The EMBO Journal Vol.16 No.22 pp.6694–6701, 1997

The C-propeptide domain of procollagen can be replaced with a transmembrane domain without affecting trimer formation or collagen triple helix folding during biosynthesis

Neil J.Bulleid[1], Jane A.Dalley and Janice F.Lees

School of Biological Sciences, The University of Manchester, 2.205 Stopford Building, Manchester M13 9PT, UK

[1]Corresponding author
e-mail: neil.bulleid@man.ac.uk

The folding and assembly of procollagen occurs within the cell through a series of discrete steps leading to the formation of a stable trimer consisting of three distinct domains: the N-propeptide, the C-propeptide and the collagen triple helix flanked at either end by short telopeptides. We have established a semi-permeabilized cell system which allows us to study the initial stages in the folding and assembly of procollagen as they would occur in the intact cell. By studying the folding and assembly of the C-propeptide domain in isolation, and a procollagen molecule which lacks the C-propeptide, we have shown that this domain directs the initial association event and is required to allow triple helix formation. However, the essential function of this domain does not include triple helix nucleation or alignment, since this can occur when the C-propeptide is substituted with a single transmembrane domain. Also the telopeptide region is not involved in triple helix nucleation; however, a minimum of two hydroxyproline-containing Gly-X-Y triplets at the C-terminal end of the triple helix are required for nucleation to occur. Thus, the C-propeptide is required solely to ensure association of the monomeric chains; once these are brought together, the triple helix is able to nucleate and fold to form a correctly aligned triple helix.

Keywords: haemagglutinin transmembrane domain/procollagen folding/trimerization

Introduction

Procollagen molecules fold and assemble through a series of distinct intermediates which can be characterized by their extent of modification, disulfide bonding status and polymeric state (Bulleid, 1996). The essential features of the folding and assembly pathway can be summarized as follows. As the individual polypeptide chains are co-translationally translocated across the membrane of the endoplasmic reticulum (ER), intra-chain disulfide bonds are formed within the N-propeptide and hydroxylation of proline and lysine residues occurs within the triple helical domain (Kivirikko *et al.*, 1992). Once the polypeptide chain is fully translocated into the lumen of the ER, the C-propeptide folds, with the final conformation being stabilized by the formation of intra-chain disulfide bonds (Doege and Fessler, 1986). The proα-chains then associate via the C-propeptides to form a trimeric molecule which can be stabilized by the formation of inter-chain disulfide bonds (Schofield *et al.*, 1974; Olsen *et al.*, 1976). The formation of inter-chain disulfide bonds is not a pre-requisite for further folding, as mutating the cysteine residues involved has no effect on assembly or triple helix formation (Bulleid *et al.*, 1996). The triple helix then nucleates and folds in a C- to N- direction, with the N-propeptides finally associating and, in some cases, forming inter-chain disulfide bonds (Bruckner *et al.*, 1978).

Clearly the C-propeptide plays a crucial role in the early stages of procollagen assembly, being responsible for the initial interactions between chains, and thereby driving chain assembly. These interactions also ensure the 'type-specific' assembly of individual procollagen chains. It has been known for some time that a single cell can synthesize several different procollagen types and yet they only ever assemble in a type-specific manner. Thus, these chains are able to discriminate between each other to enable the selective association of the various procollagen types. The specific amino acids responsible for this selective association of chains have been identified recently (Lees *et al.*, 1997), but it is not clear if these are the only residues required to stabilize the trimeric structure formed.

Although the C-propeptide clearly directs the association of the individual chains, its role in ensuring triple helix formation and conferring the correct alignment of the individual chains has still to be determined. For the non-fibrillar collagens, such as type XII, most of the C-propeptide can be removed without preventing chain association and triple helix formation (Mazzorana *et al.*, 1994). For fibrillar collagens, studies with inhibitors of prolyl 4-hydroxylase which prevent proline hydroxylation and, therefore, triple helix formation have shown that chain assembly can be broken down into two distinct steps, namely the formation of a trimeric molecule by association at the C-propeptide, followed by triple helix nucleation (Bächinger *et al.*, 1981). For type XII collagen, the nucleation and association events appear to be intimately linked, as no association of the C-propeptide domain is seen in the presence of inhibitors of prolyl 4-hydroxylase (Mazzorana *et al.*, 1992, 1996).

This study sets out to address the role of the C-propeptide domain of fibrillar collagens in the initial association of the proα-chains along with nucleation of the triple helix and alignment of the chains in the triple helical domain. To facilitate these studies, we have used a procedure to permeabilize selectively the plasma membrane of cells grown in culture, which releases the cytosol whilst retaining a functionally intact ER (Wilson *et al.*, 1995). These semi-permeabilized cells can then be included instead of microsomal membranes in a conventional cell-free translation system. Using this system, the translocation, disulfide bond formation and assembly of a

