Structure of a cDNA for the Proα2 Chain of Human Type I Procollagen. Comparison with Chick cDNA for Proα2(I) Identifies Structurally Conserved Features of the Protein and the Gene[†]

Michael P. Bernard, Jeanne C. Myers, Mon-Li Chu, Francesco Ramirez, Eric F. Eikenberry, and Darwin J. Prockop*

ABSTRACT: Nucleotide sequences were determined for cloned cDNAs encoding for more than half of the proα2 chain of type I procollagen from man. Comparisons with previously published data on homologous cDNAs from chick embryos made it possible to examine evolution of the gene in two species which have diverged for 250–300 million years. The amino acid sequence of the α-chain domain supported previous indications that there is a strong selective pressure to maintain glycine as every third amino acid and to maintain a prescribed distribution of charged amino acids. However, there is little apparent selective pressure on other amino acids. The amino acid sequence of the C-propeptide domain showed less divergence than the α-chain domain. The 5′ end or N terminus of the human C-propeptide, however, contained an insert of 12 bases coding for 4 amino acids not found in the chick C-propeptide. About 100 amino acid residues from the N terminus, two residues found in the chick sequence were missing from the human. In the second half of the C-propeptide, there was complete conservation of a 37 amino acid sequence and conservation of 50 out of 51 amino acids in the same region, an observation which suggested that the region serves some special purpose such as directing the association of one proα2(I) C-propeptide with two proα1(I) C-propeptides so as to produce the heteropolymeric structure of type I procollagen. In addition, comparison of human and chick DNAs for proα2(I) revealed three different classes of conservation of nucleotide sequence which have no apparent effect on the structure of the protein: a preference for U on the third base position of codons for glycine, proline, and alanine; a high degree of nucleotide conservation in the 51 amino acid highly conserved region of the C-propeptide; a high degree of nucleotide conservation in the 3′-noncoding region. These three classes of nucleotide conservation may reflect unusual features of collagen genes, such as their high GC content or their highly repetitive coding sequences.

Collagen provides the fibrous network which maintains the structural integrity of most tissues in vertebrates and in many other multicellular organisms. The monomer of type I collagen, the most abundant form of the protein, consists of two α1(I) chains and one α2(I) chain with differing but highly homologous primary structures [for reviews, see Piez (1976, 1980), Bornstein & Traub, (1979), and Prockop et al. (1979)]. The complete sequence of the approximately 1050 amino acid residues in the α1(I) chain was established by Edman degradation of peptide fragments of type I collagen from chicken skin (Highberger et al., 1982) and calf skin (Hofmann et al., 1978). The same technique was used to establish substantial portions of the primary structure of the α1(I) chain from rat [see Piez (1976) and Bornstein & Traub (1979)] and parts of the structure of the α2(I) chain from several species (Kang & Gross, 1970; Highberger et al., 1975; Fietzek & Kühn, 1973; Fietzek et al., 1974a,b; Fietzek & Rexrodt, 1975; Dixit et al., 1977a,b, 1979).

More recently, nucleotide sequencing of cloned DNAs (Fuller & Boedtker, 1981; Yamamoto et al., 1980; Showalter et al., 1980; Olsen & Dickson, 1981), or a combination of nucleotide and amino acid sequencing (Pesciotta et al., 1981; Dickson et al., 1981), was used to determine the structure of the C-terminal region of α1(I) and α2(I) chains of chick embryo collagen as well as the structure of the C-terminal propeptide chains which are found on the proα1(I) and proα2(I) chains of type I procollagen and which are cleaved in the conversion of procollagen to collagen. Part of the structure of the mouse proα1(I) chain was obtained by nucleotide sequencing of a genomic DNA (Monson & McCarthy, 1981).

We here report the nucleotide sequences of cloned cDNAs coding for more than half of the proα2 chain of type I procollagen from man. Comparison of these data with previously published data on homologous cDNAs from chick embryos (Fuller & Boedtker, 1981) makes it possible to examine evolution of the coding sequences of the gene in two divergent species. With this comparison, we have identified highly conserved features of both the protein and the gene.

Materials and Methods

Enzymes and Other Materials. Restriction endonucleases were purchased from New England Biolaboratories and BRL, Inc. T4 polynucleotide kinase was purchased from BRL. Labeled nucleotides were purchased from Amersham Corp.

DNA Sequence Determination. DNA sequencing was carried out essentially as described by Maxam & Gilbert (1980). The 5′ ends of restriction fragments were labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. Either the labeled fragments were restricted with a second restriction endonuclease or the strands were separated by electrophoresis. The polyacrylamide gels for DNA sequencing were 0.45 mm thick by 40 cm long and were run at 1100–1600 V. One 8% gel and one or two 5% gels were used for each determination. For determination of the first few nucleotides in a fragment, a 20% gel 0.8 mm thick was employed in order to minimize salt effects.

[†] From the Departments of Biochemistry and Obstetrics/Gynecology, University of Medicine and Dentistry of New Jersey-Rutgers Medical School, Piscataway, New Jersey 08854. *Received May 6, 1982; revised manuscript received November 12, 1982.* This work was supported in part by National Institutes of Health Grant AM-16516 and grants from The March of Dimes-Birth Defects Foundation and the Lalor Foundation.

0006-2960/83/0422-1139$01.50/0

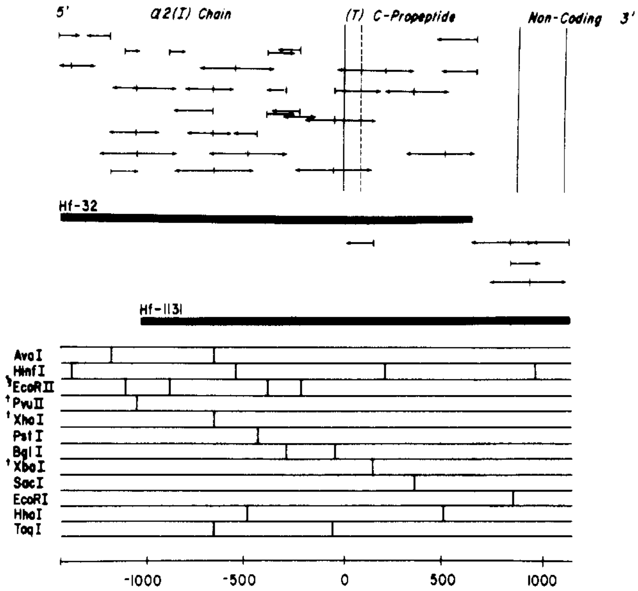

FIGURE 1: Restriction map and sequencing strategy for two clones (Hf-32 and HF-1131) of cDNA for the proα2(I) chain of normal human type I procollagen. Horizontal lines and arrows indicate fragments sequenced. Symbols: (T), C-terminal telopeptide; §, fragments obtained with a partial digestion only of Hf-32 with EcoRII; †, restriction fragments obtained with the same for endonucleases from both Hf-32 and Hf-1131. Nucleotides (scale at bottom) are numbered as suggested by Fuller & Boedtker (1981) with the last nucleotide coding for the last residue of the α-chain domain designated "-1".

