Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 2374–2378, March 1990
Immunology

A gene pair from the human major histocompatibility complex encodes large proline-rich proteins with multiple repeated motifs and a single ubiquitin-like domain

JULIAN BANERJI, JOHN SANDS*, JACK L. STROMINGER, AND THOMAS SPIES†

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138

Contributed by Jack L. Strominger, January 2, 1990

ABSTRACT **A large number of genes has been identified previously between the class I and class II gene families within the class III region of the human major histocompatibility complex. The complete sequences of two of these genes, BAT2 and BAT3 (where BAT is HLA-B-associated transcript), which are closely linked, were determined from cDNA clones. The putative BAT2 and BAT3 proteins are 228 and 110 kDa, respectively, and do not appear to be members of any known family of proteins. However, BAT3 contains an amino-terminal ubiquitin-like domain. Both BAT2 and BAT3 are very rich in proline and include short tracts of polyproline, polyglycine, and charged amino acids. In addition, these proteins contain several unrelated families of similar repeated segments. BAT2 and BAT3 are similar to other proteins with large proline-rich domains, such as some nuclear proteins, collagens, elastin, and synapsin. BAT2 also contains four Arg-Gly-Asp (RGD) motifs typical of the integrin receptor family.**

The human major histocompatibility complex (MHC) occupies 1% of chromosome 6 and encodes a number of genes that are essential for immune function. The ability to distinguish self from nonself is mediated by the polymorphic MHC class I and class II molecules that are encoded at either end of the MHC. The class I genes are telomeric to, and the class II genes are centromeric to, a central interval of 1000 kilobases (kb) called the MHC class III region. It includes a diverse set of genes encoding members of the complement cascade, the cytokines tumor necrosis factors α and β, and the heat shock protein HSP70 (1–5).

Population studies suggest that susceptibility to a number of autoimmune diseases is associated with certain MHC haplotypes (6). Although many of these genetic associations may ultimately be related to polymorphisms in MHC class I or class II molecules (7), in some cases the increased susceptibility may be due to the combinatorial effect of several gene products, some of which may yet be unidentified. Since MHC haplotypes specify allelic combinations of a number of genes linked within the MHC, it is possible that genes lying between the class I and class II gene families could contribute to disease pathophysiology.

To identify other genes within the MHC, a series of overlapping cosmids spanning 600 kb of DNA from the MHC class III region between the class I gene *HLA-B* and the complement gene *C2* has been isolated (8–10). These cosmids have been used to identify a large number of transcription units. The high density of these transcription units correlates with the frequent occurrence of unmethylated CpG dinucleotides within this region of the genome (9). Corresponding cDNA clones for most of these "HLA-B-associated transcripts" (BATs) have been isolated. Two of these genes, BAT2 and BAT3, are located 45 kb from the closely linked genes *TNFA* and *TNFB* and 260 kb from *HLA-B* (8). BAT2 and BAT3 genes are encoded on opposite strands of DNA and terminate within a few kilobases of each other. Their mRNAs are 6.7 and 3.5 kb long, respectively, and have been found in a limited panel of cell lines examined, including HeLa, Raji (B cell), HPB-ALL (T cell), U937 (monocyte), and HepG2 (hepatoma) (8, 9).

This report presents the complete sequence of BAT2 and BAT3 derived from overlapping cDNA clones. In addition, promoter and partial intron sequences for the BAT2 and BAT3 genes have been obtained from genomic clones.‡ The BAT2 and BAT3 cDNA sequences encode large proline-rich proteins of approximately 228 and 120 kDa, respectively. Both are characterized by the repeated occurrences of several different sets of related sequence motifs. BAT2 and BAT3 do not appear to be members of any known family of proteins. However, BAT3 contains an amino-terminal domain homologous to ubiquitin, a property shared with a small group of other proteins.

