainly encoded by its B- and A-domains; the C- and D-domains are highly flexible.

Although insulin and IGF-1 share high sequence homology, similar three-dimensional structure, and weakly overlapping biological activity, their folding behaviors are significantly different: insulin/PIP folds into one unique thermodynamically controlled structure (*26*), while IGF-1 folds into two isomers (native and swap) with different disulfide linkages but similar thermodynamic stability (*29*−*36*). The different folding behavior of insulin/PIP and IGF-1 is mainly controlled by their B-chain/domain that controls the different energetic state of the intra-A-chain/domain disulfide (*37*, *38*).

The amphioxus insulin-like peptide (ILP) deduced from the cDNA sequence is a polypeptide that displays structural characteristics of both mammalian insulin and IGF-1 (*39*). The deduced ILP contains B-, C-, A-, and D-domains similar to those of IGF-1 while there are potential prohormone convertase cleavage sites at the two ends of the C-domain similar to proinsulin, and its A- and B-domains bear considerable sequence homology to the A- and B-chains of insulin as well as to the A- and B-domains of IGF-1. Therefore, amphioxus ILP is proposed as a common ancestor of insulin and IGF-1. Since ILP is the common ancestor of insulin and IGF-1, how did insulin and IGF-1 acquire different folding behaviors from a common ancestor during evolution? Previously, a recombinant single-chain ILP that shows an insulin-like conformation (*40*) had been expressed successfully, and when its five or six residues were replaced by the corresponding residues of insulin, the ILP analogues acquired moderate insulin activity (*41*). To investigate the origin of the different folding behaviors of insulin/PIP and IGF-1, the folding behavior of the recombinant single-chain amphioxus ILP was investigated (the residues of ILP were numbered according to the numbering system of insulin; therefore, the first residue of ILP was numbered as B2). Similar to the folding behavior of IGF-1, two thermodynamically controlled disulfide isomers of ILP have been identified, but the equilibrium is much more favored to the native structure. Therefore, the folding behavior of ILP shows the characteristics of both insulin/PIP and IGF-1, which suggests a bifurcating evolution of the different folding behaviors of insulin and IGF-1 from a common ancestor. Moreover, the N-terminal Gln residue of ILP can spontaneously cyclize and form a pyroglutamate residue (it is named [pE]ILP), and the cyclization favors the formation of the isomer with swap disulfide linkages.

MATERIALS AND METHODS

Materials. Saccharomyces cerevisiae XV700-6B (leu2, ura3, pep4) was kindly provided by Professor Michael Smith (University of British Columbia, Vancouver, Canada). The chemical reagents used in experiments were of analytical grade. The Vydac reverse-phase columns (C8, 5 μm, 4.6 × 250 mm, and C8, 10 μm, 10 × 250 mm), Gilson 306 HPLC system, and Gilson 115 UV detector were used. In the HPLC analysis, a gradient elution was used. Solvent A was 0.15% aqueous TFA; solvent B was 60% acetonitrile containing 0.125% TFA. The elution gradient is listed as follows: 0 min, 0% solvent B; 1 min, 0% solvent B; 5 min, 40% solvent B; 35 min, 80% solvent B; 36 min, 100% solvent B; 38 min, 100% solvent B; 40 min, 0% solvent B; 45 min, 0% solvent B.

Purification and Characterization of ILP and [pE]ILP. The expression vector of ILP (*40*) was transformed into *S. cerevisiae* XV700-6B, then the transformed yeast cells were cultured in a 16 L fermenter for 3 days, and the secreted target protein in the media supernatant was purified as follows. First, target protein in supernatant was precipitated by trichloroacetic acid at the final concentration of 5% (w/v). Second, the pellet containing the target protein was dissolved with 1 M acetic acid for several times, and each time the supernatant was applied to a Sephadex G-50 column, and the column was eluted with 1 M acetic acid. Third, the lyophilized product separated by gel filtration was applied to an ion-exchange column (DEAE-Sepharose CL-6B) prebalanced with 50 mM Tris-HCl and 40% (v/v) isopropyl alcohol, pH 7.8, and eluted with a linear gradient of NaCl from 0 to 0.2 M. The eluted product was lyophilized and then desalted with a Sephadex G-25 column. Fourth, the product was further purified by C8 reverse-phase HPLC and eluted with the gradient listed in the Materials section.

Purification and Characterization of the Disulfide Isomers. The native isomer (isomer 2) of ILP or [pE]ILP was purified from the media supernatant as described above. The swap isomer (isomer 1) of ILP or [pE]ILP was converted from isomer 2 as described in the Disulfide Rearrangement Assay section at 37 °C with a protein concentration of about 1 mg/mL due to its low content in the fermentation mixture. Their molecular masses were determined by electrospray mass spectrometry. Their purity was analyzed by native pH 8.3 PAGE and analytical C8 reverse-phase HPLC.

