J. Biochem. 122, 71-82 (1997)

Molecular Cloning and Sequencing of cDNAs Encoding Three Heavy-Chain Precursors of the Inter-α-Trypsin Inhibitor in Syrian Hamster: Implications for the Evolution of the Inter-α-Trypsin Inhibitor Heavy Chain Family[1]

Tatsuya Nakatani, Yasuyuki Suzuki, Takashi Yamamoto, and Hyogo Sinohara

Department of Biochemistry, Kinki University School of Medicine, Osaka-Sayama, Osaka 589

Received for publication, January 24, 1997

Complementary DNAs encoding precursors of the three heavy chains (HC1, HC2, HC3) of the inter-α-trypsin inhibitor in Syrian hamster liver were sequenced. The deduced amino acid sequence of the HC1 precursor was 87, 82, and 79% identical with those of the HC1 precursors from mouse, man and pig, respectively. The HC2 and HC3 precursors showed similar degrees of sequence identity with the corresponding human and mouse HC precursors. When the hamster HC1 precursor was compared with its own HC2 and HC3 precursors, however, even the most highly conserved segment consisting of 565 amino acid residues, *i.e.*, about 2/3 of the whole molecule, showed only about 35 and 65% sequence identity, respectively. Essentially the same results were obtained on the intra-species comparisons of three subfamilies in man and mouse. Thus, the interspecies conservation of a given HC subfamily is much greater than the similarity between the three different HC subfamilies within a given species. These results suggest that (i) higher vertebrates possess three HC genes which have been evolving independently of each other under purifying selection; (ii) the diversification of the three HC subfamilies, for which the middle regions of the molecules were mainly responsible, occurred before eutherian radiation; and (iii) each HC subfamily may have unique function(s), although at present virtually nothing is known about the functional differences between the three HC subfamilies.

Key words: cDNA sequencing of inter-α-trypsin inhibitor heavy chains, molecular evolution of inter-α-trypsin inhibitor family, Syrian hamster.

In recent years, the incidence of pancreatic cancer has been increasing in all parts of the world for which valid statistics are available (*1*). Survival diminishes rapidly during the year after the diagnosis is made, and many reviews have indicated a 5-year survival rate of less than 10% (*1*). Little is known, however, about the biochemical mechanisms underlying pancreatic carcinogenesis, rapid tumor growth, infiltration, and metastasis. Since humans cannot be used as experimental material, it is necessary to establish an appropriate animal model to gain further insight into these issues. It was shown that the Syrian hamster is particularly useful for such study, since BOP administration readily induces pancreatic ductal adenocarcinoma, the most common type of human pancreatic tumor (*2, 3*). Moreover, the detected mutations in the Ki-*ras* genes of human and hamster pancreatic tumors are identical, that is, both types of cancer cells carry a GGT to GAT point mutation in codon 12 of this gene (*4, 5*). BOP was shown to alkylate the DNA of almost all tissues in many experimental animals, but there were striking species differences regarding the main tissues undergoing carcinogenesis. In hamsters, BOP is a potent and selective carcinogen for the pancreas and liver (*3, 6-8*), whereas in rats, BOP causes tumors mainly in the colon, lung, thyroid, urethra, and liver, but not in the pancreas (*9*). Since little difference was observed in the degree of DNA alkylation between hamster and rat pancreases (*6*), DNA injury itself is not the major factor in the preferential pancreatic carcinogenesis in hamsters. Furthermore, BOP metabolism was similar in rats and hamsters, suggesting that the species difference was not due to difference in the carcinogen metabolism (*10*). These results suggest that events after BOP initiation may play roles in the preferential pancreatic carcinogenesis in hamsters. This interpretation is also in agreement with the finding that liver cancers induced by BOP are mainly cholangiocarcinomas in hamsters but hepatocellular carcinomas in rats (*8, 9*). We showed previously that during pancreatic carcinogenesis, hamsters secrete into the urine large quantities of two trypsin inhibitors, *i.e.*, bikunin and trypstatin (*8*). In humans, bikunin (also called HI-30 or urinary trypsin inhibitor) is present at a 100- to 500-fold

[1] This research was supported in part by grants from the Japan Private School Promotion Foundation and the Environmental Science Research Institute of Kinki University. The nucleotide sequences of heavy chain precursors 1, 2, and 3 of the Syrian hamster inter-α-trypsin inhibitor reported in this paper have been submitted to the GenBank/EMBL/DDBJ Data Bank under accession numbers D89285, D89286, and D89287, respectively.

Abbreviations: α_1M, α_1-microglobulin; AMBP, α_1-microglobulin/bikunin precursor; BOP, *N*-nitrosobis(2-oxopropyl)amine; HC, heavy chain; ITI, inter-α-trypsin inhibitor; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; RT, reverse transcriptase.

elevated concentration in the urine of patients with various malignant tumors (*11-13*) as well as in lung cancer tissue extracts (*14*) and the tumor fluid of ovarian cancers (*15*). An immunohistochemical study (*16*) also showed that bikunin is widely distributed in almost all malignant tumors, whereas it was detected in only a limited number of normal tissues, such as renal proximal tubules, cerebral glial cells and bronchial epithelial cells. Recently, Kobayashi *et al.* (*17*) showed that bikunin exhibits inhibitory activity toward tumor cell invasion *in vitro*. On the other hand, McKeehan *et al.* (*18*) showed that proteinase inhibitors including bikunin stimulated the proliferation of cultured human endothelial cells. These results suggest that bikunin may modulate the development of a cancer or its infiltration into surrounding tissues.

Bikunin consists of two tandemly arranged domains of the Kunitz-type trypsin inhibitor (*19*), whereas trypstatin is identical with the C-terminal domain of bikunin (*8, 20*). The biosynthesis and posttranslational processing of bikunin are very unique among those of the proteins known to date. It is synthesized as a fusion protein with α_1M in all animals so far studied, including two species of fish (*19, 21-24*), although the two proteins are structurally and functionally unrelated. In humans, the AMBP gene comprises six α_1M-encoding exons and four bikunin-encoding exons separated by a large (7 kilobases) intron (*24*). This suggests that the AMBP gene resulted from the fusion of two ancestral α_1M and bikunin genes, and has been conserved during vertebrate evolution. In mice, AMBP is expressed exclusively in the liver, and undergoes extensive co- and posttranslational modifications to become the mature protein circulating in the plasma (*19*). These processing steps include (i) removal of the signal peptide, (ii) the addition of oligosaccharide chains to several *N*-glycosylation sites, (iii) the addition of a chondroitin 4-sulfate chain to a serine residue near the N-terminal region of mature bikunin, (iv) cleavage between α_1M and bikunin, and (v) cross-linking to either one or two of HCs of ITI through a chondroitin sulfate chain. A large C-terminal peptide (about 235 amino acid residues) of each HC precursor is removed concomitantly with the formation of an ester linkage between Asp of HCs and the C-6 hydroxyl group of the *N*-acetylgalactosamine residue of the chondroitin sulfate chain. Although there are many steps to be further substantiated in this proposed pathway (*19*), it is generally thought that this complex pathway is common to many mammals, and that free bikunin present in extrahepatic tissues is not synthesized *in situ*, but taken up from the bloodstream. This concept was also consistent with the recent findings that in mice the gene expression of three HCs and bikunin changes in parallel during development (*19*), and that ITI bound to tumor cells is cleaved into HCs and bikunin on the cell surface (*26*).

In human plasma, four different forms of HC, *i.e.*, HC1, HC2, HC3, and HC4, have been identified (*8, 19, 27, 28*). The first three HCs are present in the plasma as several complexes with bikunin, *i.e.*, ITI (HC1/HC2/bikunin), pre-α-trypsin inhibitor (HC3/bikunin), inter-α-like inhibitor (HC2/bikunin), and HC1/bikunin. The linkage between the HCs and bikunin in these proteins is essentially the same as that in ITI (*19*). It has also been noted that there are species differences in the molecular components of the ITI complexes. In bovine plasma, another form of ITI, *i.e.*, HC2/HC3/bikunin, occurs instead of human type ITI (HC1/HC2/bikunin) (*29*). We showed previously (*8*) that (i) the hamster possesses both human and bovine types of ITI, (ii) during pancreatic carcinogenesis the hamster secretes bikunin into the urine at concentrations several-fold greater than in normal controls, and (iii) trypstatin is secreted in approximately the same quantity as bikunin. On the other hand, trypstatin has only been detected in rats and hamsters, and careful examination of human urine failed to reveal trypstatin (*25*). These results indicate that thorough elucidation of the molecular species of the ITI family is a prerequisite for studying the possible role played by bikunin in the induction, growth, or metastasis of pancreatic ductal adenocarcinomas. This paper describes the cloning and sequencing of three HCs of hamster ITI, and comparison of the sequences with those of known two species, man and mouse.

MATERIALS AND METHODS

Materials—Restriction enzymes, DNA ligase, an RNA LA PCR kit, and other DNA-modifying enzymes were purchased from Takara Shuzo. An Isogen kit for RNA extraction and a pMOS*blue* T-vector Kit were obtained from Nippon Gene and Amersham, respectively. cDNA Synthesis and ^{T7}Sequencing kits were obtained from Pharmacia. Other materials were essentially the same as those used in the previous experiments (*30, 31*).

