Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 4814–4818, July 1988
Immunology

Identification and sequencing of cDNA clones encoding the granule-associated serine proteases granzymes D, E, and F of cytolytic T lymphocytes

(cytotoxicity/serine esterases/subcellular localization/perforin)

DIETER JENNE*, CLAUDINE REY*, JACQUES-ANTOINE HAEFLIGER*, BAO-YI QIAO†, PETER GROSCURTH†, AND JÜRG TSCHOPP*‡

*Institut de Biochimie, Université de Lausanne, Chemin des Boveresses, CH-1066 Epalinges, and †Institute of Anatomy, Division of Cell Biology, University of Zürich, CH-8057 Zürich, Switzerland

Communicated by Hans J. Müller-Eberhard, January 7, 1988 (received for review October 30, 1987)

ABSTRACT **Cytoplasmic granules of cytolytic T lymphocytes contain at least six related serine esterases (granzymes) that are released together with perforin, a pore-forming protein related to complement component C9, during target-cell lysis. Polyclonal antibodies were used to isolate a large number of cDNA clones from an expression library derived from cytolytic-T-cell mRNA. Three distinct full-length cDNA clones coding for granzymes D, E, and F were identified by restriction site mapping and nucleotide sequencing. The three deduced amino acid sequences are highly similar to one another (between 72% and 90% amino acid identities) and to the sequences of granzymes B and C, cathepsin G, and rat mast-cell proteases I and II (between 43% and 57% amino acid identities). Cysteine residues capable of forming intramolecular disulfide bonds are conserved, as are the catalytic-site residues characteristic of serine proteases. Comparison of the cDNA-derived protein sequences with the amino termini of the isolated granzymes provides evidence that they are stored in a fully processed, activated form after removal of the signal peptide and two additional residues (propeptide) at the amino terminus. Immunoelectron microscopic studies demonstrated that granzymes D, E, and F are present in the same morphologically distinct cytoplasmic granules in which perforin has been found previously.**

The mechanisms by which large granular lymphocytes and cytolytic T lymphocytes (CTLs) lyse their appropriate target cells are as yet poorly defined. Several lines of evidence suggest, however, that proteins contained in cytoplasmic granules of the effector cells are involved in the cytolytic process: isolated granules are lytic for a variety of target cells in a nonrestricted manner (1–3), and the granule proteins are specifically released upon target-cell recognition (4–7).

The major proteins of the granules have been isolated and characterized; these include a pore-forming protein, designated perforin (cytolysin) (8–10), and a family of at least six serine esterases, called granzymes A, B (G, H), C, D, E, and F (11–15), which are stored in association with chondroitin sulfate A (5). Purified perforin is cytolytic and shows ultrastructural, immunological, and functional similarities to complement component C9 (16–18). In contrast, little is known about the function of granzymes. Granzyme A is a disulfide-linked homodimer with trypsin-like activity (12–15). No substrates have been identified for the other granzymes, and only granzymes A and D react strongly with the serine esterase affinity label di[^{3}H]isopropyl fluorophosphate (11). Three of them, granzyme A (also named H factor or CTLA-3), granzyme B (CCP I, CTLA-1), and granzyme C (CCP II), have been cloned and structurally characterized as serine proteases (19–22).

In this paper we present cDNA sequences for granzymes D, E, and F.§ The protein structures deduced from the full-length cDNA clones show all the sequence characteristics of genuine serine proteases, suggesting that effector T cells elicit diverse proteolytic events during their interactions with target cells.

MATERIALS AND METHODS

Antibodies. Rabbit polyclonal antibodies raised against purified granzyme D (11) were used for the immunological identification of expressed antigens. This antiserum recognizes crossreacting epitopes of granzymes D, E, and F. Its affinity for purified granzyme C is low, whereas binding to other granzymes is not observed (11).

cDNA Cloning. The cDNA expression library for the H-2K^d-specific CTL line B6.1 of mouse strain C57BL/6 in the bacterial expression vector pEX1 was kindly provided by H. Haymerle (23). All methods relating to the screening, cloning, and sequencing of granzymes D, E, and F have been described (22, 24–27).

Immunoelectron Microscopy. Cells of the interleukin 2-dependent, cloned murine CTL line HY3-Ag3 were fixed with 2% glutaraldehyde plus 0.8% paraformaldehyde and embedded in Epon. Ultrathin sections (≈100 nm) were immunolabeled as described (28). The specimens were incubated first with anti-granzyme D antibody [diluted 1:50 or 1:100 in phosphate buffered saline (0.15 M NaCl/0.01 M phosphate, pH 7.2)] and then with protein A–gold complex (1:50) with a gold-particle size of 10 nm. Subsequently the sections were stained with uranyl acetate/lead citrate or left uncontrasted. Controls included replacement of specific antibody by normal rabbit serum.

RESULTS

Isolation and Characterization of cDNA Clones. Two rabbit antisera have been prepared that distinguish between two subsets of CTL-specific serine esterases, one comprising granzymes A and B (G, H) and the other comprising granzymes C, D, E, and F (11). cDNA sequences coding for the former subset of serine esterases (19, 21) and for granzyme C (22) have been reported. In order to isolate cDNA clones for the remaining three members of the latter subset, polyclonal

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CTL, cytolytic T lymphocyte: RMCP, rat mast-cell protease; HLP, human lymphocyte protease.

