Amphioxus insulin-like peptide (AILP) belongs to the insulin superfamily and is proposed as the common ancestor of insulin and insulin-like growth factor 1. Herein, the studies on oxidative refolding and reductive unfolding of AILP are reported. During the refolding process, four major intermediates, P1, P2, P3, and P4, were captured, which were almost identical to those intermediates, U1, U2, U3, and U4, captured during the AILP unfolding process. P4 (U4) has the native disulfide A20-B19; P1 (U1), P2 (U2), and P3 (U3) have two disulfide bonds, which include A20-B19. Based on the analysis of the time course distribution and properties of the intermediates, we proposed that fully reduced AILP refolded through 1SS, 2SS, and 3SS intermediate stages to the native form; native AILP unfolded through 2SS and 1SS intermediate stages to the fully reduced form. A schematic flow chart of major oxidative refolding and reductive unfolding pathways of AILP was proposed. Implications for the folding behavior of insulin family proteins was discussed. There may be seen three common folding features in the insulin superfamily: 1) A20-B19 disulfide is most important and formed during the initial stage of folding process; 2) the second disulfide is nonspecifically formed, which then rearranged to native disulfide; 3) in vitro refolding and unfolding pathways may share some common folding intermediates but flow in opposite directions. Furthermore, although swap AILP is a thermodynamically stable final product, a refolding study of swap AILP demonstrated that it is also a productive intermediate of native AILP during refolding.

Since Anfinsen and co-workers (1) first demonstrated in the 1960s that the amino acid sequence of a protein determines its three-dimensional structure, significant advances have been made in the understanding of protein folding through experimental and theoretical approaches (2–4). However, the folding mechanism, such as how a linear polypeptide strand starts and walks along what pathway to fold into a protein molecule, is still poorly understood. In the investigations of the protein folding pathway, the proteins with disulfide linkages are frequently chosen as models, because the formation of disulfide bridges is always coupled with folding and assembly, which can be used as a unique probe for the study of protein folding and assembly (5–7). Oxidative folding studies on the disulfide-coupled folding of some small globular proteins, such as bovine pancreatic trypsin inhibitor (8, 9), RNase A (10), hirudin (11), etc., have revealed the sequences of preferred kinetic intermediates of such proteins, by which the in vitro folding pathway of the proteins was defined.

The insulin superfamily members are good model proteins for the folding studies of disulfide-containing proteins. Insulin and insulin-like growth factor 1 (IGF-1) are two extensively studied members of the superfamily. They have similar three-dimensional structure and identical disulfide pairing (12, 13). Both proteins mainly consist of three α-helical segments (B9–B19, α-helix 1; A2–A8, α-helix 2; and A13–A19, α-helix 3) in insulin; 8–18, α-helix 1; 42–49, α-helix 2; and 54–61, α-helix 3 in IGF-1) in the A- and B-chains/domains; the three α-helical segments are stabilized by three identical disulfide bonds (A7-B7 and A20-B19, interchain disulfides; A6-A11, intra-A-chain disulfide in insulin; 6–48 and 18–61, interchain disulfides; 47–52, intra-A-domain disulfide in IGF-1); when IGF-1 and the insulin T-form are superimposed on the Cα positions of their respective helical segments, the root mean square deviation is only 0.47 Å (12, 13). The contributions of disulfide pairing to the native structure of insulin and IGF-1 were also investigated in analogues lacking selected disulfide bridges (14–19). In insulin, A20-B19 appears to be integral to the overall structure and necessary for the biosynthetic expression. The removal of A7-B7 or A6-A11 in insulin results in segmental unfolding. The stabilization of helix 2 needs the correct disulfide pairing of A7-B7 (6–48 in IGF-1) and A6-A11 (47–52 in IGF-1). Breaking either disulfide results in unfolding of helix 2 in both insulin and IGF-1. Although folding studies of insulin and IGF-1 revealed that they share some common features in their folding pathways (20–27), their folding behaviors are significantly different; insulin/PIP folds into one thermodynamic structure (20–22), whereas IGF-1 folds into two isomers.

1 The abbreviations used are: IGF-1, insulin-like growth factor 1; AILP, amphioxus insulin-like peptide; rAILP, a recombinant AILP, in which the D-domain has been deleted and a dipeptide Ala-Lys was used instead of the C-domain; [Gly]rAILP, rAILP mutant, in which the N terminus Gln residue of rAILP was replaced by a Gly residue; fr-[Gly]rAILP, fully reduced [Gly]rAILP; PIP, recombinant porcine insulin precursor, in which the B-chain C terminus and A-chain N terminus of porcine insulin are linked together by a dipeptide, Ala-Lys; HIP, human preinsulin; ESI-MS, electrospray ionization mass spectrometry; DTT, reduced dithiothreitol; HPLC, high performance liquid chromatography; IAA, sodium salt of iodoacetic acid; PN, native [Gly]rAILP; PS, swap [Gly]rAILP.