Experiments involving recombinant DNA were performed in P1-EK1 containment in accordance with the National Institutes of Health guidelines.

Results

Sequencing Strategy. Most of the nucleotide sequences reported here were obtained from Hf-32, a 2.2-kb cloned cDNA for the 3'- or carboxy-terminal portion of the proα2(I) chain from normal human skin fibroblasts (Myers et al., 1981). A fraction of the sequences, including short overlaps of Hf-32, was obtained from additional cloned cDNA (Hf-1131) prepared with the same procedures and with RNA from the same normal human skin fibroblasts.

A restriction map was developed by cleaving the cloned cDNAs with a series of restriction endonucleases (Figure 1). Both clones gave homologous fragments after cleavage with PvuII, XhoI, and XbaI. Also, the same sequences were obtained by analysis of two overlapping fragments from the two clones (Figure 1). Therefore, there was no evidence of rearrangements of the nucleotide sequences in the course of preparing the clones as was found with several cloned cDNAs for chick proα1(I) and proα2(I) chains (Lehrach et al., 1978, 1979).

The two cloned cDNAs were sequenced with the Maxam & Gilbert (1980) procedure. As indicated in Figure 1, both strands were sequenced for about 73% of the nucleotides contained in the two clones. The repetitive nature of sequences coding for the α-chain domain made it difficult to obtain fragments of appropriate size for sequencing both strands coding for this region. Therefore, many of the fragments were sequenced several times in the same direction.

Comparison of the Amino Acid Sequence of the α-Chain Domain in Man and Chick. The 5' end of Hf-32 extends well into the α-chain domain of the proα2(I) chain (Figure 1). The amino acid sequence for most of the chick α2(I) chain was determined by protein analysis (Dixit et al., 1977a,b, 1979). The amino acid sequence for the last 300 residues of the chick α2(I) chain was also obtained from cloned cDNAs (Fuller & Boedtker, 1981). Therefore, it was possible to compare almost all the amino acid sequence of the human α2(I) chain encoded by Hf-32 (Figures 2 and 3) with the corresponding amino acid sequences for chick α2(I).

Table I: Comparison of Amino Acid and Nucleotide Differences in Proα2(I) between Man and Chick

		nucleotides (uncorrected values)	
	amino acid replacements	replacement sites	silent sites
α2(I)[a]	108 (490)	65 (489)	74 (224)
α2(I) *less* glycines[b]	103 (317)	59 (317)	43 (141)
telopeptide	6 (15)		
C-propeptide	36 (243)	44 (628)	72 (250)
C-propeptide conserved region[c]	0 (37)	0 (97)	9 (38)
3'-noncoding region			61 (218)

[a] Includes telepeptide. [b] Excludes Gly residues in every third position of the α-chain domain and the 15 amino acid residues of the C-terminal telopeptide. [c] Amino acid residues 157–194 (nucleotides 472–582) in Figure 3.

Alignment with the amino acid sequences of the chick α2(I) chain (Dixit et al., 1977a,b, 1979) indicated that the Met residue marking the end of α2-CB3[1] corresponds to nucleotides −956 to −954 (Figure 2). The order of the first four CB peptides of chick α2(I) is 1-0-4-2-3, and the lengths are 14, 3, 321, 30, and 338 residues, respectively (Dixit et al., 1979). The first peptide, α2-CB1, includes 11 amino acids of the telopeptide. Therefore, if the residues of the α2 chain are numbered beginning with the first Gly in a -Gly-X-Y- triplet, the Met residue at the end of α2-CB3 is number 695 and the first three coding nucleotides of Hf-32 correspond to amino acid residue 533 (Figure 2).

Comparison of the α2(I) chains indicated that 108 of 490 defined amino acid residues differed between the two species (Table I). None of the differences involved Gly in the first position of the repeating triplets of -Gly-X-Y-. Most of the differences represented conservative mutations in the sense that they did not alter the charge distribution along the chain. However, there were two differences in charged residues: Glu in position 569 of the human amino acid sequence (nucleotides −1334 to −1332 in Figure 2) was Ala in the chick, and Glu in position 803 of the human sequence (nucleotides −632 to −630 in Figure 2) was Gln in the chick sequence.

The rate of amino acid replacement was 22% (Table II), a value intermediate between the value of 31% seen in comparing β-globin between chick and man and the value of 14% seen in insulin between the same two species. The amino acid replacement rate increased from 22 to 32% (Table II) when the Gly residues in every third position of the α chain and the telopeptide were excluded from the comparison.

The amino acid sequence obtained from Hf-32 includes the cleavage site for vertebrate collagenase (between nucleotides −714 and −715 or amino acid residues 775 and 776 in Figure 2). The sequence -Gly-Thr-Pro-Gly-Pro-Gln-Gly- to the left of the cleavage site is identical with the equivalent sequences in chick α2(I) (Dixit et al., 1979), in chick α1(I) (Highberger et al., 1975), and in calf α1(I) (Wendt et al., 1972) chains. The sequence to the right of the cleavage, -Leu-Leu-Gly-Ala-Pro-, is similar to the equivalent sequence of -Ile-Leu-Gly-Ala-Pro- in the chick α2(I) chain (Dixit et al., 1979) but differs from the sequence of -Ile-Ala-Gly-Gln-Arg- found in