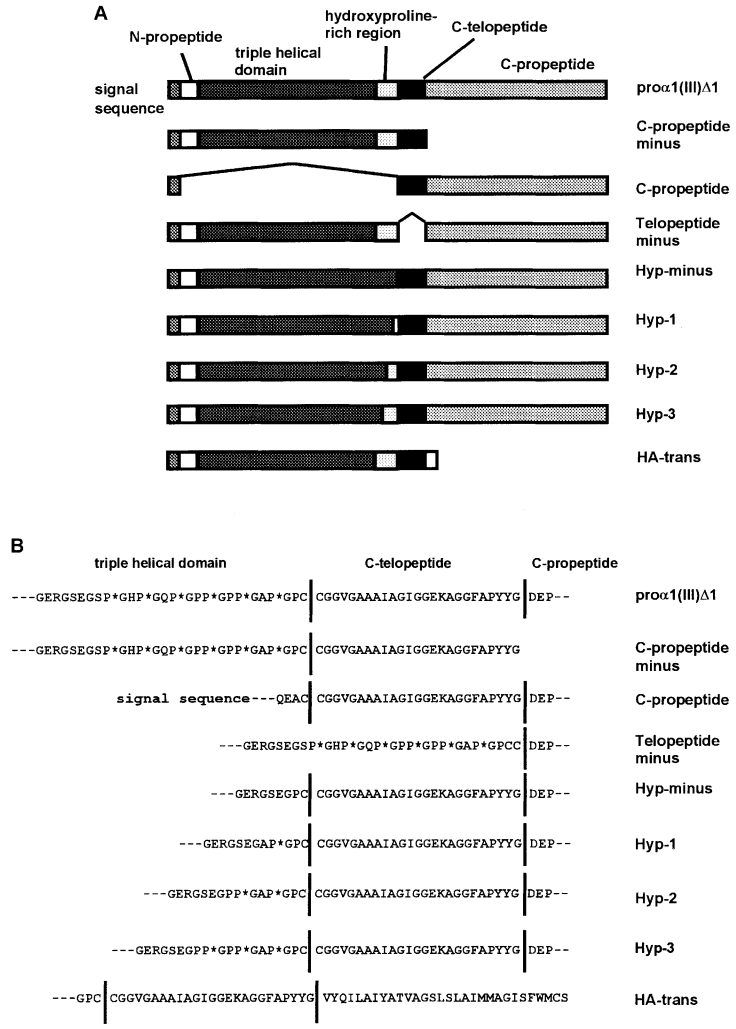

Fig. 1. (**A**) Schematic diagram indicating the domains present within the various recombinant molecules described in the text. The domains identified include the authentic N-propeptide, C-propeptide and C-telopeptide from the proα1(III) chain and a truncated triple helical domain encoding 192 amino acids (Lees and Bulleid, 1994). A region rich in hydroxyproline residues is also identified towards the C-terminal end of the triple helical domain. Chevrons indicate deleted domains in the C-propeptide and telopeptide-minus constructs. (**B**) Amino acid sequence at the C-terminal region of the triple helical domain of recombinant molecules used in this study. Proline residues with an asterisk are in the Y position of a Gly-X-Y triplet and, therefore, have the potential to become hydroxylated. Solid vertical bars denote the boundaries of the C-telopeptide region. The C-proteinase cleavage site is between Gly and Asp.

type III procollagen 'mini-gene' into a correctly aligned thermally stable, triple helical molecule has been reconstituted, mimicking the processes as they would occur within an intact cell (Bulleid *et al.*, 1996). By constructing procollagen molecules lacking either the C-propeptide, or containing the C-propeptide alone, we have demonstrated that this domain is necessary and sufficient to ensure chain association. However, we show that a transmembrane domain can substitute for the C-propeptide, allowing the formation of correctly aligned triple helical molecules. This demonstrates that the C-propeptide is not required for the nucleation and alignment of the individual chains. We also show that the C-telopeptide region is not involved in chain interactions. Rather, at least two hydroxyproline residues are required at the C-terminal end of the triple helical domain in order to ensure nucleation.

Results

A fibrillar procollagen chain can be divided into three individual folding domains, i.e. the N- and C-propeptides and the triple helical region (Figure 1A). This study utilized a procollagen type III 'mini-chain', designated proα1(III)Δ1, which encodes the N-propeptide, including the amino-terminal signal sequence, a truncated triple helical domain and the C-propeptide (Figure 1A). The triple helical region in proα1(III)Δ1 contains 192 residues compared with 1023 amino acids in the full-length protein. The deletion in the triple helical domain preserves the Gly-X-Y triplet consensus and is not predicted to alter the folding of the chains. Previous experiments have shown that procollagen 'mini-chains' translated in the presence of SP cells are efficiently translocated, modified and, in the case of proα1(III)Δ1, assembled into a correctly aligned triple helix (Bulleid *et al.*, 1996).

The role of the C-propeptide in procollagen folding and assembly

Our initial aim was to investigate whether the C-propeptide of procollagen is sufficient to drive the initial association of the individual proα-chains. To address this aim, we

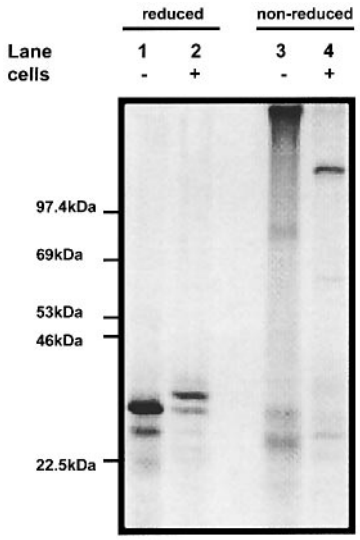

Fig. 2. Synthesis of the C-propeptide domain in the presence and absence of SP cells. RNA coding for the C-propeptide domain was translated in a rabbit reticulocyte lysate in the absence (lanes 1 and 3) or presence (lanes 2 and 4) of semi-permeabilized HT1080 cells (SP cells). SP cells were isolated by centrifugation and solubilized prior to electrophoresis. Products of translation were separated by SDS–PAGE on a 10% polyacrylamide gel either with (lanes 1 and 2) or without (lanes 3 and 4) reduction with 50 mM DTT. Commercial radiolabelled molecular weight markers used were phosphorylase B (97.4 kDa), bovine serum albumin (69 kDa), γ-globulin subunit 1 (53 kDa), ovalbumin (46 kDa) and γ-globulin subunit 2 (22.5 kDa).