55224 This paper is available on line at http://www.jbc.org

Yan Chen‡, Rui Jin‡, Hong-Yun Dong, and You-Min Feng§

From the Key Laboratory of Proteomics, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Graduate School of the Chinese Academy of Sciences Shanghai 200031, China

Received for publication, August 6, 2004, and in revised form, October 12, 2004 Published, JBC Papers in Press, October 22, 2004, DOI 10.1074/jbc.M409030200

In Vitro Refolding/Unfolding Pathways of Amphioxus Insulin-like Peptide

IMPLICATIONS FOR FOLDING BEHAVIOR OF INSULIN FAMILY PROTEINS*
residues are shown as dark circles. The dipeptide, AK, in the rectangle represents the connecting peptide linking the B-chain and the A-chain. For convenience of description, the sequence number refers to insulin A- and B-chains. The solid lines indicate native disulfides (A6-A11, A7-B7, and A20-B19); the dotted lines indicate nonnative disulfides (A7-A11 and A6-B7). The potential cleavage sites of protease V8 are indicated with arrows. Peptide fragments obtained after enzyme cleavages are designated A-G.

**Experimental Procedures**

Materials—The Escherichia coli strain was used was DH12S. Saccharomyces cerevisiae strain XY700–6B (leu2, ura3, pep4) was kindly provided by Michael Smith (University of British Columbia, Vancouver, Canada). Plasmid pVT102-U/aMFL-AILP was previously constructed in our laboratory (29). The chemical reagents used in experiments were analytical grade. The Amersham Biosciences reverse-phase column (Sephasil Peptide C8 5µ ST 4.6×250), Gilson 306 HPLC system, and Gilson 115 UV detector were used. In HPLC analysis, a gradient elution was used. Solvent A was 0.15% aqueous trifluoroacetic acid; solvent B was 60% acetonitrile containing 0.125% trifluoroacetic acid. The elution gradient was as follows: 0 min, 0% solvent B; 1 min, 0% solvent B; 5 min, 45% solvent B; 32 min, 90% solvent B; 33 min, 100% solvent B; 35 min, 100% solvent B; 38 min, 0% solvent B; 42 min, 0% solvent B. During analysis, the flow rate is 0.5 ml/min, and the detection wavelength is 230 nm.

**DNA Manipulation**—The expression plasmid encoding [Gly]rAILP was constructed using a gapped duplex DNA approach (32) for site-directed mutagenesis on the basis of pVT102-U/aMFL-AILP. The expected mutation was confirmed by DNA sequencing. The plasmid constructed was designated as pVT102-U/aMFL-[Gly]rAILP.

**Expression, Purification, and Identification of [Gly]rAILP**—The plasmid pVT102-U/aMFL-[Gly]rAILP was transformed into the yeast cells XY700–6B (leu2, ura3, pep4). The transformed yeast cells were cultured in a 16-liter fermenter, and the secreted target protein was purified from the medium with four steps as previously reported (29).

**Putative Refolding/Unfolding Pathways of AILP**

![Diagram of AILP structure](image)

**FIG. 1. Amino acid sequence, disulfide linkage, and purification of [Gly]rAILP.** A, amino acids are shown in one-letter code, and cysteine residues are shown as dark circles. The dipeptide, AK, in the rectangle represents the connecting peptide linking the B-chain and the A-chain. For convenience of description, the sequence number refers to insulin A- and B-chains. The solid lines indicate native disulfides (A6-A11, A7-B7, and A20-B19); the dotted lines indicate nonnative disulfides (A7-A11 and A6-B7). The potential cleavage sites of protease V8 are indicated with arrows. Peptide fragments obtained after enzyme cleavages are designated A-G. **B**, the C8 reverse-phase HPLC analysis and native pH 8.3 PAGE of [Gly]rAILP were performed as described under “Experimental Procedures.” HPLC profile and PAGE pattern demonstrated that [Gly]rAILP was purified to homogeneity.

AILP belongs to insulin superfamily and is proposed as a common ancestor of insulin and IGF-1. AILP deduced from the cDNA sequence is a polypeptide that displays structural characteristics of both mammalian insulin and IGF-1. The deduced AILP contains B-, C-, A-, and D-domains similar to those of IGF-1, whereas 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 (28). Previously, a recombinant single-chain AILP (rAILP) has been constructed (29), in which the D-domain of AILP was deleted; the C-domain was replaced by Gly. Here we report the investigation of the in vitro refolding/unfolding pathways of [Gly]rAILP. Based on our current results, we sketched out a possible folding/unfolding pathway of [Gly]rAILP and proposed some common features of the insulin superfamily folding pathway.
and redox ratio, respectively, were estimated by reverse-phase HPLC analysis.

