Molecular cloning of a multifunctional chicken protein acting as the prolyl 4-hydroxylase β -subunit, protein disulphide-isomerase and a cellular thyroid-hormone-binding protein

Comparison of cDNA-deduced amino acid sequences with those in other species

Tarja PARKKONEN, Kari I. KIVIRIKKO and Taina PIHLAJANIEMI*

Collagen Research Unit, Biocenter and Department of Medical Biochemistry, University of Oulu, SF-90220 Oulu, Finland

Several recent studies indicate that a single polypeptide may act as the β -subunit of prolyl 4-hydroxylase, the enzyme protein disulphide-isomerase and a cellular thyroid-hormone-binding protein. We report here the isolation and characterization of cDNA clones encoding this multifunctional protein in the chicken. All the coding sequences were determined on the basis of nucleotide sequencing of five cDNA clones and amino acid sequencing of the *N*-terminal end of the chicken β -subunit. The processed polypeptide contains 493 amino acid residues, the size of the respective mRNA being about 2.7 kb. The chicken β -subunit cDNA sequences were 78 % homologous to the previously reported human β -subunit cDNA sequences at the nucleotide level and 85 % homologous at the amino acid level. The homology of the chicken β -subunit sequences to those reported for bovine thyroid-hormone-binding protein and rat protein disulphide-isomerase was also 85 % at the amino acid level. Primary-structure comparisons between the four species indicated that the two proposed active sites of protein disulphide-isomerase, the two Trp-Cys-Gly-His-Cys-Lys sequences, are located within highly conserved regions, which are also homologous to the active sites of a number of thioredoxins. The middle of the polypeptide has an additional conserved region 100 amino acid residues in length in which the degree of homology between the four species is 94 % at the amino acid level. This long conserved region may also be important for some of the multiple functions of the protein. The four extreme *C*-terminal amino acids of the polypeptide in all four species are Lys-Asp-Glu-Leu, a sequence that has been suggested to function as a signal for the retention of a protein in the endoplasmic reticulum.

INTRODUCTION

Prolyl 4-hydroxylase [procollagen-L-proline,2-oxoglutarate:oxygen oxidoreductase (4-hydroxylating), EC 1.14.11.2] catalyses the formation of 4-hydroxyproline in collagens and other proteins with collagen-like amino acid sequences by the hydroxylation of proline residues in peptide linkages. The active enzyme is a tetramer ($\alpha_2\beta_2$) consisting of two different types of monomer with M_r of about 64 000 (α -subunit) and about 60 000 (β -subunit). The enzyme protein is present in cells in two forms: active tetramers and a form inactive as prolyl 4-hydroxylase that corresponds to the β -subunit on a number of criteria (for reviews see Cardinale & Udenfriend, 1974; Prockop *et al.*, 1976; Kivirikko & Myllylä, 1980, 1985, 1987).

The recent isolation and characterization of cDNA clones for the human prolyl 4-hydroxylase β -subunit indicates that the polypeptide is 491 amino acid residues in length (Pihlajaniemi *et al.*, 1987). These amino acid sequences of the human β -subunit were found to be highly homologous to those previously reported for rat protein disulphide-isomerase (Edman *et al.*, 1985), and the data suggested that the β -subunit of prolyl

4-hydroxylase and the enzyme protein disulphide-isomerase are both products of the same gene (Pihlajaniemi *et al.*, 1987). Protein disulphide-isomerase (EC 5.3.4.1) is a homodimeric enzyme that catalyses the rearrangement of disulphide bonds *in vitro* and is believed to be the catalyst of disulphide bond formation *in vivo* (see Freedman & Hillson, 1980; Hillson *et al.*, 1984; Freedman, 1987). Additional studies involving determination of the M_r values of the polypeptide chains, analysis of peptide maps and experiments with polyclonal and monoclonal antibodies all indicated a structural identity between the prolyl 4-hydroxylase β -subunit and protein disulphide-isomerase (Koivu *et al.*, 1987). Furthermore, the free β -subunit prepared by dissociation of the prolyl 4-hydroxylase tetramer was found to have the same amount of protein disulphide-isomerase activity as protein disulphide-isomerase itself, and even the β -subunit present in the intact prolyl 4-hydroxylase tetramer had one-half of this activity (Koivu *et al.*, 1987).

Surprisingly, the recent cloning of cDNA, coding for a protein identified as a major human and bovine cellular thyroid-hormone-binding protein present in endoplasmic reticulum indicated that this protein is also identical with

Abbreviation used: 1 $\times$ SSC, 0.15 M-NaCl/15 mM-sodium citrate buffer, pH 6.8.

* To whom correspondence and reprint requests should be addressed at: Department of Medical Biochemistry, Kajaanintie 52 A, SF-90220 Oulu, Finland.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number X13110 P4HB.

the β -subunit of human prolyl 4-hydroxylase and highly homologous to rat protein disulphide-isomerase (Cheng *et al.*, 1987; Yamauchi *et al.*, 1987). These molecular cloning studies thus suggest that a single polypeptide may act as the β -subunit of prolyl 4-hydroxylase (Pihlajaniemi *et al.*, 1987), the enzyme protein disulphide-isomerase (Edman *et al.*, 1985) and a cellular thyroid-hormone-binding protein (Cheng *et al.*, 1987; Yamauchi *et al.*, 1987).