Isolation of Total RNA from Hamster Liver and Sequencing Strategy for Individual cDNAs—Total RNA was prepared from hamster liver by the method of Chomczynski and Sacchi (*32*) using the Isogen kit. The complete cDNA sequence encoding each HC was determined by the following three step strategy, (i) RT-PCR of the region specific for each HC, (ii) 3′-RACE, and (iii) 5′-RACE. RT-PCR was carried out essentially by the method of Lynas *et al.* (*33*) using the RNA LA PCR kit, and the RACEs were performed by the method of Frohman *et al.* (*34*). The PCR products were ligated into the pMOS*blue* T-vector and sequenced by the dideoxynucleotide chain termination method (*35*) using the ^{T7}Sequencing kit. The protocols and primers are briefly described below.

Sequencing of cDNA Encoding the HC1 Precursor—Step (i): A portion (1 μg) of total RNA was incubated with avian myeloblastosis virus reverse transcriptase in the presence of an antisense primer (see below) for 25 min at 50°C, and then heated for 5 min at 99°C prior to cooling at 5°C for 5 min. Thereafter, the reaction mixture was subjected to PCR amplification using Takara La Taq polymerase in the presence of a 20-mer sense primer (see below) according to the manufacturer's instructions. The antisense primer was a 20-mer, 5′-AA aCC CTG TAG CTG CTG GGT-3′, which is complementary to nucleotides 1402-1421 (Fig. 1, double underlines), and the sense primer was a 20-mer, 5′-CAA GGA AGC CAC CCa GAa CT-3′, corresponding to positions 1162-1181 (Fig. 1, double underlines). These primers were designed based on the amino acid sequence obtained on direct sequencing of a fragment of hamster HC1 (*8*), as well as on the finding that the corresponding sequences in human and mouse HC1s are highly conserved (*8, 19*). The nucleotides given in lower case letters were found to differ from those identified experimentally. The sequence between the two primers was specific for HC1, as judged from the

corresponding sequences of the human and mouse HC1s. The PCR product was analyzed by electrophoresis on a 1% agar gel and staining with ethidium bromide. The single band material separated on the gel was excised, eluted, ligated into the pMOS*blue* T-vector, and then sequenced with the ^{T7}Sequencing kit according to the manufacturer's instructions.

Step (ii): The 3′ half of the HC1 precursor transcript was amplified by 3′-RACE (*34*) using a sense primer, 5′-AG AAC GTC CGC AAC GCT ATC-3′, positions 1262-1281 (Fig. 1, single underlines), which had been determined through the above RT-PCR. The PCR product obtained in the first round was separated on an agar gel and then sequenced using the ^{T7}Sequencing kit as described above. With this method, the sequence of about 500 nucleotides located downstream of the primer was determined, and five more rounds of similar sequencing were carried out using different primers located downstream of the previous primers (Fig. 1, dotted underlines). These experiments gave the overlapping sequences of the 3′-half of the cDNA.

Step (iii): The 5′ half of the HC1 precursor transcript was amplified by 5′-RACE by essentially the same method as that for the 3′-half. The first antisense primer was 5′-TT CTC CAT GGA CAT GAC CTC-3′, complementary to positions 1342-1361 (Fig. 1, single underlines). The PCR products were sequenced essentially as described above except for the use of different sequencing primers (Fig. 1, dotted underlines).

Sequencing of cDNAs Encoding the Precursors of HC2 and HC3—Complete cDNA sequences were determined essentially by the same method as described for the HC1 precursor with the exception that the primers used for amplification, RACEs, and sequencing were different. These primers are indicated in Figs. 2 and 3 by double, single and dotted underlines, respectively, as in Fig. 1.

Analysis of Substitution Rates—The numbers of nucleotide substitutions per synonymous site (K_S) and nonsynonymous site (K_A) between each two HCs were calculated by the method of Li *et al.* (*36*).

Analysis of the Secondary Structure—The secondary structures of the HCs were analyzed with a computer program (Genetyx Ver 8; Software Development, Tokyo) based on the algorithm of Chou and Fasman (*38*).

RESULTS AND DISCUSSION

Nucleotide and Deduced Amino Acid Sequences of HC Precursor cDNAs—The longest cDNA of HC1 precursor was 2,896 bases long and contained an open reading frame consisting of 914 amino acid residues (Fig. 1). The deduced amino acid sequence contained two regions which are specific for HC1 (*8*) and identical with those determined on direct sequencing of partially degraded plasma HC1 (*8*) (Fig. 1, shaded sequences). The longest cDNA of the HC2 precursor was 3,102 bases in length and contained an open reading frame of 946 residues (Fig. 2). The deduced amino acid sequence included six regions determined by direct sequencing of partially degraded plasma HC2 (*8*) (Fig. 2, shaded sequences). The longest cDNA of the HC3 precursor was 2,777 bases in length and contained an open reading frame of 889 amino acid residues (Fig. 3), which included three regions identical with those determined on direct sequencing of partially degraded plasma HC3 (*8*) (Fig. 3, shaded sequences).

The deduced amino acid sequences were compared among the three HC precursors (Fig. 4), as well as among those from mouse, man, and pig (*27, 28, 39-43*). The sequence of the HC1 precursor was 87, 82, and 79% identical with those of the HC1 precursors from mouse, man, and pig, respectively. The sequences of the HC2 and HC3 precursors showed similar degrees of sequence identity with the respective HCs of mouse and man (the pig sequences for these HCs were not available). On the other hand, the sequences of the three hamster HC precursors were too different from each other to be aligned unambiguously. This was mainly due to large differences in the N-terminal (positions 1-60, HC2 numbering) and middle (positions 630-700, HC2 numbering) regions of the molecules. These findings indicate that the interspecies differences in the same subfamily are much smaller than the intraspecies differences between the different subfamilies. In other words, the hamster HC1 gene is orthologous for the human, mouse, and pig HC1 genes, but more distantly related to the hamster HC2 and HC3 genes. The same holds true for the evolutionary history of the HC2 and HC3 subfamilies. Chan *et al.* (*42*) pointed out that the overall sequences of the HC precursors of mouse and man could be roughly divided into three segments: (i) an N-terminal segment (excluding the signal peptide region) in which relatively conserved residues are evenly distributed along some 400 amino acid residues, (ii) a middle segment, which contains about 200 residues and is very divergent not only in length but also in the amino acid substitutions, and (iii) a C-terminal segment consisting of about 240 residues which shows conservation to a degree similar to that of the N-terminal segment. The sequences of hamster HCs share these features in common. In this paper, the HC sequences were divided into N (N-terminal), M (middle), and C (C-terminal) segments, the borders of which are indicated by open upward arrows (Fig. 4), and are slightly different from those defined by Chan *et al.* (*42*). This division is more suitable for the estimation of evolutionary data (see below).