Circular Dichroism Analysis. The isomers of ILP and [pE]-ILP were dissolved in 1 mM HCl, respectively. The protein concentration was determined by the absorbance at 276 nm using an extinction coefficient of 1.2 mL mg^{-1} cm^{-1}, and their final concentration was adjusted to 0.2 mg/mL, respectively. The measurements were performed on a Jasco-715 circular dichroism spectropolarimeter at room temperature. The spectra were scanned from 200 to 250 nm using a cell with 0.1 cm path length in the far-UV region and scanned from 300 to 245 nm using a cell with 1.0 cm path length in the near-UV region. The data were expressed as molar ellipticity. The software "J-700 for windows secondary structure estimation, Version 1.10.00" was used for secondary structural content estimation from the CD spectra.

V8 Digestion Analysis. The isomers of ILP and [pE]ILP were dissolved in 0.1 M phosphate buffer (pH 7.8), respectively, and then V8 endoproteinase was added to the solution at about a mass ratio of 1:20 to the sample. Isomer 1 was digested at 37 °C for 6 h; isomer 2 was digested at 37 °C for 16 h. After digestion, the solution was adjusted to pH 2.0 with TFA and separated by C8 reverse-phase HPLC eluted with the gradient listed in the Materials section. The

Table 1: Molecular Masses of the Isomers of ILP and [pE]ILP

	ILP		[pE]ILP	
	isomer 1	isomer 2	isomer 1	isomer 2
measured value	5701.0	5701.0	5685.0	5684.0
theoretical value	5702.4	5702.4	5685.4	5685.4

fractions were collected manually and lyophilized. Their molecular masses were measured by electrospray mass spectrometry.

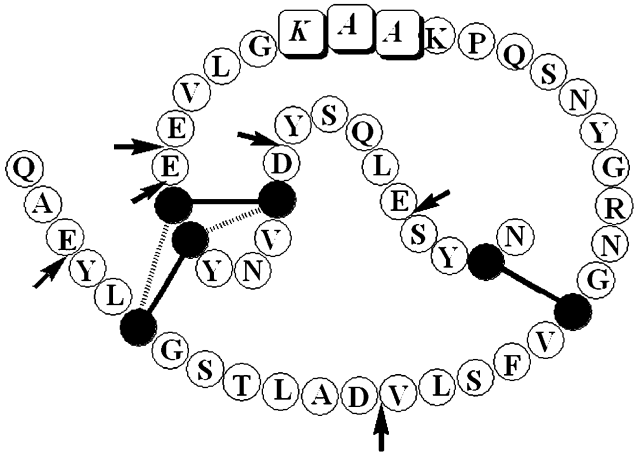
Disulfide Rearrangement Analysis. The two isomers of ILP or [pE]ILP were dissolved in buffer (0.1 M Tris-HCl, 1 mM EDTA, pH 8.7) containing 0.2 mM 2-mercaptoethanol at the final concentration of 0.05 mg/mL, respectively. The disulfide rearrangement reaction was carried out at 15 or 37 °C, respectively. At the indicated reaction times 100 μL of sample was removed and immediately adjusted to pH 2.0 with TFA to terminate the disulfide rearrangement and analyzed by the analytical C8 reverse-phase HPLC. The gradient elution described in the Materials section was used with a flow rate of 0.5 mL/min and detected at 230 nm.

In Vitro Refolding Analysis. The isomers of ILP or [pE]-ILP were respectively dissolved in buffer (0.1 M Tris-HCl, 1 mM EDTA, pH 9.0) containing 10 mM DTT at the final concentration of 0.5 mg/mL, respectively. The reduction was carried out at 15 °C for 1 h. Thereafter, 10 μL of the reaction sample was removed and carboxymethylated by 3 μL of freshly prepared 0.5 M sodium iodoacetate and then analyzed by native pH 8.3 PAGE to confirm if the sample was fully reduced. The refolding was initiated by 10-fold dilution of the fully reduced sample into the prewarmed refolding buffer. The final composition of the refolding solution was 0.1 M Tris-HCl, 1 mM EDTA, 1 mM DTT, and 6 mM GSSG, pH 9.0. The refolding reaction was carried out at 15 °C for 4 h. After incubation 100 μL of refolding solution was removed and acidified to pH 2.0 by TFA and then analyzed by C8 reverse-phase HPLC eluted by the gradient listed in the Materials section with a flow rate of 0.5 mL/min and detected at 230 nm.