In the present work cDNA clones coding for this multifunctional polypeptide in the chicken were isolated and characterized, with special emphasis on comparing cDNA-derived amino acid sequences in chicken, rat, bovine and human polypeptides in order to locate regions that are conserved between species. Such regions are likely to be important for the various functions of this protein.

MATERIALS AND METHODS

Isolation and nucleotide sequencing of cDNA clones

A chick-embryo λ gt11 cDNA library (Clontech, Palo Alto, CA, U.S.A.) was screened with 32 P-labelled nick-translated cDNA probes HB-88 and HB-210 coding for the human prolyl 4-hydroxylase β -subunit (Pihlajaniemi *et al.*, 1987). The library was plated on 150 mm plates on a lawn of *Escherichia coli* Y1090 cells at a density of 2×10^4 bacteriophages per plate. Duplicate copies on nitrocellulose filters (Schleicher und Schüll, Dassel, Germany) were prepared as described by Maniatis *et al.* (1982). The filters were hybridized with 1.3 ng of each human cDNA probe/ml in 35% (v/v) formamide in $6 \times$ SSC containing 1% (w/v) bovine serum albumin, 1% (w/v) Ficoll, 1% (w/v) polyvinylpyrrolidone, 0.250 μ g of denatured salmon DNA/ml and 0.1% (w/v) SDS at 37 °C for 20 h. The filters were washed with $2 \times$ SSC and 0.1% SDS at 60 °C. Several positive clones were identified, and one of them (CB-4) was used to re-screen the chicken cDNA library. The hybridization conditions were otherwise as above except that 50% (v/v) formamide and $5 \times$ SSC at 43 °C were used and the washing was at 65 °C. The positive bacteriophage recombinants were isolated and the inserts ligated into the *Eco*RI site of plasmid pBR322 (Maniatis *et al.*, 1982).

The inserts and various restriction fragments of them were also subcloned into M13 vectors (Messing, 1983). Nucleotide sequences were obtained from the M13 or pBR322 clones by the dideoxynucleotide sequencing method (Sanger *et al.*, 1977) with the use of the Klenow fragment of DNA polymerase I (Amersham International, Amersham, Bucks., U.K.) and the Sequenase enzyme (United States Biologicals, Cleveland, OH, U.S.A.). For some of the sequences dITP was used instead of dGTP in the reactions to enable reading through the compression areas. The primers used for the sequencing were either the M13 universal primer or specific oligomers (17 nucleotides in length) synthesized in an Applied Biosystems DNA synthesizer at the Department of Biochemistry, University of Oulu. The following seven oligomers were synthesized: O-1, 5' TGCGCCGC-CAGGGCCTG 3' (complementary to nucleotides 62–46 of the chicken β -subunit cDNA sequence); O-2, 5' GGCCAGGCGGATCTCGG 3' (complementary to nucleotides 180–164); O-3, 5' GGACCACTCCATCT-TGG 3' (complementary to nucleotides 553–537); O-4, 5' ACTGAGCAGACTGCTCC 3' (nucleotides 661–

677); O-5, 5' CTCTGTAGGTCTCTCCC 3' (complementary to nucleotides 1186–1170); O-6, 5' GCTC-ATCATCATCACCT 3' (complementary to nucleotides 1447–1431); O-7, 5' ACTGATCTACATCACCC 3' (nucleotides 1481–1497). Nucleotide sequences were fully determined by sequencing both strands of the cDNAs. The cDNA clone CB-13 contains an internal *Eco*RI site, and therefore subcloning of the λ gt11 insert into the *Eco*RI site of the M13 vector resulted in two subclones. Sequences fully covering the internal *Eco*RI site of cDNA clone CB-13 were obtained by the M13 universal primer, which allowed sequencing through the *Eco*RI sites of subclones CB-13-1 and CB-13-2.

Transfer blot analysis of polyadenylated RNA

Calvarias were isolated from 60 17-day chick embryos, and total RNA was prepared using the guanidinium isothiocyanate/CsCl method (Maniatis *et al.*, 1982). Polyadenylated RNA was isolated by using an oligo(dT)-cellulose affinity column (Pihlajaniemi & Myers, 1987). Electrophoresis of the RNA was performed in a 0.7% agarose gel containing 2 M-formaldehyde, and the RNA was transferred to a nitrocellulose filter and hybridized as described previously (Thomas, 1980).