[1] Abbreviation: CB, cyanogen bromide.

```
                                                                            -1442

                                                                            GAA CCT GGT GTG GTT GGT GCT GTG GGC ACT GCT
                                                                            GLU PRO GLY VAL VAL GLY ALA VAL GLY THR ALA
                                                                                        ASN         PRO ALA     ALA PRO

-1412

GGT CCA TCT GGT CCT AGT GGA CTC CCA GGA GAG AGG GGT GCT GCT GGC ATA CCT GGA GGC AAG GGA GAA AAG GGT GAA CCT GGT CTC AGA
GLY PRO SER GLY PRO SER GLY LEU PRO GLY GLU ARG GLY ALA ALA GLY ILE PRO GLY GLY LYS GLY GLU LYS GLY GLU PRO GLY LEU ARG
        ALA         PRO     ILE                     VAL         VAL                                     ALA

-1322

GGT GAA ATT GGT AAC CCT GGC AGA GAT GGT GCT CGT GGT GCT CAT GGT GCT GTA GGT GCC CCT GGT CCT GCT GGA GCC ACA GGT GAC CGG
GLY GLU ILE GLY ASN PRO GLY ARG ASP GLY ALA ARG GLY ALA HIS GLY ALA VAL GLY ALA PRO GLY PRO ALA GLY ALA THR GLY ASP ARG
    ASP THR     ALA THR                             LEU PRO         ILE                                 GLY ALA

-1232

GGC GAA GCT GGG GCT GCT GGT CCT GCT GGT CCT GCT GGT CCT CGG GGA AGC CCT GGT GAA CGT GGC GAG GTC GGT CCT GCT GGC CCC AAC
GLY GLU ALA GLY ALA ALA GLY PRO ALA GLY PRO ALA GLY PRO ARG GLY SER PRO GLY GLU ARG GLY GLU VAL GLY PRO ALA GLY PRO ASN
        GLY     PRO                                 PHE         ILE                         PRO         VAL         SER

-1142

GGA TTT GCT GGT CCG GCT GGT GCT GCT GGT CAA CCG GGT GCT AAA GGA GAA AGA GGA GCC AAA GGG CCT AAG GGT GAA AAC GGT GTT GTT
GLY PHE ALA GLY PRO ALA GLY ALA ALA GLY GLN PRO GLY ALA LYS GLY GLU ARG GLY ALA LYS GLY PRO LYS GLY GLU ASN GLY VAL VAL
                    PRO                                                     PRO                         THR     PRO THR

-1052

GGT CCC ACA GGC CCC GTT GGA GCT GCT GGC CC* *** GGT CCA AAT GGT CCC CCC GGT CCT GCT GGA AGT CGT GGT GAT GGA GGC CCC CCT
GLY PRO THR GLY PRO VAL GLY ALA ALA GLY PRO *** GLY PRO ASN GLY PRO PRO GLY PRO ALA GLY SER ARG GLY ASP GLY GLY PRO PRO
    ALA ILE         ILE         SER         PRO         VAL     ALA ALA                 PRO             ALA

-962

GGT ATG ACT GGT TTC CCT GGT GCT GCT GGA CGG ACT GGT CCC CCA GGA CCC TCT GGT ATT TCT GGC CCT CCT GGT CCC CCT GGT CCT GCT
GLY MET THR GLY PHE PRO GLY ALA ALA GLY ARG THR GLY PRO PRO GLY PRO SER GLY ILE SER GLY PRO PRO GLY PRO PRO GLY PRO ALA
                                            VAL     THR             ALA         THR

-872

GGG AAA GAA GGG CTT CGT GGA CC* CGA GG* GAC CAA GGA CCA GCA GGC CGA CCT GGA GAA GTA GGA GCA CCG GGT CCC CCT GGC TTC GCT
GLY LYS GLU GLY LEU ARG GLY PRO ARG GLY ASP GLN GLY PRO ALA GLY ARG PRO GLY GLU VAL GLY ALA PRO GLY PRO PRO GLY PHE ALA
        ASP     PRO         LEU             VAL         VAL         THR         GLN     ILE ALA

-782

GGT GAG AAG GGT CCC TCT GGA GAG GCT GCT ACT GCT GGA CCT CCT GGC ACT CCA GGT CCT CAG GGT CTT CTT GGT GCT CCT GGT ATT CTC
GLY GLU LYS GLY PRO SER GLY GLU ALA GLY THR ALA GLY PRO PRO GLY THR PRO GLY PRO GLN GLY LEU LEU GLY ALA PRO GLY ILE LEU
                                        ALA                                             ILE ALA
                                                                                       ↑
-692

GGT CTC CCT GGC TCG AGA GGT GAA CGT GGT CTA CCT GGT GTT GCT GGT GCT GTG GGT GAA GCT GGT CCT CTT GGC ATT GCC GGC CCT CCT
GLY LEU PRO GLY SER ARG GLY GLU ARG GLY LEU PRO GLY VAL ALA GLY ALA VAL GLY GLU PRO GLY PRO LEU GLY ILE ALA GLY PRO PRO
                                                    ILE             THR     GLN
```

FIGURE 2: Comparison of the amino acid sequences for proα2(I) between man and chick. The first line indicates the nucleotide sequence obtained from the clone of human cDNA (Hf-32). The second line is the amino acid sequence encoded by the human cDNA beginning with amino acid residue 533 and ending with amino acid residue 813. The third line indicates amino acid residues in the sequence for chick α2(I) where the amino acid residue obtained by protein analysis (Dixit et al., 1977a,b, 1979) differs from the human amino acid sequence. The arrow indicates the site at which the α2(I) chain is cleaved by vertebrate collagenase (between amino acid residues 775 and 776). Underlining indicates six amino acids at the end of the sequence have not yet been determined for chick α2(I). Asterisks indicate nucleotides not clearly defined. One additional nucleotide, G, was present at the 5′ end of the clone (Hf-32).

Table II: Comparison of Amino Acid and Nucleotide Differences in Proα2(I) between Man and Chick

	amino acid replacements (% difference)	nucleotide changes (% corrected divergence)[a]	
		replacement sites	silent sites
α2(I) chain[b]	22	16	60
α2(I) *less* glycines[c]	32	23	59
C-propeptide	15	8	65
C-propeptide conserved region	0	0	75
3′-noncoding region[d]			35 ± 5 (SD)
β-globin[e]	31	23	70
insulin[e]	14	8	122

[a] Since divergence from a common sequence is assumed, the maximal value for corrected divergence is 200%. Values calculated as described by Perler et al. (1980). [b] Includes telopeptide. [c] Excludes Gly residues in every third position in the α-chain domain and C-terminal telopeptide. [d] As noted in the text, corrected divergence for the 3′-noncoding region was calculated after alignment of the human and chick nucleotides. The standard deviation (SD) was calculated by the method of Kimura & Ohta (1972). [e] Values from Efstratiadis et al. (1980).

the α1(I) chain of both chick (Highberger et al., 1975) and calf (Wendt et al., 1972).

Comparison of Nucleotide Sequences for the α-Chain Domain in Man and Chicken. About two-thirds of the nucleotide sequences obtained here overlapped nucleotide sequences determined for chick proα2(I) (Fuller & Boedtker, 1981). Therefore, the nucleotide sequences in the two species were compared (Figure 3).

For comparison of the nucleotide sequences, we employed the procedure developed by Perler et al. (1980) for comparison of proinsulin and globin genes among different species. The procedure, in brief, involves evaluation of each nucleotide in terms of whether a change in the base present at a given site will be silent in its effect on the structure of the protein or whether it will produce a replacement of an amino acid. Then the nucleotide sequences of two species are compared to determine the number of silent site and replacement site mutations which have occurred during evolution of these two species. Finally, the fractional silent site and replacement site mutations are corrected for the average probability that more than one mutation occurred at a given nucleotide site.