**Trapping and Fractionation of Refolding Intermediates**—Refolding of fr[Gly]rAILP was carried out in the optimized refolding condition (0.1 mg/ml fr[Gly]rAILP, 50 mM Tris-HCl, 1 mM EDTA, 1 mM GSSG, and 1 mM DTT, pH 9.5, 16 °C). The refolding reaction was quenched by adding one-fourth volume of freshly prepared 0.5 M IAA and incubated for 10 min at 30 °C. To monitor the refolding process, at different time points, 100 μl of sample was removed and quenched and then applied to the analytical C8 reverse-phase HPLC column under the conditions described under “Materials.” For large scale preparation of intermediates, we quenched the folding reaction by IAA after 15 min of refolding. The sample was first dialyzed against 50 mM NH₄HCO₃ at 4 °C and then applied to a semiprepared C8 RP-HPLC column (Sephasil peptide, ST 4.6/250 mm) (Amersham Biosciences). The HPLC conditions were as described under “Materials” except that the flow rate was 1.5 ml/min. The individual fractions were collected manually and then lyophilized for further use.

**Identification of the Disulfide Linkage of Folding Intermediates**—To illustrate their disulfide linkage, the intermediates were dissolved in 0.1 M phosphate buffer (pH 7.8) and applied to V8 endoproteinase digestion at 25 °C for 16 h. The ratio of enzyme to sample was around 1:20. The reaction was terminated by making the solution in 2% trifluoroacetic acid. The sample was then loaded onto a C8 reverse-phase HPLC column for analysis and separation. The fractions were collected manually and lyophilized. Their molecular masses were measured by ESI-MS.

**Circular Dichroism Analysis of Refolding Intermediates**—The samples were each dissolved in 10 mM HCl. The protein concentration was determined by UV absorbance at 276 nm using an extinction coefficient of 1.2 ml mg⁻¹ cm⁻¹ and diluted to 0.2 mg/ml, respectively. The analysis was performed on a Jasco-715 CD spectropolarimeter at room temperature. The spectra were scanned from 200 to 250 nm using a cell with a 0.1-cm path length in the far-UV region and scanned from 300 to 245 nm using a cell with a 1.0-cm path length in 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.

**Reductive Unfolding of [Gly]rAILP**—[Gly]rAILP was dissolved in an alkaline buffer (100 mM Tris·HCl, 2 mM EDTA, pH 8.7) at a final concentration of 1 mg/ml. Reductive unfolding was initiated by adding the same volume of 10 mM DTT. The final reaction solution contains 0.5 mg/ml [Gly]rAILP, 50 mM Tris·HCl, 1 mM EDTA, and 5 mM DTT. The unfolding reaction was quenched by adding one-fourth volume of
freshly prepared 0.5 mM IAA and incubated for 10 min at 30 °C. To monitor the unfolding process, at different time points, 100 μl of sample was removed and quenched and then applied to the analytical C8 reverse-phase HPLC column under the conditions described under “Materials.” For large scale purification of intermediates, the reaction was quenched by IAA after reductive unfolding for 4 min and then applied to a semiprepared C8 HPLC column (Sephasil peptide, trivial B) (Amersham Biosciences). The HPLC conditions were as described under “Materials” except that the flow rate was 1.5 ml/min. The individual fractions were collected manually and then lyophilized. Native pH 8.3 PAGE and V8 digestion were applied, respectively, to the purified unfolding intermediates for identification.

Oxidative Refolding and Reductive Unfolding of Swap [Gly]rAILP—Refolding of swap [Gly]rAILP was carried out at 16 °C. Refolding buffer was composed of 0.1 mg/ml swap [Gly]rAILP, 50 mM Tris-HCl, 1 mM EDTA, 1 mM DTT, 2 mM GSSG, pH 9.5. At different time points, the refolding was quenched by IAA and analyzed by HPLC analysis. Unfolding of swap [Gly]rAILP was carried out at 16 °C with the unfolding buffer composed of 0.5 mg/ml swap [Gly]rAILP, 50 mM Tris-HCl, 1 mM EDTA, and 5 mM DTT, pH 8.7. At different time points, unfolding was quenched by IAA and analyzed by HPLC analysis.

RESULTS

Expression and Purification of [Gly]rAILP—The plasmid pVT102U/a-MFL-[Gly]rAILP was constructed with correct DNA sequence (data not shown), transformed into the yeast cells, and expressed in a 16-liter fermenter. The expression product of the single-chain [Gly]rAILP was purified from the fermentation supernatant as described under “Experimental Procedures.” C8 reverse phase HPLC profile and pH 8.3 PAGE analysis show that the purified [Gly]rAILP is homogenous (Fig. 1B). The molecular mass of [Gly]rAILP measured by ESI-MS is 5631, agrees with the expected value.

Oxidative Refolding of [Gly]rAILP in Vitro—Refolding of [Gly]rAILP was carried out as described under “Experimental Procedures.” To optimize the refolding condition, we tested different concentration of GSSG and values of pH. There was a big jump in refolding yields of [Gly]rAILP when the concentration of GSSG increased from 0.5 to 1 mM and the refolding yields stayed at around 80% after the concentration of GSSG reached 1 mM. However, the refolding yields increased slightly as the pH value increased from 8.5 to 9.5 (data not shown). Therefore, the oxidative folding was carried out under the optimized condition: 50 mM Tris-HCl, 1 mM EDTA, and 1 mM GSSG, pH 9.5, 16 °C. In the present study, we use the IAA quenching method to trap the folding intermediates. We found that there was no change in the distribution of intermediates when the concentration of IAA used in trapping free thiols varied from 50 to 200 mM (data not shown), thus indicating that the quenching method with 100 mM IAA was suitable for our work. On the other hand, the acid-quenching method was also applied in comparison with the IAA quenching method. However, acid quenching refolding intermediates cannot be resolved well in HPLC (data not shown). In fact, here the IAA quenching method is convenient for separation, purification, and further characterization of folding intermediates.