Protein purification and amino acid sequence determination

Prolyl 4-hydroxylase was isolated from whole chicken embryos by using a purification procedure consisting of an affinity chromatography on poly-(L-proline), DEAE-cellulose chromatography and gel filtration (see Kivirikko & Myllylä, 1987). For determination of the N-terminal sequence of the β -subunit, 25 μ g of prolyl 4-hydroxylase was electrophoretically fractionated on an SDS/8%-polyacrylamide gel. The separated subunits were electrophoretically blotted on to a trichloroxy-(3-aminopropyl) silane-activated glass-fibre filter (Aebersold *et al.*, 1986), and the proteins were detected by staining with 3,3'-dipentylloxycarbocyanine iodide. The electroblotted β -subunit was degraded directly on the activated glass-fibre filter after cutting out the stained band corresponding to it (Aebersold *et al.*, 1986), and the amino acid sequences were determined with an Applied Biosystems model 477A on-line 120A liquid-pulsed sequencer with the use of narrow-bore reversed-phase analysis.

For determination of the additional amino acid sequences, *Staphylococcus aureus*-V8-proteinase-digest peptides were prepared from purified chicken β -subunit as described previously (Koivu *et al.*, 1987). The electrophoretically fractionated peptides were blotted on to an activated glass-fibre filter and subjected to automated Edman degradation as described above.

RESULTS AND DISCUSSION

Isolation and characterization of cDNA clones coding for the β -subunit of chicken prolyl 4-hydroxylase

A chick-embryo λ gt11 cDNA library was screened under cross-hybridization conditions with two cDNA clones covering most of the coding sequences of the human prolyl 4-hydroxylase β -subunit. Five positive recombinants were identified among 4×10^5 plaques, and two of them, CB-4 and CB-9 (Fig. 1), were characterized further. Transfer blot analysis of polyadenylated RNA isolated from cultured human skin fibroblasts indicated

Fig. 1. Restriction map and sequencing strategies for cDNA clones coding for chicken prolyl 4-hydroxylase β -subunit

The strategy for sequencing is indicated with arrows representing the directions and lengths of the sequencing runs. Synthetic oligonucleotide-primed DNA syntheses are indicated by O-1 to O-7. Sequences of the internal *Eco*RI site were obtained as described in the Materials and methods section. All *Eco*RI (E), *Kpn*I (K) and *Pst*I (Ps) restriction sites are shown. The *Pvu*II (P) and *Rsa*I (R) sites used for M13 cloning are also shown. TAA indicates the end of translation. Nucleotide (nt) and amino acid (aa) scales are shown at the bottom.

that the clone CB-4 cross-hybridized with the same human mRNA species (results not shown) that had previously been detected with the human β -subunit cDNA clones (Pihlajaniemi *et al.*, 1987). The mRNA species detected in 17-day chick-embryo calvaria was 2.7 kb in length (Fig. 2). Since RNA blot analysis suggested that CB-4 coded for the β -subunit of prolyl 4-hydroxylase, this clone was used for rescreening the chicken cDNA library under stringent conditions. An

Fig. 2. RNA transfer blot analysis

A 4 μ g portion of chicken calvaria polyadenylated RNA was hybridized with 32 P-labelled nick-translated cDNA clone CB-4. The migration of RNA marks (9.5–1.4 kb) is indicated.

additional 11 cDNA clones were obtained, of which clones CB-13 and CB-23-1 were characterized further, CB-13-1 and CB-13-2 being both derived from the same λ gt11 recombinant (Fig. 1). Owing to the presence of an internal *Eco*RI site in the λ gt11 cDNA insert, subcloning into the *Eco*RI site of pBR322 resulted in two different clones, CB-13-1 extending in the 5' direction of the internal *Eco*RI site and CB-13-2 extending in the 3' direction.

The identity of the cDNA clones was established by comparing the cDNA-derived amino acid sequences with protein sequences of the isolated chicken β -subunit. The first nine amino acid residues of the intact β -subunit were obtained by protein sequencing and were found to be identical with the cDNA-derived amino acid sequences (Fig. 3), except that the cDNA clone CB-13-1, coding for the most 5' sequences, lacked codons for the first three amino acid residues. The *N*-terminal amino acid sequence determined for one V8-proteinase-digest peptide of the chicken β -subunit was also identical with the cDNA-derived sequences (Fig. 3). Furthermore, the amino acid composition calculated from the cDNA-derived amino acid sequences corresponded well to the amino acid composition reported for the chicken prolyl 4-hydroxylase β -subunit (Berg *et al.*, 1979). The protein and cDNA-derived amino acid sequences show the chicken β -subunit to consist of 493 amino acid residues, the calculated M_r of the peptide being 55200. The cDNA clones covered 240 nucleotide residues of the 3' untranslated sequences.