Deduced N-Terminal Amino Acid Sequences of HCs—The exact border of the signal peptide of HC1 is not clear, since the N-terminal amino acid of hamster mature HC1 is probably blocked (*8*), as in the cases of the human and mouse HC1s (*19*). Direct N-terminal sequencing of mature HC2 and HC3 (*8*) indicated the sequences of SLPEESGEMT (corresponding to residues 55-64, Fig. 2) and SLPEGVVDGVVYSTKISCK (corresponding to residues 34-53, Fig. 3), respectively. However, these mature HC2 and HC3 sequences were both preceded by Arg (Figs. 2 and 3), which does not correspond to the specificity of the signal peptidase (*44*). At present, the N-terminal processing mechanism for the three HCs is controversial. Gebhard *et al.* (*39*) suggested that HC precursors are synthesized as preproproteins, and then processed to mature HC precursors *via* proproteins. On the other hand, Chan *et al.* (*42*) cast doubt on the existence of propeptide sequences for these HC precursors. In any case, the first 60 amino acid residues (HC2 numbering) show remarkable diversity not only among the three hamster HCs but also among other seven HCs from man, mouse, and pig. Accordingly, the first 60 residues were omitted for the calculation of nucleotide substitution rates (see below).

```
                                                                tgggaagtgagagcctggagacc -1
   ATG GAC GGC GCC GCG GTG GGG CTT CGG GTG CTG CTG GGC CTG GGC CTG GTG TCC CTC CTC 60
 1  M   D   G   A   A   V   G   L   R   V   L   L   G   L   G   L   V   S   L   L
   ACT CTC GAG GCC ATG CCT GCT GCG TGG GGC TTG GCC ACA ACG GGC AGA CCC AGG GCC AGA 120
21  T   L   E   A   M   P   A   A   W   G   L   A   T   T   G   R   P   R   A   R
   GAG AAA CGG CAG GCC GTG GAT ACA ACG CCT GAT GGT GTG CTG GTC AAG AGC TTG AAA GTC 180
41  E   K   R   Q   A   V   D   T   T   P   D   G   V   L   V   K   S   L   K   V
   AAC TGC AAA GTC ACC TCT CGC TTC GCC CAC TAC ATC ATC ACC AGC CAA GTG GTC AAC AGG 240
61  N   C   K   V   T   S   R   F   A   H   Y   I   I   T   S   Q   V   V   N   R
   CAG CCC AAT GAA GCC AGG GAG GTG GCC TTC GAT GTG GAA ATC CCC AAG ACG GCC TTC ATC 300
81  Q   P   N   E   A   R   E   V   A   F   D   V   E   I   P   K   T   A   F   I
   AGT GAC TTC GCC ATC ACA GCA GAT GGG AAC ACA TTC ATC GGA GAC ATA AAG GAC AAA GCC 360
101 S   D   F   A   I   T   A   D   G   N   T   F   I   G   D   I   K   D   K   A
   AGT GCA TGG AAA CAG TAC CGC AAA GCC ATT TCA GGG GAG AAC GCC GGC CTT GTC AGG ACC 420
121 S   A   W   K   Q   Y   R   K   A   I   S   G   E   N   A   G   L   V   R   T
   TCG GGC AGA AAT ATG GAA CAG TTC ACC ATC CAC ATC ACC GTT GGA GCC CAG AGC AAG GCC 480
141 S   G   R   N   M   E   Q   F   T   I   H   I   T   V   G   A   Q   S   K   A
   ACA TTC CAA CTC ACC TAC GAG GAG GTG CTG AAG CGG AGA CTT ACG CAG TAC GAC ATT GTC 540
161 T   F   Q   L   T   Y   E   E   V   L   K   R   R   L   T   Q   Y   D   I   V
   ATC AAA GTC AAG CCC AAG CAG CTG GTG CAA CAT TTC GAG ATC GAT GTG GAC ATC TTT GAG 600
181 I   K   V   K   P   K   Q   L   V   Q   H   F   E   I   D   V   D   I   F   E
   CCC CAG GGG ATC AGT AAG CTG GAT GCT CAG GCC TCC TTC CTC AGC AAG GAA CTC GCT GCT 660
201 P   Q   G   I   S   K   L   D   A   Q   A   S   F   L   S   K   E   L   A   A
   CAA ACC ATC AAG GAG TCT TTC TCA GGG AAA AAG GGC CAC GTG CTC TTC CGC CCC ACG GTG 720
221 Q   T   I   K   E   S   F   S   G   K   K   G   H   V   L   F   R   P   T   V
   AGC CAG CAG CAG CAG CCC TGC CCC ACA TGC TCC ACA TCC TGG CTG AAT GGG GAC TTC AAG 780
241 S   Q   Q   Q   Q   P   C   P   T   C   S   T   S   W   L   N   G   D   F   K
   GTG ACT TAC GAC GTC AAC CGG GAC AAG CTC TGT GAC CTC TTG GTA GCC AAC AAC TAC TTT 840
261 V   T   Y   D   V   N   R   D   K   L   C   D   L   L   V   A   N   N   Y   F
   GCA CAC TTC TTT GCC CCC AAA AAC CTG ACC AAC ATG AGC AAG AAC CTG GTT TTC GTG ATT 900
281 A   H   F   F   A   P   K   N   L   T   N   M   S   K   N   L   V   F   V   I
   GAC ATC AGC GGC TCA ATG GAA GGC CAG AAA GTG AAG CAG ACC AAG GAG GCT CTA CTC AAG 960
301 D   I   S   G   S   M   E   G   Q   K   V   K   Q   T   K   E   A   L   L   K
   ATC CTG GGG GAC GTG AAG CCA GGG GAC AGC TTT GAT CTG GTC CTC TTT GGG TCT CGA GTG 1020
321 I   L   G   D   V   K   P   G   D   S   F   D   L   V   L   F   G   S   R   V
   CAA TCC TGG AAG GGC TCC CTG GTC CCC GCG ACT CAG GCC AAC CTG CAA GCA GCT CAG GAC 1080
341 Q   S   W   K   G   S   L   V   P   A   T   Q   A   N   L   Q   A   A   Q   D
   TTC GTG CGA CGC TTT TCC CTG GCT GGA GCC ACA AAC CTG AAT GGA GGC TTG CTC CGA GGA 1140
361 F   V   R   R   F   S   L   A   G   A   T   N   L   N   G   G   L   L   R   G
   ATC GAG ATC TTA AAC AAA GCT CAA GGA AGC CAC CCG GAG CTC AGC AGC CCC GCC TCA ATT 1200
381 I   E   I   L   N   K   A   Q   G   S   H   P   E   L   S   S   P   A   S   I
   CTC ATC ATG TTG ACA GAT GGA GAG CCC ACG GAG GGG GAG ACG GAC CGT TCC CAG ATC CTC 1260
401 L   I   M   L   T   D   G   E   P   T   E   G   E   T   D   R   S   Q   I   L
   AAG AAC GTC CGC AAC GCT ATC CGG GGC AGA TTC CCG CTC TAC AAC CTC GGC TTT GGC CAC 1320
421 K   N   V   R   N   A   I   R   G   R   F   P   L   Y   N   L   G   F   G   H
   GAC CTA GAC TTT AAC TTC CTG GAG GTC ATG TCC ATG GAG AAC AGT GGA TGG GCC CAG AGG 1380
441 D   L   D   F   N   F   L   E   V   M   S   M   E   N   S   G   W   A   Q   R
   ATT TAT GAG GAC CAT GAT GCC ACC CAG CAG CTA CAG GGC TTC TAC AAT CAA GTA GCC AAC 1440
461 I   Y   E   D   H   D   A   T   Q   Q   L   Q   G   F   Y   N   Q   V   A   N
   CCC CTG CTG ACC GAC GTG GAG CTG CAG TAT CCC CAG GAT TCG GTC TTA AGT CTA ACG CAG 1500
481 P   L   L   T   D   V   E   L   Q   Y   P   Q   D   S   V   L   S   L   T   Q
   CAC CGA CAC AAA CAG TAC TAT GAT GGC TCC GAG ATC GTG GTG GCC GGA CGT ATT GCT GAC 1560
501 H   R   H   K   Q   Y   Y   D   G   S   E   I   V   V   A   G   R   I   A   D
   CAC AAG CTG AGC ACT TTT AAG GCT GAC GTT CGG GCT CGT GGG GAG AGG CAA GAG TTC AAG 1620
521 H   K   L   S   T   F   K   A   D   V   R   A   R   G   E   R   Q   E   F   K
   GCA ACC TGC CTG GTG GAT GAG GAA GAG ATG AAG AAG CTG CTC CGA GAG CGT GGG CAC ATG 1680
541 A   T   C   L   V   D   E   E   E   M   K   K   L   L   R   E   R   G   H   M
   CTA GAG AAC CAC GTG GAG CGG CTG TGG GCC TAC CTC ACC ATC CAG GAG CTG CTG GCG AAG 1740
561 L   E   N   H   V   E   R   L   W   A   Y   L   T   I   Q   E   L   L   A   K
   CGG ATG AAG ATG GAG GGG GAA GAG AGG GCC AAC CTG TCA TCC CAG GCC CTG AAG ATG TCG 1800
581 R   M   K   M   E   G   E   E   R   A   N   L   S   S   Q   A   L   K   M   S
   CTG GAC TAT CAG TTC GTG ACG CCG CTG ACC TCT ATG ACG ATC AGA GGC CTG ACG GAC GAG 1860
601 L   D   Y   Q   F   V   T   P   L   T   S   M   T   I   R   G   L   T   D   E
   GAT GGG CTG GAG CCC ACC ATC GAC AAG ACC CCA GAG GAT TCT CAG CCC TTA GTG AAG GTG 1920
621 D   G   L   E   P   T   I   D   K   T   P   E   D   S   Q   P   L   V   K   V
   GGA CCC AGA AGG ACG TTC GTG CTG TCG GCC ACG CAG CCT TCT CCT ACA GCC CGC AGC TCC 1980
641 G   P   R   R   T   F   V   L   S   A   T   Q   P   S   P   T   A   R   S   S
   GTG GTC TCA AAG CTG CCG AAC CAA GTG ACA GGC GTG GAC ACC GAC CCC CAC TTC ATC ATC 2040
661 V   V   S   K   L   P   N   Q   V   T   G   V   D   T   D   P   H   F   I   I
   TAT GTG CCC CAG AAA GAG GAC AGC CTG TGC TTC AAC ATC AAT GAG GAA CCC GGG GTG ATC 2100
681 Y   V   P   Q   K   E   D   S   L   C   F   N   I   N   E   E   P   G   V   I
```