MATERIALS AND METHODS

DNA Sequence Analysis. DNA sequences were obtained by the dideoxynucleotide chain-termination procedure from both strands of restriction fragments subcloned into M13 using ^{35}S-labeled dATP and T7 polymerase (Sequenase; United States Biochemical) (11). Both dGTP/ddGTP and dITP/ddITP (where dd is dideoxy) reactions were carried out. The BAT2 and BAT3 cDNA clones have been isolated from a T-cell HPB-ALL library (8, 12). Additional BAT2 cDNAs were obtained from the same library as described (8). Computer homology searches of the National Biomedical Research Foundation Protein Sequence Data Base employed the Genetics Computer Group (GCG) and Protein Sequence Query (PSQ) programs (13).

RNase Mapping. RNA samples were prepared by the guanidinium thiocyanate method from control HeLa cells and cells heat shocked for 10 min at 45°C and then incubated for 4 hr at 37°C (14, 15). SP6 RNA polymerase was used *in vitro* to synthesize [^{32}P]UTP-labeled probes from restriction fragments of BAT2 and BAT3 cDNAs subcloned into pSP72. RNase protection of probes hybridized to total HeLa cell RNA was carried out essentially as described (16). The human HSP70 probe was a gift of R. Morimoto (Northwestern University, Evanston, IL) (17).

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: MHC, major histocompatibility complex; aa, amino acid(s); RGD, Arg-Gly-Asp; BAT, HLA-B-associated transcript.

*Present address: Department of Immunology, Scripps Clinic and Research Foundation, La Jolla, CA 92037.
†Present address: Division of Tumor Virology, Dana–Farber Cancer Institute, Harvard Medical School, Boston, MA 02115.

‡The sequences reported in this paper have been deposited in the GenBank and Natl. Biomed. Res. Found. data banks (accession nos. M31293 and M31294).

RESULTS

Sequence of BAT3 cDNA. A series of overlapping restriction fragments from the 3.5-kb insert of the cDNA clone BAT3-15 (8) was subcloned into M13 and both strands were sequenced. The sequence is shown in Fig. 1 along with the predicted coding region of the major long open reading frame. Both the 5′ and 3′ untranslated regions of the BAT3 cDNA include stop codons in all reading frames. A canonical AATAAA motif is 25 base pairs (bp) from the poly(A) tail. The putative BAT3 protein contains 1132 amino acid (aa) residues. Starting 17 residues from the amino terminus of BAT3, a stretch of 75 aa is 35% homologous to the 76 aa of ubiquitin (Fig. 2*A*) (18). The remainder of BAT3 lacks significant homology to any protein in the National Biomedical Research Foundation data bank and is unusually rich in proline. It contains a segment of 12 sequential proline residues as well as nine proline triplets. The 607-aa region after the ubiquitin-like domain contains 18% proline and includes four dispersed repeated motifs of 29 aa (Fig. 2*E*). The carboxyl-terminal 202 aa contain 15% proline. These two regions flank a 231-aa central interval containing only 3% proline. Within this segment is a cysteine/histidine-rich region (between residues 851 and 884), which is an imperfect copy of the canonical zinc finger motif encoded in many genes for nucleic acid-binding proteins. These regions are thought to serve as metal coordination centers and are also found in other ubiquitin fusion genes (Fig. 1) (18–20).

BAT3 Gene Introns and Promoter. To characterize the genomic organization of the ubiquitin-like domain at the 5′ end of the BAT3 gene, restriction maps derived from cDNA and corresponding subclones of genomic DNA from the cosmid K19A (8) were aligned. This comparison showed that at least one intron interrupted the ubiquitin-like domain in the BAT3 gene, although sequences coding for ubiquitin itself lack introns (18). Partial sequencing of genomic DNA subclones from the 5′ end of the BAT3 gene showed one intron of about 2 kb after the 20th codon and one intron of 114 bp within the 60th codon of the ubiquitin-like domain (Fig. 1). This result is paralleled by the presence of introns, albeit at different locations, within the ubiquitin-like domains of other genes (18, 21, 22).