When the α2(I) chains were compared, there were 65 nucleotide mutations in replacement sites and 74 mutations in silent sites (Table I). The corrected value for divergence in replacement sites was 16% and the corrected value for silent sites was 60% (Table II). If only the non-Gly residues in the α-chain domain were considered, the values were 23 and 59, respectively. The latter values are similar to those for β-globin from the same two species.

Comparison of the Telopeptide Domains between Man and Chick. The C-terminal telopeptide of the human proα2(I) chain was similar in length to the chick proα2(I) (Fuller &

```
-602
   T   T           C T       C           TC        T CCT   T                                           T       A   T
 GGG GCC CGT GGT CCT CCT GGT GCT GTG GGT AGT CCT GGA GTC AAC GGT GCT CCT GGT GAA GCT GGT CGT GAT GGC AAC CCT GGG AAC GAT
-GLY-ALA-ARG-GLY-PRO-PRO-GLY-ALA-VAL-GLY-SER-PRO-GLY-VAL-ASN-GLY-ALA-PRO-GLY-GLU-ALA-GLY-ARG-ASP-GLY-ASN-PRO-GLY-ASN-ASP
                     SER     PRO                     PRO

-512
       T       C   T         GCT   T   C TT        T       T     GCT       T   C CCA         AG          TTG       T
 GGT CCC CCA GGT CGC GAT GGT CAA CCC GGA CAC AAG GGA GAG CGC GGT TAC CCT GGC AAT ATT GGT CCC GTT GGT GCT GCA GGT GCA CCT
-GLY-PRO-PRO-GLY-ARG-ASP-GLY-GLN-PRO-GLY-HIS-LYS-GLY-GLU-ARG-GLY-TYR-PRO-GLY-ASN-ILE-GLY-PRO-VAL-GLY-ALA-ALA-GLY-ALA-PRO
                             ALA         PHE                     ALA             PRO         SER         LEU

-422
                  AA   T         T     A   G  C                    T C           GT                                  T
 GGT CCT CAT GGC CCC GTG GGT CCT GCT GGC AAA CAT GGA AAC CGT GGT GAA ACT GGT CCT TCT GGT CCT GTT GGT CCT GCT GGT GCT GTT
-GLY-PRO-HIS-GLY-PRO-VAL-GLY-PRO-ALA-GLY-LYS-HIS-GLY-ASN-ARG-GLY-GLU-THR-GLY-PRO-SER-GLY-PRO-VAL-GLY-PRO-ALA-GLY-ALA-VAL
                 GLN             SER         PRO                 ASP PRO         VAL                                 PHE

-332
                  TC GC                T CCA       T   G   A   T   A   T       T       A  AT           G         C G
 GGC CCA AGA GGT CCT AGT GGC CCA CAA GGC ATT CGT GGC GAT AAG GGA GAG CCC GGT GAA AAG GGG CCC AGA GGT CTT CCT GGC TTC AAG
-GLY-PRO-ARG-GLY-PRO-SER-GLY-PRO-GLN-GLY-ILE-ARG-GLY-ASP-LYS-GLY-GLU-PRO-GLY-GLU-LYS-GLY-PRO-ARG-GLY-LEU-PRO-GLY-PHE-LYS
                 LEU-ALA                 PRO         GLU                     ASP         HIS                     LEU

-242
                       G       T         C T       C   A                     C         T AA  AAC               C   A
 GGA CAC AAT GGA TTG CAA GGT CTG CCT GGT ATC GCT GGT CAC CAT GGT GAT CAA GGT GCT CCT GGC TCC GTG GGT CCT GCT GGT CCT AGG
-GLY-HIS-ASN-GLY-LEU-GLN-GLY-LEU-PRO-GLY-ILE-ALA-GLY-HIS-HIS-GLY-ASP-GLN-GLY-ALA-PRO-GLY-SER-VAL-GLY-PRO-ALA-GLY-PRO-ARG
                                         LEU         GLN                     PRO         ASN-ASN

-152
   T     C                 T     C     T   G              A    T  TC       A C C A     C           T G A   T   A T     T
 GGC CCT GCT GGT CCT TCT GGC CCT GCT GGA AAA GAT GGT CGC ACT GGA CAT CCT GGT ACG GTT GGA CCT GCT GGC ATT CGA GGC CCT CAG
-GLY-PRO-ALA-GLY-PRO-SER-GLY-PRO-ALA-GLY-LYS-ASP-GLY-ARG-THR-GLY-HIS-PRO-GLY-THR-VAL-GLY-PRO-ALA-GLY-ILE-ARG-GLY-PRO-GLN
         PRO                     PRO                     ASN     LEU         PRO-ILE                 VAL         SER-HIS

-62
                                                                                    |
     AG                        T       C   T       T   C  C    C   C  CT     CCC  AT       C   A       A G     C  TT
 GGT CAC CAA GGC CCT GCT GGC CCC CCT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT GGT GGT TAT GAC TTT GGT TAC GAT
-GLY-HIS-GLN-GLY-PRO-ALA-GLY-PRO-PRO-GLY-PRO-PRO-GLY-PRO-LEU-GLY-PRO-LEU-GLY-VAL-SER-GLY-GLY-GLY-TYR-ASP-PHE-GLY-TYR-ASP
     SER                                                 PRO         PRO     PRO-ASN                 GLU-VAL     PHE

  28
  C    A                                              A      C         T                                   T
 GGA GAC                   △                         TTC TAC AGG GCT GAC △   CAG CCT TCT CTC AGA CCC AAG GAC TAT GAA GTT