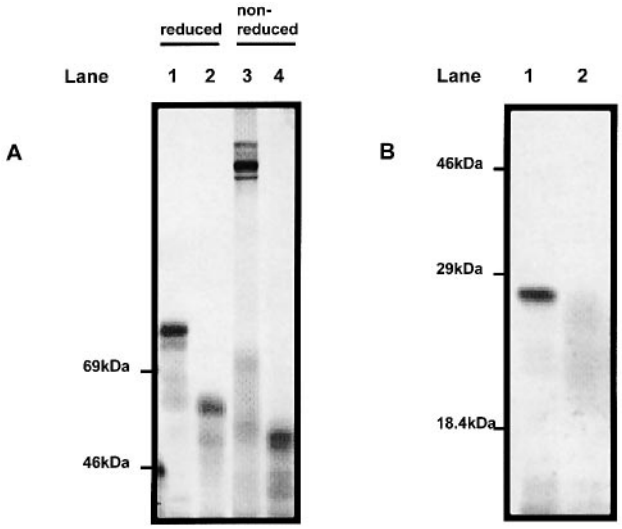

Fig. 3. Comparison of the folding and assembly of proα1(III)Δ1 and the C-propeptide-minus chains. (**A**) RNA coding for proα1(III)Δ1 (lanes 1 and 3) and the C-propeptide-minus construct were translated in a rabbit reticulocyte lysate in the presence of SP cells. After translation, SP cells were isolated by centrifugation, solubilized and the translation products separated by SDS–PAGE through a 7.5% gel either under reducing (lanes 1 and 2) or non-reducing (lanes 3 and 4) conditions. Radiolabelled molecular weight markers used were bovine serum albumin (69 kDa) and ovalbumin (46 kDa). (**B**) Translations were carried out as described for (A). Isolated SP cells were solubilized and digested at 30°C with a combination of chymotrypsin (250 μg/ml) and trypsin (100 μg/ml). The reactions were stopped by the addition of soybean trypsin inhibitor, and acidified by the addition of HCl to a final concentration of 100 mM. Samples were incubated with pepsin (100 μg/ml), the reactions being stopped by neutralization with Tris base (100 mM). Digestion products were separated by SDS–PAGE on a 12.5% gel. Lane 1, digestion products from the proα1(III)Δ1 translation. Lane 2, digestion products from the C-propeptide-minus translation. Radiolabelled molecular weight markers used were ovalbumin (46 kDa), carbonic anhydrase (29 kDa) and β-lactoglobulin (18.4 kDa).

made a cDNA construct derived from the proα1 chain of type III procollagen coding for the entire C-propeptide with the C-telopeptide fused to the signal sequence for targeting to the ER (Figure 1A and B). We also included the first three amino acids (QEA) from the authentic sequence to ensure efficient signal peptide cleavage after translocation of the polypeptide across the ER membrane. An RNA transcript was prepared and expressed in a cell-free system comprising a rabbit reticulocyte lysate optimized for the formation of disulfide bonds supplemented with semi-permeabilized HT1080 cells (SP cells). We have shown previously that SP cells can carry out the initial stages in the folding, post-translational modification and assembly of procollagen (Bulleid *et al.*, 1996).

When the RNA coding for the C-propeptide construct was translated, a single major product was observed with an apparent mol. wt of 32 kDa as deduced by SDS–PAGE (Figure 2, lane 1). Minor lower molecular weight products are due to initiation of translation at internal sites within the coding sequence and are often seen after *in vitro* translation. When SP cells were included in the translation reaction, a translation product of reduced mobility with an apparent mol. wt of 35 kDa was synthesized (Figure 2, lane 2). This difference in mobility is due to a combination of the addition of an oligosaccharide side chain and the removal of the signal sequence, as was demonstrated by an increase in mobility of the translation product following endoglycosidase H digestion (results not shown). When the same translation products were separated by SDS–PAGE without prior reduction, higher molecular weight products were observed (Figure 2, lanes 3 and 4). In the absence of cells, a large aggregate that does not enter the separating gel can be seen, which suggests that the translation product synthesized under these conditions forms non-native inter-chain disulfide bonds. However, in the presence of cells, a distinct band with an apparent mol. wt of 114 kDa corresponding to a trimer was observed. This indicates that when the polypeptide chain is translocated into the ER the C-propeptide domain folds correctly and can associate to form native inter-chain disulfide bonds. This demonstrates that the C-propeptides, in the absence of any sequence other than the C-telopeptide, can associate and form disulfide-bonded trimers.

To determine whether the C-propeptide domain was necessary for assembly of the proα-chains, we constructed a molecule which lacked the C-propeptide but which retains the signal sequence, N-propeptide, a truncated triple helical domain and the C-telopeptide with the sequence terminating at the procollagen C-proteinase cleavage site (Figure 1A and B). When RNA coding for the parental molecule, proα1(III)Δ1, was translated in the presence of cells and separated by SDS–PAGE under reducing conditions, a single translation product was seen (Figure 3A, lane 1). This product migrated as an inter-chain disulfide-bonded trimer when separated without prior reduction (Figure 3A, lane 3). We have demonstrated previously that this represents a correctly folded triple helical molecule with inter-chain disulfide bonds stabilizing the structure in both the N- and C-propeptides;

minor products seen represent inter-chain disulfide-bonded trimers linked only at the C-propeptide (Bulleid *et al.*, 1996). Further confirmation that a triple helix was formed can be shown by treating the translation products with a combination of trypsin, chymotrypsin and pepsin. The resistance to digestion of the triple helical domain demonstrates correct folding (Bruckner and Prockop, 1981) and is evidenced by a protease-resistant polypeptide with an apparent mol. wt of 28 kDa which corresponds to the truncated triple helical domain present in proα1(III)Δ1 (Figure 3B, lane 1). When a similar analysis was carried out on the C-propeptide-minus construct, no inter-chain disulfide-bonded trimers were seen (Figure 3A, lane 4), even though the cysteine residues known to form inter-chain disulfide bonds at the C-terminal end of the triple helical domain were present in the construct (Figure 1B). However, a faster migrating polypeptide was observed when the translation products were separated without prior reduction (Figure 3A, compare lanes 2 and 4). This is due to the formation of intra-chain disulfide bonds, probably in the N-propeptide (Bruckner *et al.*, 1978), which result in a more compact structure with a faster mobility than the fully reduced polypeptide. When the C-propeptide-minus translation products were treated with proteases, no resistant peptide remained (Figure 3B, lane 2), confirming that in the absence of a C-propeptide, the chains do not associate and, therefore, the triple helical domain does not fold correctly.

What are the requirements for triple helix nucleation?