The total length of the BAT3-15 cDNA sequence is in good agreement with the length of the BAT3 mRNA as estimated by RNA blot hybridization (8). The proposed initiator methionine at nucleotide position 251 complies well with consensus sequences for vertebrate translation initiation sites (Fig. 1) (23). To determine whether the BAT3-15 cDNA contained the complete 5′ untranslated region, an RNase mapping experiment was carried out. An ≈300-bp *Xho* I–*Hin*dIII genomic fragment sharing 130 bp of overlap with the 5′ end of the cDNA was subcloned into pSP72 to generate a labeled probe. After hybridization of this probe to total HeLa cell RNA, only a single fragment of 130 nucleotides was protected from RNase digestion (data not shown). Thus, barring the presence of a mini-exon further upstream, the genomic DNA immediately adjacent to the 5′ end of the BAT3-15 cDNA is the BAT3 promoter.

The putative BAT3 promoter is very G+C-rich and includes many closely spaced *Hpa* II restriction sites. Within the putative BAT3 promoter, a heat shock element (24) was identified at position −125. Another heat shock element occurred within the first intron of the ubiquitin-like domain of the BAT3 gene (data not shown). These observations suggested that the BAT3 gene might be regulated by heat shock in a manner similar to the stress-response genes encoding

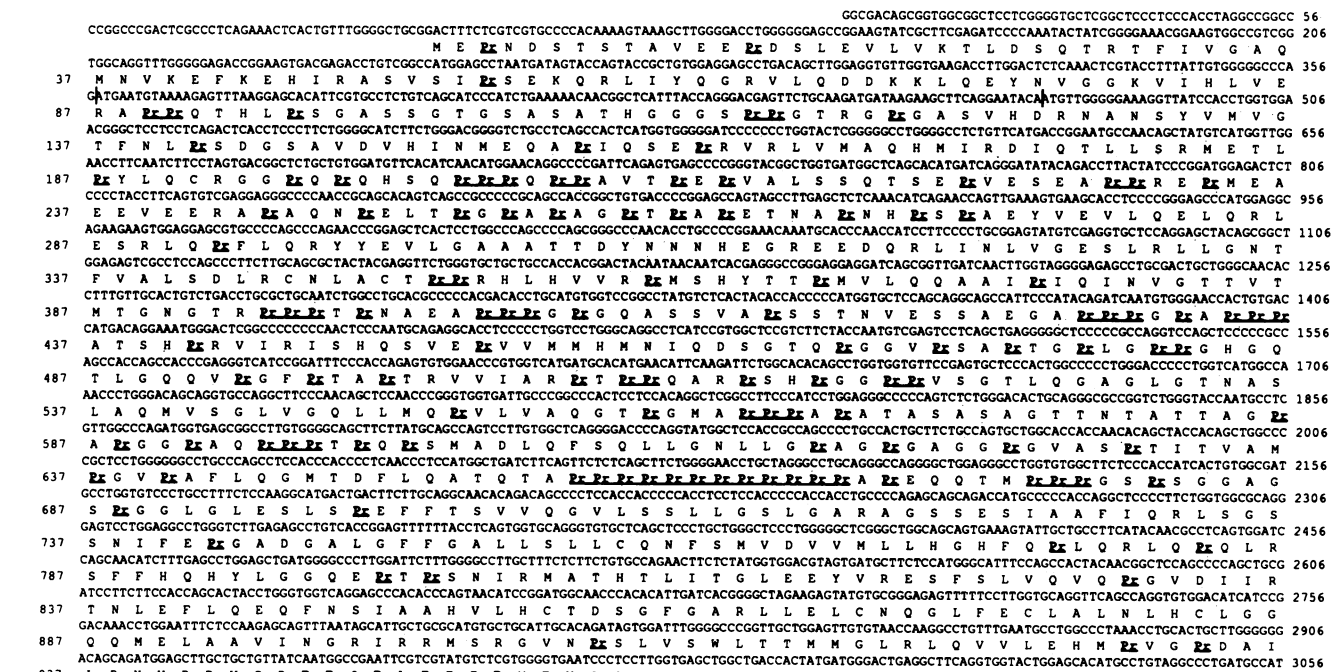

FIG. 1. BAT3 cDNA and predicted aa sequences. Residues are shown in the single-letter aa code; proline residues are represented as **Pr** and are underlined to emphasize their distribution. Known exon boundaries in the cDNA sequence are indicated by vertical bars. Their positions were determined by sequencing genomic clones by using the following three synthetic oligonucleotide primers (from 5′ to 3′): GCTTGGAGGTGTTGGTGAAG, GCATGCTGACAGAGGCACG, and CGTGCCTCTGTCAGCATCC. The numbers on the left and on the right refer to aa and nucleotide positions, respectively.