Fig. 1. (continued on next page)

```
    CTG AGC CTG GTG CAG GAC CCT GAC ACA GGC TTC TCG GTG AAT GGA CAG CTC ATT GGG AGC 2160
701 L   S   L   V   Q   D   P   D   T   G   F   S   V   N   G   Q   L   I   G   S
    AAG CCC AGC AGG CCT GGG CAG CAT GAG GCC ACA TAC TTC GGG AGA CTC GGG ATC TCA AAT 2220
721 K   P   S   R   P   G   Q   H   E   A   T   Y   F   G   R   L   G   I   S   N
    CCC CCA TCA GAT TTT CAG CTG GAA GTG ACG CCT CGG AAC ATT ACA CTG AAC CCC AGC TCT 2280
741 P   P   S   D   F   Q   L   E   V   T   P   R   N   I   T   L   N   P   S   S
    GGT GGG CCT GTG TTC TCC TGG AGG GAC CAG GCC ACG CCG CAG AAA GAC GGG GTC CTG GTG 2340
761 G   G   P   V   F   S   W   R   D   Q   A   T   P   Q   K   D   G   V   L   V
    ACC ATC AAC AAG AAA AGG AAC CTG GTG GTG TCT GTG GAA GAC GGA GCC ACC TTT GAG ATT 2400
781 T   I   N   K   K   R   N   L   V   V   S   V   E   D   G   A   T   F   E   I
    GTC CTG CAC AGG ACG TGG AAG GGG AGT GCC GCC CAC CAG GAC TTT CTG GGC TTC TAT GTG 2460
801 V   L   H   R   T   W   K   G   S   A   A   H   Q   D   F   L   G   F   Y   V
    CTG GAT AGC TCC CGG ATG TCG GCC CGG ACG CGG GGG CTG CTG GGG CAA TTC TTC TGC CCC 2520
821 L   D   S   S   R   M   S   A   R   T   R   G   L   L   G   Q   F   F   C   P
    CTG GAT TTT GAA GTA TCT GAC ATC CGC CCA GGC TCT GAC CCT ATG AAG TTA GAT GCC ACG 2580
841 L   D   F   E   V   S   D   I   R   P   G   S   D   P   M   K   L   D   A   T
    ATG CGT GTG AAG AAT CGG CAG CTG GCA GTC ACC AGG GGC TTA CAA AGA GAC TAC AGC AAG 2640
861 M   R   V   K   N   R   Q   L   A   V   T   R   G   L   Q   R   D   Y   S   K
    GAC CCC AGA CAT GGA ACA GAG GTG TCT TGC TGG TTC ATC CAC AAC AAT GGA GCT GGA CTG 2700
881 D   P   R   H   G   T   E   V   S   C   W   F   I   H   N   N   G   A   G   L
    ATT GAT GGA GTT CAC ACT GAC TAC ATT GTC CCG GAC ATC TTC TGA gccctccagctagtccgca 2764
901 I   D   G   V   H   T   D   Y   I   V   P   D   I   F   *
    cctgagactgcatctgaggagggagaggagggcatcgaattaaccccaccctcctgagcgtcctggccctttgtgattt 2843
    cattaaagagaggctgtgtcc
```

Fig. 1. **Nucleotide and deduced amino acid sequences of the cDNA encoding the HC1 precursor of hamster ITI.** The nucleotide and predicted amino acid residues are numbered on the right and left, respectively. Nucleotides preceding the start codon are presented as lower case letters, and numbered negatively. Amino acid residues that were determined by direct protein sequencing of the mature HC1 are shaded. The nucleotide sequences corresponding to the primers used for the RT-PCR are doubly underlined. The primers used for the 5′- and 3′-RACEs are singly underlined. The primers used for sequencing are indicated by dotted underlines. An asterisk shows the stop codon limiting the open reading frame and the ensuing untranslated nucleotides are shown as lower case letters.

Conservation of Cysteine Residues of HCs—At present, no data are available on the intramolecular disulfide bridges of HCs. However, the present results show that four cysteine residues, *i.e.*, Cys^{247}, Cys^{250}, Cys^{690}, and Cys^{890} (HC1 numbering), are completely conserved in all 10 HCs examined to date (Fig. 4). In marked contrast, these cysteine residues have not been conserved in human HC4, which does not form a complex with bikunin (*27, 28*). These results suggest that the major disulfide bridges are similar in all HC precursors capable of forming a complex with bikunin. The C segments containing Cys^{690} (HC1 numbering) and Cys^{890} are cleaved off when the HC precursors are linked to bikunin. All precursors of HC1 and HC3 except that of human HC3 contain an additional cysteine residue, Cys^{62} (HC1 numbering), the structural and functional significance of which remains unclear.

Deletions and Insertions in the N Segments—The N segment of hamster HC1 is unique among all other HCs in that it contains one codon shift, *i.e.* one codon insertion at or around position 81 and one codon deletion at position 132 (Fig. 4). These events took place after the separation of the hamster from the murine lineage. All members of the HC2 subfamily have a two-codon deletion at position 228 (HC1 numbering) and one-codon insertion at position 283 (HC2 numbering). It is known that loop regions connecting regular secondary structures tend to be less highly conserved and often correspond to regions of an insertion or deletion event in evolutionary history. The secondary structures around the insertion and deletion positions in hamster HC1, as predicted with the Chou-Fasman algorithm, are consistent with this generalization. Hamster HC1 was predicted to have the β structure (positions 57 to 80) and an α-helix (positions 84 to 88), and the putative insertion site lies in a short sequence between these two secondary structures. The deletion position of hamster HC1 lies at the end of the 12-13-residue-long α-helix. Chou-Fasman prediction also suggested that the deletion and insertion in the HC2 superfamily are located at similar positions (data not shown). Thus, the deletion and insertion in the N segment would have little effect on the overall conformation of the molecule, suggesting that the N segment has been conserved under relatively constant selective pressure throughout much of its evolutionary history and thus can be used for the estimation of evolutionary data.

Glutamine Repeat of Hamster HC1—Hamster HC1 has a sequence consisting of four repeats of the CAG codon encoding glutamine (positions 242-245, Fig. 1). In contrast, the corresponding regions in the human, pig, and mouse HC1s as well as in human HC4 (*27, 28*) consist of only three glutamine residues. The corresponding glutamine repeats in HC2 and HC3 of all known species consist of two residues (data not shown). It was recently shown that some genes contain a number of CAG repeats, which have a tendency to be longer in man (*45*). For instance, the Huntington's disease gene has 6-34 CAG repeats in healthy humans, 7-12 repeats in other primates, 7 repeats in mice, and 4 repeats in the pufferfish (*45*).

Molecular Evolution of Mammalian HCs—It is generally thought that synonymous nucleotide substitutions, which do not affect amino acid sequences, are selectively neutral, and accumulate at a constant rate during evolution (*46*), and that the number of nucleotide substitutions per synonymous site (K_S) reflects the evolutionary relationships among orthologues. However, it is becoming clear that the selective pressure imposed on some genes differs greatly from region to region. For example, α-macroglobulins contain a bait region, which varies greatly in length and shows much greater sequence diversity than the directly flanking introns on both sides (*47*). This indicates that the bait region has evolved under strong selective pressure to change, so it is impossible to estimate the evolutionary distance of this gene family using the entire sequence of cDNA. As pointed out above, the M segment of HCs is included in this hypervariable region. On the other hand, it is also known that the C segments of human HCs are cleaved off concomitantly with the formation of the intermolecular linkage between HCs and bikunin (*19*). This cleavage takes place between Asp^{675} and Pro^{676} (HC1 numbering) of a hexapeptide, DPHFII, which has been