-GLY-ASP-  -   -   -   -   -   -   -   -   -   -    -PHE-TYR-ARG-ALA-ASP-   -GLN-PRO-SER-LEU-ARG-PRO-LYS-ASP-TYR-GLU-VAL
 ALA-GLU                                             TYR          ▲
                                                                             CGC TCA GCA CCT
                                                                             ARG-SER-ALA-PRO
                                                                             (NOT IN CHICK)

  82
       C           A A A T G           A               G   G   C   A           C  A            G           C
 GAT GCT ACT CTG AAG TCT CTC AAG AAC CAG ATT GAG ACC CTT CTT ACT CCT GAA GGC TCT AGA AAG AAC CCA GCT CGC ACA TGC CGT GAC
-ASP-ALA-THR-LEU-LYS-SER-LEU-ASN-ASN-GLN-ILE-GLU-THR-LEU-LEU-THR-PRO-GLU-GLY-SER-ARG-LYS-ASN-PRO-ALA-ARG-THR-CYS-ARG-ASP
                     THR                                                         LYS

 172
 C C       T               A                  T                T               C         GCA   T       T  G   CC
 TTG AGA CTC AGC CAC CCA GAG TGG AGC AGC GGT TAC TAC TGG ATT GAC CCC AAC CAA GGA TGC ACT ATG GAA GCC ATC AAA GTA TAC TGT
-LEU-ARG-LEU-SER-HIS-PRO-GLU-TRP-SER-SER-GLY-TYR-TYR-TRP-ILE-ASP-PRO-ASN-GLN-GLY-CYS-THR-MET-GLU-ALA-ILE-LYS-VAL-TYR-CYS
                                             PHE                                         ALA-ASP         ARG-ALA

 262
   C   T G     T   T   G   T   C      AT   T AGC  T      G T   T   G A T      C          GTC AGC  A   A  C             A
 GAT TTC CCT ACC GGC GAA ACC TGT ATC CGC GCC CAA CCT GAA AAC ATC CCA GCC AAG AAC TGG TAT   △     AGG AGC TCC AAG GAC AAG
-ASP-PHE-PRO-THR-GLY-GLU-THR-CYS-ILE-ARG-ALA-GLN-PRO-GLU-ASN-ILE-PRO-ALA-LYS-ASN-TRP-TYR-  -    -ARG-SER-SER-LYS-ASP-LYS
         ALA                         HIS     SER LEU     ASP         THR     THR         VAL-SER-LYS-ASN-PRO
                                                                                      (NOT IN HUMAN)

 352
   G     A A     T C   T                 G     T  CT               C      G        T       C A A       C       C
 AAA CAC GTC TGG CTA GGA GAA ACT ATC AAT GCT GGC AGC CAG TTT GAA TAT AAT GTT GAA GGA GTG ACT TCC AAG GAA ATG GCT ACC CAA
-LYS-HIS-VAL-TRP-LEU-GLY-GLU-THR-ILE-ASN-ALA-GLY-SER-GLN-PHE-GLU-TYR-ASN-VAL-GLU-GLY-VAL-THR-SER-LYS-GLU-MET-ALA-THR-GLN
         ILE     PHE                     GLY     THR                     GLY                 THR     ASP

 442
       T           T                 C                                                         C
 CTT GCC TTC ATG CGC CTG CTG GCC AAC TAT GCC TCT CAG AAC ATC ACC TAC CAC TGC AAG AAC AGC ATT GCA TAC ATG GAT GAG GAG ACT
-LEU-ALA-PHE-MET-ARG-LEU-LEU-ALA-ASN-TYR-ALA-SER-GLN-ASN-ILE-THR-TYR-HIS-CYS-LYS-ASN-SER-ILE-ALA-TYR-MET-ASP-GLU-GLU-THR
                                     HIS

 532
   A       T               T   A   C       A   C                   A CGA       A               A          T   G
 GGC AAC CTG AAA AAG GCT GTC ATT CTA CAG GGC TCT AAT GAT GTT GAA CTT GTT GCT GAG GGC AAC AGC AGG TTC ACT TAC ACT GTT CTT
-GLY-ASN-LEU-LYS-LYS-ALA-VAL-ILE-LEU-GLN-GLY-SER-ASN-ASP-VAL-GLU-LEU-VAL-ALA-GLU-GLY-ASN-SER-ARG-PHE-THR-TYR-THR-VAL-LEU
                                                                     ARG                                 PHE-SER

 622
   G                          AC   C A         C   A   G           G      G                G   T       T     A         C
 GTA GAT GGC TGC TCT AAA AAG ACA AAT GAA TGG GGA AAG ACA ATC ATT GAA TAC AAA ACA AAT AAG CCA TCA CGC CTG CCC TTC CTT GAT
-VAL-ASP-GLY-CYS-SER-LYS-LYS-THR-ASN-GLU-TRP-GLY-LYS-THR-ILE-ILE-GLU-TYR-LYS-THR-ASN-LYS-PRO-SER-ARG-LEU-PRO-PHE-LEU-ASP
                             ASN     LYS                                 ARG                                 ILE

 712
                       T       C           A         GG  T   C                                G
 ATT GCA CCT TTG GAC ATC GGT GGT GCT GAC CAT GAA TTC TTT GTG GAC ATT GGC CCA GTC TCT TTC AAA TAA
-ILE-ALA-PRO-LEU-ASP-ILE-GLY-GLY-ALA-ASP-HIS-GLU-PHE-PHE-VAL-ASP-ILE-GLY-PRO-VAL-CYS-PHE-LYS-•••
                                         GLN         GLY-LEU-HIS
```

