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[†,‡]

Dieter E. Jenne,[*,§] Danièle Masson,[§] Michael Zimmer,[∥] Jacques-Antoine Haefliger,[§] Wen-Hsiung Li,[⊥] and Jürg Tschopp[§]

Institut de Biochimie, Université de Lausanne, Chemin des Boveresses, CH-1066 Epalinges s/Lausanne, Switzerland, Imperial Cancer Research Fund Laboratories, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, Great Britain, and Center for Demographic and Population Genetics, University of Texas, P.O. Box 20334, Houston, Texas 77225

Received March 9, 1989; Revised Manuscript Received June 8, 1989

ABSTRACT: A cDNA clone that is closely related to the granule-associated serine proteases of cytolytic T lymphocytes (CTL), called granzymes A–F, was isolated from a CTL expression library. The encoded serine protease, granzyme G, shows 70%–89% nucleotide identities to the granzymes C–F and, like those, consists of 228 amino acids preceded by the short propeptide Glu-Glu and a 18 residue long signal peptide. Granzyme G was identified by amino-terminal sequence analysis as a correctly processed and sorted protein stored in lysosome-like granules. The phylogenetic history of the granzyme multigene family was reconstructed by two tree-making methods and by Southern blot analyses of human, rat, and mouse DNA. Our results indicate differences in the evolutionary pathway between these species. The murine granzymes C–G descended from a progenitor present at the time of mammalian radiation. Granzyme C branched off first after the primate–rodent split and was involved in a recombination event with granzyme B before the rat–mouse divergence. Granzymes D and E have diverged after the mouse–rat speciation. However, no experimental evidence for the existence of a granzyme C-D-E-F-G equivalent was found in humans, and loss of the ancestral gene in the primate lineage is discussed. In view of the species differences in the number of granzyme gene copies during recent evolution, we propose that the murine granzymes B–G play several distinct roles in CTL-mediated effector functions as a response to quite recent changes of the biochemical environment.

Killer cell mediated cytotoxicity has been shown to play an important role in surveillance against viral infections, in the rejection of allografts, and in the elimination of tumor cells (Mullbacher & Ada, 1987; Mueller et al., 1988). Recent biochemical studies have provided compelling circumstantial evidence in favor of a granule exocytosis pathway as an important mechanism of killer cell mediated cytotoxicity (Podack, 1986; Henkart & Yue, 1987; Young & Cohn, 1987; Sitkovsky, 1988; Tschopp & Jongeneel, 1988). It has been demonstrated that cytoplasmic granules of CTLs[1] are secreted toward the target cell membrane in a vectorial fashion when CTL adhere to and lyse target cells (Zagury, 1982). The purified cytoplasmic granules have been shown to lyse a variety of nucleated cells including tumor cells and erythrocytes without specificity (Henkart et al., 1984; Podack & Königsberg, 1984). The cytolytic, hemolytic, and BLT-hydrolytic activities of CTLs released during target cell attack have been found to reside within the dense secretory granules of CTLs (Dennert & Podack, 1983; Henkart et al., 1984; Schmidt et al., 1985; Masson et al., 1985, 1986; Pasternack & Eisen, 1985; Pasternack et al., 1986; Young et al., 1986b; Henkart et al., 1987; Krähenbühl et al., 1988). One potent lytic protein called perforin/cytolysin has been identified as a granule component which can damage target cell membranes by forming transmembrane lesions (Blumenthal et al., 1984; Criado et al., 1985; Masson & Tschopp, 1985; Podack et al., 1985; Young et al. 1986a; Zalman et al., 1986). Other major constituents of cytolytic granules belong to a family of highly related serine proteases, designated as lymphocyte granzymes whose physiological functions are still unknown (Bleackley et al., 1988a; Jenne & Tschopp, 1988). So far, the six granzymes A–F have been distinguished in cultured murine CTL lines at the protein level (Pasternack et al., 1986; Young et al., 1986b; Simon et al., 1986; Masson & Tschopp, 1987; Redmond et al., 1987) as well as by cDNA sequencing (Gershenfeld et al., 1986; Lobe et al., 1986, 1988; Brunet et al., 1986; Jenne et al., 1988a,b; Bleackley et al., 1988b). The names granzyme G and H (Masson & Tschopp, 1987) have been retracted by the authors (Jenne & Tschopp, 1988), since the thereby identified serine esterases turned out to be indistinguishable from granzyme B (see Table I for a synopsis on current designations for lymphocyte serine proteases). A thorough analysis of the granule proteins and the cDNA clones previously isolated (Jenne et al., 1988b) pointed to the existence of a seventh granzyme. Here we report the characterization of this serine protease, which we call granzyme G, and also the genealogical relationships among all seven murine granzymes and the two known human lymphocyte granzymes A (Fruth et al., 1987;

[†] Supported by BSF Grant 87-00359 and NIH Grant GM 30998 (to W.-H.L.) and by grants from the Swiss National Science Foundation and Hoffmann-La Roche, Switzerland.

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02872.

[*] Author to whom correspondence should be addressed.

[§] Université de Lausanne.

[∥] Imperial Cancer Research Fund Laboratories.

[⊥] University of Texas.

[1] Abbreviations: CTL, cytolytic T lymphocyte; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; HLP, human lymphocyte protease; DFP, diisopropyl fluorophosphate; bp, base pairs; kbp, kilobase pairs; BLT, N^{α}-(benzyloxycarbonyl)-L-lysine thiobenzyl ester; UPG method, unweighted pair group method; RMCP, rat mast cell protease; PTH, phenylthiohydantoin.

0006-2960/89/0428-7953$01.50/0 © 1989 American Chemical Society

Table I: Nomenclature of Lymphocyte Granzymes[a]

human lymphocyte granzymes		murine lymphocyte granzymes	
synonyms	reference	synonyms	reference
granzyme A	Krähenbühl et al., 1988	granzyme A	Masson & Tschopp, 1987
HuTSP	Fruth et al., 1987		Jenne & Tschopp, 1988
H factor	Gershenfeld et al., 1988	H factor	Gershenfeld et al., 1986
CTL tryptase	Poe et al., 1988	TSP-1	Simon et al., 1986
granzyme 1	Hameed et al., 1988	BLT-esterase	Pasternack et al., 1986
		CTLA-3	Brunet et al., 1986
		SE 1	Young et al., 1986b
granzyme B	Krähenbühl et al., 1988	granzyme B	Masson & Tschopp, 1987
HLP	Schmid & Weismann, 1987	CCP1	Lobe et al., 1986
HSE26.1	Trapani et al., 1988		Lobe et al., 1988
SECT	Caputo et al., 1988	CTLA-1	Brunet et al., 1986
granzyme 2	Hameed et al., 1988	granzyme C	Jenne et al., 1988a
		CCP2	Lobe et al., 1986
			Lobe et al., 1988
		granzyme D	Jenne et al., 1988b
		granzyme E	Jenne et al., 1988b
		CCP3	Bleackley et al., 1988b
		granzyme F	Jenne et al., 1988b
		CCP4	Bleackley et al., 1988b
		granzyme G	Jenne et al., this paper

[a] Proper assignment of synonyms to granzymes B–G is based on cDNA sequence or partial protein sequence data. Other names for human and murine granzyme A are also listed when the identity could be ascertained from the unique physicochemical, biochemical, and functional properties (Jenne & Tschopp, 1988).