RESULTS

Purification and Characterization of ILP and [pE]ILP. The expression product of the single-chain ILP was purified from the fermentation supernatant as described in Materials and Methods. When purified by gel filtration and ion-exchange chromatography, the product shows a single peak. But when purified by a C8 reverse-phase column, there are two major peaks: the content of peak 1 is about 70% and the content of peak 2 is about 30%, but the two peaks show identical mobility rates on the native pH 8.3 PAGE (data not shown). The molecular mass measured by electrospray mass spectrometry of peak 1 is consistent with the value deduced from the amino acid sequence, while the molecular mass of peak 2 is 17 less than the expected value (Table 1). The amino acid composition analysis shows that the two components have almost identical amino acid composition (data not shown). We analyzed the sequence of ILP and found that its N-terminus is a glutamine residue. In previous reports, the N-terminal Gln residue can spontaneously form a pyroglutamate residue (*42*, *43*). Therefore, we deduced that the unexpected component of ILP was caused by the cyclization of the N-terminal Gln residue (formation of a pyroglutamate residue). So, peak 2 was named [pE]ILP. The above deduction is consistent with the following observations. (1) The unexpected molecular mass of peak 2: The cyclization of the N-terminal Gln residue will release one NH_3 molecule with a molecular mass of 17. (2) The identical amino acid composition of peak 1 and peak 2: During hydrolysis both the glutamine and pyroglutamate residues will be converted to glutamate. (3) Different retention times of peak 1 and peak 2 on reverse-phase column: Under the acidic buffer (pH 2.0) the analysis used, the N-terminal Gln residue has a positive charge but the pyroglutamate residue does not carry net charge; therefore, [pE]ILP is more hydrophobic than ILP and more difficult to be eluted from the reverse-phase column. (4) Identical mobility rate of peak 1 and peak 2 on the native pH 8.3 PAGE: Under the pH value used for the PAGE (pH 8.3), the α-NH_2 of the N-terminal Gln is not ionized; therefore, neither the N-terminal Gln nor pyroglutamate carries net charge. (5) As shown in Peptide Mapping, except for the fragment of the N-terminal three residues, the other fragments of peak 2 all have the expected molecular masses. Therefore, the unexpected molecular mass of peak 2 is caused by its N-terminal three residues. Among the three residues (Gln-Ala-Glu), cyclization of the N-terminal residue has most likely happened. Moreover, when purified ILP is incubated in 1 M acetic acid, it can be converted to [pE]ILP gradually, but [pE]ILP is stable under the same condition (data not shown). This suggests that peak 2 is converted from peak 1 during fermentation and purification.

Purification and Characterization of the Disulfide Isomers of ILP and [pE]ILP. Despite the high sequence homology, insulin/PIP and IGF-1/mini-IGF-1 have acquired quite different folding behaviors. We have investigated the sequence determinant of their different folding behaviors (*37*, *38*) but are still puzzled by its origin. To answer this question, we have expressed a single-chain amphioxus ILP that is deduced to be the common ancestor of insulin and IGF-1 (*40*). The sequence and disulfide pairing of the single-chain amphioxus ILP are shown in Figure 1 and compared with the sequences of PIP and mini-IGF-1 (*37*). Here, the folding properties of the single-chain ILP and [pE]ILP were investigated. Similar to mini-IGF-1, two isomers of ILP or [pE]ILP with identical molecular mass were identified. The purified isomers of ILP and [pE]ILP were analyzed by analytical reverse-phase HPLC and native PAGE as shown in Figure 2. On the C8 reverse-phase column, the two isomers of ILP or [pE]ILP have different retention times: the isomer with shorter retention time was named isomer 1, and the isomer with longer retention time was named isomer 2. The different retention time implies that the two isomers have different conformations. On native PAGE, isomer 2 runs a little faster than isomer 1, but cyclization has no effect on the mobility rate. This implies that the conformation of isomer 1 is likely somewhat looser than that of isomer 2. The molecular masses of the two isomers of ILP and [pE]ILP were analyzed by electrospray mass spectrometry and are listed in Table 1. These results suggest that the two isomers have identical amino acid sequence but different conformation. The different conformation of the two isomers was probably caused by the different disulfide linkages, which has been observed in IGF-1/mini-IGF-1.



PIP: FVNQHLCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLYQLENYCN

ILP: QAEYLCGSTLADVLSFVCGNRGYNSQPKAAKGLVEECCYNVCDYSQLESYCN

mini-IGF-1: GPETLCGAELVDALQFVCGDRGFYFNKPKAKGIVDECCFRSCDLRRLEMYCA

B

FIGURE 1: (A) Amino acid sequence of the single-chain amphioxus ILP. The six Cys residues were represented by filled circles; the native disulfides were represented by solid bars, the swap disulfides were represented by dot bars; the arrows showed the potential cleavage sites of V8 endoproteinase. (B) Primary sequence comparison of PIP, mini-IGF-1, and amphioxus ILP.

