Indication of possible post-translational formation of disulphide bonds in the β -sheet domain of human lysozyme

Eiko KANAYA,* Kouji ISHIHARA,† Susumu TSUNASAWA,† Kiyoshi NOKIHARA‡ and Masakazu KIKUCHI*§

*Protein Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka, 565, †Institute for Protein Research, Osaka University, 3-2, Yamadaoka, Suita, Osaka, 565, and ‡Biotechnology Instruments Departments, Shimadzu Corporation, Nishino-kyo, Kuwabaracho 1, Nakagyo-ku, Kyoto, 604, Japan

Lysozyme has two distinct folding domains, and in most molecules the α -helical domain folds more quickly than the β -sheet domain *in vitro* [Radford, Dobson and Evans (1992) *Nature* (London) 358, 302–307]. In order to investigate the relationship between the formation of disulphide bonds and protein folding *in vivo*, we carried out cysteine scanning mutagenesis to shift positions of the disulphide bonds in both the α -helical and β -sheet domains of human lysozyme. Of the constructed mutants (nine in the β -sheet domain and 13 in the α -helical domain), the mutant L79CC81A, in which Leu-79 and Cys-81 in the β -sheet domain were replaced by Cys and Ala respectively, was secreted by yeast. The rest of the mutants were retained in the insoluble fraction of the cell, probably because of a failure of folding. The distance between the two α -carbons at positions 79 and 95 in the wild-type

protein is too far to form a disulphide bond, but analysis of the primary structure revealed that the major part of L79CC81A was secreted with a non-native disulphide bond Cys⁷⁹–Cys⁹⁵ and two free cysteine residues at positions 65 and 77 in the β -sheet domain. These results suggest that the β -sheet domain of human lysozyme can tolerate the shift of locations of disulphide bonds, and the non-native folding of mutated polypeptide chains in *in vivo* folding. The free residues Cys-65 and Cys-77 formed a disulphide bond *in vitro* by air oxidation, yielding two isomers. On the basis of our previous results and present study it is suggested that the formation of Cys⁶–Cys¹²⁸ is the first step of the *in vivo* correct folding of human lysozyme, and disulphide bonds in the β -sheet domain are post-translationally formed *in vivo*.

INTRODUCTION

The formation of disulphide bonds in a secretory protein is thought to play an important role in the process of protein folding *in vivo*. However, it remains unclear whether the protein folding *in vivo* is initiated sequentially from the N-terminus or after the complete synthesis of the polypeptide chain, and whether disulphide bonds are formed during or after translation *in vivo* [1–4]. Previous reports suggest that disulphide bonds are formed *in vivo* during co-translational folding [5,6]. This situation would decrease the probability of formation of non-native disulphide bonds during folding of molecules such as immunoglobulin light chains [5], in which the cysteine partners are located either sequentially or close to one another in the amino acid sequence. However, the probability of formation of non-native bonds may increase when the proper cysteine partners are separated in the polypeptide chain by many amino acids, such as in human lysozyme (h-lysozyme).

Hinge-bending analysis [7] and surface-area calculations [8] of lysozyme suggest that it consists of two structural domains. Miranker et al. [9] suggested that the two structural domains of hen's-egg lysozyme are folding domains that differ significantly in the extent to which protected structures accumulate early in the folding process (Figure 1). One folding domain (the α -helical domain) contains the N- and C-termini and is more protected in hydrogen exchange, while the other domain (the β -sheet domain) contains the central portion of the polypeptide and is less protected. Recently, Radford et al. [10] showed, by refolding experiments *in vitro* using NMR and CD, that the folding of hen lysozyme involves partially structured intermediates and multiple pathways, and in most molecules the α -helical domain becomes

substantially folded before formation of a stable β -sheet domain; it is therefore suggested that the α -helical domain folds more quickly than the β -sheet domain.

Human lysozyme is an appropriate model protein to study the mechanism of protein folding *in vivo* because its expression, purification [11] and structural analysis [12] are easily carried out. Further, it is structurally similar to hen's egg lysozyme, and has four disulphide bonds, Cys⁶–Cys¹²⁸, Cys³⁰–Cys¹¹⁶, Cys⁶⁵–Cys⁸¹, and Cys⁷⁷–Cys⁹⁵. A mutant C30/116A, in which both Cys-30 and Cys-116 are replaced by Ala, was secreted inefficiently [13], but any mutants deficient in a terminal cysteine, such as the C6A, C128A, C6/128A, and C-terminal truncated mutants, were not secreted by yeast [13], probably because of a failure of folding *in vivo*. The Cys⁶–Cys¹²⁸ bond is obligatory for the formation of the SDS-resistant compact conformation of mutant h-lysozymes in *in vitro* translation [14]. A mutant C128A, in which Cys-128 was replaced by Ala, cannot form the Cys⁶–Cys¹²⁸ bond and was easily degraded in the endoplasmic reticulum [15], whereas the wild-type h-lysozyme was quantitatively secreted into the media without degradation [15]. These observations suggest that the formation of Cys⁶–Cys¹²⁸ in the α -helical domain decides success or failure of folding of h-lysozyme *in vivo*. We have also obtained several cysteine-substituted mutants in the β -sheet domain. The introduction of Cys-83 and Cys-91 into the wild type h-lysozyme by amino acid replacement affected the *in vivo* process of disulphide-bond formation [16]. Replacement of Cys-77, Cys-81 or Cys-95 by Ala did not influence mutant folding, but replacement of Cys-65 with Ala made the protein misfold [14], indicating that Cys-65 is indispensable [17]. The formation of a non-native disulphide bond, Cys⁶⁵–Cys⁷⁷, probably occurs in the mutant C81A in which Cys-81 was