‡To whom reprint requests should be addressed.

§These sequences are being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession nos. J03255, J03256, and J03257).

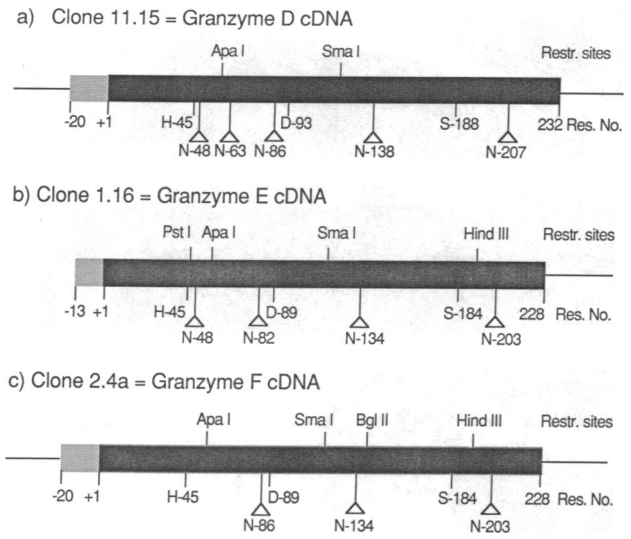

FIG. 1. Restriction sites and some features of the open reading frames of clones 11.15, 1.16, and 2.4a. The encoded hydrophobic signal peptide (light stippling) and the protein sequence of the mature enzyme (heavy stippling) are represented for each clone. Unique restriction sites that were used to discriminate the isolated heterogeneous cDNA clones are shown above each bar. The locations of the active-site residues histidine (H), aspartic acid (D), and serine (S) and the distribution of predicted asparagine (N)-linked glycosylation sites are shown below each bar. See Fig. 2 for the numbering of amino acid residues.

antibodies that recognize predominantly granzymes D, E, and F were used to screen a cDNA expression library of a cloned murine CTL line, B6.1. Upon rescreening of immunepositive colony areas, 70 expression clones were obtained and purified to homogeneity. Antibody binding to the expressed hybrid proteins was further demonstrated by immunoblot experiments using total extracts of bacteria (data not shown). The size of the open reading frame of the cloned cDNA inserts was estimated and compared to the length of the inserts. Fifteen cDNA clones contained inserts of 800 or more base pairs and expressed a fusion protein 20–30 kDa larger than β-galactosidase. Mapping with three restriction enzymes—*Pst* I, *Sma* I, and *Hin*dIII—permitted the identification of three groups of homologous cDNA clones. Group 1 included seven cDNA clones with a unique *Sma* I site; group 2 included two clones with a unique *Sma* I, *Hin*dIII, and *Pst* I sites; and group 3 contained six clones with unique *Sma* I and *Hin*dIII sites (Fig. 1). In each group, the clone having the longest extension toward the 5′ end of the mRNA was selected for sequence analysis. "Shotgun" sequencing of subclones in the M13 bacteriophage vector was performed on both DNA strands in each case.