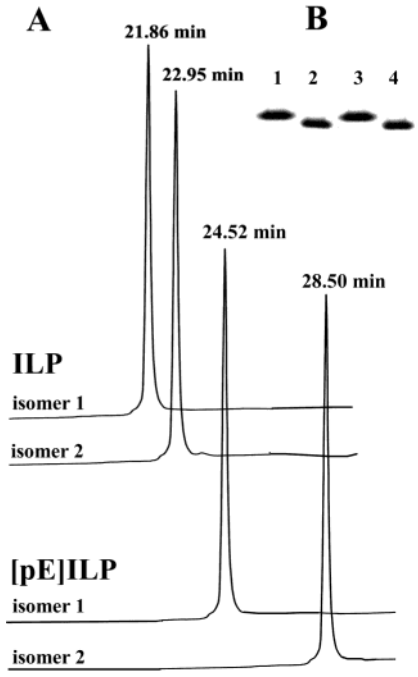


FIGURE 2: Analysis of the disulfide isomers of ILP and [pE]ILP. (A) C8 reverse-phase HPLC analysis. 5 μg of the purified sample was loaded onto a C8 reverse-phase column, eluted by the gradient listed in Materials and Methods with a flow rate of 0.5 mL/min, and detected at 230 nm. (B) Native pH 8.3 PAGE analysis. 2 μg of the purified sample was loaded onto the gel. The gel was stained by Coomassie brilliant blue R250. Lanes: 1, isomer 1 of ILP; 2, ILP isomer 2 of ILP; 3, isomer 1 of [pE]ILP; 4, isomer 2 of [pE]-ILP.

Circular Dichroism Analysis of the Disulfide Isomers of ILP and [pE]ILP. The structural differences of the isomers of ILP and [pE]ILP were analyzed by circular dichroism

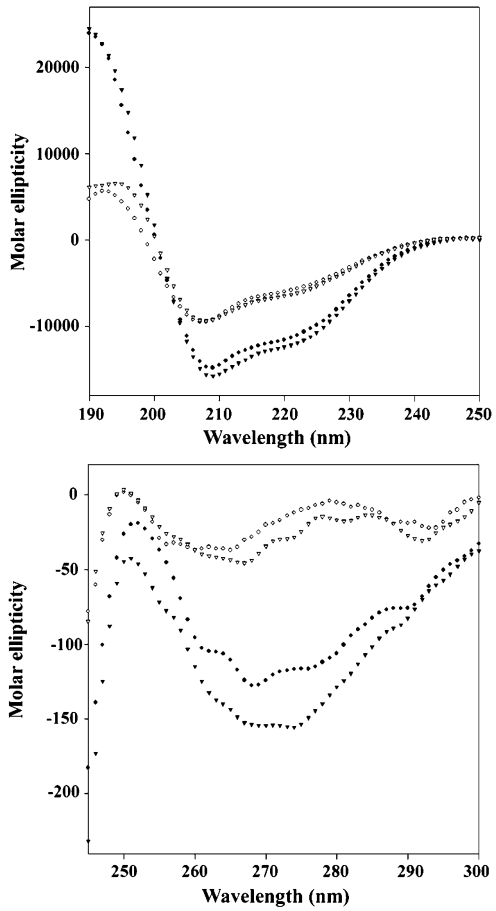


FIGURE 3: Circular dichroism analysis of the disulfide isomers of ILP and [pE]ILP in the far-UV region (upper panel) and near-UV region (lower panel). The open circle represents isomer 1 of ILP; the filled circle represents isomer 2 of ILP; the open triangle represents isomer 1 of [pE]ILP; the filled triangle represents isomer 2 of [pE]ILP.

(Figure 3). In both the far-UV and near-UV regions, the spectra of ILP and [pE]ILP (isomer 1 vs isomer 1; isomer 2 vs isomer 2) are very similar, which suggests that cyclization of the N-terminal residue has no significant effect on the conformation of ILP. However, the spectra of isomer 1 and isomer 2 are quite different in both the far-UV and near-UV regions. Therefore, the secondary and tertiary structures of the two isomers of ILP or [pE]ILP have significant differences: the conformation of isomer 2 is more ordered than that of isomer 1. The estimated α-helix content from the CD spectra of isomer 1 is significantly lower than that of isomer 2: isomer 1 of ILP, 12%; isomer 2 of ILP, 30%; isomer 1 of [pE]ILP, 16%; isomer 2 of [pE]ILP, 34%. Together with the biological activity of the ILP analogues (*41*), we deduced that the disulfides of isomer 2 are identical to those of insulin/native IGF-1, while the disulfides of isomer 1 are identical to those of swap IGF-1.