Comparison of chicken prolyl 4-hydroxylase β -subunit sequences with those of other proteins

The chicken β -subunit cDNA sequences were 78% homologous to those of the previously published human β -subunit cDNA clones (Pihlajaniemi *et al.*, 1987) at the nucleotide level and 85% homologous at the amino acid level (Fig. 4a). The homology of the chicken β -subunit to the bovine polypeptide (published as a thyroid-hormone-

1	---	---	---	GAG	GAG	GAG	GAC	GGC	GTC	CTG	GTG	CTG	CGC	GCC	GCC	AAC	TTC	GAG	CAG	GCC	CTG	GCG	GCG	CAC	CGC	CAC	CTG	CTC	GTG	GAG	TTC	TAC
1	Glu	Pro	Leu	Glu	Glu	Glu	Asp	Gly	Val	Leu	Val	Leu	Arg	Ala	Ala	Asn	Phe	Glu	Gln	Ala	Leu	Ala	Ala	His	Arg	His	Leu	Leu	Val	Glu	Phe	Tyr
88	GCG	CCG	TGG	TGC	GGG	CAC	TGC	AAG	GCG	TTG	GCC	CCG	GAG	TAC	GCG	AAG	GCG	GCG	GCG	CAG	CTG	AAG	GCG	GAG	GGT	TCC	GAG	ATC	CGC	CTG	GCC	AAG
33	Ala	Pro	Trp	Cys	Gly	His	Cys	Lys	Ala	Leu	Ala	Pro	Glu	Tyr	Ala	Lys	Ala	Ala	Ala	Gln	Leu	Lys	Ala	Glu	Gly	Ser	Glu	Ile	Arg	Leu	Ala	Lys
184	GTG	GAC	GCC	ACG	GAG	GAG	GCG	GAG	CTG	GCG	CAG	CAG	TTC	GGC	GTG	CGC	GGG	TAC	CCC	ACC	ATC	AAG	TTC	TTC	CGC	AAC	GGG	GAC	AAG	GCC	GCG	CCC
65	Val	Asp	Ala	Thr	Glu	Glu	Ala	Glu	Leu	Ala	Gln	Gln	Phe	Gly	Val	Arg	Gly	Tyr	Pro	Thr	Ile	Lys	Phe	Phe	Arg	Asn	Gly	Asp	Lys	Ala	Ala	Pro
280	GCG	GAG	TAC	ACA	GCT	GGC	AGG	GAA	GCA	GAT	GAT	ATT	GTC	AGC	TGG	CTG	AAG	AAG	CGC	ACG	GGA	CCG	GCT	GCC	ACT	ACC	CTG	ACA	GAT	GCT	GCT	GCA
97	Arg	Glu	Tyr	Thr	Ala	Gly	Arg	Glu	Ala	Asp	Asp	Ile	Val	Ser	Trp	Leu	Lys	Lys	Arg	Thr	Gly	Pro	Ala	Ala	Thr	Thr	Leu	Thr	Asp	Ala	Ala	Ala
376	GCA	GAG	ACG	TTG	GTG	GAT	TCC	AGT	GAA	GTG	GTT	GTG	ATT	GGT	TTC	TTT	AAG	GAT	GTG	ACA	TCA	GAT	GCT	GCG	AAG	GAG	TTT	TTG	TTG	GCA	GCA	GAA
129	Ala	Glu	Thr	Leu	Val	Asp	Ser	Ser	Glu	Val	Val	Val	Ile	Gly	Phe	Phe	Lys	Asp	Val	Thr	Ser	Asp	Ala	Ala	Lys	Glu	Phe	Leu	Leu	Ala	Ala	Glu
472	TCC	GTG	GAT	GAC	ATT	CCT	TTT	GGG	ATT	TCT	TCC	AGC	GCT	GAC	GTC	TTC	TCC	AAG	TAC	CAG	CTT	AGC	CAA	GAT	GGA	GTG	GTC	CTT	TTC	AAG	AAG	TTT
161	Ser	Val	Asp	Asp	Ile	Pro	Phe	Gly	Ile	Ser	Ser	Ser	Ala	Asp	Val	Phe	Ser	Lys	Tyr	Gln	Leu	Ser	Gln	Asp	Gly	Val	Val	Leu	Phe	Lys	Lys	Phe
568	GAT	GAA	GGA	CGT	AAC	AAC	TTT	GAA	GGG	GAT	CTT	ACC	AAA	GAT	AAC	TTG	CTG	AAC	TTC	ATC	AAG	TCT	AAT	CAG	CTG	CCT	CTT	GTC	ATA	GAG	TTT	ACT
193	Asp	Glu	Gly	Arg	Asn	Asn	Phe	Glu	Gly	Asp	Leu	Thr	Lys	Asp	Asn	Leu	Leu	Asn	Phe	Ile	Lys	Ser	Asn	Gln	Leu	Pro	Leu	Val	Ile	Glu	Phe	Thr
664	GAG	CAG	ACT	GCT	CCT	AAA	ATC	TTT	GGA	GGA	GAG	ATC	AAG	ACT	CAC	ATT	CTG	CTG	TTC	TTA	CCA	AAA	AGT	GTG	TCT	GAT	TAT	GAA	GGG	AAA	TTG	GAC