Nucleotide Sequences and Identification of cDNA Clones. The nucleotide sequences and derived amino acid sequences of the cDNA clones 13.3, 1.16, and 2.4a are shown in Fig. 2. The three clones are highly homologous to one another and crosshybridize even under highly stringent conditions (15mM NaCl/1.5 mM trisodium citrate, pH 7.0, at 65°C; data not

```
11.15                                                GGCCCCTGAGCCCACATCTCCTTCCTCGCCTTCCAAGGGCTTTGTCTGCTCTGCCCTCCTCCATCCTGAGCAGCCTTCCTGGGAAG   86
2.4a                                                 CCCCTTCCAAGGGCAGTGTCTCCTCTGCCCTCCTCCATCCTGAGCAGCCTTCCTGGGAAG   60

11.15  ATGCCACCAATCCTGATTCTCCTGACCCTACTTCTGCCTCTCAGAGCTGGAGCAGAGGAGATCATCGGCGGCCATGTGGTGAAGCCACACTCCCGCCCCTACATGGCATTCGTTATGTCTGTGGAT  212
Gra D  M  P  P  I  L  I  L  L  T  L  L  L  P  L  R  A  G  A  E  E  I  I  G  G  H  V  V  K  P  H  S  R  P  Y  M  A  F  V  M  S  V  D   22
1.16   -20
                    TCCTGACCCTACTTCTGCCTCTTGGAGCTGGAGCAGAGGAGATCATCGGCGGCCATGTGGTGAAGCCACACTCCCGCCCCTACATGGCGTTTGTTAAGTCTGTGGAT  107
Gra E                  -13  L  T  L  L  L  P  L  G  A  G  A  E  E  I  I  G  G  H  V  V  K  P  H  S  R  P  Y  M  A  F  V  K  S  V  D   22
2.4a   ATGCCACCAATCCTGATTCTCCTGACCCTTCTTCTGCCTCTCAGAGCTGGAGCAGAGGAGATCATCGGGGGCCATGAGGTCAAGCCCCACTCCCGCCCTTACATGGCACGTGTGAGGTTTGTGAAA  186
Gra F  M  P  P  I  L  I  L  L  T  L  L  L  P  L  R  A  G  A  E  E  I  I  G  G  H  E  V  K  P  H  S  R  P  Y  M  A  R  V  R  F  V  K   22
                                                                                +1
11.15  ATTAAGGGGAACAGGATATACTGTGGAGGCTTCCTGATTCAAGATGACTTTGTGCTGACGGCTGCTCACTGCAAAAACAGCTCTGTCCAAAGCTCAATGACAGTCACACTGGGGGCCCACAACATC  338
Gra D  I  K  G  N  R  I  Y  Ⓒ  G  G  F  L  I  Q  D  D  F  V  L  T  A  A  Ⓗ  Ⓒ  K  ▶N  S  S  V  Q  S  S  M  T  V  T  L  G  A  H  ▶N  I   64
1.16   ATTGAAGGTAATAGGAGATACTGTGGAGGCTTCTTGGTTCAAGATGACTTTGTGCTGACTGCTGCTCACTGCAGGAACAGG----------ACAATGACAGTCACACTGGGGGCCCACAACATC  221
Gra E  I  E  G  N  R  R  Y  Ⓒ  G  G  F  L  V  Q  D  D  F  V  L  T  A  A  Ⓗ  Ⓒ  R  ▶N  R  -  -  -  -  T  M  T  V  T  L  G  A  H  N  I   60
2.4a   GATAATGGAAAAAGACATTCCTGTGGAGGCTTCCTGGTTCAAGACTACTTTGTGCTGACGGCTGCTCACTGCACTGGAAGC----------TCAATGAGAGTCATACTGGGGGCCCACAACATC  300
Gra F  D  N  G  K  R  H  S  Ⓒ  G  G  F  L  V  Q  D  Y  F  V  L  T  A  A  Ⓗ  Ⓒ  T  G  S  -  -  -  -  S  M  R  V  I  L  G  A  H  N  I   60

11.15  ACGGCTAAGGAGGAGACACAGCAGATCATCCCTGTGGCAAAAGACATTCCCCATCCAGATTATAATGCTACTATCTTCTACAGTGACATCATGCTGTTAAAGCTGGAGAGTAAGGCCAAGAGAACT  464
Gra D  T  A  K  E  E  T  Q  Q  I  I  P  V  A  K  D  I  P  H  P  D  Y  ▶N  A  T  I  F  Y  S  Ⓓ  I  M  L  L  K  L  E  S  K  A  K  R  T  106
1.16   AAGGCTAAGGAGGAGACACAGCAGATCATCCCTGTGGCAAAAGCCATTCCCCATCCAGATTATAATGCCACTGCCTTCTTCAGTGACATCATGCTGTTAAAGCTGGAGAGTAAGGCCAAGAGAACT  347
Gra E  K  A  K  E  E  T  Q  Q  I  I  P  V  A  K  A  I  P  H  P  D  Y  ▶N  A  T  A  F  F  S  Ⓓ  I  M  L  L  K  L  E  S  K  A  K  R  T  102
2.4a   AGGGCTAAGGAAGAGACACAGCAGATCATCCCTGTGGCAAAAGCCATTCCCCACCCAGCTTATGATGATAAGGACAACACCAGTGACATCATGCTATTAAAGCTGGAGAGTAAGGCCAAGAGAACT  426