Peptide Mapping of the Disulfide Isomers of ILP and [pE]-ILP. To elucidate the disulfide linkages of the isomers of ILP and [pE]ILP, V8 endoproteinase digestion was carried out. The digested fragments were separated by C8 reverse-phase HPLC (Figure 4), and the measured molecular masses of the digested fragments are listed in Table 2. As shown in Figure 4, the chromatography profiles of ILP and [pE]ILP (isomer 1 vs isomer 1; isomer 2 vs isomer 2) are almost identical, while the two isomers show different characteris-

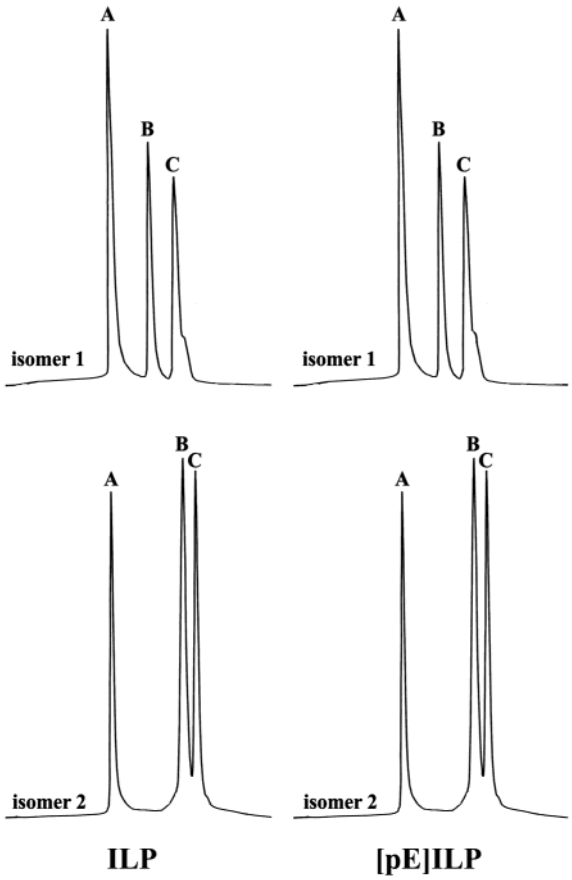


FIGURE 4: HPLC profiles of the V8-digested isomers of ILP and [pE]ILP. The digested mixture was loaded onto a C8 reverse-phase column, eluted with the gradient listed in Materials and Methods with a flow rate of 0.5 mL/min, and detected at 230 nm.

tics. Both of the two isomers of ILP or [pE]ILP have three major peptide fragments. Among them, peak A of the two isomers shows almost identical retention time; peak B of isomer 1 and peak C of isomer 2 also have similar retention times. As shown in Table 2, peak A of the two isomers is the fragment of A13Tyr−A17Glu; peak C of isomer 1 and peak B of isomer 2 are fragments containing the disulfide A20−B19; peak B of isomer 1 and peak C of isomer 2 are the fragments containing B7Cys, A6Cys, A7Cys, and A11Cys. The peptide mapping results suggest that the two isomers of ILP or [pE]ILP both contain the common disulfide A20−B19, and the B7Cys, A6Cys, A7Cys, and A11Cys form one intrachain and one interchain disulfide bond. It is the different disulfide pairing that results in the different retention time of peak B of isomer 1 and peak C of isomer 2. Together with the result of CD analysis, the folding behavior of IGF-1/mini-IGF-1 (*37*), and the results of ILP and its mutants (*40*, *41*), it can reasonably be presumed that the disulfides of isomer 2 are identical to those of insulin/native IGF-1, while the disulfides of isomer 1 are identical to those of swap IGF-1. Moreover, all of the three major fragments of [pE]ILP have the expected molecular masses; therefore, the different molecular masses between ILP and [pE]ILP must be caused by the difference of the three N-terminal residues, Gln-Ala-Glu. This observation suggests that the unexpected component, [pE]ILP, is caused by the cyclization of the N-terminal glutamine residue.

Disulfide Rearrangement of the Disulfide Isomers of ILP and [pE]ILP. If the two isomers of ILP or [pE]ILP were thermodynamically controlled folding products, they could be converted each other and finally reach an equilibrium in the presence of thiol catalyst. As expected, whether starting from isomer 1 or isomer 2, an equilibrium can be reached under a low concentration of thiol catalyst (Figure 5). For ILP, the molar ratio of isomer 1 to isomer 2 is 1−2:100 at 15 °C and 3−4:100 at 37 °C; for [pE]ILP, the molar ratio of isomer 1 to isomer 2 is 3−4:100 at 15 °C and 6−7:100 at 37 °C. From the equilibrium constant, the folding free energy difference ($\Delta\Delta G$) of isomer 2 to isomer 1 can be calculated: for ILP, $\Delta\Delta G$ is −9.4 to −11.0 kJ/mol at 15 °C and −8.3 to −9.0 kJ/mol at 37 °C; for [pE]ILP, $\Delta\Delta G$ is −7.7 to −8.4 kJ/mol at 15 °C and −6.9 to −7.3 kJ/mol at 37 °C. These results suggest that, first, as the temperature increases, isomer 1 becomes more favored, and, second, cyclization of the N-terminal residue favors the formation of the isomer with swap disulfides.