FIG. 2. Homology between BAT3 protein and ubiquitin and alignment of related segments occurring repeatedly within BAT2 and BAT3 proteins. Matching aa residues in adjacent lines are shown in bold type. Bold-faced letters above the line refer to residues that appear in three rows. Bold-faced and underlined residues appear in all rows. The single-letter aa code is used. (*A*) The aa sequence in BAT3 (bottom line) homologous to ubiquitin (top line) extends between BAT3 aa 17 and 92. The entire sequence of human ubiquitin is shown in this alignment. (*B–D*) BAT2 contains three families of homologous sequences, the type A, type B, and type C repeats. Sequences were aligned in an order maximizing aa matches between adjacent lines. Of the six BAT2 type A repeats (see text), only the four most conserved repeats are shown in *B*. They include an invariant tryptophan residue at aa 327 (first line), 142 (second line), 85 (third line), and 1784 (fourth line). The two less conserved type A repeats (data not shown) also include the invariant tryptophan residue, at positions 208 and 1847. (*C*) The two homologous BAT2 type B repeats begin at aa 237 (top line) and 478 (bottom line). (*D*) The BAT2 type C repeat family comprises three related sequences starting at aa 1899, 1965, and 2040. (*E*) BAT3 contains four similar sequences at aa 415, 242, 574, and 609, which are aligned in this order from top to bottom.

ubiquitin or HSP70 (15, 17, 18), the latter encoded 150 kb from the BAT3 gene (5). However, in an RNase protection experiment, BAT3 mRNA levels were not higher in heat-shocked HeLa cells than in untreated controls, although the level of HSP70 mRNA was increased by several orders of magnitude (data not shown).

Sequence of BAT2. The initial set of BAT2 cDNA clones formed two groups. A set of eight cDNAs of 4.6–5.0 kb corresponded to, and was coterminal with, the 3′ end of the BAT2 gene. Two additional cDNAs of 1 kb were further upstream. From these two groups, cDNAs BAT2-5 and BAT2-12, respectively, were sequenced. These two cDNAs shared a 324-bp overlap and together coded for a 6-kb open reading frame extending from the very 5′ end of the BAT2-12 cDNA. The 179-bp 3′ untranslated region included an AATAAA polyadenylylation signal and stop codons in all reading frames (Fig. 3).

To obtain the complete coding sequence of the BAT2 gene, the HPB-ALL cDNA library was screened with a probe from the 5′ end of the BAT2-12 cDNA. Thus, cDNA BAT2-17 was isolated and yielded 304 bp of additional upstream sequence. About 4 kb of genomic DNA containing this region were subcloned from the cosmid K19A (8) and sequenced. The comparison of genomic and cDNA sequences showed that the 5′ terminal 13 bp of cDNA BAT2-17 were within a separate exon. This exon was shown to include a total of 177 nucleotides by protection from RNase digestion of an appropriate genomic probe after its hybridization with total HeLa cell RNA (data not shown). Two additional cDNA clones were isolated by rescreening the library with a genomic probe containing the 177-bp exon. The sequence of the clone BAT2-18 included all these 177 bp and two additional upstream exons of 179 and 35 bp. This extended the open reading frame to a methionine that was 60 bp from the 5′ end of the cDNA. This 60-bp untranslated region contained several stop codons in the same translational frame as the initiator methionine, which lies in a sequence context consistent with the established consensus (23). Thus, the 5′ end of the BAT2 gene was defined (Fig. 3).

