

Structure and evolutionary origin of the human granzyme H gene

Patrick Haddad, Dieter Jenne¹, Jürg Tschopp¹, Marie-Véronique Clément¹, Danièle Mathieu-Mahul², and Marilynne Sasportes

Unité 93 INSERM, 1 avenue C. Vellefaux, ²Unité 301 INSERM, 27 rue J. Dodu, 75010 Paris, France

¹Institute of Biochemistry, University of Lausanne, CH-1066 Epalinges, Switzerland

Key words: cytoplasmic granules, cytotoxicity, evolution, gene structure, promoter region, serine proteinase, T lymphocytes

Abstract

Among the molecules proposed to be involved in cytotoxic T lymphocyte (CTL), natural killer (NK) and lymphokine activated killer (LAK) cell-mediated lysis are the granzymes, a family of serine proteases stored in the cytoplasmic granules of CTLs, NK and LAK cells. In addition to the granzymes A and B, a third member of this family has been cloned in man and designated granzyme H. We present the complete gene sequence including the 5' promoter region and demonstrate that the granzyme H sequence represents a functional gene expressed in activated T cells. Granzyme H shows the highest degree (>54%) of amino acid sequence homology with granzyme B and cathepsin G and, like these genes, consists of five exons separated by introns at equivalent positions. The evolutionary history of granzyme H has been analyzed by reconstructing an evolutionary tree for granzyme sequences. We provide evidence that interlocus recombination between the ancestral genes of granzyme B and granzyme H occurred about 21 million years ago, leading to a replacement of exon 3, intron 3 and part of exon 4 in human granzyme H by human granzyme B sequences. Our results suggest that the ancestral gene of granzyme H is more closely related to cathepsin G and granzyme B than to the murine granzymes C to G. Thus, granzyme H does not represent a human counterpart of the known murine granzymes A to G. It diverged from cathepsin G before mammalian radiation and should, therefore, exist in other mammalian lineages as well.

Introduction

In spite of intensive research on the mechanisms of cytotoxic T lymphocyte (CTL), natural killer (NK) and lymphokine activated killer (LAK) cell-mediated target cell lysis, the molecular mechanism responsible for the 'lethal hit' delivery is still poorly understood. Much experimental evidence indicates that the dense cytoplasmic granules (1–4) are directly implicated in the cytotoxic event. Effector cell–target cell interactions stimulate their exocytosis and the release of granule contents including a Ca²⁺-dependent lytic activity. Among the molecules purified from the granules, the pore-forming lytic protein, perforin, is likely to be the protein involved in the lytic activity of CTL/NK cells (5–9). The role of the other major granule-associated proteins, including a family of serine proteases named granzymes, remains

to be elucidated. In the mouse, seven granzymes named A to G have been identified (10–15) and the respective cDNAs have been isolated (16–22). In human CTLs, only the granzymes A and B have been characterized at the protein and gene level (23–28). In previous studies the granzyme B (*hCTLA-1*) gene was mapped to band q11–q12 on human chromosome 14 (14q11–q12) (29), which also contains the α and δ chains of T cell antigen receptors, and more recently the complete genomic sequence of granzyme B has been reported (30,31). In the course of isolating the human granzyme B gene, we also identified a related genetic locus which has been tentatively designated *hCTLA-1b* (30).

Here we report the complete genomic sequence of this new

locus including the 5' promoter and 3' noncoding regions of the transcription unit and show that it contains a new functional granzyme gene which we have designated granzyme H.

Methods

Isolation of the λ B34 clone and sequence analysis

The genomic λ B34 clone isolated previously (30) contains a 17 kb insert which can be excised from the λ EMBL3 vector by *SalI* digestion. A smaller 7.6 kb fragment was derived from this *SalI* fragment by *EcoRI* digestion. Overlapping subfragments of the *EcoRI* fragment were cloned into M13 vectors and sequenced. Sequences were determined by the dideoxy chain termination method as described (32).

Molecular probes

S1 protection experiments. A *PstI* – *HaeIII* 304 bp fragment with the 3' end located 22 nucleotides downstream from the ATG start codon (determination of the transcription initiation site), a *BamHI* – *PvuII* 422 bp fragment containing 129 bp of exon 2 (probe E2 in Fig. 3) and a *PvuII* – *BamHI* 147 bp fragment of exon 4 (probe E4 in Fig. 3) were excised from the 4.1 kb fragment containing the granzyme H gene and subcloned into M13 vectors to generate uniformly radiolabeled single-stranded probes.

Southern blot. The *PvuII* – *BamHI* 147 bp fragment derived from exon 4 of the granzyme H gene (44% nucleotide identities with the homologous granzyme B sequence) was used as a highly specific, internal probe in Southern blots of reverse-transcribed mRNA from different tissues and cell lines after amplification by the polymerase chain reaction (PCR).

S1 protection assays

Total RNA (10 μ g) was hybridized overnight with the radiolabeled single-stranded probes described above and digested 1 h at 37°C with S1 nuclease. The samples were analyzed on a 6% polyacrylamide – 7 M urea gel.

cDNA synthesis and amplification by the polymerase chain reaction

Poly(A)⁺ RNA (1 μ g) was converted to cDNA by using an oligonucleotide primer (5'-*gtcg*ACCCCGACTGCCACCCCTT-3') complementary to the non-coding sequence of the granzyme H exon 5 (nucleotide positions 3197 – 3216 in Fig. 2) and MLV reverse transcriptase (200 units) in PCR buffer, 1 mM of each dNTP, 20 units of RNasin (in a 20 μ l reaction volume) at 42°C for 60 min. Reaction volume was increased to 50 μ l, Taq polymerase (2.5 units) and a second primer (5'-*gc*ATGC-AGCCATTCCTCCTCCTGT-3') corresponding to the sense strand of the granzyme H exon 1 (beginning at position 123 in Fig. 2) were added and amplification was carried out for 30 cycles in a thermocycler (1 min denaturation at 95°C, 1 min annealing at 60°C, 2 min elongation at 72°C). Reaction products were analyzed on a 1% agarose gel. Southern blot filters were hybridized with the internal granzyme H probe E4 described above.

Statistical analyses and construction of the phylogenetic tree from nucleotide sequences

The number of substitutions at non-synonymous sites in protein coding regions were determined as described (22,33,34) and

were taken as the distance between each pair of homologous cDNA sequences. The number of substitutions between homologous intron sequences was calculated according to the method of Jukes and Cantor (35), and the phylogenetic trees were reconstructed according to the unweighted pair group (UPG) (36) and the modified UPG methods (37).