Figure 1 β -Sheet domain and the α -helical domain in the refined structure of wild-type h-lysozyme

replaced by Ala [17]. Furthermore, Cys-77 and Cys-95, which link the α - and β -domains as do Cys-76 and Cys-94 in hen's-egg lysozyme [10], could be the last residues to cross-link in h-lysozyme [18], and the slow refolding phase of the wild-type protein disappears in a mutant that lacks Cys-77 and Cys-95 [19]. The region in which these results were obtained is in the β -sheet domain. These facts suggest the possibility that either or both disulphide bonds in the β -sheet domain are formed *in vivo* after the complete synthesis of the polypeptide chain. Further, Cys-6 and Cys-128 in the α -domain and Cys-65 in the β -domain are indispensable, and the Cys⁶–Cys¹²⁸ bond may be the first native disulphide bond for the correct folding of h-lysozyme *in vivo*. To examine this possibility further, we have performed a cysteine scanning mutagenesis in both the α - and β -domains, to search for residues that could form non-native disulphide bonds with Cys-65, Cys-6, or Cys-128. If the mutant lysozymes with a non-native disulphide bond in the specified region are obtainable, which are correctly folded and secreted by yeast, it would suggest that the region carries the structural tolerance. Analysis of the structure around the non-native disulphide bond would offer information about the structural rigidity, flexibility, change and its importance around the region, and further, the relation between the formation of disulphide bonds and protein folding *in vivo*.

Here we report that h-lysozyme is able to tolerate the change in positions of disulphide bonds and the packing of mutated polypeptides in the β -sheet domain in *in vivo* folding. The Cys⁶–Cys¹²⁸ bond could not be replaced by other non-native disulphide bonds in the α -helical domain by cysteine scanning mutagenesis, probably because of the necessity of conformational rigidity of the α -helical domain. On the basis of our previous results and the present study we discuss the possibility of post-translational formation of disulphide bonds in the folding of h-lysozyme *in vivo*.

EXPERIMENTAL

Materials

Enzymes for recombinant DNA technology were purchased from Takara Shuzo (Kyoto, Japan) and Bethesda Research Laboratories (Gaithersburg, MD, U.S.A.). Enzyme reactions were carried out under the conditions recommended by the suppliers. *Micrococcus lysodeikticus* cells and trypsin were purchased from Sigma (St. Louis, MO, U.S.A.). Lysyl endopeptidase was from Wako Pure Chemicals (Osaka, Japan) and elastase was from Elastin Products (Owensville, MO, U.S.A.). Chemicals for DNA synthesis and gas-phase peptide sequencing were obtained from Applied Biosystems (Foster City, CA, U.S.A.). All other chemicals were of reagent-grade quality.

Strains and media

Saccharomyces cerevisiae AH22 (*a leu2 his4 can1 pho80*) [20] was cultured in modified Burkholder minimal medium supplemented with 8% sucrose [21].

DNA species and mutations

Mutations were produced in the h-lysozyme gene by the PCR with the Gene Amp PCR Reagent Kit purchased from Takara Shuzo, according to the procedure of Landt et al. [22]. A representative primer for introducing a mutation is as follows: Leu⁷⁹–Cys⁸¹ to Cys⁷⁹–Ala⁸¹, 5-TGAGGCAGAGCAGTGACAGGC-3. Underlining indicates mismatches. The sequence of the mutated gene was confirmed with a ³²P-CTP sequencing kit (Pharmacia, Uppsala, Sweden). To construct the expression plasmids, the genes encoding the signal sequence and the mutated h-lysozymes were combined with the large *Xho*I–*Sma*I fragment from pERI8602 [17]. The DNA oligomer was chemically synthesized using an automated DNA synthesizer (model 380B; Applied Biosystems).

Purification and characterization of a mutant h-lysozyme

The mutant L79CC81A was purified as described [21], using Indion Ion Exchange Cellulose (Phoenix Chemicals, Christchurch, New Zealand) and an Asahipak ES502C cation-exchange column (Asahikasei Co., Kanagawa, Japan). Lytic activity was measured as described [21].

Peptide sequence analysis

Peptide sequence analysis was performed by stepwise Edman degradation, using a gas-phase automated sequencer (Applied Biosystems, model 477A). The positions of disulphide bonds were determined by the method previously described [23], using a Shimadzu (Kyoto, Japan) PSQ 1 protein sequencer.

Amino-acid-composition analysis

Amino acid compositions were determined with a Hitachi model 835 amino acid analyser. Analyses were performed on protein samples hydrolysed for 15 h, *in vacuo*, at 105 °C in 6 M HCl, using the Picotag work station.

S-carboxymethylation of L79CC81A

S-carboxymethylation of L79CC81A was started by adding 0.6 mg of iodoacetic acid, dissolved in 2 μ l of 1 M NaOH, to 100 μ l of 0.4 M Tris/HCl, pH 8.2, containing 2.5 mM EDTA.

The reaction mixture, which lacked denaturants and reducing agents, was vigorously shaken at room temperature for 30 min. The protein was desalted on a PD-10 column. The excluded fraction was collected.

Proteinase digestion

The carboxymethylated L79CC81A (CM-L79CC81A) and the oxidized proteins L79CC81A-a and -b were digested with trypsin in the absence of denaturants and reducing agents. The resultant peptides were separated by reverse-phase h.p.l.c. using an Aquapore RP-300 column (4.6 × 250 mm; Brownlee Laboratories, Santa Clara, CA, U.S.A.), as described previously [24]. Peptides linked by disulphide bonds were elastase-digested by dissolving 10 µg of elastase in 100 µl of protein solution, and incubating the mixture at 37 °C for 30 min, as described previously [25]. The resultant peptides were separated by reverse-phase h.p.l.c. Reduction and S-carboxymethylation of L79CC81A, digestion with lysyl endopeptidase and separation of the resultant peptides by reverse-phase h.p.l.c. were performed as described previously [25].