The putative BAT2 protein consists of 2142 aa residues, of which 409 (19.6%) are proline. This is a higher total proline content than in any protein in the National Biomedical Research Foundation data bank. In contrast to BAT3, in BAT2 the proline residues are distributed throughout the sequence (Fig. 3). In 17 instances, 3 or more proline residues are consecutive. BAT2 contains an unusual distribution of charged aa residues. Within a 49-aa segment bounded by cysteines (aa 426–475), 27 aa are charged. A region between positions 1009 and 1034 contains 8 arginine-glycine pairs and there are 14 glutamine residues within the 22 aa after position 635. Neither a hydrophobic leader nor an obvious transmembrane region are apparent in either BAT2 or BAT3.

BAT2 Domain Structure. At the BAT2 5′ end, the presence of two similarly sized adjacent exons of 177 and 172 bp suggested a repeated domain structure. Alignment of the corresponding 59- and 57-aa segments showed that these two exons matched each other at 15 positions. Moreover, a computer-assisted search of the BAT2 protein sequence identified four additional regions sharing various degrees of similarity with these two exons (Fig. 2*B*). Thus, BAT2 contains a family of six related regions. Of these, four are within the amino-terminal 337 aa. The other two homologous regions are tandemly repeated near the BAT2 carboxyl terminus. In addition to this family of related regions (type A repeats), BAT2 contains two other nonhomologous sets of related sequences (type B and type C repeats). The two type B repeats of 88 and 82 aa residues are identical at 26 positions when a single 8-residue gap is introduced in the alignment (Fig. 2*C*). These two type B repeats immediately follow the four amino-terminal type A repeats and are separated from each other by 56 aa residues that include the cysteine-bound charged domain described above. After the two carboxyl-terminal type A repeats are three type C 50-aa repeats (Fig. 2*D*). The BAT2 type A, type B, and type C repeat families lack significant homology to each other, to the 29-aa BAT3 repeat family, or to any protein in the National Biomedical Research Foundation data bank. However, at least one copy of the sequence Pro-Ala-Pro-Pro-Pro-Ala is present in a