Results

Cloning and sequencing of the granzyme H gene

The isolation of a recombinant phage (λ B34) containing sequences homologous to the granzyme B/*hCTLA-1* gene has previously been reported (30). The restriction map of this clone clearly differed from that of the granzyme B gene (Fig. 1A). Initially, without having any sequence information, we designated this new locus *hCTLA-1b* (30), but from now on we shall use the name granzyme H, which is consistent with our phylogenetic sequence comparison (see below). A 4098 bp genomic fragment of the λ B34 clone, which cross-hybridized with granzyme B sequences, was sequenced over its entire length (Fig. 1B). Together with the 354 bp flanking region located upstream of this 4098 bp fragment we have established a continuous genomic sequence of 4452 bp. Alignment of the genomic sequence for human granzyme H with that of granzyme B revealed the position and size of exons and the respective exon – intron boundaries. The transcribed nucleotide sequence of the granzyme H gene is completely covered by the 4452 bp fragment sequenced (Fig. 2). This fragment contains five exons (Fig. 1B) and includes the 633 bp region upstream of the ATG translation start site and the 808 bp region downstream of the TAA translation stop codon (Fig. 2). Recently, the sequence of the *hCCPX* gene, which is identical to the granzyme H sequence, has been determined (38). This sequence begins at the ATG translation start and ends at the TAA translation stop. Our genomic sequence extends over the 3011 bp genomic fragment of the *hCCPX* gene and Fig. 2 displays only the sequences of the 5' and 3' flanking regions which have not yet been reported.

Analysis of the 5' and 3' sequences of the granzyme H gene

To define the transcription start site, a *PstI* – *HaeIII* fragment of 304 bp ending 22 bp downstream of the ATG start codon was used in S1 nuclease protection assays. This probe was annealed to total mRNA from CTLs and activated PBLs (mRNA from REX and DU528 cells served as negative controls). After S1 nuclease digestion, a poorly resolved band having a size of 146 – 147 bp was observed (data not shown). The length of the protected fragments suggests that transcription starts at nucleotide position 512 or 513 (A or G residue; see Fig. 2). The A residue was supposed to be the first nucleotide of the mRNA transcript and was subsequently numbered +1. A TATA motif is located 25 bp upstream of the transcription start site. As in granzyme B (30, 31), no CAAT box is present in the promoter region.

Potential regulatory elements are found in the 5' flanking sequence. The sequence TTA₁CTCAGC at position –104 matches closely the consensus criteria of TPA responsive elements (TRE) found in several TPA-inducible genes (39). From the IL-2 gene and other T cell specific genes a consensus sequence has been derived which may account for T cell specific gene expression (40,41); a similar sequence at position –223, AGAAAAAGAGCC-

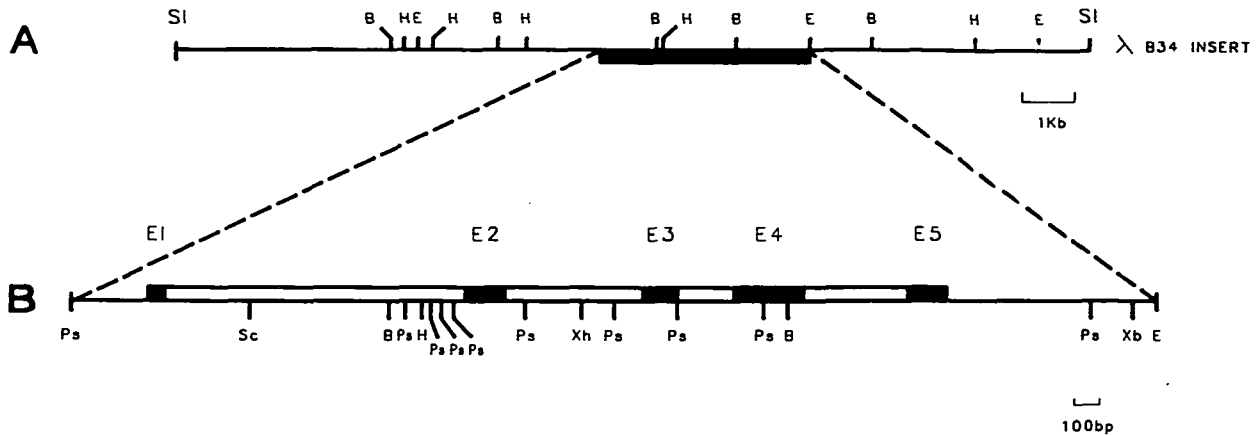


Fig. 1. Genomic organization of the granzyme H gene. (A) Restriction map of the λ B30 clone. The terminal *Sall* sites belong to the λ EMBL3 vector. (B) Expanded representation of the *Pst*I – *Eco*RI 4.1 kb fragment which has been sequenced in its entire length. Black boxes represent the translated regions of the five exons, open boxes indicate the four introns of the granzyme H gene. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; Ps, *Pst*I; Sc, *Sac*I; SI, *Sall*; Xb, *Xba*I; Xh, *Xho*I.

AAGAACTG, could be located in the promoter region of the granzyme H gene. In addition, the 5' flanking region contains the sequence CCCACCA at position –368 which resembles the AP-2 binding site in the pig plasminogen activator and the rat tyrosine aminotransferase genes (42). The relevance of these sequence elements, however, has to be explored experimentally.

In the 3' untranslated region of the granzyme H gene there are two AATAAA polyadenylation signals located at nucleotide position 3228 and 3505 (underlined in Fig. 2). However, only the first polyadenylation signal is followed by a sequence of GACTG 21 bp downstream of the AATAAA sequence which probably represents the second signal (43) required for the cleavage of the pre-mRNA transcript. Since previous hybridization studies with granzyme B and H cross-reacting probes have only identified 1.1 kb transcripts in activated peripheral blood lymphocytes (24,25), it is likely that the first AATAAA signal sequence directs polyadenylation. The isolation of polyadenylated cDNA clones will permit the definite identification of the polyadenylation signal and the poly(A) addition site, which is typically located 15–25 bp downstream of the AATAAA signal (43).

Granzyme H gene expression

S1 nuclease RNA protection assays were performed to analyze granzyme H gene expression and to overcome the problem of cross-hybridization with the granzyme B gene. Because the probe used to determine the transcription start site (see above) gave unsatisfactory signals, we used two other single-stranded probes spanning parts of granzyme H exon 2 (probe E2 in Fig. 3; see Methods) and exon 4 (probe E4). A 151 bp band (corresponding to exon 4) and a 129 bp band (exon 2) were specifically protected by mRNA from phytohemagglutinin (PHA)- or IL-2-activated PBL, but not by mRNA from monocytic THP1 cells or promonocytic U937 cells (Fig. 3).

Additionally, a 60 bp band was identified in activated PBL RNA using the probe E2 (not shown in Fig. 4). Since granzyme B and H sequences are identical over a region of 60 nucleotides in exon 2, this band is likely to represent a protected fragment of the granzyme B sequence. The intensity of this 60 bp band was much

higher than the granzyme H specific 129 bp band, indicating that the expression of granzyme B in activated T cells is much higher than that of granzyme H.