Figure 2 Separation of tryptic digests of CM-L79CC81A using reverse-phase h.p.l.c.

Trypsin digests of CM-L79CC81A were directly loaded on to an Aquapore RP-300 reverse-phase column (4.6 mm × 250 mm), equilibrated with 5% (v/v) solvent B in solvent A. Elution was performed by a linear increase in the concentration of solvent B in solvent A from 5 to 45% (v/v) for 60 min. Solvent A was aq. 0.1% (v/v) trifluoroacetic acid and solvent B was acetonitrile containing 0.05% (v/v) trifluoroacetic acid. The flow rate was 1.0 ml/min.

Figure 3 Alignment of tryptic peptides of L79CC81A

The peptide bond between T9 and T10 was only partially digested (T9/T10).

RESULTS

Cysteine scanning mutations in the β -sheet domain and expression of the mutated h-lysozyme L79CC81A in yeast

To search for positions that are able to form a non-native disulphide bond with Cys-65, the cysteine residue at 81 was replaced by alanine and then a cysteine residue was placed in a range of positions from 74 to 84. These replacements were confirmed by DNA sequencing. The protoplasts prepared from washed yeast cells expressing the mutated lysozymes were lysed, then fractionated into insoluble and soluble fractions by centrifugation according to the method by Yoshimura et al. [11]. These two fractions and the culture supernatants were boiled in Laemmli's buffer without 2-mercaptoethanol and were subjected to SDS/15% PAGE, and the gels were analysed by western blotting. Among the nine mutants, V74CC81A, N75CC81A, A76CC81A, H78CC81A, L79CC81A, S80CC81A, C81AS82C, C81AS83C, and C81AL84C, the mutant L79CC81A was secreted by yeast, with a yield of 0.2 mg/l. The eight other lysozyme mutants were retained in the insoluble fraction of the cell (results not shown). Purification of L79CC81A was performed as described previously [21]. The last step of purification was carried out by using reverse-phase h.p.l.c. In the same manner, the position of Cys-6 or Cys-128 was also shifted continuously from positions 3 to 10 or 124 to 130 respectively, in order to create a non-native disulphide bond with Cys-128 or Cys-6. However, no secretable mutant lysozymes were obtained among 13 mutated proteins. All mutants were retained in the insoluble fraction of the cell (results not shown).

Determination of the primary structure of L79CC81A

After denaturation and reduction, S-carboxymethylation of L79CC81A was performed as described previously [18]. Reduced and S-carboxymethylated L79CC81A (RCM-L79CC81A) was digested by lysyl endopeptidase and the digest was analysed by reverse-phase h.p.l.c. The peptides from five peaks were collected and their amino acid compositions were analysed (results not shown). Two substitutions, namely Cys replacing Leu-79 and Ala replacing Cys-81, were confirmed by protein sequence.

L79CC81A contains eight cysteine residues, and analysis with Ellman's reagent [26] indicated that L79CC81A had two free cysteine residues, which suggests that L79CC81A has three disulphide bonds. To determine the locations of the two free cysteine residues and the three disulphide bonds, L79CC81A was carboxymethylated by iodoacetic acid treatment, without denaturants or reducing agents (CM-L79CC81A). CM-L79CC81A was digested with trypsin in the absence of denaturants and reducing agents. The tryptic digest of CM-L79CC81A was analysed by reverse-phase h.p.l.c. (Figure 2). Seven peaks in Figure 2 were observed in the reverse-phase h.p.l.c. elution profile of the tryptic digest (Figure 2). Each peptide from the seven peaks was identified, as shown in Figure 3. The mutant was expected to be digested into 17 peptides by the trypsin treatment (Figure 3). The peptides T1, T3, T7, T11, T12, T14 and T16, which do not contain cysteine residues, were probably not retained in the column under the conditions used for elution of the tryptic peptides, that is, a linear increase in the concentration of solvent B in solvent A from 5% to 45% (v/v) for 60 min. The whole structure of L79CC81A has been confirmed by amino-acid-composition analysis and DNA sequence determination of the corresponding gene. Thus the sequences of these peptides were not assigned. The peptide bond between positions 69 and 70 was not completely digested in L79CC81A, so an undigested peptide (T9/T10) was also

Figure 4 Schematic structure of L79CC81A

The positions of the cysteine residues in the molecule are circled.

Figure 5 Elution profile of purified L79CC81A (a) and oxidized products (b) on reverse-phase h.p.l.c.

(a) The elution was performed with a 75 min linear gradient increasing from 20 to 50% acetonitrile containing 0.05% trifluoroacetic acid. L79CC81A was eluted at 40 min. (b) After storage of L79CC81A at 4 °C and pH 6.5 for several days, two new peaks were observed at 38 min (designated as peak a) and 39.5 min (designated as peak b) in addition to the peak of L79CC81A at 40 min.

obtained. The disulphide bond Cys³⁰–Cys¹¹⁶ was in peptide T5 + T15, Cys⁶–Cys¹²⁸ was in peptide T2 + T17, and the residues Cys-65, Cys-77, Cys-79, and Cys-95 were in peptides T9 + T10

and T9/T10. The following sequence was found in peptide T9/T10 (Figure 2):

Y-(W)-(CM)C⁶⁵-N-D-G-K-T-P-G-A-V-N-A-
(CM)C⁷⁷-H-(C)⁷⁹-S-A-S-A-L-L-Q-D-N-I-A-D-A-V-A-
(C)⁹⁵-A-K (residues 63–97)

Carboxymethylated cysteine residues were detected at the steps 3 and 15 of Edman degradation, but no residues were observed at the steps 17 and 33. This result indicates that L79CC81A has two free cysteine residues at positions 65 and 77 and a non-native disulphide bond, Cys⁷⁹–Cys⁹⁵ (Figure 4).