Having established that the C-propeptide is sufficient and necessary for the correct association of procollagen chains, we then addressed the sequence requirements for nucleation of the triple helical domain. It has been postulated previously that the C-telopeptide may play a role in ensuring both the nucleation of the helical domain and chain alignment (Doege and Fessler, 1986; Dion and Myers, 1987). As this region of 25 amino acids was also present in our C-propeptide construct, it is conceivable that it may play a role in the initial association of the chains. To address this point, we constructed a cDNA which encodes the same sequence as the parental proα1-(III)Δ1 chain apart from the C-telopeptide region which is deleted (Figure 1A and B). RNA coding for this telopeptide-minus construct was translated in the presence of SP cells and analysed for the ability to form inter-chain disulfide-bonded trimers (Figure 4, lanes 1 and 3). The results clearly show that under non-reducing conditions a trimer is formed (Figure 4, lane 3). A protease-resistant fragment corresponding to the triple helical domain remained after protease digestion, demonstrating that a correctly aligned triple helix had formed (Figure 4, lane 2). When the same protease-treated sample was separated without prior reduction, a slower migrating peptide was seen (Figure 4, lane 4) which corresponds to an inter-chain disulfide-bonded trimer of the protease-resistant fragments. The formation of inter-chain disulfide bonds at the C-terminal end of the triple helix is a clear indication that the triple helix not only has folded but also that the chains are aligned correctly. Thus, these results demonstrate that the C-telopeptide plays no role in either the initial association of the chains or in triple helix nucleation and alignment.

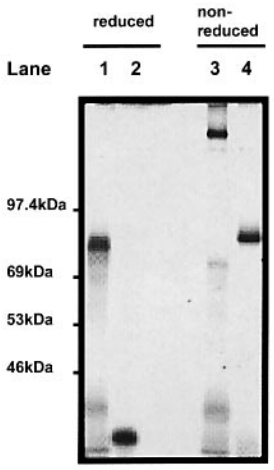

Fig. 4. Folding and assembly of the telopeptide-minus chains. RNA coding for the telopeptide-minus construct was translated in a rabbit reticulocyte lysate in the presence of SP cells. SP cells were isolated and solubilized and either separated by SDS–PAGE directly (lanes 1 and 3) or treated with proteases (lanes 2 and 4) prior to separation. Samples were separated on a 10% gel either under reducing (lanes 1 and 2) or non-reducing (lanes 3 and 4) conditions. Commercial radiolabelled molecular weight markers used were phosphorylase B (97.4 kDa), bovine serum albumin (69 kDa), γ-globulin subunit 1 (53 kDa) and ovalbumin (46 kDa).

It is clear that hydroxylation of proline residues in the collagen triple helix is a necessary prerequisite for triple helix formation (Berg and Prockop, 1973). Analysis of C-terminal Gly-X-Y triplets at the C-terminal end of collagen triple helices reveals that they are rich in proline residues in the Y position and, therefore, have the potential to be hydroxylated by prolyl 4-hydroxylase. The proα-chain of type III collagen has six proline residues in this region that potentially can become hydroxylated (Figure 1B). To determine the role that these residues play in nucleation of the chains, we first deleted this region to create a construct we termed hyp-minus. It should be noted that the truncated triple helical domain of proα1(III)Δ1 contains 29 potential hydroxyproline residues (15% of total amino acids) and that the hyp-minus construct still contains 23 of these residues (13.2% of total amino acids). This percentage of potential hydroxyproline residues in the hyp-minus construct should still be sufficient to form a stable triple helix. Thus, when a procollagen chain with a sequence with an even lower percentage of hydroxyprolines was synthesized in SP cells it was still able to form a thermally stable triple helix (Lees *et al.*, 1997). When RNA coding for the hyp-minus polypeptide was translated in the presence of SP cells, an inter-chain disulfide-bonded trimer was formed (Figure 5A, compare lanes 2 and 7). However, this trimer migrated with a slower mobility than the parent molecule (Figure 5A, compare lanes 6 and 7), which suggests that inter-chain disulfide bonds had formed in the C- but not the N-propeptide. This would give rise to a trimer covalently linked at only one end and, therefore, having a higher hydrodynamic volume (Fessler *et al.*, 1981). Thus it would appear that the triple helical domain did not nucleate and fold in this case, a conclusion that was confirmed when the translation products were treated with proteases. As

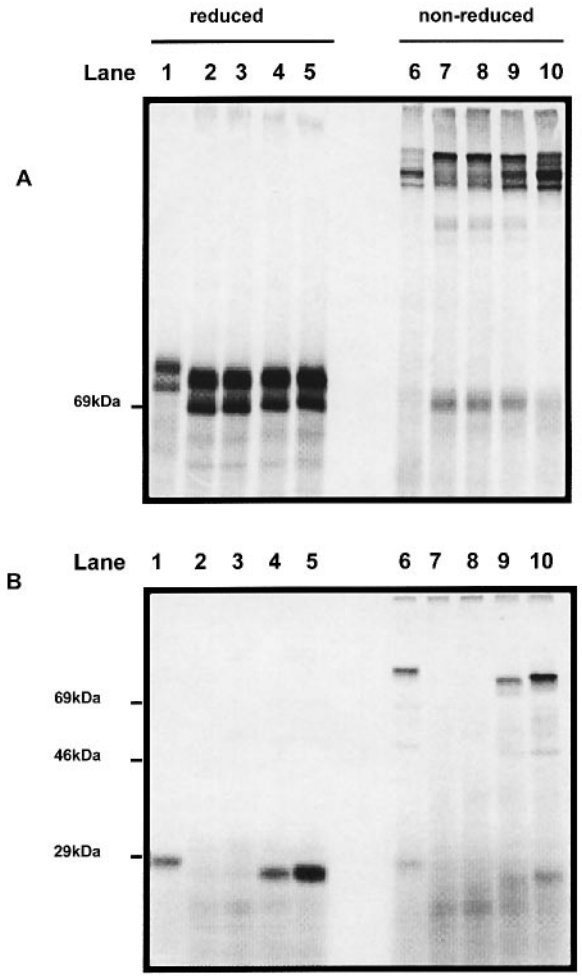

Fig. 5. Analysis of the folding and assembly of hydroxyproline recombinants. RNAs coding for proα1(III)Δ1 (lanes 1 and 6), hyp-minus (lanes 2 and 7), hyp-1 (lanes 3 and 8), hyp-2 (lanes 4 and 9) and hyp-3 (lanes 5 and 10) were translated individually in a rabbit reticulocyte lysate in the presence of SP cells. Cells were isolated by centrifugation and either separated by SDS–PAGE directly on a 7.5% gel (**A**) or treated with proteases prior to separation on a 12.5% gel (**B**). Samples were separated under reducing (lanes 1–5) or non-reducing conditions (lanes 6–10). Radiolabelled molecular weight markers were as used in Figure 3.

can be seen (Figure 5B, lanes 2 and 7), no protease-resistant fragments remained after digestion. This demonstrates that the six Gly-X-Y triplets at the C-terminal end of the triple helical domain are not required for the initial association of the chains but are necessary to allow triple helix nucleation and folding.