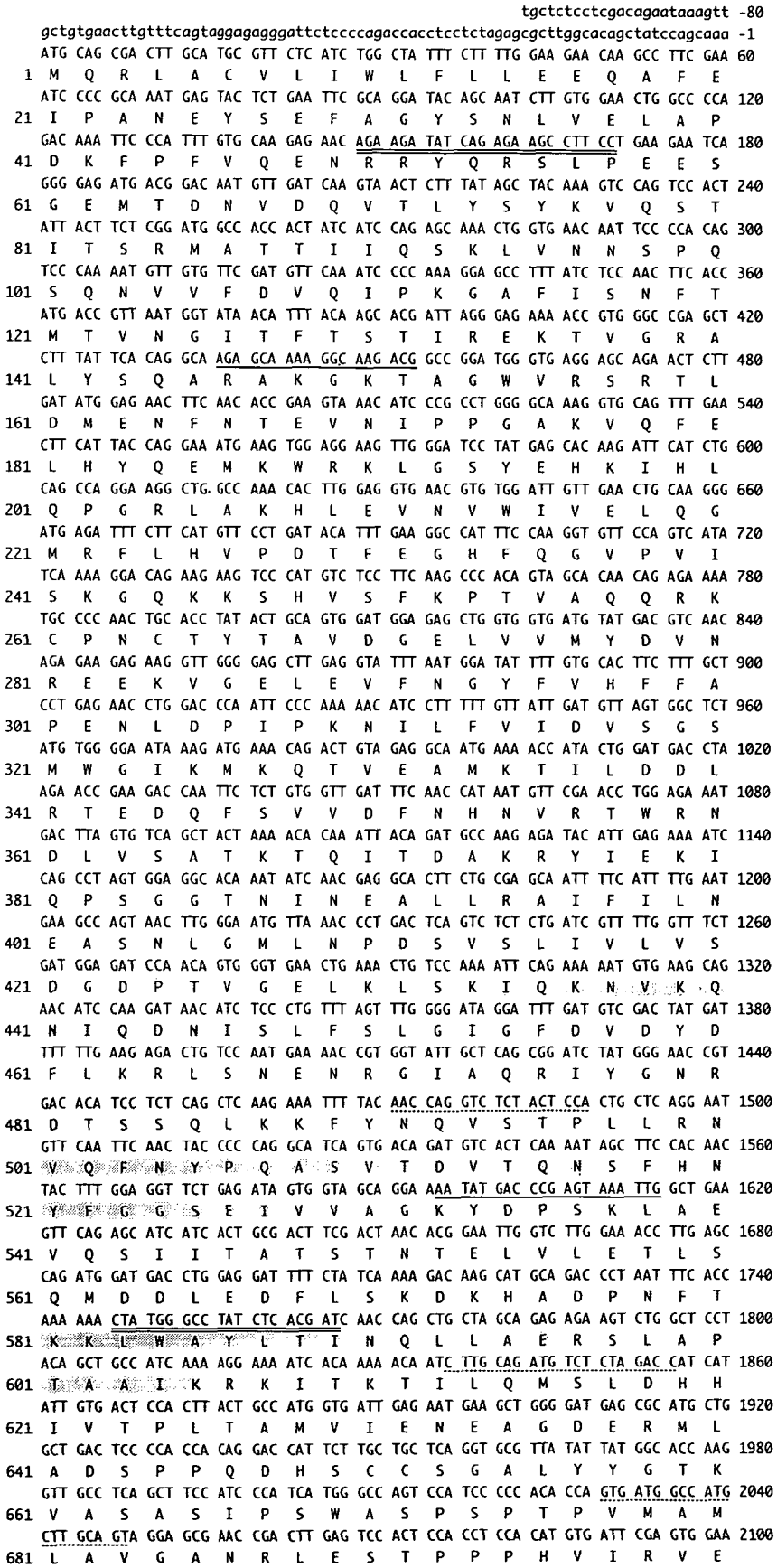

Fig. 2. (continued on next page)

```
    AAT GAC CCC CAC TTC ATC ATT TAC CTA CCA AAA AGC CAA AAG AAT ATC TGT TTC AAT ATT 2220
701 N   D   P   H   F   I   I   Y   L   P   K   S   Q   K   N   I   C   F   N   I
    GAC TCA GAA CCT GGA AAA ATC CTT AGC CTG GTG TCT GAT CCA GAA TCA GGA ATC CTA GTC 2280
721 D   S   E   P   G   K   I   L   S   L   V   S   D   P   E   S   G   I   L   V
    AAT GGC CAG CTT ATT GGT GCC AAA AAG GCA GAG AAT GGA AAA CTA AGA ACC TAC TTT GGA 2340
741 N   G   Q   L   I   G   A   K   K   A   E   N   G   K   L   R   T   Y   F   G
    AAA TTG GGA TTT TAT TTC CAA AAA GAA GAC ATG AAA ATA GAA ATC AGC ACA GAG AAC ATC 2400
761 K   L   G   F   Y   F   Q   K   E   D   M   K   I   E   I   S   T   E   N   I
    ACC CTG ATC AAC GGT TCT TCT ACA ACT AGC TTG TTC TGG TCT GAT ACA GCT CAC CTT GGG 2460
781 T   L   I   N   G   S   S   T   T   S   L   F   W   S   D   T   A   H   L   G
    AAT CAG AGG GTG CTT ATC TCC GTG AAG AAA GGA AAA TCT GTG ACT CTC ACC CTA AAT AAG 2520
801 N   Q   R   V   L   I   S   V   K   K   G   K   S   V   T   L   T   L   N   K
    GAG ATG TTC TTT TCT GTT CTG CTA CAT CAT GTG TGG AAG AAG CAT CCA GTC AAT GTG GAC 2580
821 E   M   F   F   S   V   L   L   H   H   V   W   K   K   H   P   V   N   V   D
    TTT CTG GGG ATC TAC CTT CCT CCA ACA AAC AAG TTT TCA CCC AGT GCA CAT GGA CTC TTA 2640
841 F   L   G   I   Y   L   P   P   T   N   K   F   S   P   S   A   H   G   L   L
    GGG CAG TTC ATG AAT AAG CCA AAT ATC CAC ATC TTC AAT GAG AGA CCA GGA AAA GAT CCA 2700
861 G   Q   F   M   N   K   P   N   I   H   I   F   N   E   R   P   G   K   D   P
    GAA AAA CCA GAG GCA AGC ATG GAA GTG AAA GGA CAT AAG CTG ACT GTT ACC AGA GGC TTA 2760
881 E   K   P   E   A   S   M   E   V   K   G   H   K   L   T   V   T   R   G   L
    CAG AAG GAC TAC AGG ACA GAC ATA GCG TTT GGA ACA GAC GTT CCC TGC TGG TTT GTG CAC 2820
901 Q   K   D   Y   R   T   D   I   A   F   G   T   D   V   P   C   W   F   V   H
    AAC AGT GGG AAA GGA TTC ATC GAC GGG CAT TAC AAG GAT TAC CTC GTA CCT CAG CTC TAT 2880
921 N   S   G   K   G   F   I   D   G   H   Y   K   D   Y   L   V   P   Q   L   Y
    AGC TTT CTC AAA CGG CCT TAG cggtttatggttttggaaattatatgtgtacattttcttcccttgataggt 2952
941 S   F   L   K   R   P   *
    tttgcagttattcctgagctctaacaattcaaaacaaatccagatattgcagtggtctaaaaggcctgctaatccacct 3031
    gaagaaaataaatatttgc
```

Fig. 2. **Nucleotide and deduced amino acid sequences of the cDNA encoding the HC2 precursor of hamster ITI.** The numbering of nucleotides and amino acids, as well as the symbols, underlines and shading are the same as in Fig. 1.

completely conserved in all HCs known to date (see Fig. 4). This suggests that the three hamster HC precursors are cleaved off in a similar way prior to their secretion from the liver. It is thus likely the C-terminal segment has evolved through a process different from that for the rest of the molecule. Accordingly, the K_S and K_A values were calculated using the highly conserved N segments. The K_S values between hamster and mouse for each HC ranged from 0.303 to 0.333 (Table I, A-C). The corresponding value for an α_1-antiproteinase of the orthodox type (*31*) was 0.296 (Table ID). The differences between these values were not statistically significant, indicating that the N segments of these HCs and α_1-antiproteinase evolved at similar rates, and that the N segment, the size of which is greater than that of the α_1-antiproteinase cDNA, could be used for the estimation of evolutionary data. The K_S values between hamster and human for each HC subfamily ranged from 0.419 to 0.527 (Table I, A-C). By averaging these values, it was estimated that man and hamster diverged 77×10^6 years ago. The K_S values between mouse and man ranged from 0.385 to 0.567 (Table I, A-C), which gave, on average, an estimated divergence time of 82×10^6 years. These results are in good agreement with the paleontological data (*37*) on the primate-rodent divergence time ($70\text{-}80 \times 10^6$ years ago). The K_S values between pig (artiodactyls) and primates (man) or rodents (hamster and mouse) ranged from 0.340 to 0.529 (Table I), which is also in agreement with the generally accepted view that the eutherians diverged from each other within a short period of time, $70\text{-}80 \times 10^6$ years ago, relative to the total length of time over which they have been evolving independently of each other. These results show that during mammalian evolution, the three HC genes evolved independently of each other under purifying selection and supports the aforementioned concept that each HC subfamily is orthologous in different species, but paralogous for the other subfamilies in a given species as well as in other species. This is in contrast to other multigene families, such as the α-macroglobulins, serpins, and hemoglobins. In these multigene families, the loss of one or more copies is not uncommon even recently in evolutionary terms. For example, rats have maintained two active α-macroglobulin genes; one of the two genes became extinct in the mouse lineage after the separation from rats, while the other gene disappeared in the lineages leading to man and guinea pig (*47*). The hemoglobin (*48*) and serpin (*49*) superfamilies show a much more complex pattern of extinction and duplication, their family members having been interchanged with each other several times during evolution. In other words, a gene with a potential new function evolved from different branches of the same superfamily, and another gene with a similar function became extinct in some gene lines. In marked contrast, the individual HC genes of the ITI family have been conserved under purifying selection during mammalian evolution, suggesting that each HC may have unique function and thus cannot be replaced by other HCs. At present, however, virtually nothing is known about the functional differences between the three HCs (*19*).

Evolution of the M Segment during the Mammalian Evolution—As pointed out above, all HCs contain a very diverse M segment in the middle of their molecules. This M segment, however, is highly conserved between the members of the same (but not different) subfamilies, and it was possible to determine the K_S for this M segment of the HC1 subfamily (Table II). The results indicate that the differences in the K_S values for the N and M segments are not statistically significant among the members of the same subfamilies; for example, the K_S values between hamster and mouse for the M and N segments of HC1 are 0.280 (Table II) and 0.332 (Table I), respectively. Similar results were obtained for the HC2 and HC3 subfamilies (data not shown).