225	Glu	Gln	Thr	Ala	Pro	Lys	Ile	Phe	Gly	Gly	Glu	Ile	Lys	Thr	His	Ile	Leu	Leu	Phe	Leu	Pro	Lys	Ser	Val	Ser	Asp	Tyr	Glu	Gly	Lys	Leu	Asp
760	AAC	TTC	AAG	ACT	GCA	GCT	GGA	AAC	TTC	AAA	GGG	AAG	ATC	CTG	TTC	ATC	TTC	ATA	GAC	AGT	GAC	CAC	AGT	GAT	AAT	CAA	AGG	ATT	CTG	GAG	TTT	TTT
257	Asn	Phe	Lys	Thr	Ala	Ala	Gly	Asn	Phe	Lys	Gly	Lys	Ile	Leu	Phe	Ile	Phe	Ile	Asp	Ser	Asp	His	Ser	Asp	Asn	Gln	Arg	Ile	Leu	Glu	Phe	Phe
856	GGT	CTA	AAG	AAA	GAA	GAA	TGT	CCA	GCT	GTA	CGT	TTG	ATA	ACA	TTA	GAA	GAA	GAA	ATG	ACC	AAA	TAC	AAA	CCT	GAG	TCA	GAT	GAC	CTC	ACA	GCA	GAC
289	Gly	Leu	Lys	Lys	Glu	Glu	Cys	Pro	Ala	Val	Arg	Leu	Ile	Thr	Leu	Glu	Glu	Glu	Met	Thr	Lys	Tyr	Lys	Pro	Glu	Ser	Asp	Asp	Leu	Thr	Ala	Asp
952	AAG	ATC	AAA	GAA	TTC	TGT	AAT	AAA	TTC	CTG	GAG	GGC	AAG	ATC	AAG	CCT	CAC	CTG	ATG	AGC	CAG	GAT	CTC	CCT	GAA	GAC	TGG	GAC	AAG	CAG	CCT	GTC
321	Lys	Ile	Lys	Glu	Phe	Cys	Asn	Lys	Phe	Leu	Glu	Gly	Lys	Ile	Lys	Pro	His	Leu	Met	Ser	Gln	Asp	Leu	Pro	Glu	Asp	Trp	Asp	Lys	Gln	Pro	Val
1048	AAA	GTT	CTT	GTT	GGG	AAG	AAT	TTT	GAG	GAA	GTT	GCT	TTT	GAT	GAG	AAT	AAG	AAT	GTC	TTT	GTG	GAG	TTC	TAT	GCC	CCC	TGG	TGT	GGT	CAC	TGT	AAG
353	Lys	Val	Leu	Val	Gly	Lys	Asn	Phe	Glu	Glu	Val	Ala	Phe	Asp	Glu	Asn	Lys	Asn	Val	Phe	Val	Glu	Phe	Tyr	Ala	Pro	Trp	Cys	Gly	His	Cys	Lys
1144	CAG	CTG	GCT	CCG	ATC	TGG	GAC	AAG	CTG	GGA	GAG	ACC	TAC	AGA	GAC	CAT	GAG	AAC	ATT	GTC	ATT	GCC	AAG	ATG	GAT	TCA	ACA	GCC	AAT	GAA	GTA	GAG
385	Gln	Leu	Ala	Pro	Ile	Trp	Asp	Lys	Leu	Gly	Glu	Thr	Tyr	Arg	Asp	His	Glu	Asn	Ile	Val	Ile	Ala	Lys	Met	Asp	Ser	Thr	Ala	Asn	Glu	Val	Glu
1240	GCA	GTG	AAA	ATT	CAC	AGC	TTC	CCC	ACG	CTC	AAG	TTC	TCT	CCT	GCA	GGC	TCT	GGC	AGA	AAC	GTA	ATT	GAC	TAT	AAT	GGA	GAG	AGG	ACA	CTA	GAA	GGC
417	Ala	Val	Lys	Ile	His	Ser	Phe	Pro	Thr	Leu	Lys	Phe	Phe	Pro	Ala	Gly	Ser	Gly	Arg	Asn	Val	Ile	Asp	Tyr	Asn	Gly	Glu	Arg	Thr	Leu	Glu	Gly
1336	TTC	AAA	AAA	TTC	TTG	GAG	AGT	GGA	GGC	CAG	GAT	GGA	GCA	GCA	GCT	GAT	GAT	GAT	CTG	GAA	GAT	CTG	GAG	ACA	GAT	GAA	GAA	ACG	GAC	CTT	GAG	GAA
449	Phe	Lys	Lys	Phe	Leu	Glu	Ser	Gly	Gly	Gln	Asp	Gly	Ala	Ala	Ala	Asp	Asp	Asp	Leu	Glu	Asp	Leu	Glu	Thr	Asp	Glu	Glu	Thr	Asp	Leu	Glu	Glu
1432	GGT	GAT	GAT	GAT	GAG	CAA	AAA	ATC	CAA	AAG	GAT	GAA	TTG	TAA	AGAGATGACTGATCTACATCACCAAAATCAGACACTAAATGTGGTGTCTGTTACGCAACCCAATGAGTCA																	
481	Gly	Asp	Asp	Asp	Glu	Gln	Lys	Ile	Gln	Lys	Asp	Glu	Leu	***																		
1545	GAAACCCATTAAAGATTAAAGCAGAAAGGTGAATGACTGGAAATCCAGGGATTGGCTCTTAAAGTAACGTTCCGCCCTCACTTGCTTTTACACTTCACTATCTTTTTCTTTCTTTTCACTTTTGAGAA																															
1672	GGGATCTATCACTAGGTAGCCAGCCG																															