In order to verify that all five exons predicted by sequence analysis are correctly joined together by pre-mRNA splicing, cDNA synthesis in conjunction with PCR amplification was performed using two oligonucleotides derived from exon 1 and exon 5 sequences (see Methods). In agreement with the predicted size, a fragment of 0.8 kb was specifically amplified from reverse transcribed mRNA of CTLs, and IL-2- and PHA-activated PBLs (see Fig. 4; mRNA from the T cell line REX was used as a negative control).

Comparison of the human granzyme B, granzyme H and cathepsin G genes

A comparison of the human granzyme H, granzyme B and cathepsin G genes reveals highly conserved genetic and structural features. The number of exons (5), the position and the phase type of the four introns are strictly conserved among all three genes. The total length of the genes is very similar and ranges from 2.7 kb to 3.2 kb. In the mature form 71% amino acid residues are identical between granzyme H and granzyme B, 55% between granzyme H and cathepsin G, and 54% between granzyme B and cathepsin G. The alignment of the nucleotide sequences upstream from the ATG start codon (Fig. 5) also reveals high sequence conservation (~70% homology) in the non-coding and promoter regions of the three genes. Especially for the granzyme B and the cathepsin G genes, the transcription start site and the putative promoter (TAAAA for the granzyme B gene and TTAAA for the cathepsin G gene) are located at the same position relative to the translation start site; in view of the S1 nuclease protection experiment presented above, the granzyme H cap site and promoter are located further upstream.

The structural features characteristic for the granzyme family apply to granzyme H and cathepsin G as well. These proteases are synthesized as pre-propeptides with an 18 residue leader peptide and an acidic activation dipeptide either Gly – Glu (granzyme B and cathepsin G) or Glu – Glu (granzyme H). Fig. 6 shows

```

          -511  GGGATTCTAAATATCCCAAATATGAATCTG  -481
AGATAGTGGTAAAGATAGGAATGGGTTGCCATTTGCAGTATTTGTTGGGTGTGTGCAACTC  -421
TGCATACTGCTAAGGCAAGTCCAGCCTCCGTCTCACACCAACAGGCAGATTTCCCCACCA  -361
CGGCCCTTACCATCTTGTGCTCTAACCCAGATTTTCAGTCTACTCTTGCTAACCATCAA  -301
TTCTTCAAGGCAAGAGCCATGTCTCTCATATCCCTGGATCTGGCTTGCTAAATAATTATT  -241
CATGATTTTATCTGAGTAGAAAAAGAGCCAAGAAGTGAAGGAGACATTCTTGGTTC  -181
TCAGTCACTGAGGAGCTTCTGCAGCCACAGCTTCAACAGCTCCACCTCCTTGTCTTTAT  -121
TCTTGCTTTCACCTTGTTACTCAGCAGCAGGGGTGTAATGTGACAGTGCCATGTCAACT  -61
TAAACCAAAACAGATAATTTTAATAAAACCACCTTCTATAATGACACCCATCAAACCCAGA  -1
*
AGGGGACAAGTCTAGGAGGTCTCTGAGGTACTGTACCCATCCCTCCTTCATCTCCCTCC  +60
AGCATTGTGTTCTGGAAGGAGTCAACACCAACAGCTCTGACCTGGGCAGCCTTCCTGAGA  +120
AAATG  +125
          3005 nucleotides  3130

TAA CAGCAGGCATGAGACTAACCTTCTCTGGGCTGACCATCTCTGGGACAGAGGCAAG  3190
AATCCAAGGGGTGGGCAGTCGGGTTGCAGGACTGTAATAAATGGATCTCTGGTGTAATA  3250
TATGACTGAAGTCTTATTTATTTACTGAACCATCTTCAATATGCAACAACACTGCTCCAG  3310
GCAGCCGTCTGCTGTCAAGTCTTGGCTTTTGTCTTCTCCTTTCTCCCCAGCCCTCCCTCC  3370
TCCCTAAACCCCATGCTGTACAGTTTTTTCACTAATTTCTCAATCACATGTCTTTGATAC  3430
TGAACCTCTGTAGAGCTGGTTTGAACCCATTATAAAACAACAATCTTCAACCCCTCTC  3490
CCTTCACAAAAAATAAATACCTCTCTGGACGTTAGGAATTTCTTATCGTCTCTAAACCA  3550
ACTACCTACTCCCTAACAAACGTAGCAGAGAATGGAGCAGGGCAGTGGAAAGGAAATAAC  3610
TCCAGGGACTCAGAAGGCCAGGGCCATGCCATGAGTACTCAGTTTCCATTTTCAGGGCCT  3670
GTCCCTCCAATGGGGCAGGCTTCTCCACAGCCAGGTCTGCCTGCAGCTTAGGAGGTTG  3730
AGATGCTCCTCTGCACCTAGGTCCCTCTTCCCTGCCCTGGCCAGCCCTGAGCTCAGTCA  3790
ACCTTCTTCTTCTGCGCGGCTTCCAGCTCCCAAGAGTTTTTCACCAAGTGCCCTG  3850
GCTACCAGTTTCTTCCCTCTAGATCACCTGTTCTGAAGCCAGCCTCTCTCATGTC  3910
CCTAACAGAGTTCCAGCTCTGTGTGAATTC  3941

```

Fig. 2. Genomic nucleotide sequences of the human granzyme H gene upstream of the translation start site (upper panel) and downstream of the translation stop codon (lower panel). The putative TATA box and two AATAAA polyadenylation signals are underlined. The translation initiation codon ATG and the translation stop codon TAA are doubly underlined. The asterisk indicates the transcription start site as determined by S1 mapping analysis. This nucleotide residue A is numbered +1; the nucleotides 5' to the transcription start site are numbered from -1 to -511.

the alignment of the amino acid sequences of mature granzyme H, granzyme B and cathepsin G. The His, Asp and Ser residues forming the charge relay system are found in equivalent positions, strongly suggesting that the predicted granzyme H protein is a biologically active serine protease. Six cysteine residues are conserved in homologous positions and are thus expected to

form three internal disulfide bonds. In addition, granzyme H possesses a seventh free cysteine residue at position 166 (chymotrypsin numbering). Whether this unpaired Cys166 is surface-exposed and can form intermolecular disulfide bridges awaits biochemical studies. So far only human and murine granzymes A, which also contain an odd number of cysteine residues