As Fig. 2A shows, the refolding process was quenched by IAA at different times and analyzed by HPLC. Fig. 2B shows the magnified HPLC profile of [Gly]rAILP refolding quenched at the time of 15 min. Besides native (PN) and fully reduced form (R), seven obvious disulfide intermediates were detected in IAA-trapped refolding process. The first peak was proved by ESI-MS to contain three disulfides; thus, since it is the disulfide isomer of native [Gly]rAILP, we designated this peak as swap [Gly]rAILP (P3). The other six peaks were designated P1–P6, respectively. Capture and purification of folding intermediates were performed as described under “Experimental Procedures.” Their patterns in native pH 8.3 PAGE are shown in Fig. 2C, which indicates that each fraction contains mainly one component except for P6. P6 is a mixture of disulfide species and has not been further analyzed. ESI-MS analysis was then applied to elucidate the number of disulfides pairing in each fraction, because, compared with the native form, the molecular mass of IAA-modified

FIG. 3. Kinetics of oxidative refolding of [Gly]rAILP. Shown is the recovery time course of [Gly]rAILP (A) and 2SS (B) refolding intermediates in the presence of 1 mM GSSG. The recovery percentages indicate the ratios of recovered proteins relative to the [Gly]rAILP used for the refolding (0.1 mg/ml fr[Gly]rAILP, 50 mM Tris-HCl, 1 mM EDTA, 1 mM GSSG, and 1 mM DTT, pH 9.5, 16 °C). C, proportion of P1, P2, and P3 during the refolding process. For refolding carried out under 1 mM GSSG, the height of each bar reflects the average proportion of a specific refolding intermediate (P1, P2, or P3) in the total amount of the three 2SS refolding intermediates, which was averaged from nine different time points according to B. The error bar represents S.D. For refolding carried out under 2 mM GSSG or 4 mM GSSG, the same method was used to calculate the proportion of P1, P2, and P3 according to their time-dependent recovery of 2SS intermediates (data not shown).
intermediates with two disulfide bonds should increase by 118 Da, and those with one disulfide bond should increase by 236 Da. Therefore, both PS and PN have three disulfide bonds (3SS) without IAA modification. P1, P2, and P3 have two disulfide bonds (2SS) with two thiol groups modified by IAA. Both P4 and P5 have only one disulfide bond (1SS) with four free thiol groups modified by IAA. Fig. 3 shows the kinetics of oxidative refolding of [Gly]rAILP. When refolding was carried out in 1 mM GSSG, the modified by IAA. P5 have only one disulfide bond (1SS) with four free thiol groups without IAA modification. P1, P2, and P3 have two disulfide bonds (2SS) with four free thiol groups modified by IAA. Both PS and PN have three disulfide bonds (3SS) with four free thiol groups modified by IAA. The 1SS intermediate P4 has the disulfide linkage between fragment C and fragment D; thus, P4 contains disulfide A20-B19. The 1SS intermediate P5 contains IAA-modified fragments B and C; thus, P5 does not have any inter-AB-chain disulfide linkage. P5 might have an intra-E disulfide, which should be A6-A7, A6-A11, or A7-A11. The 2SS intermediate P1 contains disulfide A20-B19 and one intra-E disulfide, as P5 does. The 3SS species PS contains disulfide A20-B19 and one intra-E disulfide, as P5 does. The 3SS species PS contains disulfide A20-B19 and one intra-E disulfide, as P5 does. P1 has another disulfide paring between fragment B and fragment E; thus, it contains one inter-AB-chain disulfide besides A20-B19, which should be B7-A6, B7-A7, or B7-A11. Like P1, the 2SS intermediate P2 also contains A20-B19 disulfide and another disulfide among B7-A6, B7-A7, and B7-A11. The 2SS intermediate P3 contains disulfide A20-B19 and one intra-E disulfide, as P5 does. The 3SS species PS contains disulfide A20-B19 and one intra-E disulfide, as P5 does. The 3SS species PS contains disulfide A20-B19 and one intra-E disulfide, as P5 does. P1 has another disulfide paring between fragment B and fragment E; thus, it contains one inter-AB-chain disulfide besides A20-B19, which should be B7-A6, B7-A7, or B7-A11. Like P1, the 2SS intermediate P2 also contains A20-B19 disulfide and another disulfide among B7-A6, B7-A7, and B7-A11. The 2SS intermediate P3 contains disulfide A20-B19 and one intra-E disulfide, as P5 does.