Krähenbühl et al., 1988; Gershenfeld et al., 1988; Hameed et al., 1988; Poe et al., 1988) and B (human lymphocyte protease, HLP) (Schmid & Weismann, 1987; Krähenbühl et al., 1988; Caputo et al., 1988; Trapani et al., 1988; Hameed et al., 1988). The observed differences in the phylogenetic history of lymphocyte proteases in the human, mouse, and rat species are discussed as an indication for species-specific, disparate biological functions.

Materials and Methods

Gene Cloning and Sequencing. All methods relating to the cloning and sequencing of granzyme G were performed according to standard protocols (Maniatis et al., 1982) or as previously described (Jenne et al., 1988a,b). The cDNA sequence was assembled from random subclones using sonication to fragment the cDNA (Deininger, 1983). In this way both DNA strands were sequenced, and each position of the cDNA sequence was covered by at least three independent clones. Genomic granzyme clones were isolated from a human genomic library of blood leukocyte DNA, constructed in the cosmid vector pcos 2 EMBL (Poustka et al., 1984). The cosmid library was screened according to the protocol described by Herrmann et al. (1987). The nylon filters were hybridized at 50 °C overnight in hybridization buffer containing 0.5 M sodium phosphate buffer, pH 7.2, and 7% SDS. They were washed in 40 mM sodium phosphate buffer and 1% SDS, pH 7.2, at 50 °C for 30 min and exposed for 24 h at −70 °C.

Southern Blot Analysis. Total genomic DNA was extracted from the spleen of C57Bl/6 mice and Kfm Wistar rats according to a published procedure (Miller et al., 1988). Total human DNA was prepared from peripheral blood lymphocytes following standard procedures (Herrmann & Frischauf, 1987). Restriction fragments of 5 μg of total genomic DNA were size-fractionated by gel electrophoresis in 0.8% agarose and were vacuum-blotted (LKB Vacuum Blotting System) onto Zeta-Probe blotting membranes (Bio-Rad Laboratories) in 0.4 M NaOH (Reed & Mann, 1985) according to the manufacturers' instructions.

DNA Probes. The cDNA of the granzyme D clone 2.4a (Jenne et al., 1988b) was used for Southern blot hybridizations and library screenings. The cDNAs of murine granzyme D and human granzyme B (HLP) (a generous gift of C. Weissmann, Institut für Molekularbiologie, Universität Zürich, Zürich, Switzerland) were labeled with [α-^{32}P]dCTP (111 TBq/mmol, Amersham) by using random primers (Feinberg & Vogelstein, 1983), and the specific oligonucleotide for human granzyme B (antisense strand of the nucleotide sequence from position 41 to 61: CTG CTG CCC AGG GCA GAT GCA; Schmid & Weissmann, 1987) was labeled at its 5′ end by using [γ-^{32}P]ATP (185 TBq/mmol, Amersham) as described (Jenne & Stanley, 1987; Jenne et al., 1988a,b).

Biochemical Identification and Purification of Granzyme G. Isolation of murine CTL granules from the B6.1 cell line, separation of granule proteins, and [^{3}H]DFP labeling were done exactly as previously described (Masson & Tschopp, 1987). The amino-terminal sequence of purified granzyme G was determined with an Applied Biosystems 470A/120A gas-phase protein sequencer/PTH-amino acid analyzer system according to the protocols supplied by the manufacturer. One hundred picomoles of purified granzyme G was subjected to Edman degradation and the phenylthiohydantoin (PTH) derivatized amino acids were identified and quantitated by reverse-phase HPLC. The initial yield recovery amounted to 70%.

Statistical Analyses of the Phylogenetic Tree from Nucleotide Sequences. To estimate the actual number of nucleotide substitutions between two homologous genes, we used the method of Li et al. (1985). As described in Li et al. (1987), we simply express the results for 4-fold degenerate, 2-fold degenerate, and nondegenerate nucleotide sites in terms of synonymous and nonsynonymous sites by counting a nucleotide site as one synonymous site when all possible changes at that site are synonymous, and as one-third or two-thirds of a synonymous site if one or two of the three possible changes are synonymous. The number of nucleotide substitutions per nonsynonymous site (K_A), in some cases also the number of nucleotide substitutions per synonymous site (K_S), was used to estimate the sequence distances among genes. Phylogenetic trees were reconstructed according to the UPG (Sneath & Sokal, 1973) and the modified UPG methods (Li, 1981).

Results

Identification of Granzyme G cDNA Clones. A rabbit serum that specifically recognizes granzymes C–F was pre-

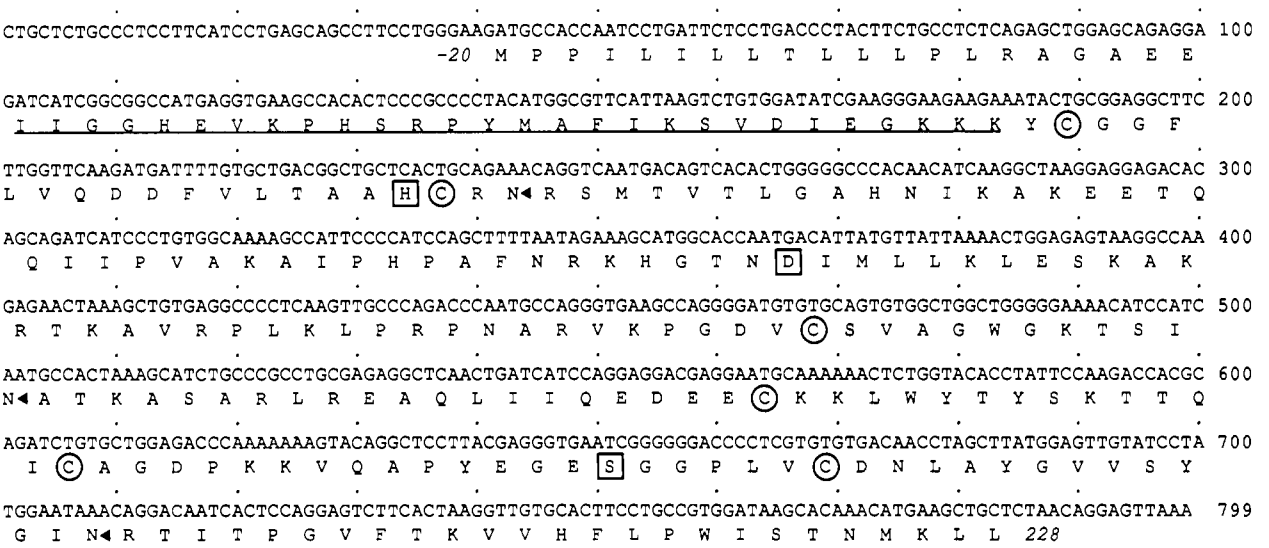

FIGURE 1: The nucleotide and amino acid sequences of the lymphocyte protease granzyme G. The numbering of the cDNA sequence begins with the first nucleotide of clone 13.4. The amino acid sequence deduced from the cDNA is shown in single-letter code under the first nucleotide of each codon. The underlined protein sequence was determined by amino acid sequencing of purified granzyme G found in fraction 60 (see Figure 2). Amino acid residues −20 to −3 represent the signal peptide; the two residues at positions −2 and −1, the propeptide of granzyme G. The residues of the mature enzyme are numbered from 1 to 228. Putative N-linked glycosylation sites are indicated by filled triangles. The six cysteines are circled and form disulfide bonds in a 1–2, 3–6, and 4–5 pattern by analogy to rat mast cell protease II. The active-site residues H-45, D-89, and S-184 are boxed.