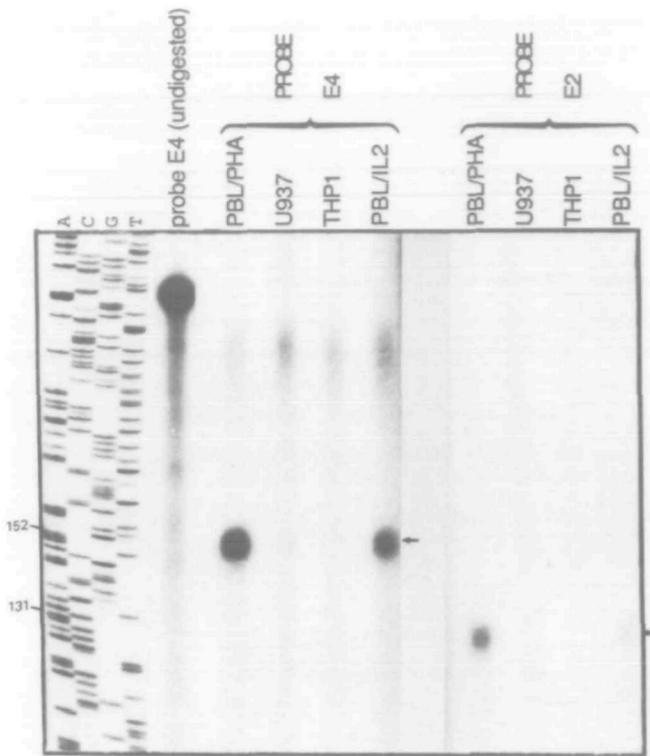


Fig. 3. S1 nuclease protection assays. Messenger RNA from PHA- and IL2-activated PBLs, and U937 and THP1 cells were hybridized to the radiolabeled single-stranded probes E2 (overlapping 131 bp of exon 2) and E4 (overlapping 152 bp of exon 4). S1 protected fragments are indicated by the arrow. The four sequencing reactions A, C, G, T of a known M13 clone were used as a size marker and an aliquot of the undigested intact probe E4 is shown in lane 5 from the left.

(28), have been shown to form homodimers through an intermolecular disulfide bridge between Cys93 (chymotrypsin numbering) residues. Three putative N-glycosylation sites are found at positions 65, 98 and 175 in granzyme H of which the two at positions 65 and 98 are also found in granzyme B; cathepsin G possesses only the N-linked glycosylation site at position 65.

The amino acid residue located six positions amino-terminal to the active site serine forms the bottom of the S1 substrate binding pocket and thus determines substrate specificity in the P1 position of a substrate. There, granzyme H, granzyme B, cathepsin G and bovine chymotrypsin contain uncharged residues: granzyme H and granzyme B the Thr189, cathepsin G the Ala189 and chymotrypsin the Met189, suggesting a chymotrypsin-like activity in all three cases. This specificity contrasts with the trypsin-like specificity of granzyme A (cleavage after lysine or arginine residues) which like trypsin contains the negatively charged Asp189 residue at position 189. However, fine specificity at the P1 position is further influenced by residues in positions 226 and 192 whose side chains could partially fill the S1 substrate binding pocket as well.

The evolutionary position of human granzyme H in the granzyme family tree

Analyzing the phylogenetic history of murine and human granzymes, we recently predicted the existence of at least one further

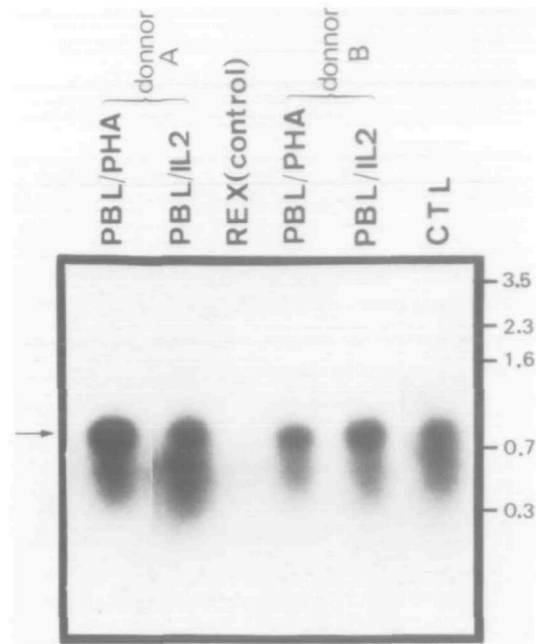


Fig. 4. Southern blot of PCR-amplified cDNA probed with a granzyme H-specific DNA fragment. Messenger RNA of IL2- and PHA-stimulated PBLs obtained from two healthy donors A and B, of a CTL clone (49), and of REX cells were converted to cDNA and a segment of the granzyme H cDNA was amplified by PCR. The amplified cDNA fragments were probed with a 152 bp fragment derived from exon 4 which is specific for the human granzyme H gene (internal probe). The arrow indicates the granzyme H-specific cDNA fragment detected by the probe. On the right, size markers are given in kb.

human granzyme that represents the homologue of the murine granzymes C – G. We therefore analyzed the evolutionary origin of the third human granzyme characterized in human lymphocytes, granzyme H, in order to clarify whether it is a descendant of the ancestral granzyme C – G progenitor in the primate lineage. We compared the entire genomic sequence of granzyme H with that of human cathepsin G and granzyme B, which are the most closely related genes in the human genome so far sequenced. We observed that the extent of nucleotide conservation between granzyme B and granzyme H was ~94% in a region that starts within intron 2 (which splits the codon of residue 61 after the second base) and ends in exon 4 after residue 142 (chymotrypsin numbering; see Fig. 6), whereas the upstream and downstream regions thereof showed only 42 – 61% conservation. We have compared the number of substitutions per 4-fold degenerate site for exon 3 and exon 4A with the number of substitutions between the aligned sequences of intron 3. These values equal each other and amount to $6.3\% \pm 1.9\%$ indicating that this complete segment of the human granzyme H gene has been replaced by granzyme B sequences recently. Due to this interlocus recombination event, the nucleotide sequences between residue positions 63 and 142 were excluded from the calculations of K_A and K_S distance values between each pair of sequences (Table 1).

Using this subsample of sequence data and including now the granzyme H sequence, the modified UPG method gave a similar tree topology and similar branch lengths for granule serine

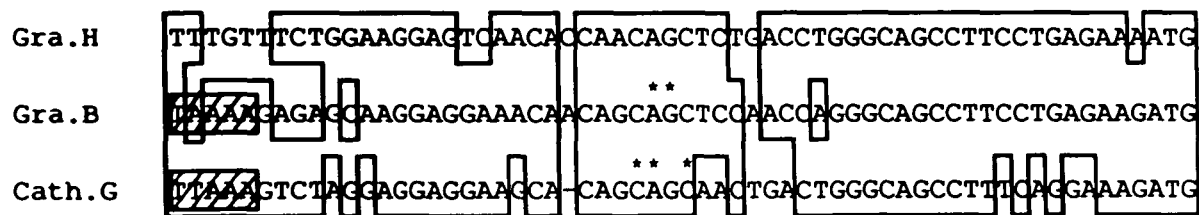


Fig. 5. Alignment of granzyme H (Gra. H), granzyme B (Gra. B) and cathepsin G (Cath. G) sequences upstream from the ATG start codon. The asterisks mark the transcription start site determined by S1-protection and/or primer extension experiments (30,31,44). Hatched boxes represent the putative TATA boxes of the granzyme B and cathepsin G genes.