Evolution of the C Segment—Since the C segments are cleaved off when the HCs are linked to bikunin, the function

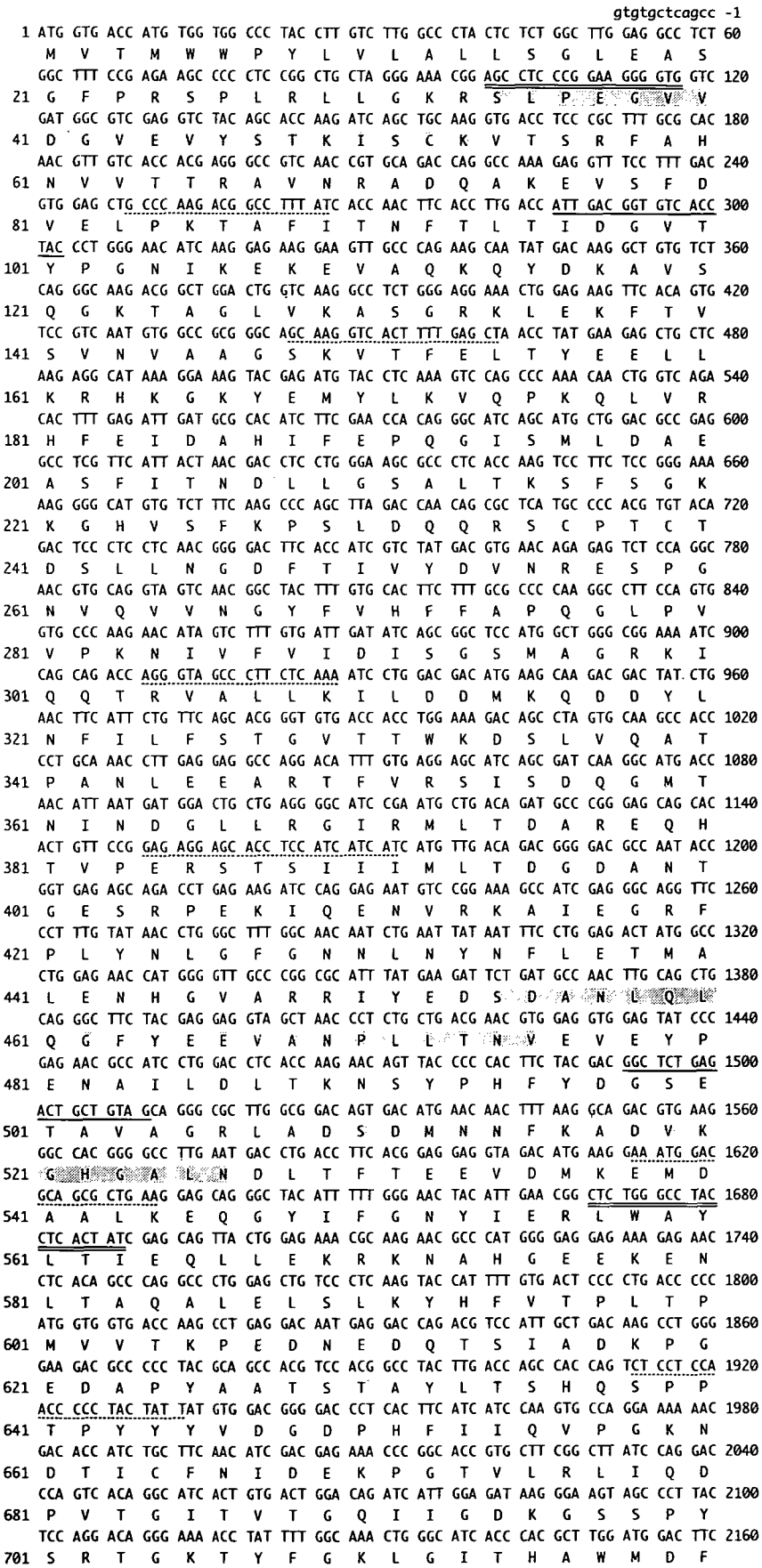

Fig. 3. (continued on next page).

```
    CGG ATT GAG GTG ACC ACA GAG AAG ATC ATC CTG GGG ACT GAA GAC GAA CTG AGC ACG TTC 2220
721 R   I   E   V   T   T   E   K   I   I   L   G   T   E   D   E   L   S   T   F
    AGT TGG CTA GAC ACG GTC ACA ATC ACA CAG ACT GGG CTG TTT GTA GCG ATC AAC AGA AAG 2280
741 S   W   L   D   T   V   T   I   T   Q   T   G   L   F   V   A   I   N   R   K
    AAG AAC ATG GTC GTG TCC TTT GGA GAT GGG GTT AAC TTC GTG ATC GTT CTA CAC CAG GTT 2340
761 K   N   M   V   V   S   F   G   D   G   V   N   F   V   I   V   L   H   Q   V
    TGG AAG AAA CAT CCG CTT CAC CAG GAC TTC CTA GGG TTC TAC GTG GTG GAC AGT CAC CGG 2400
781 W   K   K   H   P   L   H   Q   D   F   L   G   F   Y   V   V   D   S   H   R
    ATG TCA GCA CGG ACA CAC GGG CTG CTG GGG CAA TTC TTC CGC CCC TTT GAC TTT GAA GTG 2460
801 M   S   A   R   T   H   G   L   L   G   Q   F   F   R   P   F   D   F   E   V
    TCT GAC GTC CGC CCA GGC TCT GAC CCC GCA AAG CCA GAT GCC ACA ATG GTG GTG AAG AAC 2520
821 S   D   V   R   P   G   S   D   P   A   K   P   D   A   T   M   V   V   K   N
    CAT CAG CTG ACT GTC ACA AGG GGC TCC CAG AGA GAT TAC AGG AAG GAT GCC AGT GTC GGC 2580
841 H   Q   L   T   V   T   R   G   S   Q   R   D   Y   R   K   D   A   S   V   G
    ACA AAG GTC ACC TGC TGG TTT GTC CAC AAC AAT GGG GAA GGG CTG ATT GAC GGT GTG CAC 2640
861 T   K   V   T   C   W   F   V   H   N   N   G   E   G   L   I   D   G   V   H
    ACC GAC TAC ATC GTC CCC AGC TTG TTC TGA gtgacgggtcaccccgctgcggaagtggaaggcctggat 2709
881 T   D   Y   I   V   P   S   L   F   *
    ggtccggagcaaggccgcgggaacgggcttgtgggaagggccaggacaataaagccagcttgtaaacc
```

Fig. 3. **Nucleotide and deduced amino acid sequences of the cDNA encoding the HC3 precursor of hamster ITI.** The numbering of nucleotides and amino acids, as well as the symbols, underlines and shading are the same as in Fig. 1.

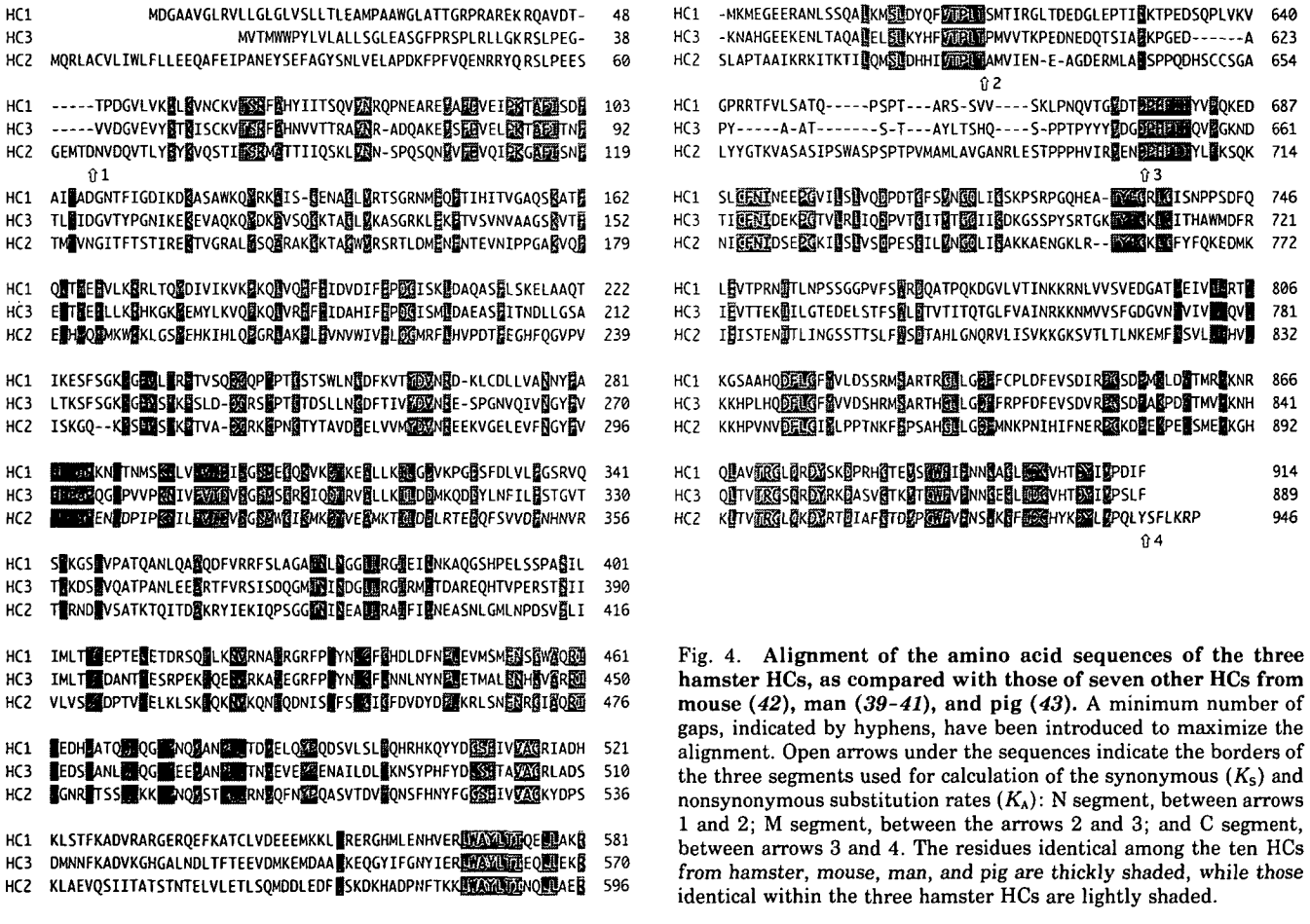

Fig. 4. **Alignment of the amino acid sequences of the three hamster HCs, as compared with those of seven other HCs from mouse (*42*), man (*39–41*), and pig (*43*).** A minimum number of gaps, indicated by hyphens, have been introduced to maximize the alignment. Open arrows under the sequences indicate the borders of the three segments used for calculation of the synonymous (K_S) and nonsynonymous substitution rates (K_A): N segment, between arrows 1 and 2; M segment, between the arrows 2 and 3; and C segment, between arrows 3 and 4. The residues identical among the ten HCs from hamster, mouse, man, and pig are thickly shaded, while those identical within the three hamster HCs are lightly shaded.