Gra F  R  A  K  E  E  T  Q  Q  I  I  P  V  A  K  A  I  P  H  P  A  Y  D  D  K  D  ▶N  T  S  Ⓓ  I  M  L  L  K  L  E  S  K  A  K  R  T  102

11.15  AAAGCTGTGAGACCCCTCAAGTTGCCCAGATCCAATGCCCGGGTGAAGCCAGGGGATGTGTGCAGTGTGGCTGGCTGGGGGTCAAGGTCCATCAATGACACTAAAGCATCTGCCCGCCTGCGAGAG  590
Gra D  K  A  V  R  P  L  K  L  P  R  S  N  A  R  V  K  P  G  D  V  Ⓒ  S  V  A  G  W  G  S  R  S  I  ▶N  D  T  K  A  S  A  R  L  R  E  148
1.16   AAAGCTGTGAGACCCCTCAAGTTGCCCAGACCCAATGCCCGGGTGAAGCCAGGGGATGTGTGCAGTGTGGCTGGCTGGGGGCCAAGGTCCATCAATGACACTAAAGCATCTGCCCGCCTGCGAGAG  473
Gra E  K  A  V  R  P  L  K  L  P  R  P  N  A  R  V  K  P  G  D  V  Ⓒ  S  V  A  G  W  G  P  R  S  I  ▶N  D  T  K  A  S  A  R  L  R  E  144
2.4a   AAAGCTGTGAGGCCCCTCAAGTTGCCCAGACCCAATGCCCGGGTGAAGCCAGGGCATGTTTGCAGTGTGGCTGGCTGGGGGAGAACATCCATCAATGCAACACAAAGATCTTCCTGCCTACGAGAG  552
       K  A  V  R  P  L  K  L  P  R  P  N  A  R  V  K  P  G  H  V  Ⓒ  S  V  A  G  W  G  R  T  S  I  ▶N  A  T  Q  R  S  S  Ⓒ  L  R  E  144

11.15  GTTCAACTGGTCATCCAGGAGGACGAGGAATGCAAAAAACGTTTCCGATACTACACTGAGACCACAGAGATTTGTGCTGGAGACTTGAAGAAAATAAAGACTCCTTTCAAGGGTGACTCGGGGGGA  716
Gra D  V  Q  L  V  I  Q  E  D  E  E  Ⓒ  K  K  R  F  R  Y  Y  T  E  T  T  E  I  Ⓒ  A  G  D  L  K  K  I  K  T  P  F  K  G  D  Ⓢ  G  G  190
1.16   GCTCAACTGGTCATCCAGGAGGATGAGGAATGCAAAAAACGTTTCCGACACTACACTGAGACCACAGAGATTTGTGCTGGAGACTTGAAGAAAATAAAGACTCCTTTCAAGGGTGACTCTGGGGGA  599
Gra E  A  Q  L  V  I  Q  E  D  E  E  Ⓒ  K  K  R  F  R  H  Y  T  E  T  T  E  I  Ⓒ  A  G  D  L  K  K  I  K  T  P  F  K  G  D  Ⓢ  G  G  186
2.4a   GCTCAACTGATCATCCAGAAGGATAAGGAATGCAAAAAATACTTCTATAAGTATTTCAAGACCATGCAGATTTGTGCTGGAGACCCAAAGAAAATACAGTCTACTTACAGTGGTGACTCCGGGGGA  678
Gra F  A  Q  L  I  I  Q  K  D  K  E  Ⓒ  K  K  Y  F  Y  K  Y  F  K  T  M  Q  I  Ⓒ  A  G  D  P  K  K  I  Q  S  T  Y  S  G  D  Ⓢ  G  G  186

11.15  CCCCTTGTGTGTCACAACCAAGCATATGGACTTTTCGCCTATGCAAAAAACGGAACAATCTCTTCAGGAATCTTCACTAAGGTTGTGCACTTCCTGCCGTGGATAAGCTGGAACATGAAGTTGCTC  842
Gra D  P  L  V  Ⓒ  H  N  Q  A  Y  G  L  F  A  Y  A  K  ▶N  G  T  I  S  S  G  I  F  T  K  V  V  H  F  L  P  W  I  S  W  N  M  K  L  L  232
1.16   CCCCTCGTGTGTGACAACAAAGCTTATGGACTTTTAGCCTATGCAAAAAACAGGACAATCTCTTCAGGAGTCTTCACTAAGATTGTGCACTTCCTGCCGTGGATAAGCAGGAACATGAAGCTGCTC  725
Gra E  P  L  V  Ⓒ  D  N  K  A  Y  G  L  L  A  Y  A  K  ▶N  R  T  I  S  S  G  V  F  T  K  I  V  H  F  L  P  W  I  S  R  N  M  K  L  L  228
2.4a   CCCCTCGTGTGTAACAACAAAGCTTATGGAGTTTAACCTATGGGCTAAACAGGACAATCGGTCCAGGAGTCTTCACTAAGGTTGTGCACTACCTGCCGTGGATAAGCAGGAACATGAAGCTGCTC  804
Gra F  P  L  V  Ⓒ  N  N  K  A  Y  G  V  L  T  Y  G  L  ▶N  R  T  I  G  P  G  V  F  T  K  V  V  H  Y  L  P  W  I  S  R  N  M  K  L  L  228

11.15  TAACAGGAGTTAAACCACCCGTGCCTGACCAGCCTGTCCGACCTCAGGCAAGAACCATGTGGAGTGGGCAGCAAAGAATGAAAATTCACAATAAATAA                              940
1.16   TAACAGGTGTTAAACCACCCGTGCCTGACCAGCCTGTCCGACCTCAGGCAAGAACCACGTGGAGTGGGCAGCAAAGAATGAAAATTCACAATAAATAACCTCCG                    829
2.4a   TAACAGGAGTTAAACCACCCGTGCCTGACCAGCCTGTCCGACCTCAGGCAAGAACCATGTGGAGTGAGCAGCAAAGAATGAAAATTCATAATAAATAACCC                       905
```