**TABLE I**

| Protein digested with V8 | Molecular mass measured | Corresponding fragments in Fig. 1A | Corresponding disulfide linkage |
|-------------------------|-------------------------|----------------------------------|---------------------------------|
| PN (3SS)                | 1885                    | B + E + D                        | Between B and E, intra-E        |
| PS (3SS)                | 1756                    | B + E                             | A20-B19                         |
|                         | 1756                    | B + G                             | A20-B19                         |
|                         | 1756                    | B + G + D                         | A20-B19                         |
| P1 (2SS)                | 1875                    | B + E                             | Between B and E                 |
|                         | 1875                    | B + G                             | A20-B19                         |
|                         | 1875                    | B + G + D                         | A20-B19                         |
| P2 (2SS)                | 3050                    | B + G                             | Between B and E                 |
|                         | 3050                    | B + G + D                         | A20-B19                         |
| P3 (2SS)                | 3050                    | B + G                             | Between B and E                 |
|                         | 3050                    | B + G + D                         | A20-B19                         |
| P4 (1SS)                | 3050                    | B + G                             | No disulfide between B and E    |
|                         | 3050                    | B + G + D                         | A20-B19                         |
| P5 (1SS)                | 2625                    | B + G + E                         | No disulfide between B and E    |
|                         | 2625                    | C + IA A                          | A20-B19                         |

**FIG. 4. Far-UV circular dichroism spectra of the intermediates.** R, fr[Gly]rAILP. At 222 nm, the spectra from bottom to top were as follows: PN, P3, P2, PS, P1, P4, and P5.
both retain more native helix contents than swap [Gly]rAILP, suggesting that they are not folding precursors of swap [Gly]rAILP. So we deduced that P2 and P3 might be direct folding precursors of native [Gly]rAILP. They should have another native disulfide besides A20-B19 disulfide. Combined with the V8-digestion mapping result above, the disulfide pairing of P2 should be A20-B19 and A7-B7; the disulfide pairing of P3 should be A20-B19 and A6-A11. The 2SS intermediate P1 retained much less helix content than P2 and P3, so the other disulfide in P1 should be nonnative. V8 disulfide mapping indicated that P1 has a disulfide between B7 and one cysteine in the A-chain. It is possible that P1 is the direct folding precursor of swap [Gly]rAILP, and P1 has disulfide A20-B19 and A6-B7.

Reductive Unfolding of [Gly]rAILP in Vitro—Reductive unfolding is the reverse of oxidative refolding. Here we investigate the reductive unfolding pathway of [Gly]rAILP under 5 mM DTT in the absence of denaturants. Reductive unfolding of [Gly]rAILP was carried out as described under “Experimental Procedures.” As Fig. 5A shows, the unfolding process was quenched by IAA at different times and analyzed by HPLC. Fig. 5B shows the magnified HPLC profile of [Gly]rAILP unfolding quenched at the time of 4 min. Besides native (N) and fully reduced form (R), six obvious intermediates were detected in the IAA-trapped unfolding process, which were designated U1–U6, respectively. Most interestingly, by the comparison of Fig. 2B with Fig. 5B, it could be revealed that the HPLC profile of reductive unfolding is very similar to that of oxidative refolding. U1–U4 have HPLC retention times almost identical to those of P1–P4, respectively. Different from the refolding process, the peak corresponding to swap [Gly]rAILP was not identified during the unfolding process, suggesting that swap [Gly]rAILP is thermodynamically unstable in the presence of 5 mM DTT. Fig. 5C shows the patterns of the intermediates U1–U6 in native pH 8.3 PAGE, in which U5 is a mixture of oligomers, possibly formed because of aggregation; U6 is a mixture of intermediates, so we did not analyze further, and it is noticeable that U1–U4 share a similar PAGE pattern with P1–P4. Based on the evidence of C8 HPLC profile comparison and native pH 8.3 PAGE patterns, we suggested that the intermediates U1–U4 captured during reductive unfolding of [Gly]rAILP are almost the same as intermediates P1–P4 captured during oxidative refolding. Kinetics of reductive unfolding of [Gly]rAILP was shown in Fig. 6. The unfolding time course of [Gly]rAILP was shown in Fig. 6A. Time course curves of 2SS unfolding intermediates U1, U2, and U3 were shown in Fig. 6B. Similar to the refolding process, the proportion of U1, U2, and U3 to their total amount seemed to remain unchanged.
during the unfolding process. The relative content of U1/U2/U3 during the unfolding process in 5 mM DTT was 29:8:63, shown in Fig. 6C.

Refolding and Unfolding of Swap [Gly]rAILP—To investigate the role of swap [Gly]rAILP in the refolding/unfolding processes of [Gly]rAILP, the oxidative refolding and reductive unfolding of swap [Gly]rAILP were performed under the same conditions used in that of [Gly]rAILP. In the refolding process, swap [Gly]rAILP can fold efficiently to native [Gly]rAILP without accumulating apparent intermediates (Fig. 7). This indicates that swap [Gly]rAILP is an on-pathway intermediate in the folding process of [Gly]rAILP. In the unfolding process of swap [Gly]rAILP (Fig. 8), intermediates U1, U2, U3, and U4 were also identified as in the unfolding process of native [Gly]rAILP. Furthermore, native [Gly]rAILP was identified to exist during the unfolding process of swap [Gly]rAILP, which suggested that native [Gly]rAILP is much more stable than swap [Gly]rAILP under the reducing environment.