Table II: Similarities between Granzyme Serine Protease Sequences[a]

	1 human Cat G	2 human Gra A	3 mouse Gra A	4 human Gra B	5 mouse Gra B	6 mouse Gra C	7 mouse Gra D	8 mouse Gra E	9 mouse Gra F	10 mouse Gra G	11 rat RMCP II
1		47.2	47.5	67.5	64.9	62.2	59.9	59.7	59.7	59.9	55.2
2	35.7		76.8	48.2	48.5	47.9	48.5	49.9	46.6	48.0	48.8
3	37.3	68.5		50.1	47.6	48.0	47.2	48.3	47.3	46.2	47.4
4	56.1	39.4	39.8		75.6	70.1	65.3	64.0	66.4	65.2	54.6
5	54.7	42.5	43.3	67.9		75.2	64.5	64.3	64.6	64.4	54.2
6	51.6	37.1	40.7	59.8	66.0		70.4	69.4	69.9	70.3	54.9
7	46.4	38.1	38.5	54.9	57.1	58.5		94.5	84.1	87.6	52.6
8	46.9	39.0	39.8	53.1	56.0	58.1	90.0		84.6	88.7	52.6
9	46.4	38.3	40.3	52.9	57.1	59.3	73.8	75.5		86.3	54.7
10	46.8	37.1	38.3	54.5	57.9	60.9	78.2	81.7	79.4		53.0
11	49.4	38.9	36.8	46.3	48.2	46.6	46.2	44.0	47.0	44.9	

[a] Numbers of nucleotide identities (above diagonal) and numbers of amino acid identities (below diagonal) per hundred residues between cathepsin G (Cat G), human and mouse granzyme (Gra) A, human and mouse granzyme B (Gra B), mouse granzymes (Gra) C–G, and rat mast cell protease (RMCP) II aligned according to Figure 3 have been calculated. The number of perfect matches over the region from position −28 to 245A (see Figure 3) has been divided by the length of the shorter sequence without counting gaps.

viously used to isolate a large number of expression clones from a murine CTL-specific expression library (Jenne et al., 1988b). A thorough examination of the cDNA clones obtained provided evidence that a further functional gene closely related to the granzymes D–F was expressed in the murine CTL line B6.1. By restriction site analysis, we found a subset of cDNA clones that differed from those coding for granzymes C–F (Jenne et al., 1988a,b). The molecular weight of the expressed fusion protein obtained from one of the clones (13.4) agreed with that of granzymes D–F.

The cDNA-Derived Amino Acid Sequence of Granzyme G. Clone 13.4 exhibited the largest cDNA insert and was sequenced on both DNA strands by using the shotgun strategy for sequencing subclones in M13. The nucleotide sequence is 799 bp long and contains 41 bp of untranslated region at the 5′ end and 14 bp of untranslated region at the 3′ end. The start site of the open reading frame is found at positions 42–44 (ATG) within a sequence that agrees with the consensus features of translation initiation sites (Kozak, 1984). The open reading frame (Figure 1) codes for 248 amino acid residues, of which the first 18 amino acids represent a typical signal peptide (Von Heijne, 1986). The amino acid sequence of the predicted protein shows high sequence similarities to previously published CTL serine proteases, especially to granzymes D–F (see Table II). As described for other granzymes (Jenne & Tschopp, 1988), the 18 residue long leader peptide of granzyme G is followed by a putative two residue long, negatively charged propeptide which may function as a spacer between the hydrophobic signal peptide and the hydrophobic amino terminus of the mature, active form of granzyme G. The three key residues forming the charge relay system in the serine protease superfamily (Kraut, 1977) are conserved at homologous positions in granzyme G as well. We, therefore, suggest that granzyme G is a further granule-associated, functional serine protease of cytolytic T cells.

Biochemical Identification of Granzyme G in CTL Granules. Soluble proteins of CTL granules were isolated from the mouse H2-D^d-specific cell line B6.1 and fractionated by Mono S cation-exchange chromatography as described (Masson & Tschopp, 1987). The fractions eluted from the column with NaCl concentrations in the range of 50–300 mM NaCl were analyzed by SDS–PAGE (Figure 2, upper panel) and tested for their reactivity with the serine esterase catalytic site specific affinity label [^{3}H]DFP (Figure 2, lower panel).

As noted previously (Masson & Tschopp, 1987), granzyme D is heterogeneous in size and charge due to differential

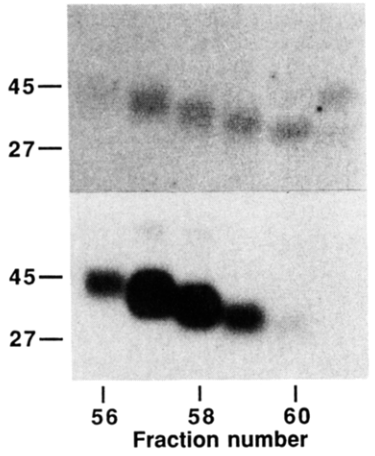

FIGURE 2: Separation of granzymes D and G by Mono S cation-exchange chromatography and labeling with [^{3}H]DFP. Upper panel: Fractions 56–60, obtained by eluting solubilized total granule proteins from the FPLC Mono S column with increasing salt concentrations, were analyzed by SDS–PAGE under nonreducing conditions followed by Coomassie blue staining. Lower panel: Fluorogram showing strong [^{3}H]DFP labeling of granzyme D in the fractions 56–59, but faint [^{3}H]DFP labeling of granzyme G in fraction 60. The fluorograph was exposed at −50 °C for 1 day.

glycosylation. All isoforms are eluted at the beginning of the salt gradient (fractions 56–59). Granzyme D reacts strongly with the [^{3}H]DFP label in contrast to the granzymes B, C, E, and F. The protein (fraction 60) that eluted just after the last granzyme D isoform, however, bound much less [^{3}H]DFP than would be expected for an equivalent amount of granzyme D. This protein was subjected to amino acid sequence analysis, and the first 28 amino-terminal residues were determined (Figure 1, underlined sequence). The amino terminus agreed completely with the cDNA-derived protein sequence for granzyme G and differed from the amino-terminal sequences established for the known granzymes A–F.

Sequence Comparisons of the Granzyme Multigene Family. The granzyme sequences were first aligned at the amino acid level. Regions of high sequence variability were aligned by minimizing the number of gaps, the number of nonconservative amino acid replacements, and the number of nucleotide substitutions (see Figure 3). The results of all pairwise comparisons at the amino acid and nucleotide level are summarized in Table II. The nucleotide identities are highest among granzymes D–G, ranging from 84% to 94%. Granzyme G shows, e.g., 86%, 88%, and 89% nucleotide identities with granzymes F, D, and E, respectively. The most closely related granzymes are granzymes D and E, having 94% nucleotide identities. Other serine protease sequences from the sequence data libraries which are orthologous genes in the human and mouse species have about 75% nucleotide identities. These are the serine protease domain of complement factor B (EMBL accession codes: HSMHBGEN, MMMHBDA), that of urokinase (HSUKM1, MMURKR), the pancreatic proelastase II (HSELAP2A, HSELAP2B, MMELA2R) (human