Formation of disulphide bonds *in vitro* in the β -sheet domain of L79CC81A

The purity of L79CC81A was analysed by reverse-phase h.p.l.c. (Figure 5a). A main peak from L79CC81A was observed at 40 min and two minor peaks were also observed at 38 min and 39.5 min after the start of a 75 min linear gradient of 20–50% acetonitrile/0.1–0.05% trifluoroacetic acid (Figure 5a). After storage of L79CC81A at 4 °C and pH 6.5 for several days, two minor peaks at 38 min (designated as peak a) and 39.5 min (designated as peak b), which were observed previously, increased (Figure 5b). The height of the two peaks at 38 min and 39.5 min increased, and that of the peak at 40 min decreased, with time. The proteins from peaks a and b (designated L79CC81A-a and L79CC81A-b) were collected separately and analysed for their primary structures. It was confirmed by the use of Ellman's reagent [26] that L79CC81A-a and -b have no free cysteine residues. In addition, there were no differences in amino acid composition among L79CC81A-a, L79CC81A-b and L79CC81A. These results suggest that L79CC81A changed into two new proteins by air oxidation and formed two alternative combinations of disulphide bonds.

Intramolecular exchange of disulphide bonds in the β -sheet domain

In order to determine the primary structures, the two mutants L79CC81A-a and -b were digested with trypsin in the absence of denaturants and reducing agents. The resultant peptides were separated by reverse-phase h.p.l.c. with a 60 min linear gradient of 5–45% acetonitrile (Figure 6). The elution profile of the digests from L79CC81A-a is shown in Figure 6(a), and that from L79CC81A-b is shown in Figure 6(b). The only difference between Figures 6(a) and 6(b) is the retention time of the peptide T9 + T10. The peak containing T9 + T10 from L79CC81A-a was eluted at 37.6 min, while that from L79CC81A-b was eluted at 37.0 min was co-eluted out with T5 + T15. The two co-eluted peptides, T9 + T10 and T5 + T15, from L79CC81A-b in Figure 6(b) were separable using a 90 min linear gradient of 5–45% acetonitrile. It was confirmed that both T9 + T10 peptides from L79CC81A-a and -b had the same amino acid compositions and the same peptide sequences, that is:

Y⁶³-W-(C)-N-D-G-K⁶⁹

and

T⁷⁰-P-G-A-V-N-A-(C)-H-(C)-S-A-S-A-L-L-Q-D-N-I-
A-D-A-V-A-(C)-A-K⁹⁷

It was presumed that the difference in the retention time of T9 + T10 between L79CC81A-a and L79CC81A-b (Figures 6a and 6b) was due to the difference in the positions of disulphide bonds among Cys-65, Cys-77, Cys-79, and Cys-95.

Figure 6 Elution profiles of tryptic digests of L79CC81A-a (a) and L79CC81A-b (b)

The tryptic digests of L79CC81A-a and -b were separated by reverse-phase h.p.l.c. with a 60 min linear gradient from 5 to 45% (v/v) acetonitrile, containing 0.05% trifluoroacetic acid. The peak T9 + T10 from L79CC81A-a was eluted at 37.6 min and that from L79CC81A-b was co-eluted with T5 + T15 at 37.0 min. The latter T9 + T10 peak was separated from T5 + T15 by a 90 min linear gradient from 5 to 45% (v/v) acetonitrile, containing 0.05% trifluoroacetic acid.

Figure 7 Elution profile of elastase digests of peaks T9 + T10 on reverse-phase h.p.l.c.

The two T9 + T10 peptides were further digested with elastase. The digests were analysed using a 20 min linear gradient increasing from 0 to 30% acetonitrile. The elastase digest of T9 + T10 from L79CC81A-a produced three main peaks (a), and that from L79CC81A-b produced two main peaks (b). The peptide sequence of each peak was confirmed as described in the Figure.

Scheme 1 Two possible locations of disulphide bonds in the elastase digest from L79CC81A-a and L79CC81A-b

The two T9 + T10 peptides were further digested with elastase. The digests were analysed by reverse-phase h.p.l.c. with a 20 min linear gradient of 0–30% acetonitrile (Figure 7). The elastase