**DISCUSSION**

Reversible in Vitro Refolding/Unfolding Pathways and Identical Intermediates—AILP belongs to the insulin superfamily and was proposed as the common ancestor of insulin and IGF-1 (28). Its in vitro folding behavior exhibits the characteristics of both insulin and IGF-1 (31). Here we report the detailed investigation on the oxidative refolding and reductive unfolding of [Gly]rAILP. In the oxidative refolding, fr[Gly]rAILP refolds through the 1SS, 2SS, and 3SS stages to the native form. During the reductive unfolding, native [Gly]rAILP unfolds through the 2SS and 1SS stages to fr[Gly]rAILP. Based on the characterization of C8 HPLC profiles and native pH 8.3 PAGE patterns, the major unfolding intermediates, U1, U2, U3, and U4 are almost identical to those intermediates, P1, P2, P3, and P4, captured in the refolding process. From the analysis of the refolding/unfolding intermediates, we propose that the refolding and unfolding may share the same intermediates and follow a similar pathway but reverse. Similarly, in the in vitro refolding/unfolding process of HPI, an almost identical intermediate was also observed (21, 22).

Formation/Holding of the A20-B19 Disulfide Bridge Is the Most Important Step in Both Oxidative Refolding and Reductive Unfolding Pathways of [Gly]rAILP—The 1SS intermediate P4(U4) that contains the A20-B19 disulfide bridge is an essen-
The studies of other insulin family members also provide examples for the importance of A20-B19 disulfide. The pairwise substitution of cysteine residues of disulfide A20-B19 with serine in PIP resulted in an undetectable expression level of the derivative (19), whereas a PIP mutant retaining only one disulfide, A20-B19, has been successfully expressed in yeast (34), suggesting that the A20-B19 bond is important in the folding pathway and in maintaining its native conformation. CD results showed that it retained certain secondary structure, and the in vitro refolding experiment showed that cysteine A20 and cysteine B19 could pair rapidly with high yield (34). In both PIP and HPI, formation of disulfide A20-B19 is a key step during their in vitro refolding pathways (20–22). In the case of IGF-1, when the disulfide 18–61 (corresponding to the disulfide A20-B19 of insulin) is deleted, the mutant IGF-1 cannot be expressed, but the IGF-1 mutant with the single disulfide 18–61 acquires a compact partially folded conformation (35). Moreover, during in vitro refolding, the disulfide 18–61 is formed first (23–27). And this disulfide bond was retained in all of the intermediates identified. These results suggest that the disulfide 18–61 is the most essential disulfide bond in maintaining the native structure of IGF-1. Considering together the important role of A20-B19 disulfide bond in the folding of insulin, IGF-1, and [Gly]rAILP, it could be suggested that the in vitro folding pathway of insulin family members in the initial stage is conservative, in which the formation of the A20-B19 disulfide bond is most essential.

When A20-B19 Disulfide Is Formed (Held), the Formation (Holding) of Secondary Disulfide Is Nonspecific—Both native and nonnative disulfide could be formed (held) during the 2SS stage. 2SS intermediates P1(U1), P2(U2), and P3(U3) retain dynamic equilibrium during refolding and unfolding process, indicating that the disulfide reshuffling among these 2SS intermediates is much faster than transferring to 3SS or 1SS species by oxidative (reductive) reaction. The nonspecific formation of the secondary disulfide during the refolding process could reduce the entropy quickly, which leads to a decrease in the time needed for conformational searching. Compared with the 1SS intermediate P4, P1 does not have more native conformation, which indicates that the formation of the second disulfide in P1 was not coupled with conformational folding. Under oxidative environment, P1 can either be oxidized to form swap [Gly]rAILP or reshuffle to P2 or P3 by means of conformational folding. P2 and P3 maintain more native conformation than the final native product. P1 can also be reduced to form native [Gly]rAILP by disulfide formation coupled with conformational folding.

Nonspecific formation of secondary disulfide was also noticeable in both the folding pathway of insulin/PPI and IGF-1 (22–27). Previously, we have analyzed four two-disulfide PIP models that retain A20-B19 disulfide (36), and we found that all of the model peptides adopted a partially folded conformation; the refolding rates of the four model peptides were similar, but their refolding yields were different. This indicates that the two-disulfide intermediates corresponding to the model peptides all can be formed during PIP refolding, but their probabilities are different.

Role of Swap [Gly]rAILP—Swap [Gly]rAILP might be a kinetic trap from the oxidation of unstructured 2SS intermediate P1. However, swap [Gly]rAILP should be viewed as a productive intermediate, because it can reshuffle to native [Gly]rAILP efficiently without accumulating apparent intermediates. The breakage of nonnative disulfide in swap [Gly]rAILP might be accompanied with quickly conformational folding; thus, the breakage of nonnative disulfide was immediately followed by the formation of native disulfide. Swap [Gly]rAILP is also a
final refolding product because it exists when oxidative refolding was performed overnight to achieve equilibrium. Native [Gly]rAILP can also reshuffle back to form swap [Gly]rAILP, which suggested that swap and native [Gly]rAILP are under thermodynamic equilibrium in a redox environment.