To ascertain the minimum number of Gly-X-Y triplets required to nucleate the triple helix, we made a series of constructs where we deleted five, four or three of the original six triplets. We called these hyp-1, hyp-2 and hyp-3, respectively, to indicate the number of potential hydroxyproline residues in this region (Figure 1A and B). When RNA coding for these individual constructs was translated in the presence of SP cells, inter-chain disulfide-bonded trimers were formed (Figure 5A, compare lanes 3–5 with 8–10). However, only hyp-2 and hyp-3 were able to form the faster migrating trimer seen with the parent molecule, which is indicative of inter-chain disulfide bonds having formed in the N-propeptide. This would suggest that only these constructs were able to nucleate the triple helix, thereby allowing triple helix folding. This was confirmed when the translation products were treated with proteases. Protease-resistant fragments remained after digestion of the hyp-2 and hyp-3 polypeptides but not the hyp-1 polypeptide (Figure 5B, lanes 3–5). A comparison of the proportion of the translation products remaining after digestion (Figure 5A and B, compare lanes 4 and 5) would suggest that the efficiency of triple helix formation was lower with the hyp-2 than the hyp-3 polypeptide. This indicates that the additional hydroxyproline residue present in the latter construct increases the interaction of the polypeptide chains at the C-terminal end of the triple helical domain. Importantly, the protease-resistant fragment formed after digestion of these polypeptides is disulfide bonded (Figure 5B, lanes 9 and10), indicating that the helix is correctly aligned. Thus, at least two hydroxyproline-containing Gly-X-Y triplets are required at the C-terminal end of the triple helical domain to ensure nucleation, folding and the correct alignment of the collagen triple helix.

Can the triple helical domain nucleate without the C-propeptide?

Our results demonstrated that the triple helical domain requires a nucleation sequence, hydroxyproline-containing Gly-X-Y triplets, in order to assemble and fold. We also showed that, in the absence of the C-propeptide, this nucleation event could not occur. This could be due to the C-propeptide also having a role in triple helix nucleation or it could simply be due to the C-propeptide being required to bring the chains together in close proximity to allow the nucleation event to occur. We decided to address this question by replacing the C-propeptide with the transmembrane domain from the trimeric molecule influenza virus haemagglutinin. We would predict that this would anchor the collagen-derived triple helix at the ER membrane and would not interfere with the folding of this domain. We replaced the entire C-propeptide domain up to the C-proteinase cleavage site with this transmembrane region to give rise to a construct which we termed HA-trans (Figure 1A and B). When RNA coding for this HA-trans polypeptide was translated in the presence of SP cells, a proportion of the translation products formed inter-chain disulfide bonds (Figure 6A, lane 4). The chains which did not form inter-chain disulfide bonds did, however, form intra-chain disulfides, as evidenced by a faster migration of the monomer (Figure 6A, compare lanes 2 and 4). This indicates that the N-propeptide of HA-trans folded correctly and that some of the monomers were able to associate to form inter-chain disulfide bonds. The efficiency of this association was less than that seen with the parent molecule (Figure 6A, lanes 3 and 4). To determine if the triple helical domain of HA-trans had formed, the translation products were treated with proteases. The results clearly demonstrate that a correctly aligned triple helix is formed and that it is aligned correctly, as evidenced by a protease-resistant fragment, which is inter-chain disulfide-bonded, remaining after digestion (Figure 6B, lane 4). This fragment has a mobility identical to that of the fragment produced from the parent molecule, indicating that a trimer is formed. Again the efficiency of trimerization was less than that seen with the parent molecule;

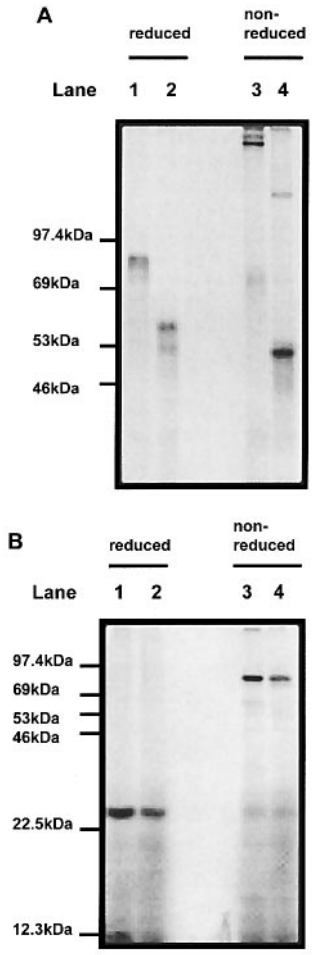

Fig. 6. Analysis of the folding and assembly of HA-trans recombinant. RNAs coding for proα1(III)ΔΔ1 (lanes 1 and 3) and HA-trans (lanes 2 and 4) were translated in a rabbit reticulocyte lysate in the presence of SP cells. Cells were isolated by centrifugation and either separated by SDS–PAGE directly on a 7.5% gel (**A**) or treated with proteases prior to separation on a 12.5% gel (**B**). Samples were separated under reducing (lanes 1 and 2) or non-reducing conditions (lanes 3 and 4). Radiolabelled molecular weight markers were as used in Figure 2.

this relative efficiency was, however, variable between different preparations of SP cells. Thus, the triple helical domain can nucleate to form a correctly aligned helix in the absence of the C-propeptide. The essential function of the C-propeptide in chain association during biosynthesis can, therefore, be reproduced by a transmembrane domain, albeit at a lower efficiency, indicating that the main role of the C-propeptide during triple helix formation is to ensure that the chains are brought close enough together to allow nucleation of the helix to occur.