In Vitro Refolding of the Disulfide Isomers of ILP and [pE]ILP. To further demonstrate that both of the two disulfide isomers of ILP or [pE]ILP are thermodynamically controlled folding products, in vitro refolding was carried out. After reduction an aliquot was removed, and the free thiols were modified by sodium iodoacetate and then analyzed by native PAGE (data not shown). PAGE indicates that the three disulfides of ILP or [pE]ILP are fully reduced under the condition used. The refolding was initiated by dilution of the fully reduced isomer into the refolding buffer; after incubation the refolding product was analyzed by C8 reverse-phase HPLC (Figure 6). During refolding each isomer folds into a mixture of the two isomers with a ratio similar to that obtained from the disulfide rearrangement. Calculated from the peak area, the refolding yields of the two isomers of ILP or [pE]ILP are both over 90% under the refolding conditions used. This result further demonstrates that both of the two isomers of ILP or [pE]ILP are thermodynamically controlled disulfide isomers.

DISCUSSION

From the fermentation supernatant of ILP, two major products were obtained: one with the expected molecular mass and the other with unexpected molecular mass. Our present results suggest that the unexpected component is caused by the cyclization of the N-terminal glutamine residue, and its cyclization has no significant effect on the conformation of ILP. For both ILP and [pE]ILP, two disulfide isomers have been identified, but the percentage of isomer 1 is much lower than that of isomer 2. Peptide mapping shows that the two isomers of ILP or [pE]ILP have the common disulfide A20−B19, but the pairing of the other two disulfides is different. The CD analysis shows that the conformation of isomer 2 is much more ordered than that of isomer 1. The results of disulfide rearrangement and refolding suggest that both of the two isomers are thermodynamically controlled folding products. Together with the folding behavior of IGF-1/mini-IGF-1 as well as the biological activity of ILP analogues (*41*), it can reasonably be presumed that the two isomers are caused by different disulfide linkages: the disulfides of isomer 1 are identical to those of the swap IGF-1, while the disulfides of isomer 2 are identical to those of insulin/native IGF-1.

The present results reveal that the folding behavior of ILP shares the characteristics of both insulin/PIP and IGF-1. On

Table 2: Molecular Masses of the V8-Digested Fragments of the Isomers of ILP and [pE]ILP

sample	letter in Figure 4	retention time on HPLC (min)	measured molecular mass	corresponding fragment
isomer 1 of ILP	A	13.81	638.0	A13Tyr−A17Glu
	B	15.77	1756.0	B5Tyr−B13Asp + A6Cys−A12Asp
	C	17.05	3051.0	B14Val−A5Glu + A18Ser−A21Asn
isomer 2 of ILP	A	13.84	638.0	A13Tyr−A17Glu
	B	17.21	2921.0	B14Val−A4Glu + A18Ser−A21Asn
	C	17.86	1885.0	B5Tyr−B13Asp + A5Glu−A12Asp
isomer 1 of [pE]ILP	A	13.89	638.0	A13Tyr−A17Glu
	B	15.77	1756.0	B5Tyr−B13Asp + A6Cys−A12Asp
	C	16.99	3051.0	B14Val−A5Glu + A18Ser−A21Asn
isomer 2 of [pE]ILP	A	13.85	638.0	A13Tyr−A17Glu
	B	17.24	2921.0	B14Val−A4Glu + A18Ser−A21Asn
	C	17.89	1885.0	B5Tyr−B13Asp + A5Glu−A12Asp

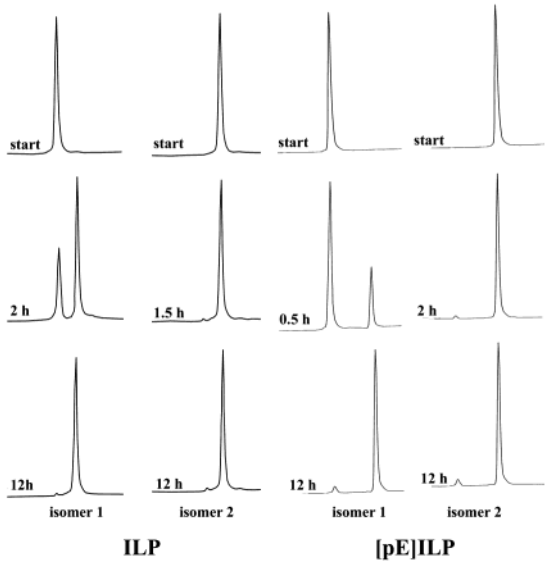


FIGURE 5: Disulfide rearrangement of the isomers of ILP and [pE]-ILP at 15 °C. At indicated reaction times 100 μL (5 μg) of the sample was removed and acidified by TFA to pH 2.0 and immediately loaded onto a C8 reverse-phase column. The elution gradient described in Materials and Methods was used with a flow rate of 0.5 mL/min and detected at 230 nm.