Fig. 3. Nucleotide and deduced amino acid sequence of cDNAs for chicken prolyl 4-hydroxylase β -subunit

First line, nucleotide sequence of the cDNAs for the chicken β -subunit; second line, the cDNA deduced amino acid sequence. The amino acid residues obtained by sequencing of the *N*-terminal end and one *Staphylococcus aureus*-V8-proteinase-digest peptide are underlined. Numbering of the nucleotides begins with the extreme 5' nucleotide of the β -subunit cDNA clones, and numbering of the amino acid residues with the first residue of the processed β -subunit. Symbols: ---, codons not present in the cDNA; ***, stop codon of translation.

binding protein; Yamauchi *et al.*, 1987) and rat polypeptide (published as protein disulphide-isomerase; Edman *et al.*, 1985) was also 85 % at the amino acid level (Fig. 4a). The polypeptides were of slightly different size in the four species, being 491 amino acid residues in the human and bovine proteins, 493 residues in the chicken and 489 residues in the rat.

Primary structure comparisons indicated that 82 amino acid residues in the chicken sequence are different from those in one or more of the other species (Fig. 4a). Of these residues 48 are different only in the chicken, being identical in the human, bovine and rat proteins.

The two proposed active sites of protein disulphide-isomerase (Hillson & Freedman, 1979; Edman *et al.*, 1985), the two Trp-Cys-Gly-His-Cys-Lys sequences, are located in highly conserved regions (Figs. 4a and 4b). The first of these is found in the middle of a stretch of 24 amino acid residues that are identical in all four species, and the second is in the middle of 29 identical amino acid residues.

Thioredoxins have been found to be homologous to regions a and a' of the human and rat proteins (Edman *et al.*, 1985; Pihlajaniemi *et al.*, 1987). Comparison of the chicken β -subunit with one of the bacterial thioredoxins,

Fig. 4. Comparison of cDNA derived amino acid sequences for the chicken prolyl 4-hydroxylase β -subunit with comparable sequences from human, bovine and rat proteins

(a) First, second and fourth lines, bovine (Yamauchi *et al.*, 1987), human (Pihlajaniemi *et al.*, 1987) and rat (Edman *et al.*, 1985) amino acid sequences respectively where they differ from the chicken sequence. Some minor corrections have been introduced into the human sequences during characterization of the corresponding gene (Tasanen *et al.*, 1988). Third line, chicken amino

Vol. 256

Prolyl 4-hydroxylase β -subunit

	Residues	
Human, region a	(27-46)	L L V E F Y A P W C G H C K A L A P E Y
Bovine, region a	(27-46)	L L V E F Y A P W C G H C K A L A P E Y
Rat, region a	(26-45)	L L V E F Y A P W C G H C K A L A P E Y
Chicken, region a	(27-46)	L L V E F Y A P W C G H C K A L A P E Y
Human, region a'	(371-390)	V F V E F Y A P W C G H C K Q L A P I W
Bovine, region a'	(371-390)	V F V E F Y A P W C G H C K Q L A P I W
Rat, region a'	(370-389)	V F V E F Y A P W C G H C K Q L A P I W
Chicken, region a'	(371-390)	V F V E F Y A P W C G H C K Q L A P I W

Thioredoxins

	Residues	
Rat liver		X X X X X X X X X C G P C K M X X X X X
Spinach chloroplasts	(28-47)	S M V D F W A P W C G P C K L I A P V I
Yeast thioredoxin I		X X X X X Y A T W C G P C K X X X X X
Yeast thioredoxin II		X X X X X F A T W C G P C K M I A P M I
<i>Chlorobium thiosulfatophilum</i>	(23-42)	V L V D F W A S W C G P C M M I L G P V L
<i>Chromatium vinosum</i>	(23-42)	V L V D Y W A D W C G P C K M I L A P V L
<i>Corynebacterium nephridii</i> C-1	(21-40)	V V V D F W A E W C G P C K M I A P A L
<i>Corynebacterium nephridii</i> C-2	(39-58)	V L V D F W A G W C A P C K A I A P V L
<i>Cyanobacterium anabaena</i>	(22-41)	V L V D F W A P W C G P C R M V A P V V
<i>Escherichia coli</i> M10424	(23-42)	I L V D F W A E W C G P C K M I A P I L
<i>Escherichia coli</i> K02845	(23-42)	I L V D F W A E W C G P C K M I A P I L
<i>Escherichia coli</i> M12779	(23-42)	I L V D F W A E W C G P C K M I A P I L
Bacteriophage T4	(5-24)	Y G Y D S N I H K C V Y C D N A K R L L