Discussion

The folding and assembly of procollagen molecules in the ER has been shown to involve an initial association of the individual proα-chains at the C-propeptide, followed by a nucleation of the triple helical domain at its C-terminal end, ensuring correct alignment of the chains. Evidence to support this assembly pathway comes from experiments carried out on the biosynthesis of procollagen under conditions which prevent sufficient hydroxylation of proline residues to enable a thermally stable triple helix to form (Berg and Prockop, 1973). Under these conditions, an intermediate in the folding pathway is observed which is a non-helical trimeric molecule stabilized by the formation of inter-chain disulfide bonds in the C-propeptide (Bächinger *et al.*, 1981; Fessler *et al.*, 1981). Once this intermediate becomes hydroxylated, the triple helix folds, demonstrating the requirement for proline hydroxylation in order for the nucleation event to occur. As a consequence of the nucleation event occurring at the C-terminal end of the triple helix, this domain folds in a C- to N- direction, the directionality of the process being confirmed by the studies on the refolding of denatured type III collagen (Bächinger *et al.*, 1980). This general mechanism of helix formation can also be applied to the non-fibrillar collagens such as type XII, which also requires an association event to occur prior to helix formation (Mazzorana *et al.*, 1994).

Our results demonstrate that chain association and helix nucleation during assembly are directed by distinct domains within the polypeptide chain. The essential function of the C-propeptide is to drive the initial association event to form a trimer, a function which can be complemented to a certain extent by the replacement of this complete domain with the transmembrane domain from influenza virus haemagglutinin. The role of this transmembrane domain in the assembly of the trimeric haemagglutinin molecule is not clear, but it is not essential to ensure trimer formation (Singh *et al.*, 1990). Interestingly, the assembly of this molecule seems to be concentration dependent and occurs as the protein is packaged into transport vesicles destined for exit from the ER to the intermediate compartment (Tatu *et al.*, 1995). If the trimerization of our HA-trans construct is dependent upon the lateral diffusion of the molecule within the phospholipid bilayer, rather than a more positive association via the C-propeptide, this would explain the modest efficiency of trimerization of this construct. The ability of the HA-trans polypeptide to assemble to form correctly aligned triple helices demonstrates that the C-propeptide domain is not required specifically to direct the nucleation event or to ensure chain alignment, a function that has often been ascribed to this domain in the past (Doege and Fessler, 1986; Dion and Myers, 1987).

Thus, our data suggest that as long as the triple helical domains are brought together in the same vicinity, and are tethered at the carboxy-terminus, they are able to nucleate and fold. This is in contrast to the situation when collagen lacking the C-propeptides is refolded after denaturation *in vitro*. Unless inter-chain disulfide bonds are present at the C-terminal end of the chain (as is the case with type III collagen), then these chains are unable to fold into correctly aligned helices (Engel and Prockop, 1991). Instead, they associate to form high molecular weight aggregates consisting of overlapping misaligned helices which are susceptible to digestion with proteases (Bruckner and Prockop, 1981). Thus, the specific association of the C-propeptides also functions to tether the chain at one end, thereby preventing misaligned helices from forming.

Our results also demonstrate that the second assembly event, the nucleation and alignment of the triple helix, is clearly directed by the most C-terminal Gly-X-Y triplet of the triple helix and not by the C-propeptide or the C-telopeptide. The strict requirement for at least two

triplets containing hydroxyproline residues within this region of the helix raises the question as to why other adjacent Gly-X-Y triplets containing hydroxyproline residues are unable to nucleate the triple helix. As helix formation is a co-operative process, one would predict that a sequence containing at least three hydroxyproline-containing triplets would be able to nucleate the helix and allow it to fold. During biosynthesis this clearly is not the case since the hyp-minus triple helix used here was unable to form a helix. Although the proline residues in the Y position of the Gly-X-Y sequences throughout the domain are hydroxylated, they were unable to nucleate the triple helix. The ability of the C-terminal triplets to nucleate the helix could be due simply to the fact that the chains are also brought closer together at the C-propeptide, thus allowing nucleation to occur. Alternatively, other molecules such as prolyl 4-hydroxylase (Chessler and Byers, 1992) or Hsp47 (Nakai *et al.*, 1992) could prevent the internal residues from nucleating the chains by binding to the helix. Once the C-propeptide associates, these proteins could be displaced, allowing the most C-terminal triplets to nucleate the chains. Clearly further work will need to be carried out to address this point.