of the mature ITI is not directly related to this segment, suggesting the possibility that its evolutionary rate is different from that of the other segments. As shown in Table III, however, the K_S values of the C segment of HC1 are not significantly different from those of the N (Table IA) and M segments (Table II). Similar results were obtained for HC2 and HC3 (data not shown).

Alteration in the Evolutionary Rate of the M Segment—The evolutionary rates of the three segments described above suggest that (i) all three segments accumulated nucleotide replacements at similar (near neutral) rates during mammalian evolution, (ii) the major differences among the HC subfamilies arose before mammalian evolution, during which the M segment evolved under strong diversifying selection pressure (since this segment differs remarkably in length and nucleotide sequence between the different subfamilies, making it impossible to estimate the K_S values), and (iii) the evolutionary rate of the M segment slowed down sometime before mammalian radiation (since the K_S values for the M, N, and C segments of the individ-

TABLE I. **Pairwise comparison of the N segment of orthologous HCs among hamster, mouse, man, and pig.** Above the diagonal, number of nucleotide substitutions per synonymous site (K_S); below the diagonal, number of nucleotide substitutions per nonsynonymous site (K_A). The corresponding values of α_1-antiproteinase of the orthodox type were calculated from the data in Ref. *31* for comparison. The numbers of codons compared are 565 and 366 for HCs and α_1-antiproteinase, respectively, and the standard error of each value is less than 10%.

(A) HC1

	Hamster	Mouse	Human	Pig
Hamster		0.332	0.447	0.529
Mouse	0.0528		0.385	0.556
Human	0.0831	0.0957		0.340
Pig	0.1145	0.1038	0.1050	

(B) HC2

	Hamster	Mouse	Human
Hamster		0.333	0.527
Mouse	0.0286		0.567
Human	0.0831	0.0773	

(C) HC3

	Hamster	Mouse	Human
Hamster		0.303	0.419
Mouse	0.0424		0.395
Human	0.0724	0.0690	

(D) α_1-Antiproteinase

	Hamster	Mouse	Human
Hamster		0.296	0.536
Mouse	0.156		0.600
Human	0.194	0.230	

TABLE II. **Pairwise comparison of the M segment of HC1 among hamster, mouse, man, and pig.** Above the diagonal, K_S; below the diagonal, K_A; the number of codons compared is 65, and the standard error is shown in parenthesis.

	Hamster	Mouse	Human	Pig
Hamster		0.280 (0.082)	0.313 (0.093)	0.395 (0.113)
Mouse	0.120 (0.032)		0.597 (0.158)	0.568 (0.145)
Human	0.234 (0.048)	0.211 (0.044)		0.366 (0.112)
Pig	0.183 (0.041)	0.172 (0.039)	0.120 (0.033)	

TABLE III. **Pairwise comparison of the C segment of HC1 among hamster, mouse, man, and pig.** Above the diagonal, K_S; below the diagonal, K_A; the number of codons compared is 240, and standard error is shown in parenthesis.

	Hamster	Mouse	Human	Pig
Hamster		0.302 (0.049)	0.458 (0.067)	0.478 (0.070)
Mouse	0.093 (0.015)		0.488 (0.073)	0.449 (0.065)
Human	0.112 (0.016)	0.108 (0.016)		0.366 (0.112)
Pig	0.134 (0.018)	0.138 (0.018)	0.0776 (0.033)	

ual subfamilies are close to each other among the three mammals studied). Thus, the selective pressure for the M segment should have switched from the diversifying to the unifying direction prior to the emergence of mammals. Goodman *et al.* (*48*) demonstrated that rapid rates occurred when the monomeric hemoglobin of the primitive vertebrates evolved into an allosteric tetramer whose subunits interacted cooperatively. During this period, an accelerated rate of amino acid substitutions occurred in particular regions, such as the monomer contact site and 2,3-diphosphoglycerate-binding sites, which are responsible for the subunit cooperativity. Once tetrameric hemoglobin was established, however, the evolutionary rates of these sites decreased, possibly to maintain the subunit cooperativity. A similar evolutionary tendency was observed in the evolution of the serpin superfamily (*49*), the acceleration and slowdown of the rate occurring in the reactive site region, which is responsible for the specificity of the target serine proteinases (*49*). The present results indicate that from the molecular evolutionary point of view, the M segment is similar to the sites responsible for the subunit cooperativity in hemoglobin as well as for the reactive site in serpins.

The Selective Pressure Imposed on the Three Segments—The K_A/K_S ratio reflects the rate of amino acid substitution relative to the neutral rate (*37*). A ratio smaller than unity would indicate that the amino acid substitution rate is less than the neutral rate, and the smaller the ratio the greater the structural and functional constraints. The K_A/K_S ratios for the N, M, and C segments of HC1 of three mammalian species (pig HC1 was excluded from this calculation since data on its HC2, HC3, and α_1-antiproteinase are not available) were calculated to be 0.158 ± 0.016 (SE), 0.518 ± 0.117, and 0.259 ± 0.026, respectively. The corresponding ratios of HC2 and HC3 were close to the respective values of HC1, except that the ratio of the M segment of HC2 was 0.177 ± 0.018. In comparison, the K_A/K_S ratio of α_1-antiproteinase among the same species (Table ID) was calculated to be 0.329 ± 0.065. No statistically significant differences were found between the ratios of the three segments except those of the M segments of HC1 (see above) and HC3 (0.590 ± 0.176), which are significantly higher ($p<0.05$). These results indicate that all segments of mammalian HCs, except for the M segments of HC1 and HC3, evolved under nearly constant structural constraints, similar to in the case of α_1-antiproteinase. On the other hand, the selective pressure imposed on the M segments of HC1 and HC3 was somewhat less than those of the other segments.

We are grateful to Dr. W.-H. Li, Center for Demographic and Population Genetics, University of Texas, Houston, TX, USA, for the gift of the computer program for calculating the K_S and K_A values.

REFERENCES

1. Cohn, I., Jr. (1990) Overview of pancreatic cancer. *Int. J. Pancreatol.* **7** 1-11
2. Pour, P.M., Runge, R.G., Birt, D., Gingell, R., Lawson, T., Nagel, D., Wallcave, L., and Salmasi, S.Z. (1981) Current knowledge of pancreatic carcinogenesis in the hamster and its relevance to the human disease. *Cancer* **47**, 1573-1587
3. Mizumoto, K., Kitazawa, S., Ito, S., Takashima, Y., Tatsumi, M., Denda, A., and Konishi, Y. (1989) Cycles of repeated augmentation pressure in rapid production of pancreatic and cholangiocellular carcinoma in hamsters initiated with *N*-nitrosobis(2-oxopropyl)amine. *Carcinogenesis* **10**, 1457-1459
4. Ushijima, T., Tsutsumi, M., Sakai, R., Ishizaka, Y., Takaku, F., Konishi, Y., Takahashi, M., Sugimura, T., and Nagano, M. (1991) *Ki-ras* activation in pancreatic carcinomas of Syrian hamsters induced by *N*-nitrosobis(2-hydroxypropyl)amine. *Jpn. J. Cancer Res.* **82**, 965-968
5. Fujii, H., Egami, H., Chaney, W., Pour, P., and Pelling, J. (1990) Pancreatic ductal adenocarcinoma induced in Syrian hamsters by *N*-nitrosobis(2-oxopropyl)amine contains a c-Ki-*ras* oncogene with a point mutated codon 12. *Mol. Carcinogen.* **3**, 296-301
6. Bax, J., Schippers-Gillissen, C., Woutersen, R.A., and Scherer,