Fig. 5. Comparison of the chicken prolyl 4-hydroxylase β -subunit and corresponding proteins from human, bovine and rat sources with various thioredoxins

A stretch of amino acid residues from both regions a and a' of the β -subunit is compared with the thioredoxins. The human β -subunit sequences are from Pihlajaniemi *et al.* (1987) and the bovine and rat sequences homologous to the β -subunit from Yamauchi *et al.* (1987) and Edman *et al.* (1985) respectively. The amino acid sequences for thioredoxins from rat liver are from Luthman & Holmgren (1982), those for spinach chloroplasts from Maeda *et al.* (1986), those for yeast thioredoxins I and II from Hall *et al.* (1971), those for *Chlorobium thiosulfatophilum* from Mathews *et al.*, (1987), those for *Chromatium vinosum* from Johnson & Biemann (1987), those for *Corynebacterium nephridii* C-1 and C-2 from Meng & Hogenkamp (1981) and Lim *et al.* (1987) respectively, those for *Cyanobacterium anabaena* from Gleason *et al.* (1985), those for *Escherichia coli* M10424, K02845 and M12779 from Lim *et al.* (1985), Wallace & Kushner (1984) and Matsumoto *et al.* (1986) respectively, and those for bacteriophage T4 from LeMaster (1986). The amino acids are indicated with the one letter code, X denoting a non-determined amino acid and I/L the presence of either leucine or isoleucine. Amino acid residues identical with region a are stippled, and amino acid residues identical only with region a' are boxed.

that of *Corynebacterium nephridii* (Meng & Hogenkamp, 1981), which is 105 amino acid residues in length, indicated 35% homology within chicken amino acid residues 7-115 and 30% homology within chicken amino acid residues 350-453. The best-conserved regions of the β -subunit and various thioredoxins were compared (Fig. 5) between sequences from vertebrates (Luthman & Holmgren, 1982; Edman *et al.*, 1985; Pihlajaniemi *et al.*, 1987; Yamauchi *et al.*, 1987), spinach (Maeda *et al.*, 1986), yeast (Hall *et al.*, 1971) and bacteria (Meng & Hogenkamp, 1981; Wallace & Kushner, 1984; Gleason *et al.*, 1985; Lim *et al.*, 1985, 1986, 1987; Matsumoto *et al.*, 1986; Mathews *et al.*, 1987; Johnson & Biemann, 1987). A consensus sequence Leu/Val(Ile)-Leu/Phe-Val-Glu/Asp-Thr-Tyr/Trp-Ala-Pro(in β -subunit, otherwise variable)-Trp-Cys-Gly-His/Pro-Cys-Lys-variable-hydrophobic-Ala-Pro is seen (Fig. 5). This stretch of amino acid residues contains Trp-Cys-Gly-His/Pro-Cys-Lys, the presumed active site of thioredoxins and protein disulphide-isomerase (Edman *et al.*, 1985). The com-

parable sequence in bacteriophage T4 thioredoxin (LeMaster, 1986) differs extensively from the other proteins, only the two cysteine residues being conserved.

Another highly conserved region of 100 amino acid residues was found in the middle of the β -subunit polypeptide. This can possibly be divided into three subregions, a 37-amino-acid-residue sequence (residues 215-251) that is completely conserved, a 13-amino-acid-residue sequence with five differences (residues 252-264), and a 50-amino-acid-residue sequence (residues 265-314) in which a single serine residue in the chicken is replaced by a threonine residue in the other species. No data are currently available on the functional significance of this middle area of the polypeptide.

Munro & Pelham (1987) have proposed that proteins retained within the lumen of the endoplasmic reticulum have a C-terminal sequence of Lys-Asp-Glu-Leu. This sequence is found as the C-terminus of the β -subunit polypeptide in all four species studied here. Currently it is not known whether the α -subunit of prolyl 4-

hydroxylase also has this C-terminal sequence or whether one function of the β -subunit in this enzyme is to be responsible for the retention of the enzyme tetramer within the lumen of the endoplasmic reticulum. The two amino acid residues preceding this tetrapeptide sequence in the chicken polypeptide differ from those in the other species, supporting the suggestion that the sequence requirement for retention within the endoplasmic reticulum is completely fulfilled by the four amino acid residues (Munro & Pelham, 1987). No similar conservation was found within the four N-terminal amino acid residues of the polypeptide, suggesting that the N-terminus of the polypeptide may not be involved in any functionally important domain.

Note added in proof (received 30 September 1988)

While this paper was in the press the complete cDNA-derived amino acid sequence was reported for a phosphoinositide-specific phospholipase C (Bennett *et al.*, 1988). This phospholipase has a high sequence homology to the multifunctional β -subunit studied here.

We gratefully acknowledge Phil.Cand. Tarja Helaakoski for her help in isolating the β -subunit for protein sequencing. This work was supported by grants from the Research Councils for Medicine and the Natural Sciences within the Academy of Finland, and by the Emil Aaltonen Foundation and the Finnish Cultural Fund.