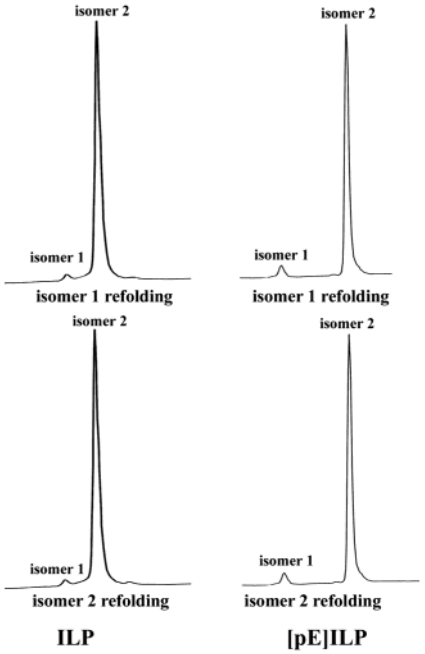


FIGURE 6: In vitro refolding of the isomers of ILP and [pE]ILP. After incubation 100 μL of the refolding mixture was removed and acidified by TFA and then analyzed by a C8 reverse-phase column. The elution gradient described in Materials and Methods was used with a flow rate of 0.5 mL/min and detected at 230 nm.

one hand, two thermodynamically controlled disulfide isomers of ILP have been identified; this behavior is similar to that of IGF-1. On the other hand, isomer 2 (with native disulfides) is much more favored than isomer 1 (with non-native disulfides); that is, more than 96% of ILP folds into the native structure. This behavior is similar to the refolding of insulin/PIP that folds into the native structure with nondetectable disulfide isomers under undenatured conditions. Since ILP is deduced to be the common ancestor of insulin and IGF-1, it can thus be deduced that the tendency of forming the disulfide isomer with non-native disulfides is diminished during evolution from ILP to insulin, while this tendency is amplified during evolution from ILP to IGF-1. As a result, insulin/PIP folds into one unique thermodynamically controlled structure, while IGF-1 folds into two disulfide isomers with similar thermodynamic stability. Therefore, the refolding behavior of amphioxus ILP suggests that the different folding behaviors of insulin and IGF-1 were evolved through a bifurcating evolution from a common ancestor.

Although the cyclization of the N-terminal residue has no significant effect on the conformation of ILP, it has a significant effect on the folding behavior. Although two disulfide isomers of [pE]ILP also have been identified, the equilibrium of the two isomers has been altered: the percentage of isomer 1 of [pE]ILP is approximately 2-fold of that of isomer 1 of ILP. Therefore, cyclization of the N-terminal residue favors the formation of the isomer with swap disulfides. Our previous work suggested that the B-chain/domain plays a critical role in determining the different folding behavior of insulin and IGF-1 (*37*, *38*) and proposed the co-evolution of IGF-1 and its binding proteins. Since the N-terminal sequence of IGF-1 is important for binding with IGF binding proteins, we deduce that the N-terminal residue probably has a significant effect on the folding behavior of insulin/PIP and IGF-1. The present results are consistent with our above deduction: the N-terminal sequence of the B-chain/domain has a significant effect on the folding behavior of insulin/PIP, IGF-1, and ILP.

ACKNOWLEDGMENT

We thank Dr. Yan Chen and Rui Jin for valuable discussion on this project.