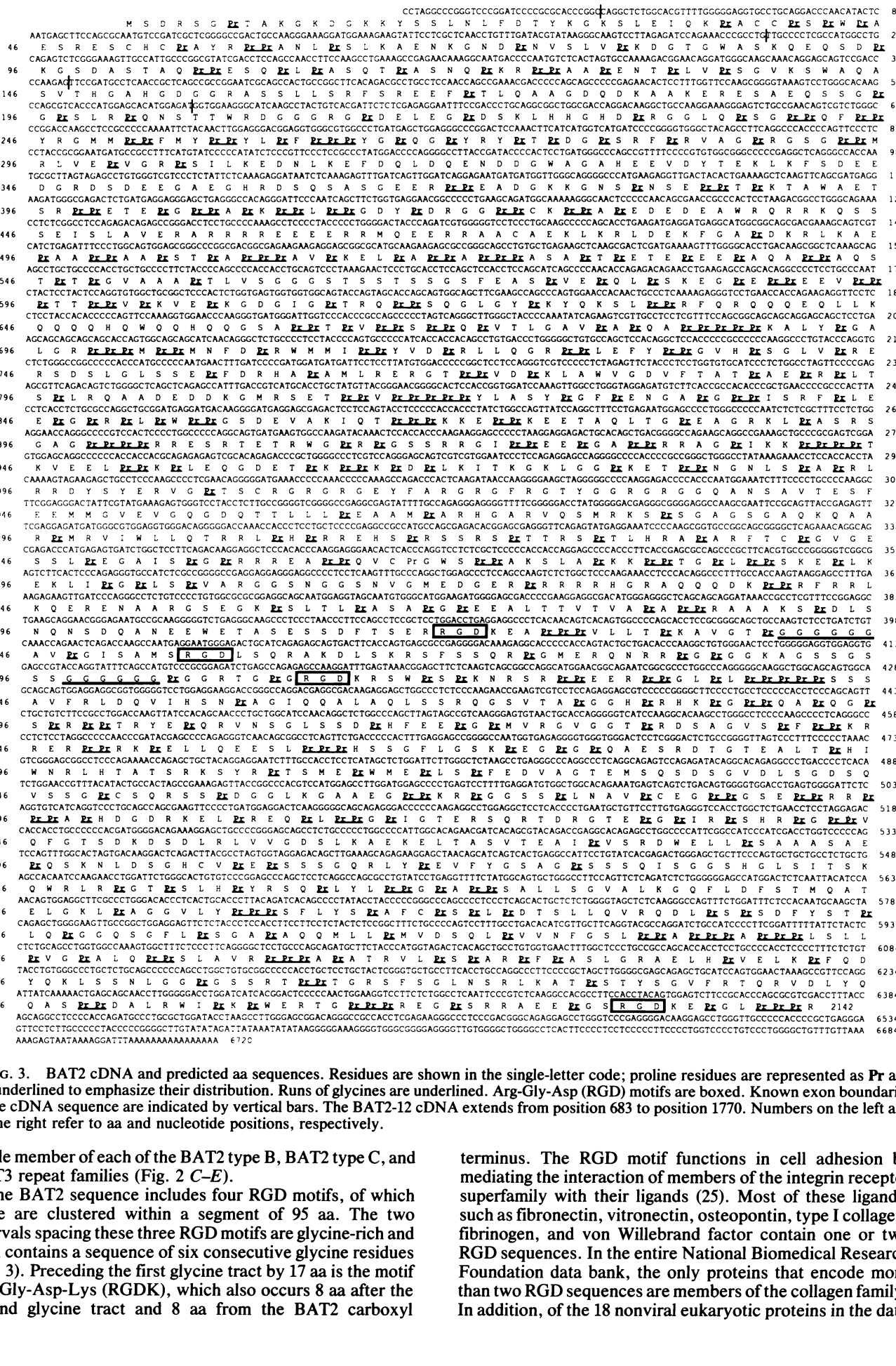

FIG. 3. BAT2 cDNA and predicted aa sequences. Residues are shown in the single-letter code; proline residues are represented as **Pr** and are underlined to emphasize their distribution. Runs of glycines are underlined. Arg-Gly-Asp (RGD) motifs are boxed. Known exon boundaries in the cDNA sequence are indicated by vertical bars. The BAT2-12 cDNA extends from position 683 to position 1770. Numbers on the left and on the right refer to aa and nucleotide positions, respectively.

single member of each of the BAT2 type B, BAT2 type C, and BAT3 repeat families (Fig. 2 *C–E*).

The BAT2 sequence includes four RGD motifs, of which three are clustered within a segment of 95 aa. The two intervals spacing these three RGD motifs are glycine-rich and each contains a sequence of six consecutive glycine residues (Fig. 3). Preceding the first glycine tract by 17 aa is the motif Arg-Gly-Asp-Lys (RGDK), which also occurs 8 aa after the second glycine tract and 8 aa from the BAT2 carboxyl terminus. The RGD motif functions in cell adhesion by mediating the interaction of members of the integrin receptor superfamily with their ligands (25). Most of these ligands, such as fibronectin, vitronectin, osteopontin, type I collagen, fibrinogen, and von Willebrand factor contain one or two RGD sequences. In the entire National Biomedical Research Foundation data bank, the only proteins that encode more than two RGD sequences are members of the collagen family. In addition, of the 18 nonviral eukaryotic proteins in the data

bank that contain the motif RGDK, 11 are collagens. Thus, the occurrence in BAT2 of four RGD motifs including three RGDK sequences, although of uncertain significance, shows certain parallels with collagen.