Implication for the Folding Behavior of Insulin Family Proteins—Energy funnel theory tells us that protein folding is something like ensembles of unfolded polypeptide flow through different trajectories of the funnel to the end point of ordered three-dimensional structure (2, 3). Here we considered the formation of certain disulfide as kinetic traps of folding trajectory. Based on the above discussions, we sketched out a schematic flow chart of both oxidative refolding and reductive unfolding of [Gly]rAILP (Fig. 9). The fr[Gly]rAILP flowed through different folded 1SS, 2SS, and 3SS stages to form native [Gly]rAILP.

Three common folding features in the insulin superfamily are proposed based on the folding studies of insulin/PIP/HPI, IGF-1, and [Gly]rAILP. 1) The A20-B19 disulfide is most important and formed during the initial stage of the folding process. 2) The second disulfide was nonspecifically formed and then rearranged to native disulfide. 3) the in vitro refolding and unfolding pathway may share some common folding intermediates but flow in the opposite direction (Fig. 9).

The different folding behavior of insulin and IGF-1 might evolve from AILP. Insulin and IGF-1 have significantly different folding behavior: insulin folds into one thermodynamically stable structure, whereas IGF-1 folds into two disulfide isomers with similar thermodynamic stability. The role of swap insulin and swap IGF-1 is different. On one hand, swap insulin could only be obtained as a kinetic trap with the existence of denaturant (37). During refolding of HPI, the intermediate with swap disulfide might be the most important folding intermediate (21). Thus, in the oxidative folding of insulin, the swap form might be viewed as a productive kinetic folding intermediate that disappeared after folding was completed. On the other hand, swap IGF-1 played as a thermodynamically stable final product as the native form (23–27). Here, our work demonstrated that the role of swap [Gly]rAILP was combination of the roles of swap insulin and swap IGF-1 (i.e., it plays as both productive kinetic folding intermediate and thermodynamically stable final product in [Gly]rAILP refolding, suggesting the different folding behavior of insulin and IGF-1 might be acquired through a bifurcating evolution from AILP). In addition, we have reported that the different folding behavior of insulin and IGF-1 is the result of their different B-chain/domains that, somehow, control the different energetic state of the intra-A-chain/domain disulfide (38, 39) and that exchange of the N-terminal residues 1–10/1–9 in the B-chain/domain of insulin and IGF-1 results in exchange of their folding behavior (40). Therefore, one possible explanation is that the bifurcating evolution of folding behavior in the insulin superfamily was based on mutations in the N-terminal region of B-chain/domain.

Finally, it should be noted that [Gly]rAILP used here does not have a complete C-domain. We have to admit that shortening of the C-domain to 2 residues could perturb the details of their folding pathways, if we compare the putative folding pathways of PIP (20) and HPI (21). However, in PIP and HPI, it changed the stability of folding intermediates but not the basic ideas that A20-B19 is the most essential disulfide and
formed at the initial stage of refolding or that the formation of the second disulfide has no preference. Therefore, the two-residue C-peptide should not change the implication of our current finding. Furthermore, based on the facts from mini-IGF-1 and PIP, thermodynamically, C-domain has no significant influence on their folding behaviors. For example, although C-domain in IGF-1 is shortened to a dipeptide in mini-IGF-1, both of them fold into two thermodynamically stable disulfide isomers (23, 38). In addition, both HPI and PIP fold into one thermodynamically stable product regardless of the length of the C-domain (20, 21). Since rAILP to AILP is analogous to PIP to HPI, it is reasonable to deduce that rAILP with dipeptide linker might display folding behavior similar to AILP with whole C-domain thermodynamically. Anyhow, to investigate folding behavior of AILP with a complete C-domain would absolutely be a good follow-up study for better understanding of the role of the C-domain during refolding/unfolding.