REFERENCES

1. Anfinsen, C. B. (1973) *Science 181*, 223−230.
2. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) *J. Mol. Biol. 277*, 985−994.
3. Viguera, A. R., Serrano, L., and Wilmanns, M. (1996) *Nat. Struct. Biol. 3*, 874−880.
4. Baker, D. (2000) *Nature 405*, 39−42.
5. Creighton, T. E. (1974) *J. Mol. Biol. 87*, 603−642.
6. Creighton, T. E., Darby, N. J., and Kemmink, J. (1996) *FASEB J. 10*, 110−118.
7. Darby, N. J., Morin, P. E., Talbo, G., and Creighton, T. E. (1995) *J. Mol. Biol. 249*, 463−447.
8. Weissman, J. S., and Kim, P. S. (1991) *Science 253*, 1386−1393.
9. Weissman, J. S., and Kim, P. S. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 497−526.
10. Wedemeyer, W. J., Welker, E., Narayan, M., and Scheraga, H. A. (2000) *Biochemistry 39*, 4207−4216.
11. Rothwarf, D. M., and Scheraga, H. A. (1993) *Biochemistry 32*, 2671−9216.
12. Wu, J., Yang, Y., and Watson, J. T. (1998) *Protein Sci. 7*, 1017−1028.
13. Grantcharova, V., Alm, E. J., Baker, D., and Horwich, A. (2001) *Curr. Opin. Struct. Biol. 11*, 70−81.
14. Ellgaard, L., Molinari, M., and Helenius, A. (1999) *Science 286*, 1882−1888.
15. Wickner, S., Maurizi, M. R., and Gottesman, S. (1999) *Science 286*, 1888−1893.
16. Barnhart, M. M., Pinkner, J. S., Soto, G. E., Sauer, F. G., Langermann, S., Waksman, G., Frieden, C., and Hultgren, S. J. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 7709−7714.
17. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M. C., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N., and Vijayan, N. M. (1988) *Philos. Trans. R. Soc. London, Ser. B 319*, 369−456.
18. The Peking Insulin Structure Research Group (1974) *Sci. Sin. 17*, 752−778.
19. Weiss, M. A., Hua, Q.-X., Frank, B. H., Lynch, C., and Shoelson, S. E. (1991) *Biochemistry 30*, 7373−7389.
20. Roy, M., Lee, R. W.-K., Brange, J., and Dunn, M. F. (1990) *J. Biol. Chem. 265*, 5448−5453.
21. Olsen, H. B., Ludvigsen, S., and Kaarsholm, N. C. (1996) *Biochemistry 35*, 8836−8845.
22. Wang, C. C., and Tsou, C. L. (1991) *Trends Biochem. Sci. 16*, 279−281.
23. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Bing, X. G., and Markussen, J. (1991) *J. Mol. Biol. 220*, 425−433.
24. Hua, Q.-X., Hu, S.-Q, Jia, W., Chu, Y.-C., Burke, G. T., Wang, S.-H., Wang, R.-Y., Katsoyannis, P. G., and Weiss, M. A. (1998) *J. Mol. Biol. 277*, 103−118.
25. Zhang, Y.-S., Hu, H.-M., Cai, R.-R., Feng, Y.-M., Zhu, S.-Q., He, Q.-B., Tang, Y.-H., Xu, M.-H., Xu, Y.-G., Liu, B., and Liang, Z.-H. (1996) *Sci. China (Ser. C) 39*, 225−233.
26. Qiao, Z.-S, Guo, Z.-Y, and Feng, Y.-M. (2001) *Biochemistry 40*, 2662−2668.
27. Humbel, R. E. (1990) *Eur. J. Biochem. 190*, 445−462.
28. Cook, R. M., Harvey, T. S., and Campbell, I. D. (1991) *Biochemistry 30*, 5484−5491.
29. Hober, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B. (1992) *Biochemistry 31*, 1749−1751.
30. Hober, S., Ljung, J. L., Uhlen, M., and Nilsson, B. (1999) *FEBS Lett. 443*, 271−276.
31. Miller, J. A., Narhi, L., Hua, Q.-X. Rosenfeld, R., Arakawa, T., Rodhe, M., Prestrelski, S., Lauren, S., Stoney, K. S., Tsai, L., and Weiss, M. A. (1993) *Biochemistry 32*, 5203−5213.
32. Hober, S., Ljung, J. L., Uhlen, M., and Nilsson, B. (1997) *Biochemistry 36*, 4616−4622.
33. Rosenfeld, R. D., Miller, J. A., Narhi, L. O., Hawkins, N., Katta, V., Lauren, S. L., Weiss, M. A., and Arakawa, T. (1997) *Arch. Biochem. Biophys. 342*, 298−305.
34. Milner, S. J., Carver, J. A., Ballard, F. J., and Francis, G. (1999) *Biotechnol. Bioeng. 62*, 693−703.
35. Narhi, L. O., Hua, Q. X., Arakawa, T., Fox, G. M., Tsai, L., Rosenfeld, R., Holst, P., Miller, J. M., and Weiss, M. A. (1993) *Biochemistry 32*, 5214−5221.
36. Hober, S., Hansson, A., Uhlen, M., and Nilsson, B. (1994) *Biochemistry 33*, 6758−6761.
37. Guo, Z.-Y., Shen, L., and Feng, Y.-M. (2002) *Biochemistry 41*, 1556−1567.
38. Guo, Z.-Y., Shen, L., and Feng, Y.-M. (2002) *Biochemistry 41*, 10585−10592.
39. Chan, S. J., Cao, Q. P., and Steiner, F. D. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9319−9323.
40. Shen, L., Guo, Z.-Y., Chen, Y., Liu, L.-Y., and Feng, Y.-M. (2001) *Acta Biochim. Biophys. Sin. 33*, 629−633.
41. Guo, Z.-Y., Shen, L., Gu, W., Wu, A.-Z., Ma, J.-G., and Feng, Y.-M. (2002) *Biochemistry 41*, 10603−10607.
42. Shirotani, K., Tsubuki, S., Lee, H.-J., Maruyama, K., and Saido, T. C. (2002) *Neurosci. Lett. 327*, 25−28.
43. Khandke, K. M., Fairwell, T., Chait, B. T., and Manjula, B. N. (1989) *Int. J. Pept. Protein Res. 34*, 118−123.

BI0346289