Materials and methods

Construction of recombinant plasmids

Recombinant pα1(III)Δ1 has been described previously (Lees and Bulleid, 1994). Recombinant plasmid constructs were generated by PCR overlap extension using the principles outlined by Horton (1993). PCR reactions (100 μl) comprised template DNA (500 ng), oligonucleotide primers (100 pmol each) in 10 mM KCl, 20 mM Tris–HCl pH 8.8, 10 mM $(NH4)_2SO_4$, 2 mM $MgSO_4$, 0.1% (v/v) Triton X-100, 300 μM each dNTP. Ten rounds of amplification were performed in the presence of 1 U of Vent DNA polymerase (New England Biolabs, MA). Recombinant C-propeptide-minus was generated using a 5′ oligonucleotide primer (5′ GATTACGCCAAGCGCGCA 3′) complementary to the T3 promoter sequence upstream of the initiation codon in pα1(III)Δ1 and a 3′ oligonucleotide primer (5′ TCGCTAGGTACCCTATTATCCATAATACGGGGCAAAAC 3′) complementary to the sequence in pα1(III)Δ1 up to the C-proteinase cleavage site and incorporating a *Kpn*I site to facilitate subsequent subcloning. PCR yielded a 1200 bp fragment which was cut with *Hin*dIII and *Kpn*I and subcloned into pBS-SK$^-$ (Stratagene Ltd, Cambridge, UK). Recombinants hyp-minus, hyp-1, hyp-2, hyp-3 and the construct pα1(III)Δ1-telo- (see below) were generated using a 5′ oligonucleotide primer (5′ AATGGAGCTCCTGGACCCATG 3′) complementary to a sequence 100 bp upstream of an *Xho*I site in pα1(III)Δ1 and a 3′ amplification primer (5′ TCGCAGGGTACCGTCGGTCACTTGCACTGGTT 3′) complementary to a region 100 bp downstream of the stop codon in pα1(III)Δ1. A *Kpn*I site was incorporated to facilitate subsequent subcloning. Pairs of internal oligonucleotides, of which one included a 20 nucleotide overlap, were designed to generate molecules with precise junctions as delineated in Figure 1B. Overlap extension yielded a product of ~1000 bp which was purified, digested with *Xho*I and *Kpn*I and ligated into pα1(III)Δ1 from which a 1080 bp *Xho*I–*Kpn*I fragment had been excised. Recombinant telopeptide-minus was generated by deleting 12 bp from the parental molecule pα1(III)telo- using the *Pfu*I mutagenesis method. The parent construct pα1(III)telo- included the last four amino acids of the C-telopeptide PYYG and was generated before the correct C-proteinase cleavage site was known (Figure 1B). Recombinant C-propeptide was generated using a 5′ oligonucleotide primer complementary to the T3 promoter sequence upstream of the initiation codon in pα1(III)Δ1 as described above and a 3′ amplification primer (5′ TCGCAGGGATCCGTCGGTCACTTGCACTGGTT 3′) complementary to a region 100 bp downstream of the stop codon in pα1(III)Δ1. A *Bam*HI site was incorporated to facilitate subsequent subcloning. Pairs of internal oligonucleotides, of which one included a 20 nucleotide overlap, were designed to generate molecules with precise junctions to include the signal sequence and the C-telopeptide as delineated in Figure 1B. Overlap extension yielded a product of ~1100 bp which was purified, digested with *Bam*HI and ligated into pBS-SK$^-$ which previously had been digested with *Bam*HI. Recombinant HA-trans was generated using a 5′ oligonucleotide primer (5′ AATGGAGCTCCTGGACCCATG 3′) complementary to a sequence 100 bp upstream of an *Xho*I site in pα1(III)Δ1 and two 3′ primers HA1 (5′ CAGTGACAGAGAACCTGCTACTGTAGCATAAATGGCAAGGATTTGATAAACTCCATAATACGGGGCAAA 3′), containing overlap sequence at the junction site towards the C-terminal end of the C-telopeptide in pα1(III)Δ1, and HA2 (5′ CGTTCACTCGGTACCTCAGGAGCACATCCAGAAAGAGATCCCAGCCATCATGATTGACAGTGACAGAGAACCTGC 3′) which contains an overlap with HA1 and incorporates a *Kpn*I site for ease of subcloning. The result of sequential amplification with these primers was an ~1000 bp fragment which was purified, cut with *Xho*I and *Kpn*I and ligated into pα1(III)Δ1 from which a 1080 bp *Xho*I–*Bam*HI fragment had been excised.

Transcription in vitro

Transcription reactions were carried out as described by Gurevich *et al.* (1987). Recombinant plasmids were linearized and transcribed using T3 RNA polymerase (Promega, Southampton, UK). Reactions (100 μl) were incubated at 37°C for 4 h. Following purification over RNeasy columns (Qiagen, Dorking, UK), the RNA was resuspended in 100 μl of RNase-free water containing 1 mM dithiothreitol (DTT) and 40 U of RNasin (Promega).

Translation in vitro

RNA was translated using a rabbit reticulocyte lysate (FlexiLysate, Promega) for 90 min at 30°C. The translation reaction (25 μl) contained 17.5 μl of reticulocyte lysate, 0.5 μl of 1 mM amino acids (minus methionine), 0.5 μl of 100 mM KCl, 0.25 μl of ascorbic acid (5 mg/ml), 15 μCi of L-[^{35}S]methionine, (NEN, Dreiech, Germany), 1 μl of transcribed RNA and 1 μl of (~2×10^5) semi-permeabilized cells (SP cells) prepared as previously described (Wilson *et al.*, 1995). After translation, *N*-ethylmaleimide was added to a final concentration of 20 mM. SP cells were isolated by centrifugation at 10 000 *g* for 5 min and the pellet resuspended in an appropriate buffer for subsequent enzymatic digestion or gel electrophoresis.

Proteolytic digestion

Isolated SP cells were solubilized in 50 mM Tris–HCl, pH 7.4, containing 0.15 M NaCl, 10 mM EDTA, 1% (v/v) Triton X-100 and centrifuged at 10 000 *g* for 5 min to remove cell debris. The supernatant was recovered and then digested with a combination of chymotrypsin (250 μg/ml) and trypsin (100 μg/ml) (Sigma, Poole, Dorset) for 1 min at 30°C. The reactions were stopped by the addition of soybean trypsin inhibitor (Sigma) to a final concentration of 500 μg/ml, and acidified by the addition of HCl to a final concentration of 100 mM. Samples were incubated with pepsin (100 μg/ml) for 2 h at 30°C. The reactions were stopped by neutralization with Tris base (100 mM) and prepared for electrophoresis as described below.

SDS–polyacrylamide gel electrophoresis

Samples were prepared for electrophoresis by the addition of SDS–PAGE loading buffer [0.0625 M Tris–HCl pH 6.8, SDS (2% w/v), glycerol (10% v/v) and bromophenol blue] in the presence or absence of 50 mM DTT and boiled for 5 min. After electrophoresis, gels were dried, processed for autoradiography and exposed to Kodak X-Omat AR film. Radiolabelled molecular weight markers were either obtained commercially (NEN, Dreiech, Germany) or by reductive methylation of lysyl residues using the procedure of Dottavio-Martin and Ravel (1978).

Acknowledgements

We thank Karl Kadler and Steve High for critical reading of the manuscript. This work was supported by The Royal Society and a grant from The Wellcome Trust (ref. 42411).

References

Bächinger,H.-P, Bruckner,P., Timpl,R., Prockop,D.J. and Engel,J. (1980) Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen: role of disulfide bridges and peptide bond isomerization. *Eur. J. Biochem.*, **106**, 619–632.