REFERENCES

1. Anfinsen, C. B. (1973) Science 181, 223–230
2. Dill, K. A., and Chan, H. S. (1997) Nat. Struct. Biol. 4, 10–19
3. Kuhlman, B., and Baker, D. (2004) Curr. Opin. Struct. Biol. 14, 89–95
4. Dobson, C. M. (2003) Nature 426, 884–890
5. Creighton, T. E. (1986) Methods Enzymol. 131, 83–106
6. Wedemeyer, W. J., Welker, E., Narayan, M., and Scheraga, H. A. (2000) Biochemistry 39, 4207–4216
7. Chang, J.-Y., Li, L., and Bulychev, A. (2000) J. Biol. Chem. 275, 8287–8289
8. Creighton, T. E. (1992) Science 256, 111–114
9. Weissman, J. S., and Kim, P. S. (1991) Science 253, 1386–1393
10. Scheraga, H. A., Wedemeyer, W. J., and Welker, E. (2001) Methods Enzymol. 341, 189–221
11. Chatrenet, B., and Chang, J.-Y. (1993) J. Biol. Chem. 268, 20988–20996
12. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N., and Vijayan, N. M. (1988) Philos. Trans. R. Soc. Lond. B Biol. Sci. 319, 369–456
13. Vajdos, F. F., Ultsch, M., Schaffer, M. L., Deshayes, K. D., Liu, J., Skelton, N. J., and de Vos, A. M. (2001) Biochemistry 40, 11022–11029
14. Hua, Q.-X., Nakagawa, S. H., Jia, W., Hu, S.-Q., Chu, Y.-C., Katsoyannis, P. G., and Weiss, M. A. (2003) Biochemistry 42, 12290–12311
15. Hua, Q.-X., Jia, W., Frank, B. H., Phillips, N. F., and Weiss, M. A. (2002) Biochemistry 41, 14700–14715
16. Weiss, M. A., Hua, Q.-X., Jia, W., Chu, Y.-C., Wang, R.-Y., and Katsoyannis, P. G. (2000) Biochemistry 39, 15429–15440
17. Hua, Q.-X., Narhi, L., Jia, W., Arakawa, T., Rosenfeld, R., Hawkins, N., Miller, J. A., and Weiss, M. A. (1996) J. Mol. Biol. 259, 297–313
18. Hua, Q.-X., Hu, S.-Q., Frank, B. H., Jia, W., Chu, Y.-C., Wang, S.-H., Burke, G. T., Katsoyannis, P. G., and Weiss, M. A. (1996) J. Mol. Biol. 264, 390–403
19. Guo, Z.-Y., and Feng, Y.-M. (2001) Biochem. 382, 443–448
20. Qiao, Z.-S., Guo, Z.-Y., and Feng, Y.-M. (2001) Biochemistry 40, 2662–2668
21. Qiao, Z.-S., Min, C.-Y., Hua, Q.-X., Weiss, M. A., and Feng, Y.-M. (2003) J. Biol. Chem. 278, 17800–17809
22. Min, C.-Y., Qiao, Z.-S., and Feng, Y.-M. (2004) Eur. J. Biochem. 271, 1737–1747
23. Hofer, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B. (1992) Biochemistry 31, 1749–1756
24. Miller, J. A., Narhi, L. O., Hua, Q.-X., Rosenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., Stoney, K. S., Tsai, L., and Weiss, M. A. (1993) Biochemistry 32, 5203–5213
25. Rosenfeld, R. D., Miller, J. A., Narhi, L. O., Hawkins, N., Katta, V., Lauren, S., Weiss, M. A., and Arakawa, T. (1997) Arch. Biochem. Biophys. 342, 298–305
26. Milner, S. J., Carver, J. A., Ballard, F. J., and Francis, G. L. (1999) Biochim. Biophys. Acta 55233
27. Yang, Y., Wu, J., and Watson, J. T. (1999) J. Biol. Chem. 274, 37598–37604
28. Chan, S.-J., Cao, Q.-P., and Steiner, F. D. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 9319–9323
29. Shen, L., Guo, Z.-Y., Chen, Y., Liu, L.-Y., and Feng, Y.-M. (2001) Sheng Wu Xue Yu Sheng Wu Li Xue Bao (Shanghai) 33, 629–635
30. Guo, Z.-Y., Shen, L., Gu, W., Wu, A.-Z., Ma, J.-G., and Feng, Y.-M. (2002) Biochemistry 41, 10603–10607
31. Wang, S., Guo, Z.-Y., Shen, L., Zhang, Y.-J., and Feng, Y.-M. (2003) Biochemistry 42, 9687–9690
32. Kramer, W., Drusta, V., Jansen, H. W., Kramer, B., Pfugfelder, M., and Frits, H. J. (1984) Nucleic Acids Res. 12, 9441–9454
33. Hua, Q.-X., Chu, Y.-C., Jia, W., Phillips, N. F., Wang, R.-Y., Katsoyannis, P. G., and Weiss, M. A. (2002) J. Biol. Chem. 277, 43445–43453
34. Yan, H., Guo, Z.-Y., Gong, X.-W., Xi, D., and Feng, Y.-M. (2003) Protein Sci. 12, 768–777
35. Narhi, L. O., Hua, Q.-X., Arakawa, T., Fox, G. M., Tsai, L., Rosenfeld, R., Holst, P., Miller, J. A., and Weiss, M. A. (1993) Biochemistry 32, 5214–5221
36. Jia, X.-Y., Guo, Z.-Y., Wang, Y., Xu, Y., Duan, S.-S., and Feng, Y.-M. (2003) Protein Sci. 12, 2412–2419
37. Hua, Q.-X., Gezani, S. N., Chance, H. E., Hoffmann, J. A., Frank, B. H., and Weiss, M. A. (1995) Nat. Struct. Biol. 2, 129–138
38. Guo, Z.-Y., Shen, L., and Feng, Y.-M. (2002) Biochemistry 41, 1556–1567
39. Guo, Z.-Y., Shen, L., and Feng, Y.-M. (2002) Biochemistry 41, 10585–10592
40. Chen, Y., You, Y., Jin, R., Guo, Z.-Y., and Feng, Y.-M. (2004) Biochemistry 43, 9225–9233