	16		25		35		45
Gra H	I I G G H E A K P H S R P Y M A F V Q F L Q E K S R K R						C G G I L V R K D F V L
Gra B	- -				Y L M I W D Q - - L - - - - - - - - - - - - - - -		- - - F - I Q D - - - - -
Cat G	- - - - - R - S R -				Y L - I Q S P A G Q S - - - - - - - - - - - - - - -		- - - F - - - - - E - - - - -
CHT	I V N G E E A V P G S W P W Q V S L Q D K T * * G F H F						C G G S L I N E N W V V
	55		65		75		85
Gra H	T A A H C Q G S * S I N V T L G A H N I K E Q E R T Q Q F I P V K R P I P H P A						
Gra B	- - - - - W - - * -						- - - P - - - - - - - - - - - - - - -
Cat G	- - - - - W - - * N -				Q R R - N - - - H - T A R - A - R - - Q		
CHT	T A A H C G V T T S D V V V A G E F D Q G S S S E K I Q K L K I A K V F K N S K						
	95		105		115		125
Gra H	Y N P K N F S N D I M L L Q L E R K A K W T T A V R P L R L P S S K A Q V K P G						
Gra B	- -				R - R - - Q - - - - - - - - - - - - - - -		N - - - - - - - - - - -
Cat G	- - Q R T I Q -				S - R V R T N R N - N - V A - - R A Q E G L R - -		
CHT	Y N S L T I N N D I T L L K L S T A A S F S Q T V S A V C L P S A S D D F A A G						
	135		145		155		165
Gra H	Q L C S V A G W G * * Y V S M S T L A T T L Q E V L L T V Q K D C Q C E R L F H						
Gra B	- T - - - - - - - - - - * Q T A P L G K H S H - - - - - K M - - - E - R K - - - S D L *						
Cat G	T - - T - - - - - - * R V S M R R G * T D - - R - - Q - R M - R - R - - L - I - *						
CHT	T T C V T T G W G L T R Y T N A N T P D R L Q Q A S L P L L S N T N C K K Y W G						
	175		185		195		205
Gra H	G N Y S R A T E I C V G D P K K T Q T G F K G D S G G P L V C K * * * * D V A Q						
Gra B	R H - D S T I - L - - - - - E I K K - S -						N * * * * K - - - -
Cat G	- S - D P R R Q - - - - - R R E R K A A - - - - - - - - - - - L - N * * * * N - - H						
CHT	T K I K * D A M I C A G A S * * G V S S C M G D S G G P L V C K K N G A W T L V						
	215		225		235		245
Gra H	G I L S Y G N K K G T P * * P G V Y I K V S H F L P W I K R T M K R L						
Gra B	- - V - - - R N N - M - * * - R A C T - - - S - V H - - - K - - - - - Y						
Cat G	- - V - - - K S S - V - * * - E - F T R - - - S - - - - - R T - - R S F K L L D Q M E T P L						
CHT	G I V S W G S S T C S T S T P G V Y A R V T A L V N W V Q Q T L A A N						

Fig. 6. Alignment of the amino acid sequences of human granzyme H (Gra H), human granzyme B (Gra B), human cathepsin G (Cat G) and bovine α -chymotrypsin (CHT). The residues of Gra B and Cat G are printed at those positions that differ from the Gra H sequence; the residues that agree with the Gra H sequence are indicated by a dash. The asterisk (*) represents a gap that has been inserted to maximize sequence similarity. The numbering of amino acid residues shown in the top line is that of CHT. The residues of the catalytic triad, His57, Asp102 and Ser195, are enclosed in black boxes; asparagines that represent putative glycosylation sites are printed in white on a black background. Conserved cysteines found in all four sequences are boxed. Sequence data sources are: Gra B (=HLP; ref. 25), Cat G (ref. 44), CHT, NBRF protein data base, identification code KYBOA.

proteases to those previously described (22). Both distance values, K_S and K_A , shown in Table 1 lead to the same tree topology. Fig. 7 illustrates the evolutionary position of human granzyme H and its relationships to other granule serine proteases. The ancestral granzyme H gene is most closely related

to cathepsin G and granzyme B. Apparently it existed as a separate gene at the time of mammalian radiation. This conclusion is based on the branching pattern and branch lengths of the granzymes and on the K_A values of 0.307 ± 0.035 between granzyme H and cathepsin G and 0.280 ± 0.032 between

Table 1. Numbers of substitutions per synonymous site, K_S (above diagonal) and numbers of substitutions per nonsynonymous site, K_A (below diagonal), in granzyme sequences from which the codons 63 – 142 (chymotrypsin residue numbering) were excluded

	1 Human Cat G	2 Human Gra B	3 Mouse Gra B	4 Mouse Gra D	5 Mouse Gra E	6 Mouse Gra F	7 Mouse Gra G	8 Human Gra H	9 Mouse MCP2
1		0.817 0.116	0.970 0.132	1.263 0.188	1.169 0.168	1.240 0.204	1.342 0.198	0.698 0.119	NA
2	0.331 0.029		0.608 0.091	1.059 0.147	1.106 0.156	0.927 0.133	1.201 0.178	0.774 0.142	NA
3	0.362 0.032	0.229 0.023		1.121 0.155	1.149 0.162	1.121 0.159	1.203 0.175	0.919 0.162	NA
4	0.438 0.036	0.343 0.030	0.348 0.031		0.102 0.026	0.245 0.044	0.201 0.040	1.140 0.206	NA
5	0.425 0.036	0.349 0.031	0.351 0.031	0.045 0.009		0.239 0.043	0.195 0.039	1.197 0.205	NA
6	0.450 0.037	0.344 0.030	0.345 0.030	0.161 0.018	0.152 0.018		0.256 0.046	1.119 0.228	NA
7	0.423 0.035	0.332 0.029	0.338 0.030	0.119 0.015	0.101 0.014	0.125 0.016		1.266 0.238	NA
8	0.307 0.035	0.280 0.032	0.320 0.035	0.374 0.040	0.350 0.038	0.381 0.040	0.363 0.039		NA
9	0.485 0.040	0.448 0.036	0.473 0.038	0.505 0.040	0.481 0.039	0.471 0.038	0.466 0.038	0.478 0.048	

The standard error of the mean (\pm) is shown below each value. NA: no reliable estimate has been obtained.

granzyme H and granzyme B, both of which are significantly higher than the K_A value of 0.23 ± 0.02 between the orthologous human and murine granzymes B (see Table 1). Granzyme H diverged from the cathepsin G progenitor at a time shortly before mammalian radiation, when granzyme B already existed as a separate gene. These evolutionary data are consistent with genomic mapping studies showing that granzyme H is located between the cathepsin G and granzyme B gene on the same 50 kb fragment. Our results imply that granzyme H is not a human homologue of the murine granzymes C – G, and that a granzyme H homologue in rodent species has not yet been identified.