digest of T9 + T10 from L79CC81A-a gave three peaks, at 12.5, 14.8 and 15.2 min (Figure 7a). The peptide sequences were confirmed to be: T-P-G-A-V at 12.5 min; Y-W-(C)⁶⁵-N-D-G-K, (C)⁷⁷-H-(C)⁷⁹-S-A and A-(C)⁹⁵-A-K at 14.8 min; Y-W-(C)⁶⁵-N-D-G-K, N-A-(C)⁷⁷-H-(C)⁷⁹-S-A and A-(C)⁹⁵-A-K at 15.2 min. The elastase digest of T9 + T10 from L79CC81A-b gave two peaks on reverse-phase h.p.l.c. at 12.5 and 15.6 min. The peptide sequences were as follows: T-P-G-A-V at 12.5 min; Y-W-(C)⁶⁵-N-D-G-K, N-A-(C)⁷⁷-H-(C)⁷⁹-S-A and A-(C)⁹⁵-A-K at 15.6 min. On the basis of the finding that the four cysteine residues at positions 65, 77, 79, and 95 were contained in peptides from one peak, the bonds Cys⁶⁵-Cys⁷⁷, Cys⁶⁵-Cys⁷⁹, Cys⁷⁹-Cys⁹⁵, and Cys⁷⁷-Cys⁹⁵ were highly probable. Neither the Cys⁶⁵-Cys⁹⁵ bond, nor the Cys⁷⁷-Cys⁷⁹ bond, were formed, because the two peptides, (Y-W-C⁶⁵-N-D-G-K)-(A-C⁹⁵-A-K) and C⁷⁷-H-C⁷⁹-S-A, would be separately eluted if Cys⁶⁵-Cys⁹⁵ and Cys⁷⁷-Cys⁷⁹ were formed. Edman degradation of the peptide, absorbed on a poly(vinylidene difluoride) (PVDF) membrane, was performed. If the peptide from the peak at 14.8 min (Figure 7a) has Cys⁶⁵-Cys⁷⁷ and Cys⁷⁹-Cys⁹⁵, then the two peptides shown in Scheme 1(a) would be recovered from the PVDF membrane after one step of Edman degradation in the peptide sequencer. If the peptide from the peak at 14.8 min (Figure 7a) has Cys⁶⁵-Cys⁷⁹ and Cys⁷⁷-Cys⁹⁵, then the two peptides shown in Scheme 1(b) would be recovered. Actually, by soaking the PVDF membrane in 600 μ l of the mixed solvent (formic acid/propan-2-ol/acetonitrile/water, 6:2:1:1, by vol.) at room temperature for 10 min by shaking, the peptides H-(C)⁷⁹-S-A and (C)⁹⁵-A-K were recovered, and were connected

Figure 8 Location of disulphide bonds in peaks T9 + T10 from L79CC81A-a and -b

by Cys⁷⁹–Cys⁹⁵. This indicates that T9 + T10 from L79CC81A-a has Cys⁷⁹–Cys⁹⁵. The peptides eluted at 15.2 min in Figure 7(a) or 15.6 min in Figure 7(b), which contained two sets of disulphide bonds, Cys⁶⁵–Cys⁷⁷ and Cys⁷⁹–Cys⁹⁵, or Cys⁶⁵–Cys⁷⁹ and Cys⁷⁷–Cys⁹⁵, were not adequate for analysis by Edman degradation on PVDF membrane, because three steps of Edman degradation of both peptides at 15.2 min and 15.6 min in the peptide sequencer should give same degraded peptides [N-D-G-K and (H-C⁷⁹-S-A)-(C⁶⁵) or (H-C⁷⁹-S-A)-(C⁶⁵)] shown in Scheme 1(c) and 1(d). The confirmation of the disulphide-bonded structure of L79CC81A-a and the identification of that of L79CC81A-b were carried out by using a protein sequencer (Shimadzu model PSQ-1). Routine detection of intact cystinyl residues with disulphide linkage was established using this instrument [23]. That is, no amino acid phenylthiohydantoin derivatives can be seen in the first half-cystine-residue cycle; on the other hand, of the second half-cystine the phenylthiohydantoin derivatives-residue cycle appear to be those of a cysteine and a dithiothreitol adduct of dehydroalanine (ΔSer). These are eluted at specific positions and are not superimposed with other phenylthiohydantoin derivatives or by-products under the isocratic separation mode of PSQ-1. With the use of a gas-phase sequencer, cystinyl residues as well as serine residues are relatively stable during Edman degradation. Sequence analyses of both intact and S-pyridylethylated cysteine can easily identify a disulphide form if one disulphide linkage exists in the sample. These were observed at steps 8 and 26 of the peptide sequence analysis of T9 + T10 from L79CC81A-a, and at steps 10 and 26 in that from L79CC81A-b. The locations of the disulphide bonds in the peptides T9 + T10 from L79CC81A-a and L79CC81A-b were determined as shown in Figure 8. These results indicate that the air oxidation of L79CC81A *in vitro* gave two products with different sets of disulphide bonds, that is, tandem-linked Cys⁶⁵–Cys⁷⁷ and Cys⁷⁹–Cys⁹⁵ bonds, or cross-linked Cys⁶⁵–Cys⁷⁹ and Cys⁷⁷–Cys⁹⁵ bonds, and the non-native Cys⁷⁹–Cys⁹⁵ bond in L79CC81A was partly interchanged with Cys⁷⁷–Cys⁹⁵ in the slow-folding domain to form L79CC81A-b.

The pH-dependence of the oxidation of L79CC81A was examined. The ratio of L79CC81A-a, L79CC81A-b and the starting material, L79CC81A, was 39:11:49 when L79CC81A was kept in the pH 6.0 solution at 30 °C for 2 days. This ratio changed to 22:71:7 when L79CC81A was kept in a solution at pH 7.5. This result suggests that L79CC81A-b was formed by a thiol–disulphide exchange reaction at the higher pH between the free thiols at Cys-65 and Cys-77 and the disulphide bond Cys⁷⁹–Cys⁹⁵.

Proteins with the same primary structure and different enzyme activities

The lytic activities of L79CC81A, L79CC81A-a and L79CC81A-b were compared with that of the wild-type enzyme (taken as

100%). The relative activity of L79CC81A was 86%, that of L79CC81A-a was 5% and that of L79CC81A-b was 84%. The results indicate that the formation of the Cys⁶⁵–Cys⁷⁷ bond markedly reduced the activity of the mutant enzyme, and that one gene encoding a structural protein produced three structures with different disulphide bonds and different activities.