FIG. 2. Nucleotide and amino acid sequences of granzymes (Gra) D, E, and F. The nucleotide sequences and deduced amino acid sequences of clones 11.15, 1.16, and 2.4a are aligned with one gap introduced for granzymes E and F to maximize sequence similarities both at the protein and at the nucleotide level. Nucleotide sequences are numbered from the first nucleotide of the cloned cDNA inserts. Amino acid sequences encoded by clones 11.15, 1.16, and 2.4a are shown by standard one-letter symbols below the respective DNA sequences. The numbering of the amino acid sequences starts with the amino-terminal residue (+1) of the mature enzymes as determined by amino acid sequencing. The residues forming the catalytic site of serine proteases are boxed by squares, and all cysteine residues are circled. The partial protein sequences determined for the amino termini of granzymes D, E, and F are underlined, and putative N-linked glycosylation sites are marked by filled triangles. The polyadenylylation signal sequence AATAAA is underlined in the 3′ noncoding region.

shown). The 5′ noncoding regions of 13.3 and 2.4a are conserved as well, and the polyadenylylation signal (underlined in Fig. 2) is present in all three clones at the same distance downstream from the translation stop codon TAA. Clones 1.16 and 2.4a each have one large open reading frame encoding 248 amino acids, whereas clone 13.3 codes for 252 amino acids. The proposed translation initiation codons are preceded by an adenine at the −3 position in agreement with the most conserved feature of translation start sequences (29). Translation of the mRNAs starts with a very hydrophobic stretch of 20 residues showing all the features of a cleavable signal peptide. A complete signal peptide cannot be deduced from clone 1.16, but the high sequence homologies at the nucleotide level suggests that only seven codons at the 5′ end of clone 1.16 are missing. Cleavage by signal peptidase occurs most likely between alanine at position −3 and glutamic acid at −2 or, with somewhat lesser probability, between alanine at −5 and glycine at −4 (30). The two to four residues following the signal peptide probably form a short propeptide that is removed by a second processing enzyme, a putative dipeptidyl peptidase, since the sequence of all mature granzymes starts with Ile-Ile. Since residues 1–20 of the three proteins predicted from the cDNA sequences agree completely with the amino-terminal sequences determined for granzymes D, E, and F by amino acid sequencing (11), we conclude that the clones 13.3, 1.16, and 2.4a code for the complete covalent structure of granzymes D, E, and F, respectively.

Granzymes D, E, and F Are Serine Proteases. Although previous results showed that granzyme F did not react with di[^{3}H]isopropyl fluorophosphate, and granzyme E gave only a borderline signal (11), the cDNA-derived protein sequences now reveal the structural feature of genuine serine proteases in each case. The three residues, histidine, aspartic acid, and serine, that make up the active site of serine proteases are found at homologous positions flanked by well-conserved peptide segments in the aligned sequences. Six cysteine residues occur in the same positions as in rat mast cell protease II (RMCP II) and in granzymes C and B; presumably, these residues form three internal disulfide bonds as in RMCP II, whose tertiary structure is known (31). Granzyme F has an additional cysteine residue at position 145, close to a putative glycosylation site. In contrast to granzyme A, granzyme F does not form disulfide-linked homodimers via the free cysteine residue.

In agreement with the elution pattern of granzymes from a Mono S (Pharmacia) cation-exchange column (ref. 11; with increasing salt concentration, granzyme D is eluted first, granzyme E thereafter, and granzyme F last), the granzyme D sequence has the highest number of putative N-linked glycosylation sites, namely five, followed by granzyme E with four and granzyme F with three. Glycosylation is highly heterogeneous for each granzyme and most prominent for granzyme D (11). The higher the number of N-linked glycosylation sites, the greater is the heterogeneity in molecular size and overall negative charge of the molecules observed.

Subcellular Localization of Granzymes D, E, and F. Previously, CTL clone HY3-Ag3 was studied by immunoelectron microscopy (28) to identify the subcellular localization of perforin. A peculiar feature of these cells are electron-dense, lysosome-like granules located predominantly in the cell periphery (Fig. 3*A*). Immuno-labeling with antibodies against granzymes D, E, and F and protein A–gold exclusively stains these granules (Fig. 3 *B* and *C*). Gold particles are deposited in the fine granular matrix of lysosome-like granules, which represents a morphologically distinct subcompartment of granules. The same subcompartment has been found to contain perforin (28). The vesicular compartment in the periphery of the granules is free of gold particles.

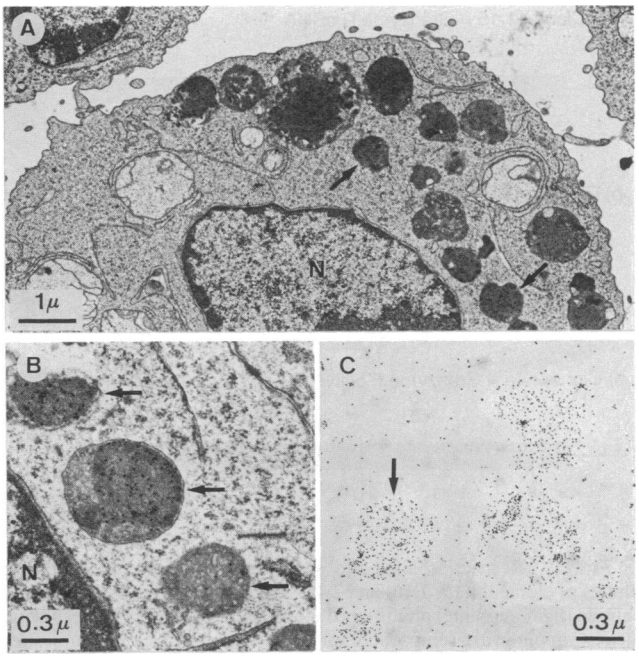

FIG. 3. Subcellular localization of granzymes D, E, and F by immunoelectron microscopy. (*A*) Ultrathin section of an Epon-embedded HY3-Ag3 cell stained with anti-granzyme D antibodies at a dilution of 1:50. A large nucleus (N), a few slightly swollen mitochondria, and typical large lysosome-like granules are visible (arrows). (*B*) Higher magnification shows deposits of gold particles within the fine granular matrix of granules. The periphery of granules and all other organelles of the cell are not stained. (*C*) Similar section to *B* before staining with uranyl acetate/lead citrate. The gold particles are within the granules and only low background staining is encountered in other areas. μ, μm.