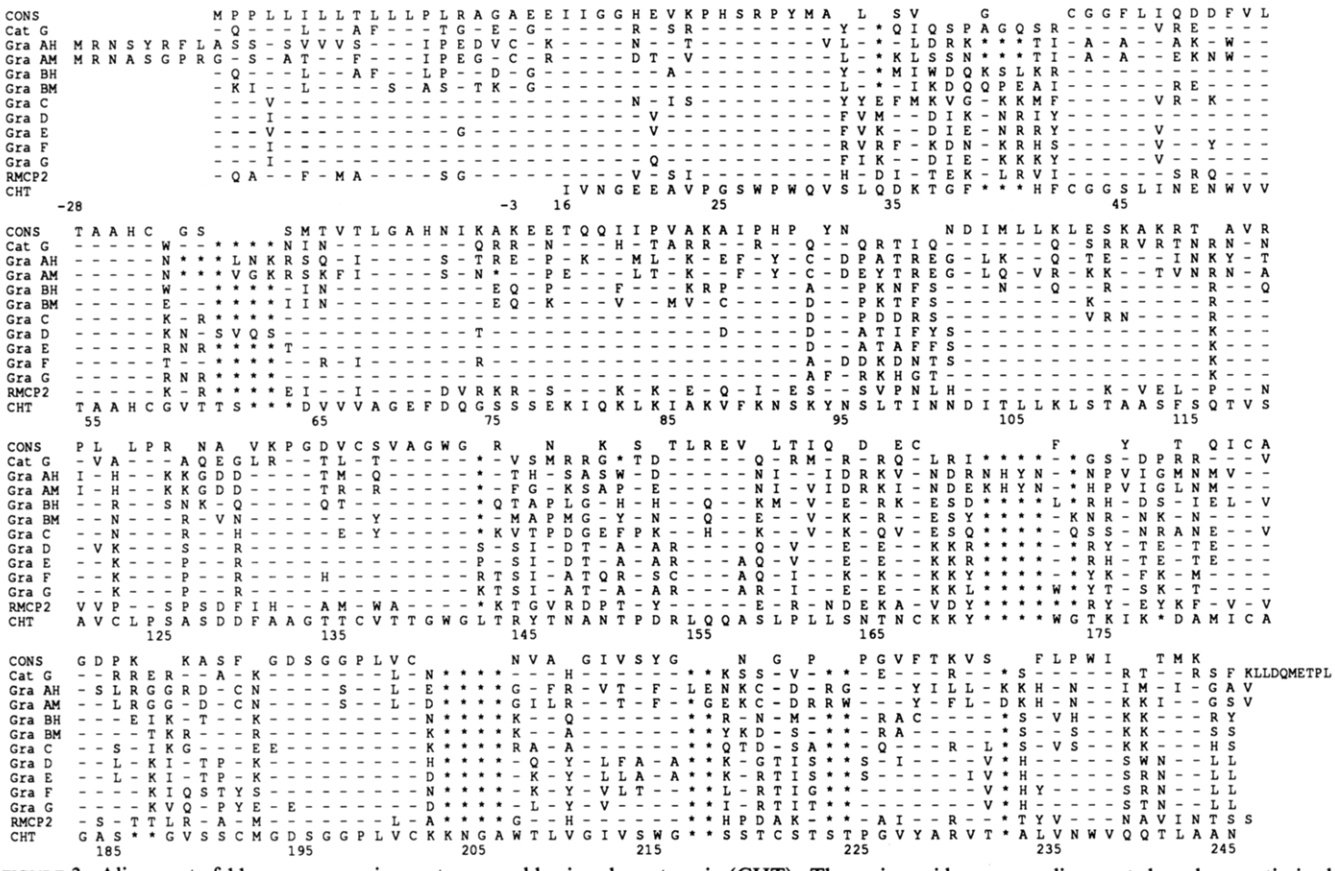

FIGURE 3: Alignment of 11 granzyme serine proteases and bovine chymotrypsin (CHT). The amino acid sequence alignments have been optimized at the codon level by minimizing the number of nucleotide replacements among all nucleotide sequences. The top line named CONS shows a consensus amino acid residue when at least five residues in the aligned 11 granzyme sequences are identical. The residues of the aligned sequences are printed only at those positions that differ from the consensus residues, whereas all residues that agree with the consensus sequence are displayed by a dash. The asterisk (*) indicates a gap inserted to maximize the sequence similarity. The numbering of amino acid residues is taken from CHT (NBRF protein identification code, KYBOA) as shown in the bottom line. Sequence data sources: cathepsin G (Cat G), Salvesen et al. (1987); human granzyme A (Gra AH), Gershenfeld et al. (1988); mouse granzyme A (Gra AM), Gershenfeld et al. (1987); human granzyme B (Gra BH), Schmid and Weissmann (1987); mouse granzyme B (Gra BM), Brunet et al. (1986) and Lobe et al. (1986); granzyme C (Gra C), Jenne et al. (1988a); granzyme D (Gra D), Jenne et al. (1988b); granzymes E and F (Gra E, Gra F), Jenne et al. (1988b) and Bleackley et al. (1988b); granzyme G (Gra G), this report; rat mast cell protease II (RMCP2), Benfey et al. (1987).

Table III: Numbers of Substitutions per Synonymous Site, K_S (above Diagonal), and Numbers of Substitutions per Nonsynonymous Site, K_A (below Diagonal)[a]

	1 human Cat G	2 human Gra A	3 mouse Gra A	4 human Gra B	5 mouse Gra B	6 mouse Gra C	7 mouse Gra D	8 mouse Gra E	9 mouse Gra F	10 mouse Gra G	11 rat RMCP II
1		NA	NA	0.886 ± 0.13	0.993 ±0.14	1.175 ±0.18	1.292 ±0.20	1.220 ±0.18	1.335 ±0.26	1.385 ±0.21	NA
2	0.583 ±0.05		0.605 ±0.09	NA	NA	NA	NA	NA	NA	NA	NA
3	0.572 ±0.05	0.181 ±0.02		NA	NA	NA	NA	NA	NA	NA	NA
4	0.329 ±0.03	0.570 ±0.05	0.551 ±0.05		0.612 ±0.10	0.880 ±0.14	1.120 ±0.16	1.171 ±0.18	0.944 ±0.14	1.189 ±0.18	NA
5	0.354 ±0.03	0.529 ±0.04	0.575 ±0.05	0.221 ±0.02		0.580 ±0.09	1.073 ±0.15	1.116 ±0.16	1.085 ±0.16	1.175 ±0.17	NA
6	0.339 ±0.04	0.574 ±0.05	0.586 ±0.05	0.277 ±0.03	0.213 ±0.02		0.719 ±0.11	0.763 ±0.11	0.762 ±0.11	0.840 ±0.13	NA
7	0.421 ±0.04	0.545 ±0.04	0.573 ±0.05	0.331 ±0.03	0.323 ±0.03	0.297 ±0.03		0.096 ±0.03	0.248 ±0.05	0.204 ±0.04	NA
8	0.407 ±0.04	0.547 ±0.04	0.567 ±0.05	0.328 ±0.03	0.326 ±0.03	0.297 ±0.03	0.039 ±0.01		0.236 ±0.04	0.201 ±0.04	NA
9	0.431 ±0.04	0.588 ±0.05	0.592 ±0.05	0.327 ±0.03	0.321 ±0.03	0.302 ±0.03	0.144 ±0.02	0.129 ±0.02		0.259 ±0.05	NA
10	0.392 ±0.04	0.561 ±0.05	0.599 ±0.05	0.315 ±0.03	0.309 ±0.03	0.274 ±0.03	0.103 ±0.02	0.090 ±0.02	0.106 ±0.02		NA
11	0.459 ±0.04	0.539 ±0.04	0.600 ±0.05	0.444 ±0.04	0.429 ±0.04	0.439 ±0.04	0.504 ±0.04	0.518 ±0.04	0.488 ±0.04	0.508 ±0.04	

[a] The standard error of the mean (±) is shown below each value. NA: no reliable estimate has been obtained.

elastase IIa and IIb are almost identical paralogous genes), and the lymphocyte granzyme A gene (H factor).