E. (1991) Cell specific DNA alkylation in target and non-target organs of *N*-nitrosobis(2-oxopropyl)amine-induced carcinogenesis in hamster and rat. *Carcinogenesis* **12**, 583-590
7. Pour, P.M., Salmasi, S.Z., and Runge, R.G. (1978) Selective induction of pancreatic ductular tumors by single doses of N-nitrosobis(2-oxopropyl)amine in Syrian golden hamster. *Cancer Lett.* **4**, 317-323
8. Yamamoto, T., Yamamoto, K., and Sinohara, H. (1996) Inter-α-trypsin inhibitor and its related proteins in Syrian hamster urine and plasma. *J. Biochem.* **120**, 145-152
9. Pour, P., Salmasi, S., Runge, R., Gingell, R., Wallcave, L., Nagel, D., and Stepan, K. (1979) Carcinogenicity of N-nitrosobis(2-hydroxypropyl)amine and N-nitrosobis(2-oxopropyl)amine in MRC rats. *J. Nat. Cancer Inst.* **63**, 181-190
10. Gingell, R., Brunk, G., Nagel, D., and Pour, P.M. (1979) Metabolism of three radiolabeled pancreatic carcinogenic nitrosoamines in hamsters and rats. *Cancer Res.* **39**, 4579-4583
11. Onitsuka, K., Sumi, H., Maruyama, M., and Mihara, H. (1985) Increase of UTI activity in cases of malignant tumor. *Jpn. J. Clin. Pathol.* **33**, 445-449
12. Rudman, D., Chawla, R.K., Hendrickson, L.J., Vogler, W.R., and Sophianopoulos, A.J. (1976) Isolation of a novel glycoprotein (EDC1) from the urine of a patient with acute myelotic leukemia. *Cancer Res.* **36**, 1837-1846
13. Rudman, D., Chawla, R.K., Nixon, D.W., Vogler, W.R., Keller, J.W., and MacDonnel, R.C. (1979) Proteinuria with disseminated neoplastic disease. *Cancer Res.* **39**, 699-703
14. Okumichi, T., Nishiki, M., Takasugi, S., Toki, N., and Ezaki, H. (1984) Isolation of urinary trypsin inhibitor-like inhibitor from human lung cancer tissue. *Cancer Res.* **44**, 2011-2015
15. Akazawa, K., Sumi, H., Maruyama, M., and Mihara, H. (1983) Acid stable protease inhibitor in ascites of ovarian carcinoma. *Clin. Chim. Acta* **131**, 87-99
16. Yoshida, E., Maruyama, M., Sugiki, M., and Mihara, H. (1994) Immunohistochemical demonstration of bikunin, a light chain of inter-α-inhibitor, in human brain tumors. *Inflammation* **18**, 589-596
17. Kobayashi, H., Gotoh, J., Hirashima, Y., Fujie, M., Sugino, D., and Terao, T. (1995) Inhibitory effect of a conjugate between human urokinase and urinary trypsin inhibitor on tumor cell invasion *in vitro*. *J. Biol. Chem.* **270**, 8361-8366
18. McKeehan, W.L., Sakagami, Y., Hoshi, H., and McKeehan, K.A. (1986) Two apparent human endothelial cell growth factors from human hepatoma cells are tumor-associated proteinase inhibitors. *J. Biol. Chem.* **261**, 5378-5383
19. Salier, J.-P., Rouet, P., Raguenez, G., and Daveau, M. (1996) The inter-α-inhibitor family: from structure to regulation. *Biochem. J.* **315**, 1-9
20. Itoh, H., Ide, H., Ishikawa, N., and Nawa, Y. (1994) Mast cell protease inhibitor, trypstatin, is a fragment of inter-α-inhibitor light chain. *J. Biol. Chem.* **269**, 3818-3822
21. Leaver, M.J., Wright, J., and George, S.G. (1994) Conservation of the tandem arrangement of α_1-microglobulin/bikunin mRNA: cloning of a cDNA from plaice (*Pleuronectes platessa*). *Comp. Biochem. Physiol.* **108B**, 275-281
22. Hanley, S., McIberney, J.O., and Powell, R. (1995) Sequence analysis and evolutionary aspects of piscine α-1-microglobulin/bikunin mRNA transcript. *Mol. Marine Biol. Biotechnol.* **4**, 105-109
23. Ide, H., Itoh, H., and Nawa, Y. (1994) Sequencing of cDNA encoding α1-microglobulin/bikunin of Mongolian gerbil and Syrian golden hamster in comparison with man and other species. *Biochim. Biophys. Acta* **1209**, 286-292
24. Diarra-Mehrpour, M., Bourguignon, J., Sesboue, R., Salier, J.-P., Leveillard, T., and Martin, J.-P. (1990) Structural analysis of the human inter-α-trypsin inhibitor light-chain gene. *Eur. J. Biochem.* **191**, 131-139
25. Yuki, Y., Nomura, K., Kirihara, M., Shimomura, M., Hiratani, H., Nishimura, R., and Kato, K. (1993) Charge isomers of urinary bikunin (trypsin inhibitor). *Biochim. Biophys. Acta* **1203**, 298-303
26. Kobayashi, H., Gotoh, J., Hirashima, Y., and Terao, T. (1996) Inter-α-trypsin inhibitor bound to tumor cells is cleaved into the heavy chains and the light chain on the cell surface. *J. Biol. Chem.* **271**, 11362-11367
27. Saguchi, K., Tobe, T., Hashimoto, K., Sano, Y., Nakano, Y., Miura, N., and Tomita, M. (1995) Cloning and characterization of cDNA for inter-α-trypsin inhibitor family heavy chain-related protein (IHRP), a novel human plasma glycoprotein. *J. Biochem.* **117**, 14-18
28. Nishimura, H., Kakizaki, I., Muta, T., Sasaki, N., Pu, P.X., Yamashita, T., and Nagasawa, S. (1995) cDNA and deduced amino acid sequence of human PK-120, a plasma kallikrein-sensitive glycoprotein. *FEBS Lett.* **357**, 207-211
29. Castillo, G.M. and Templeton, D.M. (1993) Subunit structure of bovine ESF (extracellular-matrix stabilizing factor(s). A chondroitin sulfate proteoglycan with homology to human Iαi (interα-trypsin inhibitors). *FEBS Lett.* **318**, 292-296
30. Suzuki, Y., Yoshida, K., Honda, E., and Sinohara, H. (1991) Molecular cloning and sequence analysis of cDNAs coding for guinea pig α_1-antiproteinases S and F and contrapsin. *J. Biol. Chem.* **266**, 928-932
31. Nakatani, T., Suzuki, Y., Yoshida, K., and Sinohara, H. (1995) Molecular cloning and sequence analysis of cDNA encoding plasma α-1-antiproteinase from Syrian hamster: implications for the evolution of Rodentia. *Biochim. Biophys. Acta* **1263**, 245-248
32. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159
33. Lynas, C., Cook, S.D., Laycock, K.A., Bradfield, J.W.B., and Maitland, N.J. (1989) Detection of latent virus mRNA in tissues using the polymerase chain reaction. *J. Pathol.* **157**, 285-289
34. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002
35. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
36. Li, W.-H., Wu, C.-I., and Luo, C.-C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**, 150-174
37. Li, W.-H., Tanimura, M., and Sharp, P. (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* **25**, 330-342
38. Chou, P.Y. and Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**, 45-148
39. Gebhard, W., Schreitmuller, T., Hochstrasser, K., and Wachter, E. (1989) Complementary DNA and derived amino acid sequence of the precursor of one of the three protein components of the inter-α-trypsin inhibitor complex. *FEBS Lett.* **229**, 63-67
40. Diarra-Mehrpour, M., Bourguignon, J., Bost, F., Sesboue, R., Muschio, F., Sarafan, N., and Martin, J.-P. (1992) Human inter-α-trypsin inhibitor: full-length cDNA sequence of the heavy chain H1. *Biochim. Biophys. Acta* **1132**, 114-118
41. Bourguignon, J., Diarra-Mehrpour, M., Thiberville, L., Bost, F., Sesboue, R., and Martin, J.-P. (1993) Human pre-α-trypsin inhibitor-precursor heavy chain. cDNA and deduced amino-acid sequence. *Eur. J. Biochem.* **212**, 771-776
42. Chan, P., Risler, J.-L., Raguenez, G., and Salier, J.-P. (1995) The three heavy-chain precursors for the inter-α-inhibitor family in mouse: new members of the multicopper oxidase protein group with differential transcription in liver and brain. *Biochem. J.* **306**, 505-512
43. Suzuki, H., Hamashima, N., Kimura, M., Ozawa, A., and Yasue, H. (1995) The sequence of porcine inter-alpha-trypsin inhibitor heavy chain H1 cDNA. *DDBJ/EMBL/GenBank Nucleotide Database*, D38784
44. Von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683-4690
45. Baxendale, S., Abdulla, S., Elgar, G., Buck, D., Berks, M.,

Micklem, G., Durbin, R., Bate, G., Brenner, S., Beck, S., and Lehrach, H. (1995) Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nature Genet.* **10**, 67-76
46. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge
47. Iwasaki, H., Suzuki, Y., and Sinohara, H. (1996) Cloning and sequencing of cDNAs encoding plasma α-macroglobulin and murinoglobulin from guinea pig: Implications for molecular evolution of α-macroglobulin family. *J. Biochem.* **120**, 1167-1175
48. Goodman, M., Czelusniak, J., Koop, B.E., Tagle, D.A., and Slightom, J.L. (1987) A case study in molecular phylogeny. *Cold Spring Harbor Symp. Quart. Biol.* **52**, 875-890
49. Sinohara, H. (1995) Serpin superfamily (in Japanese). *Seitai no Kagaku* **46**, 505-507