Discussion

The human genes for granzyme B and cathepsin G are located on the same chromosomal band (14q11 – q12) as the α and δ chains of the T cell antigen receptor (TCR α/δ) (29,30,44,45; Dorra Cherif, unpublished result). Recently, Hanson and collaborators established a continuous genomic restriction map for cathepsin G and two cathepsin G-like genes which were designated *CGL-1* and *CGL-2* (46). These three genes are tightly clustered in a 50 kb genomic region and have the same 5' to 3' orientation in the band 14q11.2. *CGL-1* was reported to be identical to granzyme B based on restriction site analyses. Comparing the published restriction map of *CGL-2* with that of granzyme H, we now conclude that *CGL-2* is identical to granzyme H. Thus, the gene order of the entire locus on the long arm of human chromosome 14 is given as (29): centromere – TCR α/δ – granzyme B – granzyme H – cathepsin G. Interestingly the three proximal genes, the TCR α/δ gene, the granzyme B and granzyme H genes, are specifically expressed in T cells, while the most distal gene, cathepsin G, is only transcribed in the myelocyte – monocyte lineage (44). In mouse, a complex of four

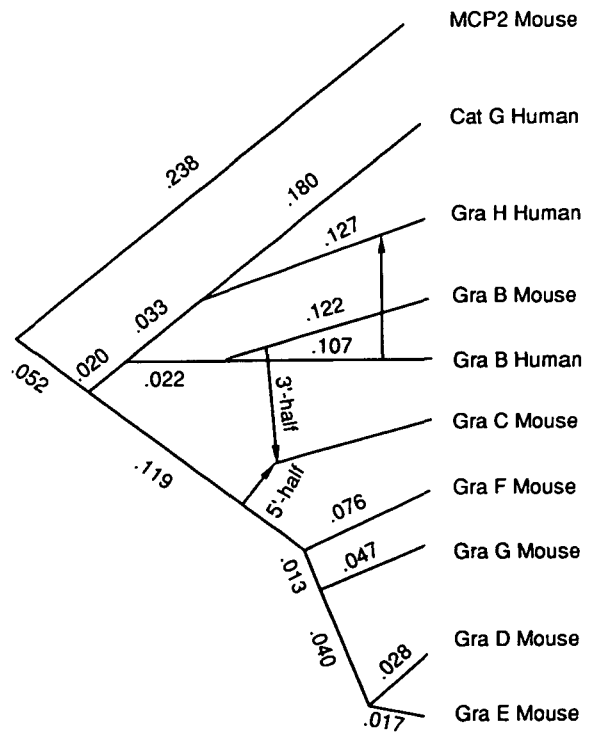


Fig. 7. Evolutionary tree of hematopoietic serine proteases showing the evolutionary position of human granzyme H. The branching order and branch lengths (not completely drawn to scale) were calculated from the distances between nucleotide sequences from which the region of sequence exchange (exon 3 and part of exon 4) between human granzyme B and granzyme H was excluded. The number on each branch indicates the number of nucleotide substitutions per nonsynonymous site. The evolutionary position of murine granzyme C is taken from Jenne *et al.* (22).

Table 2. Percentage identities of nucleotides at 4-fold degenerate sites between human granzyme H and human granzyme B and cathepsin G genes in different regions

Region	Intron 1	Intron 2	Intron 3	Intron 4	Exons* 1, 2, 4B, 5	Exon* 3, 4A
Gra B	52	54	94	42	61	89
Cat G	47	52	35	45	62	56

Only 4-fold degenerate sites have been compared in the coding regions.

serine protease genes (including the granzyme B gene) has been mapped to the D segment of mouse chromosome 14, in the vicinity of the TCR α/δ locus (29,47), suggesting homology between the q11–q12 region of human chromosome 14 and the D segment of mouse chromosome 14.

S1 protection assays which have both high sensitivity and specificity were performed to demonstrate that the granzyme H gene is specifically expressed in activated T cells. First, high sequence homology between human granzymes B and H within the converted region (99%) would have resulted in cross-hybridization with human granzyme B mRNA transcripts even under high stringency washing conditions. Secondly, expression levels of granzyme H after T cell activation are low compared to that of granzyme B. Northern blot results recently published for *hCCPX*/granzyme H (38) must be interpreted with caution, since the genomic 1.5 kb *Bam*HI fragment used as the probe (38) contained exon 3, which shows a 98.5% nucleotide homology to that of granzyme B (134 out of the 136 nucleotides of exon 3 are identical). Hybridization with such a probe, even under high stringency washing conditions, inevitably leads to hybridization with both granzyme transcripts, which are about the same size and are thus not distinguishable.

The choice of the name 'granzyme H' was made on the basis of genealogical relationships between granzyme H and other members of the granzyme family. Since granzyme H is more closely related to the neutrophil serine protease cathepsin G than to the murine granzymes C–G, the name 'human granzyme C' was avoided. The latter name should be chosen once a real human counterpart of the murine granzymes C–G has been identified. Furthermore the new name 'granzyme H' is justified by our finding that granzyme H evolved before mammalian radiation, whereas the granzymes C–G were generated after the rodent lineage had diverged from primates.

In order to exclude the hypothesis that high conservation between the granzymes B and H from position 63–142 (chymotrypsin residue numbering) is due to unusually tight functional or structural constraints in these granzymes, we compared the number of substitutions at 4-fold degenerate sites in this region with that in the remaining portion of the protein coding region and in intron 3 (Table 2). These sites are on average subjected to very weak selective constraints. The number of substitutions at 4-fold degenerate sites should be uniform over the entire length of the gene, irrespective of large differences in selective constraints at sites of replacement substitutions. However, if the substitution rate at 4-fold degenerate sites is much lower in the highly conserved region than outside of the boundaries, strong conservation of the protein sequence between residue 63 and residue 142 would indicate a recent replacement of granzyme H sequences by granzyme B sequences. This latter gene conver-

sion hypothesis is supported by our data (Table 2). The percentage of nucleotide substitutions at 4-fold degenerate sites in the converted coding region and in intron 3 is 11.0% and 6.0%, respectively (actual numbers: $12.0 \pm 6.3\%$ and $6.3 \pm 1.9\%$) as compared with 39% substitutions at 4-fold degenerate sites in the remaining protein coding sequence and 46–58% substitutions in intron sequences (Table 2).