Southern Blot Analysis of Rat, Mouse, and Human Genomic DNA. Since we assumed the existence of (an) equivalent gene(s) in the human genome, we analyzed total human DNA by Southern blot analysis using the full-length cDNA of granzyme D (Jenne et al., 1988b) as a probe. These experiments failed in identifying an equivalent gene for the granzymes D–G in the human genome (results not shown). Using moderately high stringency conditions (hybridization and washing at 58 °C), however, we readily identified cross-hybridizing restriction fragments in rat genomic DNA (Figure 4, right panel).

Furthermore, we screened a human genomic library two times using different hybridization and washing conditions and the murine granzyme D cDNA as a probe. With reduced stringency conditions (hybridization and washing at 50 °C) and otherwise following the method of Church and Gilbert (1984) we obtained four human cosmid clones which cross-hybridized to the murine granzyme D cDNA. The murine granzyme D probe, the full-length cDNA probe 1-3E of human granzyme B (Schmid & Weissmann, 1987), and an oligomer specific for human granzyme B cross-hybridized to distinct *Bam*HI, *Bgl*II, *Eco*RI, and *Hin*dIII restriction fragments whose size agreed with that of human genomic restriction fragments in total human DNA detected by the human granzyme B probe (Schmid & Weissmann, 1987). We, therefore, conclude that all four human cosmid clones represent the human granzyme B (HLP) gene whose cDNA sequence has 65% overall homology to the murine granzyme D sequence (Table II).

Reconstruction of an Evolutionary Tree for Granzymes. The evolutionary relationships among the granzymes are shown in Figure 5. The tree was obtained by the unweighted pair group (UPG) method (Sneath & Sokal, 1973). The number of nucleotide substitutions per nonsynonymous site, K_A, was used as the distance between each pair of sequences. These values were calculated by the method of Li et al. (1985) and are listed in the lower triangle of Table III. The number of substitutions per synonymous site, K_S (upper triangle of Table III), was not used because some sequence pairs have become very divergent so that estimates of their K_S values are not reliable. The evolutionary position of mouse granzyme C was determined separately because we have found evidence for sequence exchanges with murine granzyme B in the 3′ half (see below).

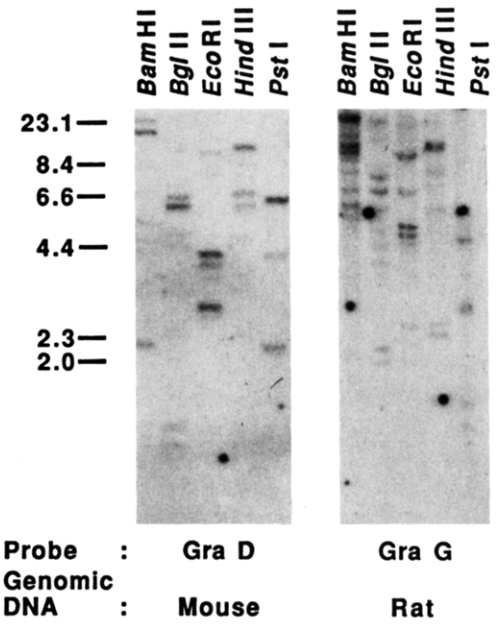

FIGURE 4: Southern blot hybridization of total genomic DNA from mouse (left panel) and rat (right panel) using a granzyme D and granzyme G cDNA probe, respectively. Five micrograms of DNA digested with the restriction enzymes *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, and *Pst*I, respectively, were separated on a 0.8% agarose gel and analyzed as described under Materials and Methods. The Zeta-Bond membrane (Bio-Rad Laboratories) was hybridized in 0.5 M NaH_2PO_4, 7% SDS, 1 mM EDTA (pH 7.2) at 58 °C for 24 hours. The final washes were done in 40 mM NaH_2PO_4, 7% SDS, 1 mM EDTA (pH 7.2) at 55 °C. λ DNA, digested with *Hin*dIII and end-labeled, served as size marker (shown in kbp).

The reliability of the tree was tested by using the modified UPG method (Li, 1981). Whereas the UPG method implicitly assumes a constant rate of nucleotide substitutions, the mod-

Table IV: Number of Nucleotide Substitutions per Site between Granzyme C and Other Highly Related Granzymes in Region I (Positions −19 to 107, See Figure 3) and Region II (Positions 108 to 245A)[a]

region	Gra B	Gra D	Gra E	Gra F	Gra G
synonymous sites					
I	0.939 ± 0.198	0.298 ± 0.079	0.415 ± 0.098	0.390 ± 0.095	0.473 ± 0.111
II	0.393 ± 0.091	1.349 ± 0.293	1.284 ± 0.268	1.263 ± 0.276	1.360 ± 0.342
nonsynonymous sites					
I	0.273 ± 0.040	0.169 ± 0.029	0.161 ± 0.028	0.156 ± 0.028	0.135 ± 0.025
II	0.181 ± 0.027	0.426 ± 0.049	0.437 ± 0.051	0.458 ± 0.051	0.425 ± 0.049

[a] The mean of substitutions at synonymous and nonsynonymous sites and the standard error (±) are given.

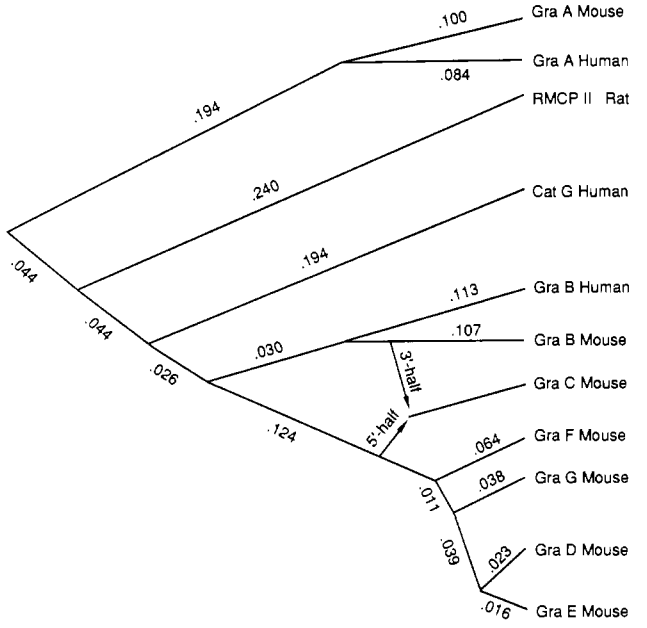

FIGURE 5: Phylogenetic tree showing the evolutionary relationships for 11 granule serine proteases (granzymes) aligned in Figure 3. The branching sequences and branch lengths (not completely drawn to scale) were calculated by applying Li's method (Li, 1981) to the distances between the nucleotide sequences in Table III (below diagonal). The number on each branch is the number of nucleotide substitutions per nonsynonymous site.

ified UPG method makes corrections for unequal rates of evolution among different lineages. Both methods gave almost identical tree topologies. The branching order only differed in that granzymes D and E and granzymes F and G were grouped in two separate clusters, which were combined by a common root (data not shown). However, when human and mouse granzyme A, the human cathepsin G, and rat mast cell protease (RMCP) II were excluded from the analysis, the modified UPG method gave the same branching as the UPG method: granzymes D, E, G, and F were joined together in that order (see Figure 5). As illustrated in Figure 5, granzyme A, rat mast cell protease II, human cathepsin G, and granzyme B existed as separate genes before the primate–rodent divergence. Granzyme A branched off first, followed by rat mast cell protease II, human cathepsin G, and granzyme B in that order. Granzyme B branched off last of all before the human–rodent split. Since granzyme B originated before the human–rodent divergence, it is paralogous to granzymes D, E, F, and G.