DISCUSSION

The BAT2 and BAT3 genes encode large proline-rich proteins with repeated domain structure. Although they do not appear to be members of any known gene family, the products of these closely linked genes may be functionally related. BAT2 and BAT3 share similarities with some transcriptional regulatory proteins containing zinc finger motifs and proline- or glutamine-rich regions (19, 26, 27). Similarities were also observed between BAT2 and BAT3 and a number of other proteins including the oncogene homolog *elk* (28), collagens, elastin, and synapsin (29–31). However, these similarities are of low statistical significance and merely reflect the common occurrence of large proline-rich domains in these structurally and functionally unrelated proteins. Polyproline, polyglycine, and individual collagen chains are, however, able to adopt a common helical structure (32), so it is possible that a similar conformation might be assumed by some of the proline- and/or glycine-rich regions in BAT2 and BAT3.

BAT3 contains an amino-terminal ubiquitin-like domain. This feature is also found in a number of other proteins. Ubiquitin itself is synthesized as a polyprotein, which is cleaved subsequently to yield ubiquitin monomers (18). Moreover, four ribosomal proteins are each synthesized with a perfect amino-terminal copy of ubiquitin, which is deleted in the mature proteins (18). The gene *An1*, whose mRNA is sequestered at the animal pole of unfertilized *Xenopus* eggs and zygotes, encodes an amino-terminal domain showing 48% homology to ubiquitin (D. Weeks and D. Melton, personal communication). The proteins encoded by the ribosomal subunit genes, An1, and to a lesser extent BAT3, contain a cysteine/histidine-rich region in their carboxyl-terminal portions. Other examples of ubiquitin fusion genes include the human genes GdX (22) and UCRP (33), of which the latter is highly inducible by γ-interferon. Moreover, in certain strains of bovine viral diarrhea virus, the presence of a ubiquitin fusion gene correlates with a cytopathic phenotype (34).

Ubiquitin, as well as other heat shock proteins are highly conserved stress-response molecules. An ability shared by these proteins is to serve as molecular chaperones (35–37), although ubiquitin is best known for its role in protein degradation. It has been proposed that the ubiquitin domain in the preprocessed ribosomal proteins serves to stabilize these proteins prior to their incorporation into multimeric complexes (38). Therefore, it is possible that the current role of the ubiquitin-like domain in BAT3 has diverged from an original chaperoning function.

A large number of genes has been mapped in the vicinity of BAT2 and BAT3 within the class III region of the MHC (8–10). In addition to the well-known genes for several complement components and *TNFA* and *TNFB*, this region may encode other, presently uncharacterized, genes involved in immune function or in MHC-associated disease susceptibility. The available data suggest that the organization of genes within this region of the MHC is highly conserved between human and mouse (8). In the mouse, a locus within this region governs the appearance of steroid-induced cleft palate (39). In addition, the *hemopoietic histocompatibility 1 locus* has been mapped to this region. This locus controls the natural killer cell-mediated rejection of bone marrow grafts (40). The possible relation of BAT2, BAT3, or both to these phenomena as well as any potential immunological relevance of these genes remain to be investigated.

We thank R. Dunbrack, J. Prendergast, J. Shin, R. Ulrich, A. Varshavsky, and D. Weeks for discussions, and B. Seed for the HPB-ALL cDNA library. J.B. was a recipient of a fellowship from the Helen Hay Whitney Foundation. This research was supported in part by a grant from the German Rheumatism Research Center (Berlin) to T.S. and by National Institutes of Health grants (DK-30241 and CA-47554) to J.L.S.