FIGURE 3: Comparison of nucleotide sequences from human and chick cDNAs for proα2(I). First line: nucleotide sequence in the chick cDNAs where this differs from the human. Second line: nucleotide sequence obtained from the human cDNAs (Hf-32 and Hf-1131). Third line: amino acids encoded by the human cDNAs. Fourth line: amino acid sequence for the chick α2(I) chain where this differs from the human. (*) Nucleotides not clearly defined. (△) Spaces inserted into the proα2(I) sequence by Fuller & Boedtker (1981) in order to align it with the proα1(I) sequence or sequences present in the chick cDNA but not in the human. (▲) Site at which the C-propeptide is cleaved during the conversion of procollagen to collagen. (• • •) First of two stop codons for translation (see Figure 4). (|) End of the α-chain domain containing 1014 amino acid residues. (—) Nucleotides immediately preceding and following the 12-base insert found in the human cDNA which are homologous with the insert itself. (Rectangle) Carbohydrate attachment site in C-propeptide.

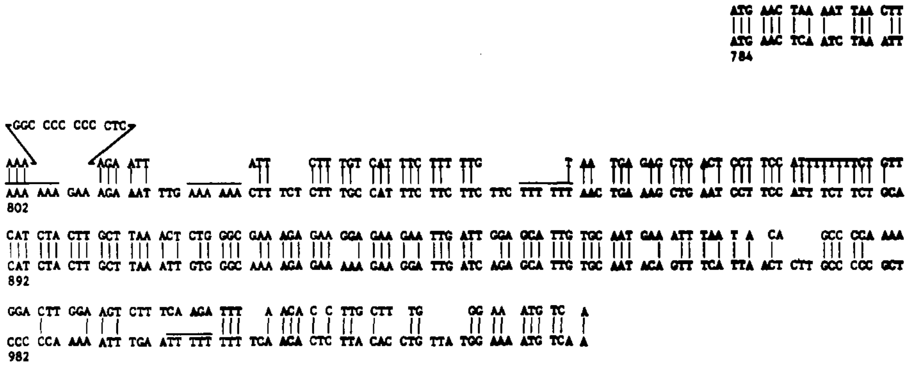
ATG AAC TAA AAT TAA CTT
ATG AAC TCA ATC TAA ATT
784
GGC CCC CCC CTC
AAA AGA ATT ATT CTT TGT CAT TTC TTT TTG T AA TGA GAG CTG ACT CCT TCC ATTTTTTTTCT GTT
AAA AAA GAA AGA AAT TTG AAA AAA CTT TCT CTT TGC CAT TTC TTC TTC TTC TTT TTT AAC TGA AAG CTG AAT CCT TCC ATT TCT TCT GCA
802
CAT CTA CTT GCT TAA ACT CTG GGC GAA AGA GAA GGA GAA GAA TTG ATT GGA GCA TTG TGC AAT GAA ATT TAA T A CA GCC CCA AAA
CAT CTA CTT GCT TAA ATT GTG GGC AAA AGA GAA AAA GAA GGA TTG ATC AGA GCA TTG TGC AAT ACA GTT TCA TTA ACT CTT GCC CCC GCT
892
GGA CTT GGA AGT CTT TCA AGA TTT A ACA C C TTG CTT TG GG AA ATG TC A
CCC CCA AAA ATT TGA ATT TTT TTT TCA ACA CTC TTA CAC CTG TTA TGG AAA ATG TCA A
982

FIGURE 4: Comparison of the 3′-noncoding region of the human and chick proα2(I) cDNAs. The two nucleotide sequences were aligned to optimize homology. Numbers indicate nucleotide bases with the same reference point as used in Figures 3 and 4. Horizontal lines indicate apparent inserts of As and Ts.

Boedtker, 1981). Therefore, it appeared to lack 11 amino acid residues present in the C-terminal telopeptides of α1(I) from several species [see Fuller (1981)]. Six of 15 amino acids in the α2(I) telopeptide differed between man and chick α2(I) chains (Table I). Four of the replacements were conservative in that Tyr replaced Phe at one site, Phe replaced Tyr at another, and Asp replaced Glu at two others. The number of nucleotides in the telopeptides was too small to make a meaningful comparison between species.

Comparison of Amino Acid and Nucleotide Sequences of the C-Propeptides in Man and Chick. In the C-propeptides, 36 of 243 amino acids, or 15%, differed between man and chick (Table I). In terms of nucleotide changes, there were 44 differences in replacement sites and 72 in silent sites. The corrected value for divergence of silent sites was 65%, or about the same as for the α2(I) chain (Table II). However, the divergence in replacement sites for the C-propeptide was 8%, or about half the value for the α2(I) chain.

Three features were of special note. The first was that 9 bases beyond the 5′ end of the C-propeptide, the human sequence contained a 12-base insert coding for -Arg-Ser-Ala-Pro-. As shown in Figure 1, the 12-base insert was found in both of the cDNA clones. As noted in Figure 3, the bases in the insert were highly homologous with 13 bases immediately preceding the insert and 12 bases following the insert.

The second feature of note was the deletion from the human sequence of two amino acid residues found in the chick sequence. The deletion corresponded to amino acid residues 110 and 111 (nucleotides 328–333) of the chick C-propeptide. The deletion of these two amino acid residues was confirmed by nucleotide sequencing of a genomic clone for human proα2(I) (L. Dickson, personal communication).

The third feature was that a highly conserved region was seen in the second half of the structure of the C-propeptide. Beginning with the Met residue at amino acid position 144 (nucleotides 430–432), there was one amino acid replacement in 51 amino acid residues. Beginning with the Ala residue at amino acid position 158 (nucleotides 472–474), there was complete conservation of 37 amino acids. Over much of the same region there was only a limited number of silent nucleotide mutations as well, e.g., 5 in the 104 nucleotide stretch between nucleotide 430 and nucleotide 533. The highly conserved region included the Asn at residue position 161 (nucleotides 481–483) through which the mannose-rich carbohydrate chain is attached to the C-propeptide of the proα2(I) chain from chick (Pesciotti et al., 1981).

There were no mutations in the C-propeptide affecting Cys residues, an observation which suggests that interchain and intrachain bonding of the C-propeptide is the same in man and chick. However, there was on change in the number of Met residues. At amino acid position 80 (nucleotides 238–240), Met was found in the human sequence whereas Ala was present in the chick. Therefore, the human C-propeptide should generate an additional cyanogen bromide fragment.

Comparison of 3′-Noncoding Region. In the 3′-noncoding region of the human cDNA for proα2(I) two stop codons of TAA were found (nucleotides 781–783 in Figure 3 and nucleotides 796–798 in Figure 4). Comparison with the chick nucleotide sequences demonstrated a surprisingly high degree of homology throughout the noncoding region.

For evaluation of the degree of divergence in the 3′-noncoding region, the human and chick sequences were aligned to obtain the best fit as was previously done in comparing 3′-noncoding regions of globin genes (Efstratiadis et al., 1980). As shown in Figure 4, the human nucleotide sequence was longer than chick and the additional length was in large part accounted for by inserts of As and Ts. After alignment, the corrected divergence for the aligned bases was calculated with a procedure (Kimura & Ohta, 1972) similar to that used to calculate corrected divergence for silent site mutations (above). There were 61 mutations in a total of 218 aligned nucleotides (Table I), and the value for corrected divergence was 35 ± 5% SD. As indicated in Table II, this value was significantly less than the corrected divergence for silent sites in both the α2(I) and the C-propeptide domains. The value was also less than the value for globin genes which have diverged more recently than the human and chick α2(I) genes, e.g., 59% for the 3′-noncoding region of mammalian β- and δ-globin genes which have diverged for 40 million years (Efstratiadis et al., 1980). Thus, we infer that this region, although not expressed in a protein structure, has evolved under a relatively high degree of selective pressure.

Codon Usage. Analysis of codon usage for the α-chain domain indicated that the same third base preference for U and C in codons for Gly, Pro, and Ala previously noted for chick proα2(I) (Fuller & Boedtker, 1981). In chick α2(I), U was used in 67% of the Gly and 75% of the Pro codons (Fuller & Boedtker, 1981). In man, these values were 56% and 66%, respectively. G was not used for these codons in chick; it was used in 4% of the Gly and 3% of the Pro codons in man. Similar analysis of the C-propeptide did not reveal any marked preference for codon usage.