The Evolutionary Position of Granzyme C. Several genetic and immunological characteristics of granzyme C appeared to be peculiar and unique. First, it was shown to be immune-reactive with antibodies purified by affinity from granzyme D–F as well as from murine granzyme B (Masson & Tschopp, 1987). Second, the granzyme C cDNA cross-hybridized with all cDNA clones isolated for murine granzymes B and D–G, whereas the murine granzyme B cDNA probe only bound to the granzyme C clone in dot-blot DNA hybridization experiments using moderately high stringency conditions (data not shown). In addition, the extent of nucleotide sequence conservation between granzymes B and C was not evenly distributed along the sequence; it was much higher in the carboxy-terminal half. Third, the nucleotide sequences of introns 1 and 2 between granzymes B and C were little conserved (16% and 25%, respectively), whereas the sequences of intron 4 were 85% conserved (Bleackley et al., 1988a).

The coding region of all granzymes was subdivided into two halves, a 5′ (region I) and a 3′ coding region (region II), at residue position 107 (chymotrypsin numbering), which corresponds to the location of the third intron in the RMCP II (Benfey et al., 1987) and granzyme B genes (Jenne et al., unpublished results). We have compared the percentage of nucleotide identities in region I (corresponding to residues −19 to 107) and region II (residues 108 to 245A) among murine granzymes B–G and human granzyme B. In region I, granzyme C shares 82% nucleotide identities with each of the granzymes D–G, whereas the percentage of nucleotide identities are 10% lower in comparison to human and murine granzyme B. The reverse relationship was found in region II: only 60% identities in comparison to granzymes D–G, but 80% identities in comparison to granzyme B.

To evaluate the significance of the data and to find the evolutionary position of granzyme C, we calculated the number of nucleotide substitutions at synonymous (K_S) and nonsynonymous sites (K_A) in both regions separately for the murine granzymes B–G and human granzyme B (Table IV) and constructed a genealogical tree (not shown) for this subcluster on the basis of the K_A values in regions I and II. In region I, the K_S and K_A values between granzyme C and each of the granzymes D–G are significantly lower than the values between granzyme C and murine granzyme B, but these are still higher than the K_S and K_A values for each pair of the granzyme D–G subset. Thus, we conclude that region I of granzyme C was derived from the granzyme D-E-F-G progenitor just before the divergence of the latter sequences. The converse is true for region II. The K_S and K_A values between murine granzyme B and granzyme C are much lower than between granzyme C and each of the granzymes D–G, indicating that region II of granzyme C originated from the granzyme B gene. Whether the gene duplication and the recombination events occurred successively in that order or simultaneously by unequal crossing-over, e.g., cannot readily be deduced from our calculations, since the K_A and K_S values between granzymes B and C in region II (0.181 ± 0.027 and 0.393 ± 0.091) are not significantly different from those calculated for granzyme C versus each of the granzymes D–G in region I (e.g., versus granzyme D, 0.169 ± 0.029 and 0.298 ± 0.79). However, considering the fact that the K_S and K_A values between the human or murine granzyme B versus each of granzymes D–G are about 50% higher in region II than in region I, we assume that the sequence transfer from murine

granzyme B to the 3′ half of granzyme C occurred after granzyme C had branched off from the granzyme D-E-F-G progenitor.

Discussion

Cytolytic T lymphocytes and natural killer cells play an important role in the effector function of the cellular immune system. In the process of target cell killing, proteins present in cytoplasmic granules of activated T lymphocytes are released toward the target cell in a strictly regulated, vectorial way and appear to be involved effectively in the killing and/or postkilling processes (Sitkovsky et al., 1988). Six major proteins in cytoplasmic granules of cloned murine CTLs are serine proteases, which we have termed granzymes A–F (Jenne & Tschopp, 1988). In this study, we have identified a novel serine protease in cytoplasmic granules at the biochemical level by amino acid sequencing and at the molecular genetic level by cDNA cloning. Since the names granzymes G and H had inadvertently been assigned to two variant forms of granzyme B (Masson & Tschopp, 1988), we have kept the designation granzyme only for the six granzymes A–F which have been unambiguously identified by cloning and analytical biochemical techniques (Jenne & Tschopp, 1988). From now on, we use the designation granzyme G for the newly identified and sequenced granzyme described above.

Granzyme G shares all sequence features with other typical serine proteases and is found like the other granzymes A–F within cytoplasmic, lysosome-like granules of murine CTL clones. Although the reactivity of DFP with granzyme G is low and no proteinaceous substrate has been found, the overall three-dimensional structure including the topology of the charge relay system is likely to be conserved. The residue at position −6 relative to the active-site serine (position 189 according to the chymotrypsin numbering) is an alanine, suggesting a chymotrypsin-like activity (Figure 3). The fact that we could not readily find synthetic substrates for granzyme G may indicate that granzyme G has a restricted, unusual substrate binding specificity or requires additional, as yet unknown factors in order to exhibit proteolytic activity.

To evaluate the relative importance of all these granzymes in murine CTLs, we have done comparative biochemical studies in human CTLs (Krähenbühl et al., 1988) and in this study also at the genomic level by using murine cDNA probes in Southern blot experiments. Our attempts to distinguish and detect other than the two already characterized human granzymes, granzyme A (Fruth et al., 1987; Gershenfeld et al., 1988; Krähenbühl et al., 1988; Poe et al., 1988) and granzyme B (Schmid & Weissmann, 1987; Caputo et al., 1988; Trapani et al., 1988; Krähenbühl et al., 1988) in the human species, did not yield any positive result. First, no specific signal was obtained in total human DNA by cross-hybridization studies with the murine cDNA probe of granzyme D. Second, mRNA transcripts differing in size from that of HLP were not observed in human cytolytic T cell clones (Krähenbühl et al., 1988) and in CD3-stimulated peripheral blood lymphocytes (Jenne et al., unpublished results) with low-stringency hybridization conditions. Third, a cDNA probe for murine granzyme D picked up only human granzyme B clones in a human genomic library. Fourth, the human granzyme B probe identified only the fragments of a single-copy gene in the human genome (Schmid & Weissmann, 1987). These lines of evidence appear to support the hypothesis that the human genome lacks (a) functional gene(s) that are equivalent to the murine granzyme D–G genes. On the basis of our evolutionary analysis (see below), however, one would have to assume the existence of at least one descendant of the granzyme D-E-F-G ancestor in the primate lineage.

The observed striking differences in the copy number of functional granzyme genes between the human and mouse species prompted us to reconstruct the phylogenetic history. The robustness of the tree topology was confirmed by applying two different methods on the whole set of sequences. When the 5′ half and 3′ half of the seven most closely related sequences (HLP, granzymes B–G) were analyzed separately on the basis of K_S and K_A values using both the UPG and the modified UPG method, respectively, we obtained only one minor change in the branching topology for three out of the eight subsample trees: in contrast to the tree presented in Figure 5, granzymes D and E and granzymes G and F were grouped in separate clusters which were then joined together. On the basis of current sequence data, we cannot definitely rule out the latter possibility.