DISCUSSION

It has not been concluded whether the folding of proteins *in vivo* is initiated sequentially from the N-terminus or after the completion of the polypeptide chain synthesis [1]. Bergman and Kuel [5] suggested that the folding of immunoglobulin *in vivo* is initiated sequentially from its N-terminus, whereas Taniuchi and co-workers [27,28] have reported that the C-terminal portion is important for the correct folding and activity of ribonuclease A. There would also be difficulties in evaluating the role of disulphide-bond formation in the folding of a protein *in vivo*.

Based on n.m.r. analyses of the refolding of hen's-egg lysozyme, Miranker et al. [9] concluded that its two structural domains are distinct folding domains. Hydrogen exchange of the amide protons revealed that highly protected regions lie in both the N- and C-termini, and the middle region of the polypeptide chain is quite exchangeable. Recently, Radford et al. [10] showed that the α-helical domain folds more quickly than the β-sheet domain. The situation of lysozyme in folding may not be unique, and similar marginal protection has been observed in the folding of apomyoglobin [29] and cytochrome *c* [30].

Our experimental results suggest flexibility and tolerance in the *in vivo* folding of the β-sheet domain of h-lysozyme. There are two important observations in this regard: (a) that the Cys⁶⁵–Cys⁸¹ bond could be shifted to Cys⁶⁴–Cys⁸¹ in the mutant W64CC65A, in which Trp-64 and Cys-65 were replaced by Cys and Ala respectively [25], and (b) that, in the mutant W64CC65/77/95A, Trp-64, Cys-65, Cys-77 and Cys-95 could be replaced by Cys, Ala, Ala and Ala respectively [24]. The formation of a disulphide bond between positions 64 and 81 appeared to be rather difficult in the wild-type h-lysozyme, since, in the wild-type protein, the side chains of residues 64 and 81 are in opposite directions from each other, and their α-carbons are too far apart to form a disulphide bond [0.836 nm (8.36 Å)] [31]. The mutant W64CC65A was crystallized under the same conditions as that of the wild-type protein, and it was shown that the crystals were not isomorphous [25]. These facts suggest that the local conformational changes have occurred in the β-sheet domains of W64CC65A and W64CC65/77/95A. Other results obtained with the cysteine-substituted mutant h-lysozymes AC83/91 [16], C81A [17], C77A and C77/95A [14] have also suggested that some of the disulphide bonds formed in the β-sheet domain are less important in the protein folding *in vivo*: (i) an artificial disulphide bond, Cys⁸³–Cys⁹¹, could be introduced in the mutant AC83/91 [16]; (ii) the non-native disulphide bond Cys⁶⁵–Cys⁷⁷ was observed in C81A [17]; and (iii) the native disulphide bond Cys⁷⁷–Cys⁹⁵ could be removed without much influence on the protein folding *in vivo* and secretion in yeast [14] and animal [32] cells. In the present study the conserved cysteine residue at position 81 in the β-sheet domain of h-lysozyme was continuously shifted from positions 74 to 84 by cysteine scanning mutagenesis. Of the nine mutants constructed, the mutant L79CC81A was secreted by yeast. The other eight mutant proteins (V74CC81A, N75CC81A, A76CC81A, H78CC81A, S80CC81A, C81AS82C, C81AA83C and C81AL84C) were retained in the insoluble fraction of the yeast (results not shown). There are several reports which suggest that non-secretion of secretory proteins is attributed to incorrect folding [33–36]. All bands of the eight

non-secreted mutants were clearly observed at the position of a monomer of h-lysozyme in Western blotting after non-reducing SDS/PAGE. Thus there would be no aggregation of molecules through intermolecular disulphides, suggesting that the proteins failed to fold *in vivo*. The reason for the misfolding is unclear at present. Several causes for non-secretion of mutant h-lysozymes can be considered, such as degradation of polypeptides, aggregation and decrease of solubility by mutation. The results of Western blotting after non-reducing SDS/PAGE (results not shown) suggest that aggregation through intermolecular disulphides could be excluded from the causes of non-secretion of the mutants. In SDS/PAGE under non-reducing conditions, the hydrodynamic volume of non-secreted mutant proteins which were retained in the insoluble fraction of yeast cells was slightly greater than that of the wild-type lysozyme which was secreted by yeast [25]. The increase of the hydrodynamic volume and decrease of solubility would suggest that the non-secreted mutants in the insoluble fraction did not fold in yeast.

The major part of the mutant L79CC81A has two free cysteine residues at positions 65 and 77, and three disulphide bonds, namely Cys⁶–Cys¹²⁸, Cys³⁰–Cys¹¹⁶, and Cys⁷⁹–Cys⁹⁵. The native disulphide bonds Cys⁶⁵–Cys⁸¹ and Cys⁷⁷–Cys⁹⁵ in the β -sheet domain were completely replaced by the two free Cys-65 and Cys-77 residues and the non-native disulphide bond Cys⁷⁹–Cys⁹⁵. The two α -carbons at positions 79 and 95 exist at a distance of 0.81 nm (8.10 Å) in the three-dimensional structure of wild-type h-lysozyme [31], and the formation of the disulphide bond Cys⁷⁹–Cys⁹⁵ was thought to be difficult. In spite of this distance, the Cys⁷⁹–Cys⁹⁵ bond was formed, suggesting that local shifts in the positions of the amino acid residues occurred within this region during folding. These results indicate the possibility that local non-native folding of the mutated polypeptide chain occurs *in vivo* in the β -sheet domain of L79CC81A. The conformational change of T4 lysozyme protein by the introduction of disulphide bonds could not be expected by theoretical calculation and computer modelling [37]. But some local conformational shifts may actually be expected in L79CC81A, W64CC65A [25], and W64CC65/77/95A [24]. Introducing disulphide bonds by cysteine scanning mutagenesis may make it possible to trap some locally modified folding structure, and that the introduction of disulphide bonds to the unexpected position would reflect the local structural flexibility and tolerance in the protein structure. Creighton [38] has stated that disulphide bonds are not determinants of the folded conformation of proteins but are formed spontaneously when the protein conformation favours them. The formation of Cys⁷⁷–Cys⁹⁵ may support this idea, because the Cys⁷⁷–Cys⁹⁵ bond is not necessary for *in vivo* folding and secretion; linking of the α - and β -domains is thought to occur in the last stage of the folding process in h-lysozyme [18].