DISCUSSION

By immunoelectron microscopy, we have identified antigenic determinants of granzymes D, E, and F within the characteristic electron-dense granules of the cytoplasm of CTLs. Since each of the granzymes D, E, and F have been isolated in similar amounts from granules sedimenting in high-density fractions of Percoll gradients, it is very likely that all of them are present within the same granules of T lymphocytes. Perforin, a CTL protein that forms pores in the membranes of target cells, has been localized in the fine granular matrix of granules (28) morphologically indistinguishable from those now shown to contain granzymes D, E, and F. Therefore, granzymes and perforin are likely to be released together through exocytosis of cytoplasmic granules.

The cDNA-derived amino acid sequences of granzymes D, E, and F are extremely similar to one another as judged by the percentage of identical residues. Furthermore, granzymes D, E, and F show a high degree of sequence similarity to the other granzymes of T lymphocytes, granzymes A, B (G, H), and C, as well as to three serine proteases present in cytoplasmic granules of cells belonging to other cell lineages—namely RMCP I (32), RMCP II (33, 34), and cathepsin G from human neutrophils (35) (Fig. 4). While granzyme D has 90% residues in common with granzyme E and 72% residues in common with granzyme F, its similarity to granzyme B (G, H) (57%) and C (55%) is only slightly higher than its similarity to RMCPI (45%), RMCP II (43%), and cathepsin G (46%). In comparison, the degree of overall positional identities between granzyme D and A is relatively low (39%). Therefore, granzymes D, E, and F may form an evolutionarily highly related subgroup of murine granzymes. Another member of the granzyme family is human lymphocyte protease (HLP), which has been characterized as a