Human and murine granzyme B have been shown to be orthologous genes. This conclusion is supported not only by the branching pattern of the phylogenetic tree based on K_A values, but also by the fact that the K_S value of 0.61 ± 0.10 between the two genes is about the same as the average K_S value between rodent and human genes [0.59, see K_{13} in Table 7 of Li et al. (1987); Li & Tanimura, 1987]. Granzyme C diverged from the granzyme D-E-F-G progenitor after the human–rodent split. The present-day copy of the gene is the product of a recombination event between murine granzyme B and the granzyme D-E-F-G progenitor. However, we cannot yet draw conclusions concerning the underlying mechanisms. A two-step generation of granzyme C appears more likely than a single unequal crossing-over event between the two genes. Both events appear to have preceded the mouse–rat divergence because almost all K_S values are significantly higher than the average value between mouse and rat genes (0.237). Hence, our analysis suggests the existence of a rat granzyme C. The number of rat genes equivalent to murine granzymes D–G is predicted to be either one, two, or at the utmost three. Granzymes D and E have diverged after the mouse–rat speciation, since the K_S value between these two genes are about half (0.096, see Table II) of the average K_S value between mouse and rat genes (0.237). Since the separation of granzyme G from the progenitor D-E and granzyme F from the progenitor D-E-G was closer to the time of speciation, we cannot estimate reliably the number of orthologous genes in the mouse and rat species. Considering the results of Southern blot analyses which identified multiple cross-hybridizing fragments in the mouse as well as in the rat genome (Figure 4, left and right panels), we suggest the existence of three rat genes which are orthologous to murine granzymes C, F, and G.

Since granzyme B separated from the granzyme C-D-E-F-G progenitor before the primate–rodent split, one would assume the existence of at least one gene copy derived from this progenitor in addition to human granzyme B (HLP) in the human species. The fact that we failed to find any gene equivalent to any of the five murine granzymes C–G may be explained by assuming elimination of the ancestor through gene conversion or unequal crossing-over in the primate lineage. Alternatively, this progenitor may have been silenced and converted into a pseudogene due to a reduction in specific selective forces during primate evolution. On the other hand, the granzyme D-E-F-G progenitor has expanded its usefulness in the rodent lineage. The selective advantage of extra gene copies in the rodent lineage must have been significant in order to understand the spread and fixation of several new granzyme gene copies within rodent populations. The functional di-

versification of rodent granzymes may be linked to relatively frequent changes in the enzyme–substrate contact residues during recent evolution, so that each of the granzymes may play a distinct, species-restricted role in the complex process of CTL-mediated target cell killing.

Added in Proof

The 3′ cDNA fragment of MCSP-1 published by Kwon et al. (1988) matches the cDNA sequence of granzyme G shown here. The amino acid sequence of MCSP-1 is identical except that 50 amino-terminal residues are missing from MCSP-1 and that Val-52 and Pro-198 are replaced by Gly and Thr, respectively, in MCSP-1 (chymotrypsin numbering).

Acknowledgments

D.E.J. was the recipient of a long-term fellowship of the European Molecular Biology Organization, Meyerhofstrasse 1, D-6900 Heidelberg, FRG. M.Z. thanks Drs. A.-M. Frischauf and H. Lehrach for their support.

References

Benfey, P. N., Yin, F. H., & Leder, P. (1987) *J. Biol. Chem. 262*, 5377–5384.

Bleackley, R. C., Lobe, C. G., Duggan, B., Ehrman, N., Fregeau, C., Meier, M., Letellier, M., Havele, C., Shaw, J., & Paetkau, V. (1988a) *Immunol. Rev. 103*, 1–19.

Bleackley, R. C., Duggan, B., Ehrman, N., & Lobe, C. G. (1988b) *FEBS Lett. 234*, 153–159.

Blumenthal, R., Millard, P. J., Henkart, M. P., Reynolds, C. W., & Henkart, P. A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5551–5555.

Brunet, J. F., Dosseto, M., Denizot, F., Mattei, M. G., Clark, W. R., Haqqi, T. M., Ferrier, P., Nabholz, M., Schmitt-Verhulst, A. M., Luciani, M. F., & Goldstein, P. (1986) *Nature 322*, 268–271.

Caputo, A., Fahey, D., Lloyd, C., Vozab, R., McCairns, E., & Rowe, P. B. (1988) *J. Biol. Chem. 263*, 6363–6369.

Church, G. M., & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1991–1995.

Criado, M., Lindstrom, J. M., Anderson, C. G., & Dennert, G. (1985) *J. Immunol. 135*, 4245–4251.

Deininger, P. L. (1983) *Anal. Biochem. 129*, 216–223.

Dennert, G., & Podack, E. R. (1983) *J. Exp. Med. 157*, 1483–1495.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

Fruth, U., Sinigaglia, F., Schlesier, M., Kilgus, J., Kramer, M. D., & Simon, M. M. (1987) *Eur. J. Immunol. 17*, 1625–1633.

Gershenfeld, H. K., & Weissman, I. L. (1986) *Science 232*, 854–858.

Gershenfeld, H. K., Hershberger, R. J., Shows, T. B., & Weissman, I. L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1184–1188.

Hameed, A., Lowrey, D. M., Lichtenheld, M., & Podack, E. R. (1988) *J. Immunol. 141*, 3142–3147.

Henkart, P., & Yue, C. H. (1988) *Prog. Allergy 40*, 82–110.

Henkart, P. A., Millard, P. J., Reynolds, C. W., & Henkart, M. P. (1984) *J. Exp. Med. 160*, 75–93.

Henkart, P. A., Berrebi, G. A., Takayama, H., Munger, W. E., & Sitkovsky, M. V. (1987) *J. Immunol. 139*, 2398–2405.

Herrmann, B. G., & Frischauf, A.-M. (1987) *Methods Enzymol. 152*, 180–183.

Herrmann, B. G., Barlow, D. P., & Lehrach, H. (1987) *Cell 48*, 813–825.

Jenne, D. E., & Stanley, K. K. (1987) *Biochemistry 26*, 6735–6742.

Jenne, D. E., & Tschopp, J. (1988) *Immunol. Rev. 103*, 53–71.

Jenne, D. E., Rey C., Masson, D., Stanley, K. K., Plaetinck, G., & Tschopp, J. (1988a) *J. Immunol. 140*, 318–323.

Jenne, D., Rey, C., Haefliger, J.-A., Qiao, B.-Y., Groscurth, P., & Tschopp, J. (1988b) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4814–4818.

Kozak, M. (1984) *Nucleic Acids Res. 12*, 857–872.

Krähenbühl, O., Rey, C., Jenne, D., Lanzavecchia, A., Groscurth, P., Carrel, S., & Tschopp, J. (1988) *J. Immunol. 141*, 3471–3477.

Kwon, B. S., Kestler, D., Lee, E., Wakulchik, M., & Young, J. D.-E. (1988) *J. Exp. Med. 168*, 1839–1854.

Li, W.-H. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 1085–1089.

Li, W.-H., & Tanimura, M. (1987) *Nature 326*, 93–96.

Li, W.-H., Wu, C. I., & Luo, C. C. (1985) *Mol. Biol. Evol. 2*, 150–174.

Li, W.-H., Tanimura, M., & Sharp, P. M. (1987) *J. Mol. Evol. 25*, 330–342.

Lobe, C. G., Finlay, B. B., Paranchych, W., Paetkau, V. H., & Bleackley, R. C. (1986) *Science 232*, 858–861.