REFERENCES

- Aebersold, R. H., Teplow, D. B., Hood, L. E. & Kent, S. B. H. (1986) *J. Biol. Chem.* **261**, 4229–4238
- Bennett, C. F., Balcarek, J. M., Varrichio, A. & Crooke, S. (1988) *Nature (London)* **334**, 268–270
- Berg, R. A., Kedersha, N. L. & Guzman, N. A. (1979) *J. Biol. Chem.* **254**, 3111–3118
- Cardinale, G. J. & Udenfriend, S. (1974) *Adv. Enzymol. Relat. Areas Mol. Biol.* **41**, 245–300
- Cheng, S.-Y., Gong, Q.-H., Parkison, C., Robinson, E. A., Appella, E., Merlino, G. T. & Pastan, I. (1987) *J. Biol. Chem.* **262**, 11221–11227
- Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A. & Rutter, W. J. (1985) *Nature (London)* **317**, 267–270
- Freedman, R. (1987) *Nature (London)*, **329**, 196–197
- Freedman, R. B. & Hillson, D. A. (1980) in *The Enzymology of Post-Translational Modification of Proteins* (Freedman, R. B. & Hawkins, H. C., eds.), pp. 167–212, Academic Press, London
- Gleason, F. K., Whittaker, M. M., Holmgren, A. & Jörnvall, H. (1985) *J. Biol. Chem.* **260**, 9567–9573
- Hall, D. E., Baldesten, A., Holmgren, A. & Reichard, P. (1971) *Eur. J. Biochem.* **23**, 328–335
- Hillson, D. A. & Freedman, R. B. (1979) *Biochem. Soc. Trans.* **7**, 573–574
- Hillson, D. A., Lambert, N. & Freedman, R. B. (1984) *Methods Enzymol.* **107**, 281–294
- Johnson, R. S. & Biemann, K. (1987) *Biochemistry* **26**, 1209–1214
- Kivirikko, K. I. & Myllylä, R. (1980) in *The Enzymology of Post-Translational Modification of Proteins* (Freedman, R. B. & Hawkins, H. C., eds.), pp. 53–104, Academic Press, London
- Kivirikko, K. I. & Myllylä, R. (1985) *Ann. N.Y. Acad. Sci.* **460**, 187–201
- Kivirikko, K. I. & Myllylä, R. (1987) *Methods Enzymol.* **144**, 96–114
- Koivu, J., Myllylä, R., Helaakoski, T., Pihlajaniemi, T., Tasanen, K. & Kivirikko, K. I. (1987) *J. Biol. Chem.* **262**, 6447–6449
- LeMaster, D. M. (1986) *J. Virol.* **59**, 759–760
- Lim, C.-J., Geraghty, D. & Fuchs, J. A. (1985) *J. Bacteriol.* **163**, 311–316
- Lim, C.-J., Gleason, F. K., Fuchs, J. A. (1986) *J. Bacteriol.* **168**, 1258–1264
- Lim, C.-J., Fuchs, J. A., McFarlen, S. C. & Hogenkamp, H. P. C. (1987) *J. Biol. Chem.* **262**, 12114–12119
- Luthman, M. & Holmgren, A. (1982) *Biochemistry* **21**, 6628–6633
- Maeda, K., Tsugita, A., Dalzoppo, D., Vilbois, F. & Schurmann, P. (1986) *Eur. J. Biochem.* **154**, 197–203
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Mathews, W. R., Johnson, R. S., Cornwell, K. L., Johnson, T. C., Buchanan, B. B. & Biemann, K. (1987) *J. Biol. Chem.* **262**, 7537–7545
- Matsumoto, Y., Shigesada, K., Hirano, M. & Imai, M. (1986) *J. Bacteriol.* **166**, 945–958
- Meng, M. & Hogenkamp, H. P. C. (1981) *J. Biol. Chem.* **256**, 9174–9182
- Messing, J. (1983) *Methods Enzymol.* **101**, 20–78
- Munro, S. & Pelham, H. (1987) *Cell* **48**, 899–907
- Pihlajaniemi, T. & Myers, J. C. (1987) *Methods Enzymol.* **145**, 213–222
- Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllylä, R., Huhtala, M.-L., Koivu, J. & Kivirikko, K. I. (1987) *EMBO J.* **6**, 643–649
- Prockop, D. J., Berg, R. A., Kivirikko, K. I. & Uitto, J. (1976) in *Biochemistry of Collagen* (Ramachandran, G. N. & Reddi, A. H., eds.), pp. 163–273, Plenum Press, New York
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Tasanen, K., Parkkonen, T., Chow, L., Kivirikko, K. I. & Pihlajaniemi, T. (1988) *J. Biol. Chem.*, in the press
- Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201–5205
- Wallace, B. J. & Kushner, S. R. (1984) *Gene* **32**, 399–408
- Yamauchi, K., Yamamoto, T., Hayashi, H., Koya, S., Takikawa, H., Toyoshima, K. & Horiuchi, R. (1987) *Biochem. Biophys. Res. Commun.* **146**, 1485–1492

Received 16 June 1988; accepted 13 July 1988