Discussion

The data obtained here provide the first extensive information about the structure of a human collagen gene. Because two-thirds of the nucleotide sequence overlapped with previously sequenced cDNAs for proα2(I) from chick embryo, the data could be used to identify conserved and divergent features in two genes which have been evolving separately for 250–300 million years (Dickerson, 1971; Moore et al., 1976; Wilson et al., 1977). As recently illustrated with globin genes and several other genes (Perler et al., 1980), such data make it possible to identify functionally important features of both the protein and the gene which are conserved through selective pressure.

The data on the amino acid sequence encoded by the α-chain domain of the human proα2(I) gene support extensive data previously developed by amino acid sequencing which indicate that there is strong selective pressure to maintain Gly as every third amino acid to maintain a prescribed distribution of charged amino acids, since these amino acids are almost invariant in type I collagens from a broad range of animal species (Piez, 1976, 1980). However, there is little apparent selective pressure on other amino acids in the α-chain domain since the divergence between man and chicken in amino acids other than Gly is 32%, or about as great as the divergence between human and chick β-globin, a protein in which only a small fraction of the total structure is essential for normal function.

The data on the amino acid sequence encoded by the C-propeptide domain of the proα2(I) gene suggested that the divergence between man and chick is less for the C-propeptide than the α2(I) chain. Moreover, the degree of conservation of amino acid sequence varied from one region of the C-propeptide to another. The 5′ end or N terminus of the human C-propeptide contained an insert of 12 bases coding for 4 amino acids. The homology with the nucleotides immediately preceding and following the insert suggests that it may have arisen though duplication of a short segment of DNA. Finerty (1982) recently pointed out two highly homologous nucleotide sequences of 36 bases, each within a single exon of the chicken α2(I) gene which was sequenced by Wozney et al. (1981). Monson & McCarthy (1981) noted extensive homology between exons of a mouse proα1(I) gene and suggested that such homologies may play a role in evolution of exons of differing sizes. Interestingly, the presence of the four amino acid insert in the human C-propeptide does not prevent cleavage of the protein by chick procollagen C-proteinase; a preparation of enzyme from chick embryo calvaria (Njieha et al., 1982) cleaves both chick embryo and human type I procollagen (F. K. Njieha, W. de Wet, and D. J. Prockop, unpublished observations). Ninety-one amino acid residues beyond the insertion in the human sequence there was a deletion from the human sequence of two amino acid residues found in the chick, an observation suggesting that this region is also not important for normal function.

The complete conservation of the 37 amino acid sequence and conservation of 50 of 51 amino acids in the same region of the second half of the C-propeptide suggests that this region has a critically important function. The highly conserved region is larger than the three amino acid recognition site usually required by carbohydrate transferases, such as the dolichol-mediated enzymes which attach a mannose-rich carbohydrate to the middle of the region (Lennarz, 1980). Therefore, the data suggest that the conserved region serves some additional purpose such as directing the association of one proα2(I) C-propeptide with two proα1(I) C-propeptides so as to produce the heteropolymeric structure of type I procollagen. Analysis of the chick C-propeptides with the Chou & Fasman (1978) technique for predicting conformation suggested that the single carbohydrate attachment site in each of the C-propeptides of the chick proα1(I) and proα2(I) chains lies within a functional domain (Fuller, 1981; Olsen & Dickson, 1981). Also, it lies in the middle of an exon (Wozney et al., 1981). However, comparison of the two C-propeptides of proα1(I) and proα2(I) from chick did not reveal any highly conserved region (Fuller & Boedtker, 1981; Fuller, 1981). There are two possible explanations for these observations. One is that the structure of the two C-propeptides underwent independent mutations for some time after α1(I) and α2(I) first arose as separate genes about 500 million years ago (Bornstein & Sage, 1980), but they began to evolve coordinately under strong selective pressure before birds and mammals diverged 250–300 million years ago (Wilson et al., 1977). A second possible explanation is that the highly conserved region in the C-propeptide of the proα2(I) chain may reflect strong selective pressure on the evolution of this structure which has not been exerted on the evolution of the proα1(I) C-propeptide. It should be possible to resolve these two possibilities by comparing the structure of the cDNAs for human proα1(I) with the previously published data on the cDNAs for chick proα1(I) (Fuller & Boedtker, 1981).

The comparison of human and chick cDNAs for proα2(I) also revealed three different classes of conservation at the level of nucleotide sequence of the genes which have no apparent effect on the structure of the protein: A preference for U or C in the third base position of codons for Gly, Pro, and Ala; a high degree of conservation of nucleotide sequences in the highly conserved region of the C propeptide; a high degree of nucleotide conservation in the 3′-noncoding region. A high degree of nucleotide conservation was noted for large regions of the tandem γ-globin genes in mammals, and it was postulated that the conservation of nucleotide indicates the presence of a mechanism for exchange of information among these globin genes within a given species (Smithies et al., 1981). At the moment there is no obvious explanation for the three, apparently distinct, classes of nucleotide conservation seen here in proα2(I) genes from two highly divergent species. The extent of the conservation, however, suggests strong selective pressure which apparently is not exerted at the level of protein function. Since the kinds of nucleotide conservation are not entirely analogous to the kinds of nucleotide conservation seen in other genes [see Smithies et al. (1981) and Efstratiadis et al. (1980)], they may reflect unusual features of collagen genes, such as their high GC content or their highly repetitive coding sequences. The significance of the nucleotide conservation may become more apparent when data are available on the entire structure of proα1(I) and proα2(I) genes.

Acknowledgments

We are grateful to Drs. Forrest Fuller, Helga Boedtker, and Leon Dickson for helpful suggestions and comments in the course of the work.

References

Bornstein, P., & Traub, W. (1979) *Proteins 4*, 417.

Bornstein, P., & Sage, H. (1980) *Annu. Rev. Biochem. 49*, 957.

Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251.

Dickerson, R. E. (1971) *J. Mol. Evol. 1*, 26.

Dickson, L. D., Ninomiya, Y., Bernard, M. P., Pesciotta, D. M., Parsons, J., Green, G., Eikenberry, E. F., de Crombrugghe, B., Vogeli, G., Pastan, I., Fietzek, P. P., & Olsen, B. R., (1981) *J. Biol. Chem. 256*, 8407.

Dixit, S. N., Seyer, J. M., & Kang, A. H. (1977a) *Eur. J. Biochem. 73*, 213.

Dixit, S. N., Seyer, J. M., & Kang, A. H. (1977b) *Eur. J. Biochem. 81*, 599.

Dixit, S. N., Mainardi, C. L., Seyer, J. M., & Kang, A. H. (1979) *Biochemistry 18*, 5417.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slighton, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., & Proudfoot, N. J. (1980) *Cell* (*Cambridge, Mass.*) *21*, 653.

Fietzek, P. P., & Kühn, K. (1973) *FEBS Lett. 36*, 289.