Lobe, C. G., Upton, C., Duggan, B., Ehrman, N., Letellier, M., Bell, J., McFadden, G., & Bleackley, R. C. (1988) *Biochemistry 27*, 6941–6946.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: a Laboratory Manual*, pp 1–545, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Masson, D., & Tschopp, J. (1985) *J. Biol. Chem. 260*, 9069–9072.

Masson, D., & Tschopp, J. (1987) *Cell 49*, 679–685.

Masson, D., Corthêsy, P., Nabholz, M., & Tschopp, J. (1985) *EMBO J. 4*, 2533–2538.

Masson, D., Nabolz, M., Estrade, C., & Tschopp, J. (1986) *EMBO J. 5*, 1595–1600.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988) *Nucleic Acids Res. 16*, 1215.

Mueller, C., Gershenfeld, H. K., & Weissman, I. L. (1988) *Immunol. Rev. 103*, 73–85.

Mullbacher, A., & Ada, G. L. (1987) *Microb. Pathogenesis 3*, 315–318.

Pasternack, M. S., & Eisen, H. N. (1985) *Nature 314*, 743–745.

Pasternack, M. S., Verret, C. R., Liu, M. A., & Eisen, H. N. (1986) *Nature 322*, 740–743.

Podack, E. R. (1986) *J. Cell. Biochem. 30*, 133–170.

Podack, E. R., & Königsberg, P. J. (1984) *J. Exp. Med. 160*, 695–710.

Podack, E. R., Young, J. D.-E., & Cohn, Z. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8629–8633.

Poustka, A., Rackwitz, H.-R., Frischauf, A.-M., Hohn, B., & Lehrach, H. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4129–4133.

Redmond, M. J., Letellier, M., Parker, J. M. R., Lobe, C. G., Havele, C., Paetkau, V., & Bleackley, R. C. (1987) *J. Immunol. 139*, 3184–3189.

Reed, K. C., & Mann, D. A. (1985) *Nucleic Acids Res. 13*, 7207–7221.

Salvesen, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C., & Travis, J. (1987) *Biochemistry 26*, 2289–2293.

Schmid, J., & Weissmann, C. (1987) *J. Immunol. 139*, 250–256.

Schmidt, R. E., MacDermott, R. P., Bartley, G., Bertovich, M., Amatao, D. A., Austen, K. F., Schlossmann, S. F.,

Stevens, R. L., & Ritz, J. (1985) *Nature 318*, 289–291.
Simon, M. M., Hoschützky, H., Fruth, U., Simon, H.-G., & Kramer, M. D. (1986) *EMBO J. 5*, 3267–3274.
Sitkovsky, M. V. (1988) *Immunol. Rev. 103*, 127–160.
Sitkovsky, M. V., & Paul, W. E. (1988) *Nature 332*, 306–307.
Sneath, P. H. A., & Sokal, R. R. (1973) in *Numerical Taxonomy*, Freeman, San Francisco, CA.
Trapani, J. A., Klein, J. L., White, P. C., & Dupont, B. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6924–6928.
Tschopp, J., & Jongeneel, C. V. (1988) *Biochemistry 27*, 2641–2646.
Von Heijne, G. (1986) *Nucleic Acids Res. 14*, 4683–4690.
Young, J. D.-E., & Cohn, Z. A. (1987) *Adv. Immunol. 41*, 269–332.
Young, J. D.-E., Hengartner, H., Podack, E. R., & Cohn, Z. A. (1986a) *Cell 44*, 849–859.
Young, J. D.-E., Leong, L. G., Liu, C.-C., Damiano, A., Wall, D. A., & Cohn, Z. A. (1986b) *Cell 47*, 183–194.
Zagury, D. (1982) *Adv. Exp. Med. Biol. 146*, 149–163.
Zalman, L. S., Brothers, M. A., Chiu, F. J., & Müller-Eberhard, H. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5262–5266.

Long-Range Electrostatic Interactions Can Influence the Folding, Stability, and Cooperativity of Dihydrofolate Reductase†

Kathleen M. Perry,‡ James J. Onuffer,§ Mitchell S. Gittelman, Lawrence Barmat,‖ and C. Robert Matthews*

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802
Received March 13, 1989; Revised Manuscript Received June 9, 1989

ABSTRACT: To test the possibility that long-range interactions might influence the folding and stability of dihydrofolate reductase, a series of single and double mutations at positions 28 and 139 were constructed and their urea-induced unfolding reactions studied by absorbance and circular dichroism spectroscopy. The α carbons of the two side chains are separated by 15 Å in the native conformation. The replacement of Leu 28 by Arg and of Glu 139 by Gln resulted in additive effects on both kinetic and equilibrium properties of the reversible unfolding transition; no evidence for interaction was obtained. In contrast, the Arg 28/Lys 139 double replacement changed the equilibrium folding model from two state to multistate and showed evidence for interaction in one of the two kinetic phases detected in both unfolding and refolding reactions. The results can be explained in terms of a long-range, repulsive electrostatic interaction between the cationic side chains at these two positions.

The complex three-dimensional structures of proteins are stabilized by a large number of noncovalent interactions such as hydrogen bonds, van der Waals interactions, and electrostatic interactions and by the hydrophobic effect (Kauzmann, 1959). Although the individual interactions are relatively weak, their propensity to act in a cooperative fashion results in the folding of newly synthesized polypeptides to unique, stable conformations. The cooperative nature of the forces that stabilize proteins raises the possibility that an amino acid at one site may interact in an energetic sense with amino acids at distant sites. Such long-range interactions play an important role in a variety of biological phenomena including allosteric regulation of enzyme function and in the cooperative binding of substrates by multisubunit enzymes.

One approach toward probing the possibility of long-range interactions in proteins is to examine the effects of amino acid replacements on the reversible unfolding transition. Specifically, a set of mutant proteins can be constructed that includes single amino acid replacements at the two sites of interest and the double mutant with simultaneous replacements at the two sites. Comparison of the sum of the effects of the single replacements on the stability and folding kinetics with the effect of the double replacement provides a test for interaction. Additivity of the free energy of folding or the activation free energies of unfolding and refolding rate constants would mean that the replacements act independently. Nonadditivity would provide direct evidence that the residues at these positions interact (Ackers & Smith, 1985).

Previous application of this approach to the α subunit of tryptophan synthase showed that the residues at positions 175 and 211 interact with each other in the native conformation (Hurle et al., 1986). Examination of the preliminary X-ray structure provides a simple explanation for this interaction in that residues 175 and 211 are adjacent in the three-dimensional structure (Hyde et al., 1988). Although similar studies on the lysozyme from phage T4 (Becktel et al., 1987) and staphylococcal nuclease (Shortle et al., 1988; Shortle & Meeker, 1986) have generally found that double replacements have an additive effect on stability, the Gly 46 → Ala/Gly → 48 Ala double mutant in the N-terminal domain of the λ repressor protein was reported to have a nonadditive effect (Hecht et al., 1986).

Because a consistent pattern for such interactions has not yet emerged, we felt that it was important to further test this

† This work was supported by the National Science Foundation, Grant DMB-8705673.
‡ Present address: Department of Biochemistry and Biophysics, University of California Medical Center, San Francisco, CA 94143.
§ Present address: Department of Biochemistry, University of California, Berkeley, CA 94720.
‖ Present address: Temple University Medical School, Philadelphia, PA 19122.

0006-2960/89/0428-7961$01.50/0 © 1989 American Chemical Society