The percentage divergence between human granzyme H and granzyme B in intron 3 can be used to estimate the time elapsed since the gene conversion event. These figures ($6.3 \pm 1.9\%$) suggest that the conversion took place at or shortly after the time that primates and Old World monkeys separated into two lineages, since the percentage of substitution at 4-fold degenerate sites in functionless DNA for human versus Old World monkey sequences amounts to $7.5 \pm 0.4\%$ (48). Assuming a divergence time of 25 million years for primates and Old World monkeys (34), the conversion dates back to 21 ± 6 million years before present.

In spite of the close association between granzyme gene expression and cytotoxic capabilities of activated T cells, NK cells, and LAK cells *in vitro*, the role of the three human granzyme genes A, B, and H remains unclear. With the cloning of the third human lymphocyte specific granzyme, further experiments at the molecular genetic level can be designed to explore this intriguing phenomenon.

Acknowledgements

We thank Jacques Soudon for help in S1 protection analysis, Eva Cesar, Annie Soulié and Jacqueline Guillet for cell cultures, Christian-Jacques Larsen, Olivier Bernard, David Grausz and W.-H. Li for helpful discussion, Robin Nancel and Jacqueline Boisse for preparing photographs, and Ligia Caro for the preparation of the manuscript. This work was supported by a grant from the Association pour la Recherche contre le Cancer (ARC) and from the Swiss National Science Foundation.

Abbreviations

CRE	cyclic AMP responsive element
CTL	cytotoxic T lymphocyte
IL2	interleukin 2
LAK	lymphokine activated killer
Myr	million year before present
NK	natural killer
PBL	peripheral blood lymphocyte
PCR	polymerase chain reaction
PHA	phytohemagglutinin
TCR	T cell antigen receptor
TPA	12-o-tetradecanoylphorbol-13-acetate
TRE	TPA responsive element
UPG	unweighted pair group