By cysteine scanning mutagenesis, Cys-6 was shifted from positions 3 to 10 to form a non-native disulphide bond with Cys-128, and Cys-128 was also shifted from position 124 to 130 to form a non-native disulphide bond with Cys-6 in the α -helical domain. However, no secreted mutant has been obtained, probably because of a failure to fold *in vivo* (results not shown). The mutant h-lysozymes with C-terminally truncated regions and terminal-cysteine-deficient mutants (C6A, C128A and C6/128A) were not secreted by yeast [14], due to lack of folding. The Cys⁶–Cys¹²⁸ bond is obligatory for the formation of the SDS-resistant compact conformation of mutant h-lysozymes after *in vitro* translation [14]. That the mutant h-lysozyme C128A was easily degraded in the endoplasmic reticulum may be due to failure of the Cys⁶–Cys¹²⁸-bond formation [13]. These observations suggest that the close packing of amino acid residues in the α -helical domain is required for the folding of h-lysozyme,

and shifting the position of the Cys⁶–Cys¹²⁸ bond is critical. Judging from these results, the formation of the Cys⁶–Cys¹²⁸ bond in the α -helical domain may be important in the *in vivo* folding of h-lysozyme. Considering these results we suggest that the Cys⁶–Cys¹²⁸ bond is the first native disulphide bond formed during the *in vivo* folding of h-lysozyme and that disulphide bonds in the β -sheet domain are post-translationally formed *in vivo*, depending on the stabilizing core of the α -helical domain. Our observations in the α -helical domain suggest that the formation of Cys⁶–Cys¹²⁸ may occur at the adhesion step which determines structure in the *in vivo* folding of h-lysozyme, and may support the ideas put forward by Kim and Baldwin [39]. They proposed that disulphide-bond formation is an adhesion step that locks a region of a protein into a particular folded state [39].

Air oxidation of L79CC81A yielded the two oxidized products of L79CC81A, namely L79CC81A-a and -b, which have two alternative combinations of disulphide bonds in the β -sheet domain and different relative activities. L79CC81A is a rare mutant in which Cys-65 and Cys-95 can form disulphide bonds with either Cys-77 or Cys-79, depending upon the circumstances. These findings seem to contradict the 'one gene, one enzyme' theory. Our results show that the structure of the β -sheet domain of h-lysozyme can tolerate the formation of two different sets of disulphide bonds (Cys⁶⁵–Cys⁷⁷ and Cys⁷⁹–Cys⁹⁵ in L79CC81A-a and Cys⁶⁵–Cys⁷⁹ and Cys⁷⁷–Cys⁹⁵ in L79CC81A-b) and can tolerate the change of disulphide bond from Cys⁷⁹–Cys⁹⁵ (L79CC81A) to Cys⁷⁷–Cys⁹⁵ (L79CC81A-b) by oxidation and the resulting conformational change, even after the completion of folding. The change in conformation among L79CC81A, L79CC81A-a and L79CC81A-b was reflected in the change in their enzymic activities (L79CC81A, 86 %; L79CC81A-a, 5 %; L79CC81A-b, 84 %). It is believed that disulphide bonds of secreted proteins are formed in the endoplasmic reticulum [39]. It is therefore puzzling that the major part of L79CC81A had two free cysteines at positions 65 and 77, and that it did not form the Cys⁶⁵–Cys⁷⁷ bond *in vivo*. We suppose that the mutant L79CC81A was secreted by yeast after *in vivo* folding, but the formation of disulphide bonds had not yet been completed in the β -sheet domain of L79CC81A. The fact that the pH of the culture medium was below 5.0 might be advantageous in keeping the two cysteines of L79CC81A free. The elucidation of the role of protein disulphide isomerase [40] in the endoplasmic reticulum will be helpful in answering this question.

At present it is difficult to draw conclusions about the relationship between disulphide-bond formation and protein folding. The possibility discussed above proposes that the linking of h-lysozyme between the N- and C-terminal regions by disulphide bonds would contribute to a decrease in the entropy of the polypeptide and become the folding seed of the folding process *in vivo* [41]. This would provide an opportunity for the amino acid residues in the β -sheet domain to arrange and form a folding intermediate similar to a molten-globule state [42]. Thus the formation of disulphide bonds *in vivo* would seem to occur even after the completion of polypeptide-chain synthesis, especially in the β -sheet domain.

We thank T. Miyazawa and F. Sakiyama for their interest and encouragement, S. Kanaya and N. Morita for helpful discussions, K. R. Rao for critical reading of the manuscript, K. Inaka for providing the refined structure of wild-type h-lysozyme, and Takeda Chemical Industries for providing the plasmid pGEL125.