1. Carroll, M. C., Katzman, P., Alicot, E. M., Koller, B. H., Geraghty, D. E., Orr, H. T., Strominger, J. L. & Spies, T. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8535–8539.
2. Dunham, I., Sargent, C. A., Trowsdale, J. & Campbell, R. D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7237–7241.
3. Carroll, M. C., Campbell, R. D., Bentley, D. R. & Porter, R. R. (1984) *Nature (London)* **307**, 237–241.
4. White, P. C., Grossberger, D., Onufer, B. J., Chaplin, D. D., New, M. I., Dupont, B. & Strominger, J. L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1089–1093.
5. Sargent, C. A., Dunham, I., Trowsdale, J. & Campbell, R. D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1986–1972.
6. Svejgaard, A., Platz, P. & Ryder, L. P. (1983) *Immunol. Rev.* **70**, 193–218.
7. Todd, J. A., Acha-Orbea, H., Bell, J. I., Chao, N., Fronek, Z., Jacob, C. O., McDermott, M., Sinha, A. A., Timmerman, L., Steinman, L. & McDevitt, H. O. (1988) *Science* **240**, 1003–1009.
8. Spies, T., Blanck, G., Bresnahan, M., Sands, J. & Strominger, J. L. (1989) *Science* **243**, 214–217.
9. Sargent, C. A., Dunham, I. & Campbell, R. D. (1989) *EMBO J.* **8**, 2305–2312.
10. Spies, T., Bresnahan, M. & Strominger, J. L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8955–8958.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
12. Aruffo, A. & Seed, B. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8573–8577.
13. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
14. Chirgwin, J., Przybyla, A., MacDonald, R. & Rutter, W. (1979) *Biochemistry* **18**, 5295–5299.
15. Fornace, A. J., Jr., Alamo, I., Jr., Hollander, C. M. & Lamoreaux, E. (1989) *Nucleic Acids Res.* **17**, 1215–1230.
16. Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) *Nucleic Acids Res.* **12**, 7035–7056.
17. Wu, B., Hunt, C. & Morimoto, R. I. (1985) *Mol. Cell. Biol.* **5**, 330–341.
18. Finley, D., Ozkaynak, E., Jentsch, S., McGrath, J. P., Bartel, B., Pazin, M., Snapka, R. M. & Varshavsky, A. (1988) in *Ubiquitin*, ed. Rechsteiner, M. (Plenum, New York), pp. 39–75.
19. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609–1614.
20. Frankel, A. D. & Pabo, C. O. (1988) *Cell* **53**, 675.
21. Graham, R. W., Jones, D. & Candido, P. M. (1989) *Mol. Cell. Biol.* **9**, 268–277.
22. Toniolo, D., Persico, M. & Alcalay, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 851–855.
23. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8132.
24. Bienz, M. & Pelham, H. R. B. (1987) *Adv. Genet.* **24**, 31–72.
25. Ruoslahti, E. (1988) *Annu. Rev. Biochem.* **57**, 375–413.
26. Mermod, N., O'Neill, E. A., Kelly, T. J. & Tjian, R. (1989) *Cell* **58**, 741–753.
27. Courey, A. & Tjian, R. (1988) *Cell* **55**, 887–898.
28. Rao, V. N., Huebner, K., Isobe, M., Rushdi, A., Croce, C. M. & Reddy, E. S. P. (1989) *Science* **244**, 66–70.
29. Bornstein, P. & Sage, H. (1980) *Annu. Rev. Biochem.* **49**, 957–1003.
30. Sandberg, L. B., Soskel, N. T. & Leslie, J. G. (1981) *N. Engl. J. Med.* **304**, 566–577.
31. Sudhof, T. C., Czernik, A. J., Kao, H.-T., Takei, K., Johnston, P. A., Horiuchi, A., Kanazir, S. D., Wagner, M. A., Perin, M. S., De Camilli, P. & Greengard, P. (1989) *Science* **245**, 1474–1480.
32. Koenig, J. L. (1972) *J. Polym. Sci. Part D*, 60–176.
33. Haas, A. L., Ahrens, P., Bright, P. M. & Ankel, H. (1987) *J. Biol. Chem.* **262**, 11315–11323.
34. Meyers, G., Rümenapf, T. & Thiel, H.-J. (1989) *Nature (London)* **341**, 491.
35. Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W. & Ellis, R. J. (1988) *Nature (London)* **333**, 330–334.
36. Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, E. A. & Schekman, R. (1988) *Nature (London)* **332**, 800–805.
37. Chirico, W. J., Waters, M. G. & Blobel, G. (1988) *Nature (London)* **332**, 805–810.
38. Finley, D., Bartel, B. & Varshavsky, A. (1989) *Nature (London)* **338**, 394–401.
39. Tyan, M. L. (1987) *J. Immunogenet.* **14**, 239–245.
40. Bennett, M. (1987) *Adv. Immunol.* **41**, 333–445.