References

- 1 Dourmashkin, R. R., Deteix, P., Simone, C. B., and Henkart, P. 1980. Electron microscopic demonstration of lesions in target cell membranes associated with antibody-dependent cellular cytotoxicity. *Clin. Exp. Immunol.* 42:554.
- 2 Henkart, P. A., Millard, P. J., Reynolds, C. W., and Henkart, M. P. 1984. Cytolytic activity of purified cytoplasmic granules from cytotoxic rat large granular lymphocyte tumors. *J. Exp. Med.* 160:75.
- 3 Podack, E. R. and Königsberg, P. J. 1984. Cytolytic T cell granules. Isolation, structural, biochemical, and functional characterization. *J. Exp. Med.* 160:695.
- 4 Tschopp, J. and Nabholz, M. 1990. Perforin-mediated target cell lysis by cytolytic T lymphocytes. *Annu. Rev. Immunol.* 8:279.
- 5 Lichtenheld, M. G., Olsen, K. J., Lu, P., Lowrey, D. M., Hameed, A., Hengartner, H., and Podack, E. R. 1988. Structure and function of human perforin. *Nature* 335:448.
- 6 Kuta, A. E., Reynolds, C. R., and Henkart, P. A. 1989. Mechanism of lysis by large granular lymphocyte granule cytotoxicity: generation of a stable cytolytic-RBC intermediate. *J. Immunol.* 142:4378.
- 7 Lichtenheld, M. G., and Podack, E. R. 1989. Structure of the human perforin gene. A simple gene organization with interesting potential regulatory sequences. *J. Immunol.* 143:4267.
- 8 Shinkai, Y., Yoshida, M. C., Maeda, K., Kobata, T., Maruyama, K., Yodoi, J., Yagita, H., and Okumura, K. 1989. Molecular cloning and chromosomal assignment of a human perforin (PFP) gene. *Immunogenetics* 30:452.
- 9 Tschopp, J., Schäfer, S., Masson, D., Peitsch, M. C., and Heusser, C. 1989. Phosphorylcholine acts as a Ca^{2+} -dependent receptor molecule for lymphocyte perforin. *Nature* 337:272.
- 10 Pasternack, M. S., and Eisen, H. N. 1985. A novel serine esterase expressed by cytolytic T lymphocytes. *Nature* 314:743.
- 11 Pasternack, M. S., Verret, C. R., Liu, M. A., and Eisen, H. N. 1986. Serine esterase in cytolytic T lymphocytes. *Nature* 322:740.
- 12 Young, J. D. E., Leong, L. G., Liu, C. C., Damiano, A., Wall, D. A., and Cohn, Z. A. 1986. Isolation and characterization of a serine esterase from cytolytic T cell granules. *Cell* 47:183.
- 13 Simon, M. M., Hoschützky, H., Fruth, U., Simon, H.-G., and Kramer, M. D. 1986. Purification and characterization of a T cell specific serine proteinase (TSP-1) from cloned cytolytic T lymphocytes. *EMBO J.* 5:3267.
- 14 Redmond M. J., Letellier, M., Parker, J. M. R., Lobe, C., Havele, C., Paetkau, V., and Bleackley, R. C. 1987. A serine protease (CCP1) is sequestered in the cytoplasmic granules of cytotoxic T lymphocytes. *J. Immunol.* 139:3184.
- 15 Masson, D. and Tschopp, J. 1987. A family of serine esterases in lytic granules of cytolytic T lymphocytes. *Cell* 49:679.
- 16 Brunet, J. F., Dossetto, M., Denizot, F., Mattei, M. G., Clark, W. R., Haqqi, T. M., Ferrier, P., Nabholz, M., Schmitt-Verhulst, A. M., Luciani, M. F., and Golstein, P. (1986). The inducible cytotoxic-T-lymphocyte-associated gene transcript CTLA-1 sequence and gene localization to mouse chromosome 14. *Nature* 322:268.
- 17 Gershenfeld, H. K. and Weissman, I. L. 1986. Cloning of a cDNA for a T cell-specific serine protease from a cytotoxic T lymphocyte. *Science* 232:854.
- 18 Lobe, C. G., Finlay, B. B., Paranchych, W., Paetkau, V. H., and Bleackley, R. C. 1986. Novel serine proteases encoded by two cytotoxic T lymphocyte-specific genes. *Science* 232:858.
- 19 Kwon, B. S., Kestler, D., Lee, E., Wakulchik, M., and Young, J. D.-E. 1988. Isolation and sequence analysis of serine protease cDNAs from mouse cytolytic T lymphocytes. *J. Exp. Med.* 168:1839.
- 20 Jenne, D., Rey, C., Masson, D., Stanley, K. K., Herz, J., Plaetinck, G., and Tschopp, J. 1988. cDNA cloning of granzyme C, a granule-associated serine protease of cytolytic T lymphocytes. *J. Immunol.* 140:318.
- 21 Jenne, D., Rey, C., Haefliger, J.-A., Qiao, B. Y., Groscurth, P., and Tschopp, J. 1988. Identification and sequencing of cDNA clones encoding the granule-associated serine proteases granzymes D, E, and F of cytolytic T lymphocytes. *Proc. Natl. Acad. Sci. USA* 85:4814.
- 22 Jenne, D. E., Masson, D., Zimmer, M., Haefliger, J.-A., Li, W.-H., and Tschopp, J. 1989. Isolation and complete structure of the lymphocyte serine protease granzyme G, a novel member of the granzyme multigene family in murine cytolytic T lymphocytes. Evolutionary origin of lymphocyte proteases. *Biochemistry* 28:7953.
- 23 Hameed, A., Lowrey, D. M., Lichtenheld, M. G., and Podack, E. R. 1988. Characterization of three serine esterases isolated from human IL-2 activated killer cells. *J. Immunol.* 141:3142.
- 24 Krähenbühl, O., Rey, C., Jenne D., Lanzavecchia, A., Groscurth, P., Carrel, S., and Tschopp, J. 1988. Characterization of granzymes A and B isolated from granules of cloned human cytotoxic T lymphocytes. *J. Immunol.* 141:3471.
- 25 Schmid, J. and Weissmann, C. 1987. Induction of mRNA for a serine protease and a β -thromboglobulin-like protein in mitogen-stimulated human leukocytes. *J. Immunol.* 139:250.
- 26 Caputo, A., Fahey, D., Lloyd, C., Vozab, R., McCairns, E., and Rowe, P. B. 1988. Structure and differential mechanisms of regulation of expression of a serine esterase gene in activated human T lymphocytes. *J. Biol. Chem.* 5:6363.
- 27 Trapani, J. A., Klein, J. L., White, P. C., and Dupont, B. 1988. Molecular cloning of an inducible serine esterase gene from human cytotoxic lymphocytes. *Proc. Natl. Acad. Sci. USA* 85:6924.
- 28 Gershenfeld, H. K., Hershberger, R. J., Shows, T. B., and Weissman I. L. 1988. Cloning and chromosomal assignment of a human cDNA encoding a T cell- and natural killer cell-specific trypsin-like serine protease. *Proc. Natl. Acad. Sci. USA* 85:1184.
- 29 Harper, K., Mattei, M. G., Simon, D., Suzan, M., Guénet, J. L., Haddad, P., Sasportes, M., and Golstein, P. 1988. Proximity of the CTLA-1 serine esterase and TCR α loci in mouse and man. *Immunogenetics* 28:439.
- 30 Haddad, P., Clément, M. V., Bernard, O., Larsen, C.-J., Degos, L., Sasportes, M., and Mathieu-Mahul, D. 1990. Structural organization of the hCTLA-1 gene encoding human granzyme B. *Gene* 87:265.
- 31 Klein, J. L., Shows, T. B., Dupont, B., and Trapani, J. A. 1989. Genomic organization and chromosomal assignment for a serine protease gene (CSPB) expressed by human cytotoxic lymphocytes. *Genomics* 5:110.
- 32 Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463.
- 33 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.
- 34 Li, W.-H., Tanimura, M., and Sharp, P. M. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* 25:330.
- 35 Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- 36 Sneath, P. H. and Sokal, R. R. 1973. *Numerical Taxonomy*. Freeman, San Francisco, CA.
- 37 Li, W.-H. 1981. Simple method for constructing phylogenetic trees from distance matrices. *Proc. Natl. Acad. Sci. USA* 78:1085.
- 38 Meier, M., Kwong, P. C., Frégeau, C. J., Atkinson, E. A., Burrington, M., Ehrman, N., Sorensen, O., Lin, C. C., Wilkins, J., and Bleackley, R. C. 1990. Cloning of a gene that encodes a new member of the human cytotoxic cell protease family. *Biochemistry* 29:4042.
- 39 Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P., and Karin, M. 1987. Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. *Cell* 49:729.
- 40 Fujita, T., Shibuya, H., Ohashi, T., Yamanishi, K., and Taniguchi, T. 1986. Regulation of human interleukin-2 gene: functional DNA sequences in the 5' flanking region for the gene expression in activated T lymphocytes. *Cell* 46:401.
- 41 Durand, D. B., Shaw, J.-P., Bush, M. R., Replogle, R. E., Belagaje, R., and Crabtree, G. R. 1988. Characterization of antigen receptor response elements within the interleukin-2 enhancer. *Mol. Cell. Biol.* 8:1715.
- 42 Roesler, W. J., Vandenbark, G. R., and Hanson, R. W. 1988. Cyclic AMP and the induction of eukaryotic gene transcription. *J. Biol. Chem.* 263:9063.
- 43 Berget, M. S. 1984. Are U4 small nuclear ribonucleoproteins involved in polyadenylation? *Nature* 309, 179 – 182.
- 44 Hohn, P. A., Popescu, N. C., Hanson, R. D., Salvesen, G., and Ley, T. J. 1989. Genomic organization and chromosomal localization of the human cathepsin G gene. *J. Biol. Chem.* 264:13412.
- 45 Lin, C. C., Meier, M., Sorensen, O., Sasi, R., Tainaka, T., and

66 *Granzyme H gene structure and evolution*

- Bleakley, R. C. 1990. Chromosome localization of two human serine protease genes to region 14q11.2→q12 by in situ hybridization. *Cytogenet. Cell. Genet.* 53:169.
- 46 Hanson, R. D., Hohn, P. A., Popescu, N. C., and Ley, T. J. 1990. A cluster of hematopoietic serine protease genes is found on the same chromosomal band as the human α/δ T-cell receptor locus. *Proc. Natl. Acad. Sci. USA* 87:960.
- 47 Crosby, J. L., Bleakley, R. C., and Nadeau, J. H. 1990. A complex of serine protease genes expressed preferentially in cytotoxic-T-lymphocytes is closely linked to the T-cell receptor α - and δ -chain on mouse chromosome 14. *Genomics* 6:252.
- 48 Wolfe, K. H., Sharp, P. M., and Li, W. H. 1989. Mutation rates differ among regions of the mammalian genome. *Nature* 337:283.
- 49 Chen, L. K., Mathieu-Mahul, D., Bach, F. H., Dausset, J., Bensussan, A., and Saspportes, M. 1986. Recombinant interferon α can induce rearrangement of T-cell antigen receptor α -chain genes and maturation to cytotoxicity in T-lymphocyte clones in vitro. *Proc. Natl. Acad. Sci., USA* 83:4887.