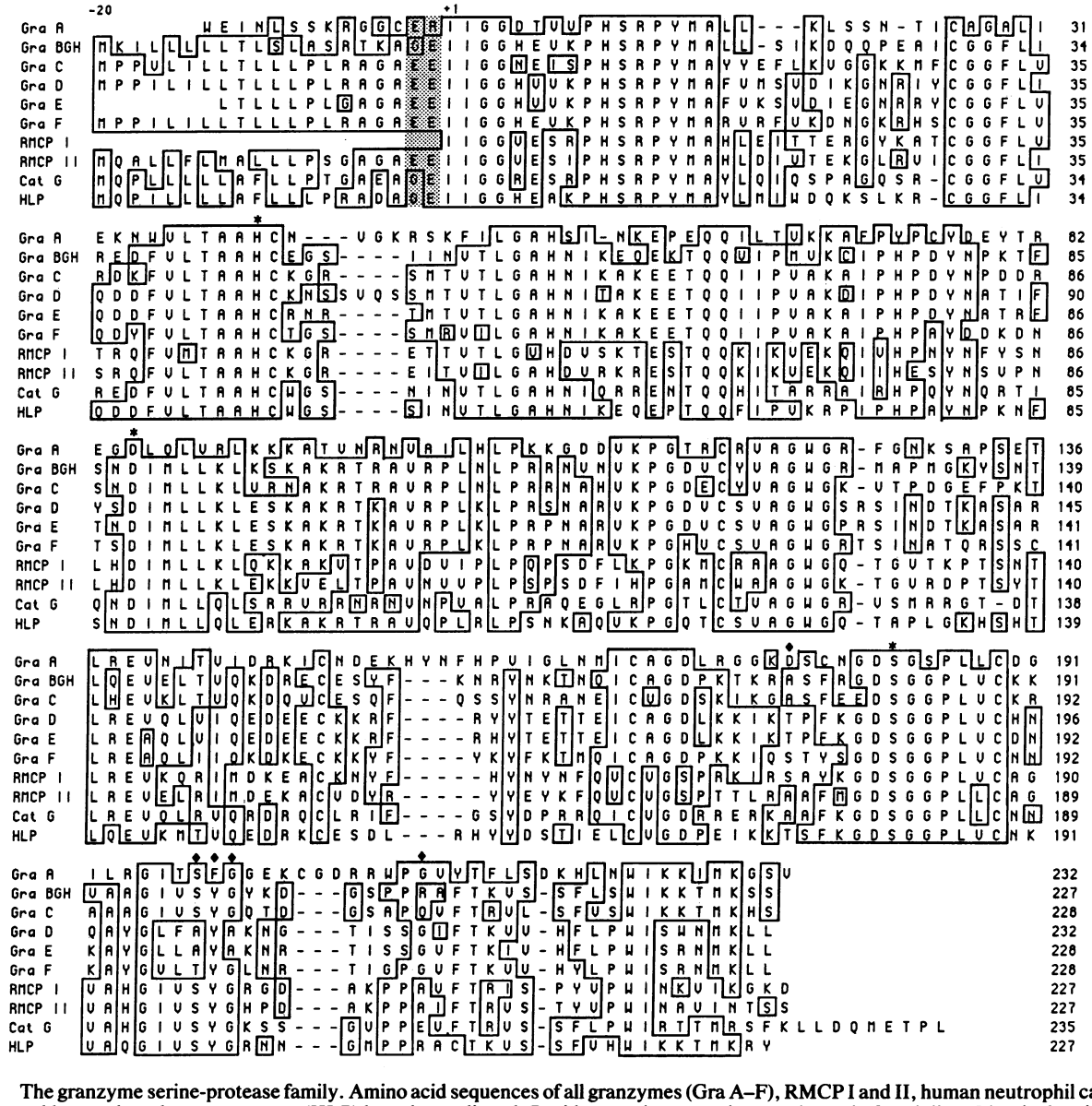

FIG. 4. The granzyme serine-protease family. Amino acid sequences of all granzymes (Gra A–F), RMCP I and II, human neutrophil cathepsin G (Cat G), and human lymphocyte protease (HLP) have been aligned. Residue numbers are given at the end of each line. Identical amino acids are boxed, where residues are conserved among at least 4 of the 10 proteins. The histidine, aspartic, and serine residues that form the catalytic site in serine proteases are indicated by asterisks. Residues lining the substrate-binding pocket are found in positions −6, +15 to 17, and +28 relative to the active serine in the aligned sequences (36) and are marked by filled diamonds. Sources for the amino acid sequences: Gra A, ref. 19; Gra BGH, ref. 21; Gra C, ref. 22; RMCP I, ref. 32; RMCP II, refs. 33 and 34; Cat G, ref. 35; HLP, ref. 37.

T-lymphocyte-specific mRNA transcript (37). HLP shows the highest degree of relatedness to granzyme B (68%), and the two are probably species homologues.

Three structural features are characteristic for the granzyme family. The amino-terminal sequence of the isolated granzymes starts with the conserved sequence Ile-Ile-Gly-Gly, followed by four variable residues in positions 5–8. Residues 9–16 are strictly conserved among all known granzymes. The commonly found acidic propeptide at the amino terminus consists of only two residues, either Gly-Glu or Glu-Glu. The prepropeptide sequence of granzyme A, which has not been completely determined, appears to be an exception in that it ends with an arginine residue.

Unlike tryspin and chymotrypsin, two serine proteases of the exocrine pancreas, granzymes of T lymphocytes are not secreted as zymogens. They are stored as fully processed and activated enzymes in the cytoplasmic granules of CTL lines and are presumably bound electrostatically to chondroitin sulfate molecules due to the basic nature of their polypeptide chains (11). The enzyme activities of granzyme D and granzyme A, for which synthetic substrates are known, are not enhanced in the presence of other granzymes (D. Masson and J.T., unpublished data). Therefore, we suggest that conversion of granzyme precursors to mature, enzymatically active forms takes place during or just before packaging into cytoplasmic granules. The short prodipeptide at the amino terminus of granzymes may protect other constitutively secreted proteins against proteolysis by granzymes during traversal of their common intracellular pathway.

Alignment of the amino acid sequences of granzymes D, E, and F with those of other members of the granzyme family also permits comparison of the structural features of the substrate-binding pocket in the various granzymes. The residues at positions −6, +15 to +17, and +28 relative to the active-site serine (indicated by filled diamonds in Fig. 4) are thought to determine the specificity of the substrate-binding pocket (36). When the corresponding residues of granzymes D, E, and F in these positions are compared to

those of the remaining granzymes and to those of RMCP I and II, HLP, and cathepsin G, amino acid replacements are observed, suggesting distinct substrate specifities due to the altered shape and size of the substrate-binding pocket. However, the corresponding residues are highly similar among granzymes D, E, and F, and in the case of granzymes D and E, identical, suggesting a very similar or identical substrate specificity. One striking structural difference between granzymes D and E is an insertion of four residues in a region that is likely to be surface-exposed, since introns are found at this position in the closely related RMCP II (34) and adipsin (38, 39) genes. Granzyme D may therefore exert a distinct biological function due to a unique surface structure despite a highly conserved substrate binding pocket.

Granzymes D, E, and F are highly glycosylated (up to 50% carbohydrate by weight) in accordance with the three to five putative N-linked glycosylation sites, whereas granzyme C appears to be devoid of carbohydrate. Its calculated molecular mass of 27 kDa agrees with the value determined by $NaDodSO_4$/PAGE and with the molecular mass after deglycosylation. These data thus suggest that structures other than carbohydrates are involved in the targeting of granzymes to lysosome-like cytoplasmic granules (40). One of the putative glycosylation sites of granzymes D and E, which is located three amino acid residues from the catalytic histidine residue, toward the amino terminus (Figs. 1 and 2), resembles a glycosylation site of granzyme B (G, H), cathepsin G, and HLP that is found seven residues away from the catalytic histidine (Fig. 4). Glycosylation at this position may influence the specificity of substrate binding (35) and may result in functional heterogeneity when this site is not constantly used for glycosylation. Since the amount of negatively charged, N-linked carbohydrates influences the net charge of granzymes, differences in the extent of glycosylation among granzymes A–F may determine their affinities for the negatively charged proteoglycans and thus affect their dissociation rates after secretion. Granzyme C, which is the most positively charged of all granzymes, would be expected to be tightly associated with the granule matrix and to act at a short distance.