Bächinger,H.-P., Fessler,L.I., Timpl,R. and Fessler,J.H. (1981) Chain assembly intermediate in the biosynthesis of type III procollagen in chick embryo blood vessels. *J. Biol. Chem.*, **256**, 13193–13199.

Berg,R.A. and Prockop,D.J. (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilising the triple helix of collagen. *Biochem. Biophys. Res. Commun*., **52**, 115–120.

Bruckner,P. and Prockop,D.J. (1981) Proteolytic enzymes as probes for the triple-helical conformation of procollagen. *Anal. Biochem*., **110**, 360–368.

Bruckner,P., Bächinger,H.-P., Timpl,R. and Engel,J. (1978) Three conformationally distinct domains in the amino-terminal segment of type III procollagen and its rapid triple helix to coil transition. *Eur. J. Biochem*., **90**, 595–603.

Bulleid,N.J. (1996) Novel approach to study the initial events in the folding and assembly of procollagen. *Semin. Cell Dev. Biol*., **7**, 667–672.

Bulleid,N.J., Wilson,R. and Lees,J.F. (1996) Type III procollagen assembly in semi-intact cells: chain association, nucleation and triple helix folding do not require formation of inter-chain disulfide bonds but triple helix nucleation does require hydroxylation. *Biochem. J*., **317**, 195–202.

Chessler,S.D. and Byers,P.H. (1992) Defective folding and stable association with protein disulfide isomerase/prolyl hydroxylase of type I procollagen with a deletion in the proα2(I) chain that preserves the Gly-X-Y repeat pattern. *J. Biol. Chem*., **267**, 7751–7757.

Dion,A.S. and Myers,J.C. (1987) COOH-terminal propeptides of the major human procollagens: structural, functional and genetic comparisons. *J. Mol. Biol*., **193**, 127–143.

Doege,K.J., and Fessler,J.H. (1986) Folding of carboxyl domain and assembly of procollagen I. *J. Biol. Chem*., **261**, 8924–8935.

Dottavio-Martin,D. and Ravel,J.M. (1978) Radiolabelling of proteins by reductive alkylation with [^{14}C]formaldehyde and sodium cyanoborohydride. *Anal. Biochem*., **87**, 562–565.

Engel,J. and Prockop,D.J. (1991) The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu. Rev. Biophys. Chem*., **20**, 137–152.

Fessler,L.I., Timpl,R. and Fessler,J.H. (1981) Assembly and processing of procollagen type III in chick embryo blood vessels. *J. Biol. Chem*., **256**, 2531–2537.

Gurevich,V.V., Pokrovskaya,I.D., Obukhova,T.A. and Zozulya,S.A. (1991) Preparative *in vitro* mRNA synthesis using SP6 and T7 polymerases. *Anal. Biochem*., **195**, 207–213.

Horton,R.M. (1993) *In vitro* recombination and mutagenesis of DNA. In White,B.A. (ed.), *PCR Protocols: Current Methods and Applications. Methods in Molecular Biology.* Humana Press Inc., Totowa, NJ, Vol. 15, pp. 251–261.

Kivirikko,K.I., Myllylä,R. and Pihlajaniemi,T. (1992) Hydroxylation of proline and lysine residues in collagens and other animal and plant proteins. In Harding,J.J. and Crabbe,M.J.C. (eds), *Post-translational Modification of Proteins*. CRC Press, Boca Raton, FL, pp. 1–51.

Lees,J.F. and Bulleid,N.J. (1994) The role of cysteine residues in the folding and association of the COOH-terminal propeptide of types I and III procollagen. *J. Biol. Chem*., **269**, 24354–24360.

Lees,J.F., Tasab,M. and Bulleid,N.J. (1997) Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. *EMBO J*., **16**, 908–916.

Mazzorana,M., Gruffat,H., Sergeant,A. and van der Rest,M. (1992) Mechanisms of collagen trimer formation: construction and expression of a recombinant minigene in HeLa cells reveals a direct effect of prolyl hydroxylation on chain assembly of type XII collagen. *J. Biol. Chem*., **268**, 3029–3032.

Mazzorana,M., Giry-Lozinguez,C. and van der Rest,M. (1994) Trimeric assembly of collagen XII: effect of deletion of the C-propeptide part of the molecule. *Matrix Biol*., **14**, 583–588.

Mazzorana,M., Snellman, Kivirikko,K.I., van der Rest,M. and Pihlajaniemi,T. (1996) Involvement of prolyl 4-hydroxylase in the assembly of trimeric minicollagen XII: study in a baculovirus expression system. *J. Biol. Chem*., **271**, 29003–29008.

Nakai,A., Satoh,M., Hirayoshi,K. and Nagata,K. (1992) Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. *J. Cell Biol*., **117**, 903–914.

Olsen,B.R., Hoffman,H.-P. and Prockop,D.J. (1976) Interchain disulfide bonds at the COOH-terminal end of procollagen synthesised by matrix-free cells from chick embryonic tendon and cartilage. *Arch. Biochem. Biophys*., **175**, 341–350.

Schofield,D.J., Uitto,J. and Prockop,D.J. (1974) Formation of interchain disulfide bonds and helical structure during biosynthesis of procollagen by embryonic tendon cells. *Biochemistry*, **13**, 1801–1806.

Singh,I., Doms,R.W., Wagner,K. and Helenius,A. (1990) Intracellular transport of soluble and membrane-bound glycoproteins: folding, assembly and secretion of anchor-free influenza hemagglutinin. *EMBO J*., **9**, 631–639.

Tatu,U., Hammond,C. and Helenius,A. (1995) Folding and oligomerization of influenza hemagglutinin in the ER and the intermediate compartment. *EMBO J*., **14**, 1340–1348.

Wilson,R., Allen,A.J., Oliver,J., Brookman,J.L., High,S. and Bulleid,N.J. (1995) Development of a semi-permeabilised cell system to study the translocation, folding, assembly and transport of secretory proteins. *Biochem J*., **307**, 679–687.

Received on July 15, 1997; revised on September 1, 1997