Fietzek, P. P., & Rexrodt, F. W. (1975) *Eur. J. Biochem. 59*, 113.

Fietzek, P. P., Breitkreutz, D., & Kühn, K. (1974a) *Biochim. Biophys. Acta 365*, 205.

Fietzek, P. P., Furthmayr, H., & Kühn, K. (1974b) *Eur. J. Biochem. 47*, 257.

Finerty, M. P. (1982) *Nature* (*London*) *295*, 362.

Fuller, F. (1981) Ph.D. Thesis, Harvard University.

Fuller, F., & Boedtker, H. (1981) *Biochemistry 20*, 996.

Highberger, J. H., Corbett, C., Kang, A. H., & Gross, J. (1975) *Biochemistry 14*, 2872.

Highberger, J. H., Corbett, C., Dixit, S. N., Yu, W., Seyer, J. M. Kang, A. H., & Gross, J. (1982) *Biochemistry 21*, 2048.

Hofmann, H., Fietzek, P. P., & Kühn, K. (1978) *J. Mol. Biol. 175*, 137.

Kang, A. H., & Gross, J. (1970) *Biochemistry 9*, 796.

Kimura, M., & Ohta, T. (1972) *J. Mol. Evol. 2*, 87.

Lehrach, H., Frischauf, A. M., Manahan, D., Wozney, J., Fuller, F., Crkvenjakov, R., Boedtker, H., & Doty, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5417.

Lehrach, H., Frischauf, A. M., Hanahan, D., Wozney, J., Fuller, F., & Boedtker, H. (1979) *Biochemistry 18*, 3416.

Lennarz, W. L. (1980) in *The Biochemistry of Glycoproteins and Proteoglycans*, p 47, Plenum Press, New York.

Maxam, A., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499.

Monson, J. M., & McCarthy, B. J. (1981) *DNA 1*, 59.

Moore, G. W., Goodman, M., Callahan, C., Holmquist, R., & Moise, H. (1976) *J. Mol. Biol. 105*, 15.

Myers, J. C., Chu, M.-L., Faro, S. H., Clark, W. J., Prockop, D. J., & Ramirez, F. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3516.

Njieha, F., Morikawa, T., Tuderman, L., & Prockop, D. J. (1982) *Biochemistry 21*, 757.

Olsen, B., & Dickson, L. (1981) in *The Chemistry and Biology of Mineralized Connective Tissues* (Veis, A., Ed.) p 143, Elsevier/North-Holland, New York.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., & Dodgson, J. (1980) *Cell* (*Cambridge, Mass.*) *20*, 555.

Pesciotta, D. M., Dickson, L. A., Showalter, A. M., Eikenberry, E. F., de Crombrugghe, B., Fietzek, P. P., & Olsen, B. R. (1981) *FEBS Lett. 125*, 170.

Piez, K. A. (1976) in *Biochemistry of Collagen* (Ramachandran, G. N., & Reddi, A. H., Eds.) p 1, Plenum Press, New York.

Piez, K. A. (1980) in *Gene Families of Collagen and Other Proteins* (Prockop, D. J., & Champe, P. C., Eds.) p 143, Elsevier/North-Holland, New York.

Prockop, D. J., Kivirikko, K. I., Tuderman, L., & Guzman, N. A. (1979) *N. Engl. J. Med. 301*, 13, 77.

Showalter, A. M., Pesciotta, D. M., Eikenberry, E. F., Yamamoto, T., Pastan, I., de Crombrugghe, B., Fietzek, P. P., & Olsen, B. R. (1980) *FEBS Lett. 111*, 61.

Smithies, O., Engels, W. R., Devereaux, J. R., Slightom, J. L., & Shen, S. (1981) *Cell* (*Cambridge, Mass.*) *26*, 345.

Wendt, P., von der Mark, K., Rexrodt, F., & Kühn, K. (1972) *Eur. J. Biochem. 30*, 169.

Wilson, A. C., Carlson, S. S., & White, T. J. (1977) *Annu. Rev. Biochem. 46*, 573.

Wozney, J., Hanahan, D., Tate, V., Boedtker, H., & Doty, P. (1981) *Nature* (*London*) *294*, 129.

Yamamoto, T., Sobel, M. E., Adams, S. L., Avvedimento, V. E., DiLauro, R., Pastan, I., de Crombrugghe, B., Showalter, A., Pesciotta, D. M., Fietzek, P. P., & Olsen, B. R. (1980) *J. Biol. Chem. 255*, 2612.

Localization of the Sites of Iodination of Human β_2-Microglobulin: Quaternary Structure Implications for Histocompatibility Antigens[†]

Kenneth C. Parker* and Jack L. Strominger

ABSTRACT: Human urinary β_2-microglobulin (β_2m) and papain-solubilized human histocompatibility antigen HLA–B7 were iodinated with iodogen and the sites of iodination determined. In the case of free urinary β_2m, four of the six tyrosines were modified to some degree. Two of these were heavily iodinated (tyrosine-63 and -67) while two were lightly iodinated (tyrosine-10 and -26). In the case of β_2m iodinated in the intact HLA–B7 complex, only one of these tyrosines was modified substantially (tyrosine-67). β_2m iodinated at either of the two major sites exchanged into the HLA–B7 complex, whereas β_2m iodinated at either of the two minor sites did not exchange at all. The relationship of these findings to the quaternary structure of HLA is discussed.

Histocompatibility antigens of the class I type contain two noncovalently associated polypeptides of molecular weights 44 000 and 12 000 [see Ploegh et al. (1981) for a recent review]. The light chain, β_2-microglobulin (β_2m),[1] is also found free in the blood and the urine. The heavy chain, which spans the plasma membrane, is found on the plasma membrane of all nucleated cells in association with β_2m and is highly polymorphic, as defined by both antibodies and cytotoxic T cells.

[†] From the Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138. *Received September 29, 1982.*

[1] Abbreviations: β_2m, β_2-microglobulin; HLA, class I human histocompatibility antigen; u-β_2m, urinary β_2m; c-β_2m, β_2m isolated from HLA–B7; HLA$_{\text{pap}}$, papain-solubilized HLA; TSAO8, 10 mM Tris, pH 8.0/140 mM NaCl/3 mM NaN$_3$/0.1 mg/mL ovalbumin; HPLC, high-performance liquid chromatography; NaDodSO$_4$, sodium dodecyl sulfate; Ig, immunoglobulin; C$_L$, Ig light chain constant region; C$_H$1, Ig heavy chain first constant region; Tris, tris(hydroxymethyl)aminomethane; TLC, thin-layer chromatography; EDTA, ethylenediaminetetraacetic acid.

0006-2960/83/0422-1145$01.50/0