The role of granzymes A–F in cell-mediated cytotoxicity is not clear. They do not appear to form a proteolytic activation cascade together with pore-forming proteins, analogous to proteins of the complement system. Rather, granzymes may fulfill diverse functions during CTL–target-cell interaction by causing a localized and time-limited proteolysis of suitable substrates after their secretion by the CTLs. The lack of measurable proteolytic activity of granzymes B, C, E, and F toward a variety of synthetic substrates (11) suggests that these granzymes bind to and cleave highly specific substrates in a very restricted fashion.

D.J. is the recipient of a long-term fellowship from the European Molecular Biology Organization. This work was supported by a grant from the Swiss National Science Foundation.

1. Henkart, P. A., Millard, P. J., Reynolds, C. W. & Henkart, M. P. (1984) *J. Exp. Med.* **160**, 75–83.
2. Podack, E. R. & Konigsberg, P. J. (1984) *J. Exp. Med.* **160**, 695–710.
3. Masson, D., Corthésy, P., Nabholz, M. & Tschopp, J. (1985) *EMBO J.* **4**, 2533–2538.
4. Young, J. D.-E, Leong, L. G., Liu, C.-C., Damiano, A. & Cohn, Z. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5668–5672.
5. 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 (London)* **318**, 289–291.
6. Pasternack, M. S. & Eisen, H. N. (1985) *Nature (London)* **314**, 743–745.
7. Garcia-Sanz, J. A., Plaetinck, G., Velotti, F., Masson, D., Tschopp, J., MacDonald, H. R. & Nabolz, M. (1987) *EMBO J.* **4**, 933–938.
8. Podack, E. R., Young, J. D.-E & Cohn, Z. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8629–8633.
9. Masson, D. & Tschopp, J. (1985) *J. Biol. Chem.* **260**, 9069–9072.
10. Zalman, L. S., Brothers, M. A. & Müller-Eberhard, H. J. (1985) *Biosci. Rep.* **5**, 1093–1100.
11. Masson, D. & Tschopp, J. (1987) *Cell* **49**, 679–685.
12. Simon, M. M., Hoschützky, H., Fruth, U., Simon, H. G. & Kramer, M. D. (1986) *EMBO J.* **5**, 3267–3274.
13. Masson, D., Nabolz, M., Estrade, C. & Tschopp, J. (1986) *EMBO J.* **5**, 1595–1600.
14. Young, J. D.-E, Leong, L. G., Liu, C.-C., Damiano, A., Wall, D. A. & Cohn, Z. A. (1986) *Cell* **47**, 183–194.
15. Pasternack, M. S., Verret, C. R., Liu, M. A. & Eisen, H. N. (1986) *Nature (London)* **322**, 740–743.
16. Zalman, L. S., Brothers, M. A., Chiu, F. J. & Müller-Eberhard, H. J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5262–5266.
17. Tschopp, J., Masson, D. & Stanley, K. K. (1986) *Nature (London)* **322**, 831–834.
18. Young, J. D.-E, Cohn, Z. A. & Podack, E. R. (1986) *Science* **233**, 184–190.
19. Gershenfeld, H. K. & Weissman, I. L. (1986) *Science* **232**, 854–858.
20. 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. & Golstein, P. (1986) *Nature (London)* **322**, 268–271.
21. Lobe, C. G., Finlay, B. B., Paranchych, W., Paetkau, V. H. & Bleackley, R. C. (1986) *Science* **232**, 858–861.
22. Jenne, D., Rey, C., Masson, D., Stanley, K. K., Plaetinck, G. & Tschopp, J. (1988) *J. Immunol.* **140**, 318–323.
23. Haymerle, H., Herz, J., Bressan, G. M., Frank, R. & Stanley, K. K. (1986) *Nucleic Acids Res.* **14**, 8615–8624.
24. Stanley, K. K. (1983) *Nucleic Acids Res.* **11**, 4077–4092.
25. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
26. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
28. Groscurth, P., Qiao, B.-Y., Podack, E. R. & Hengartner, H. (1987) *J. Immunol.* **138**, 2749–2752.
29. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
30. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
31. Reynolds, R. A., Remington, S. J., Weaver, L. H., Fischer, R. G., Anderson, W. F., Ammon, H. L. & Matthews, B. W. (1985) *Acta Crystallogr.* **B41**, 139–147.
32. Trong, H. L., Parmelee, D. C., Walsh, K. A., Neurath, H. & Woodbury, R. G. (1987) *Biochemistry* **26**, 6988–6994.
33. Woodbury, R. G., Katunuma, N., Kobayashi, K., Titani, K. & Neurath, H. (1978) *Biochemistry* **17**, 811–819.
34. Benfey, P. N., Yin, F. H. & Leder, P. (1987) *J. Biol. Chem.* **262**, 5377–5384.
35. Salvesen, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C. & Travis, J. (1987) *Biochemistry* **26**, 2289–2293.
36. Kraut, J. (1977) *Annu. Rev. Biochem.* **46**, 331–358.
37. Schmid, J. & Weissmann, C. (1987) *J. Immunol.* **139**, 250–256.
38. Phillips, M., Djian, P. & Green, H. (1986) *J. Biol. Chem.* **261**, 10821–10827.
39. Min, H. Y. & Spiegelman, B. M. (1986) *Nucleic Acids Res.* **14**, 8879–8892.
40. Kelly, R. B. (1985) *Science* **230**, 25–32.