REFERENCES

- 1 Tsou, C.-L. (1988) *Biochemistry* **27**, 1809–1812
- 2 Wright, P. E., Dyson, H. J. and Lerner, R. A. (1988) *Biochemistry* **27**, 7167–7175

- 3 Baldwin, R. L. (1989) *Trends Biochem. Sci.* **14**, 291–294
- 4 Fisher, G. and Schmid, F. X. (1990) *Biochemistry* **29**, 2205–2212
- 5 Bergman, L. W. and Kuel, W. M. (1979) *J. Biol. Chem.* **254**, 8869–8876
- 6 Ishiwata, A., Kawata, Y. and Hamaguchi, K. (1991) *Biochemistry* **30**, 7766–7771
- 7 McCammon, J. A., Gelin, B. R., Karplus, M. and Wolynes, P. G. (1976) *Nature (London)* **262**, 325–326
- 8 Janin, J. and Wodak, S. J. (1983) *Prog. Biophys. Mol. Biol.* **42**, 21–78
- 9 Miranker, A., Radford, S. E., Karplus, M. and Dobson, C. M. (1991) *Nature (London)* **349**, 633–636
- 10 Radford, S. E., Dobson, C. M. and Evans, P. A. (1992) *Nature (London)* **358**, 302–307
- 11 Yoshimura, K., Toibana, A., Kikuchi, K., Kobayashi, M., Hayakawa, T., Nakahama, K., Kikuchi, M. and Ikehara, M. (1987) *Biochem. Biophys. Res. Commun.* **145**, 712–718
- 12 Inaka, K., Taniyama, Y., Kikuchi, M., Morikawa, K. and Matsushima, M. (1991) *J. Biol. Chem.* **266**, 12599–12603
- 13 Taniyama, Y., Yamamoto, Y., Nakao, M., Kikuchi, M. and Ikehara, M. (1988) *Biochem. Biophys. Res. Commun.* **152**, 962–967
- 14 Omura, F., Taniyama, Y. and Kikuchi, M. (1991) *Eur. J. Biochem.* **198**, 477–484
- 15 Omura, F., Otsu, M., Yoshimori, T., Tashiro, Y. and Kikuchi, M. (1992) *Eur. J. Biochem.* **210**, 591–599
- 16 Taniyama, Y., Kuroki, R., Omura, F., Seko, C. and Kikuchi, M. (1991) *J. Biol. Chem.* **266**, 6456–6561
- 17 Taniyama, Y., Yamamoto, Y., Kuroki, R. and Kikuchi, M. (1990) *J. Biol. Chem.* **265**, 7570–7575
- 18 Kikuchi, M., Taniyama, Y., Kanaya, S., Takao, T. and Shimomishi, Y. (1990) *Eur. J. Biochem.* **187**, 315–320
- 19 Taniyama, Y., Ogasawara, K., Yutani, K. and Kikuchi, M. (1992) *J. Biol. Chem.* **267**, 4619–4624
- 20 Toh-e, A., Ueda, Y., Kamimoto, S. and Oshima, Y. (1973) *J. Bacteriol.* **113**, 727–738
- 21 Kikuchi, M., Yamamoto, Y., Taniyama, Y., Ishimaru, K., Yoshikawa, W., Kaisho, Y. and Ikehara, M. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9411–9415
- 22 Landt, O., Grunert, H.-P. and Hahn, U. (1990) *Gene* **96**, 125–128
- 23 Nohkura, K., Morita, N., Yamaguchi, M. and Watanabe, T. (1992) *Anal. Lett.* **25**, 513–533
- 24 Kanaya, E., Kanaya, S. and Kikuchi, M. (1990) *Biochem. Biophys. Res. Commun.* **173**, 1194–1199
- 25 Kanaya, E. and Kikuchi, M. (1992) *J. Biol. Chem.* **267**, 15111–15115
- 26 Ellman, G. L. (1959) *Arch. Biochem. Biophys.* **82**, 70–77
- 27 Taniuchi, H. (1970) *J. Biol. Chem.* **245**, 5459–5468
- 28 Andria, G. and Taniuchi, H. (1987) *J. Biol. Chem.* **262**, 2262–2270
- 29 Hughson, F. M., Wright, P. E. and Baldwin, R. L. (1990) *Science* **249**, 1544–1548
- 30 Jeng, M.-F. and Englander, S. W. (1991) *J. Mol. Biol.* **221**, 1045–1061
- 31 Artymuk, P. J. and Blake, C. C. F. (1981) *J. Mol. Biol.* **152**, 737–762
- 32 Omura, F., Otsu, M. and Kikuchi, M. (1992) *Eur. J. Biochem.* **205**, 551–559
- 33 Woychik, N. A. and Dimond, R. L. (1987) *J. Biol. Chem.* **262**, 10008–10014
- 34 Copeland, C. S., Zimmer, K.-P., Wagner, K. R., Healey, G. A., Mellman, I. and Helenius, A. (1988) *Cell* **53**, 197–209
- 35 Williams, D. B., Borriello, F., Zeft, R. A. and Nathenson, S. G. (1988) *J. Biol. Chem.* **263**, 4549–4560
- 36 Sato, T., Uemura, H., Izumoto, Y., Nakao, J., Nakamura, Y. and Matsubara, K. (1989) *Gene* **83**, 355–365
- 37 Matsumura, M., Becktel, W. J., Levitt, M. and Matthews, B. W. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6562–6566
- 38 Creighton, T. E. (1988) *BioEssays* **8**, 57–63
- 39 Kim, P. S. and Baldwin, R. L. (1990) *Annu. Rev. Biochem.* **59**, 631–660
- 40 Noiva, R. and Lennarz, W. J. (1992) *J. Biol. Chem.* **267**, 3553–3556
- 41 Seckler, R. and Jaenicke, R. (1992) *FASEB J.* **6**, 2545–2552
- 42 Kuwajima, K. (1989) *Proteins* **6**, 87–103