Brief Definitive Report

The Lymphocyte Glycoprotein CD6 Contains a Repeated Domain Structure Characteristic of a New Family of Cell Surface and Secreted Proteins

By Alejandro Aruffo,*‡§ Michael B. Melnick,‡ Peter S. Linsley,§ and Brian Seed*‡

*From the *Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114; the ‡Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115; and the §Bristol-Myers Squibb Pharmaceutical Research Institute, Seattle, Washington 98121*

Summary

The isolation, characterization, and expression of a full-length cDNA encoding the human T cell glycoprotein CD6 is described. COS cells transfected with the CD6 clone express a 90-kD protein that reacts with all available anti-CD6 monoclonal antibodies. RNA blot hybridization analysis indicates that CD6 transcripts are predominantly restricted to cells in the T lineage. The predicted CD6 sequence is 468 amino acids long, with the typical features of a type I integral membrane protein. The cytoplasmic domain of CD6 contains two serine residues, one or both of which are substrates for phosphorylation during T cell activation. The extracellular domain of CD6 is significantly related to the extracellular domain of the human and mouse T cell antigen CD5, the cysteine-rich domain of the bovine and mouse type I macrophage scavenger receptor, the extracellular domain of the sea urchin spermatozoa protein that crosslinks the egg peptide speract, the mammalian complement factor 1, and the human lung tumor antigen L3. These molecules, therefore, constitute a new gene superfamily that is well conserved across species boundaries.

CD6 is a 100–130-kD glycoprotein expressed by peripheral blood T lymphocytes, medullary thymocytes, a small number of peripheral blood B lymphocytes, as well as on some B cell chronic lymphocytic leukemias, and in some cells of the brain (1–6). Although the physiological role of CD6 is not known, recent studies suggest that CD6 plays an important role in T cell activation. The anti-CD6 mAb T12 can activate T cells in a macrophage-dependent fashion (7). A second anti-CD6 mAb 2H1, has been shown to activate T cells in conjunction with PMA or the anti-CD2 mAb T113 (2). These two mAbs recognize two distinct CD6 epitopes (7). The T12 epitope is present only under nondenaturing conditions, while 2H1 recognizes an epitope which is unaffected by denaturation. Furthermore, 2H1 binding is only partially blocked by preincubating T cells with the T12 mAb, while T12 binding is not affected by pretreatment of T cells with the 2H1 mAb (7). It has been suggested that the difference in epitope specificity of these two anti-CD6 mAbs accounts for their distinct activation properties (7). In vivo anti-CD6 mAbs (T12) have been shown to act as immunosuppressive agents for patients undergoing renal or bone marrow allograft rejection (8, 9). In addition, preliminary studies indicate that anti-CD6 mAb might have immunomodulatory effects in patients with multiple sclerosis (10).

Biochemical characterization of CD6 indicates that this protein is a highly glycosylated cell surface molecule (7, 11). The addition of N-linked carbohydrate groups accounts for at least 30 kD of the molecular mass of the protein (7, 11). Western blot analysis of the protein under reducing and nonreducing conditions suggests the presence of intrachain disulfide bonds (7). Activation of T cells results in the phosphorylation of the protein on cytoplasmic serine residues (7, 11, 12), but CD6 does not express any detectable kinase activity (7).

Here we report the isolation and expression of two human CD6 cDNA clones. RNA blot hybridization analysis with RNA samples prepared from a number of T cell, B cell, and myeloid cell lines shows that CD6 transcripts are present predominantly in T cells. Comparison of the predicted amino acid sequence of CD6 with that of other published proteins indicates that the extracellular domain of CD6 is homologous to the members of a recently identified family of proteins which contain domains defined by the cysteine-rich domain of type I macrophage scavenger receptor (13).

Materials and Methods

Preparation of cDNA Library and Recovery and Characterization of CD6 Clones. Preparation of a human peripheral blood acute lymphocytic leukemia (HPB-ALL) expression cDNA library in the shuttle vector πH3 was described previously (14, 15). The library was introduced into COS cells by the spheroplast method, and was enriched for CD6 cDNAs by panning with a panel of anti-CD6 mAbs obtained from the Second International Leukocyte Typing

 J. Exp. Med. © The Rockefeller University Press • 0022-1007/91/10/0949/04 $2.00
Volume 174 October 1991 949–952

Workshop including HB8-36, M-T502, BL-TP6a, M-T606, M-T605, M-T604, M-T603, and CD6-9C4. After three rounds of introduction into COS cells and panning, plasmid DNA was prepared from eight single colony isolates, transfected into COS cells by the DEAE-Dextran method and scored for CD6 expression by indirect immunofluorescence as described (14, 15). Additional immunofluorescence studies were carried out with mAb 2H1 which was a gift of Dr. C. Marimoto (Harvard Medical School, Boston, MA).

Immunoprecipitation, RNA Blot Hybridization, and DNA Sequencing. Radioimmunoprecipitation, RNA blot hybridization were carried out as described previously (14, 15). Sequencing was done by the dideoxy method (16).

Protein Sequence Alignment. Comparison of the predicted amino acid sequence of CD6 with other proteins in the National Biomedical Research Foundation database was done using the FASTP, Bestfit, and Compare algorithms available in the University of Wisconsin Genetics Computer Group, sequence analysis software package (Madison, WI).

Results and Discussion

Two cDNA clones encoding the lymphocyte glycoprotein CD6 (CD6-15 and CD6-13) were isolated from an expression library as previously described. COS cells transfected with these two cDNA clones reacted with all the anti-CD6 mAbs tested, as determined by indirect immunofluorescence (data not shown). Immunoprecipitation of cDNA-encoded protein from transfected COS cells showed a single band of 90 kD from COS cells transfected with the CD6-15 cDNA (Fig. 1), and a broad band ranging from 58 to 80 kD from COS cells transfected with the CD6-13 cDNA. In both cases the COS cell–derived proteins are smaller than the protein obtained from the cell line HPB-ALL. We believe that this difference in size is the result of incomplete glycosylation of the protein by COS cells, a phenomenon that we have observed in the past (15).

RNA blot hybridization analysis revealed a single RNA species of ~ 3 kb whose expression was restricted to T cells (data not shown). We found CD6 message in RNA isolated from: the T cell leukemia, HPB-ALL; and lymphokine stimulated peripheral blood T cells. No CD6 message was found in RNA isolated from: the T cell leukemia, Jurkat; the PMA stimulated or unstimulated promonocyte leukemia, U937; the human erythroleukemia, HEL; the hepatocarcinoma, HepG2; the amelanotic melanoma, C32; the EBV-transformed B lymphoblastoid cell line, CESS; the pre-B cell leukemia, IM-9; the Burkitt lymphoma, Raji; and the plasmacytoma, RPMI 8226.

The two cDNA clones contain inserts of 3 kb (CD6-15) and 1.4 kb (CD6-13), respectively. CD6-15 encodes a long open reading frame of 468 residues having the typical features of an integral membrane protein (the nucleotide sequence is available from EMBL, Genbank, and DDBJ data bases under accession number X60992). Using the method of Von Heijne (1986) (17) we predict a 24 amino acid NH_2-terminal signal sequence. Removal of the NH_2-terminal signal sequence results in a mature protein of 444-residues with a predicted M_r of 47.4 kD. In the extracellular domain there are eight N-linked glycosylation sites (Asn-X-Ser/Thr). This is in agreement with the biochemical characterization of CD6 by Swack et al. (1989, 1991) (7, 11), who showed that at least 30 kD of the molecular mass can be attributed to N-linked glycosylation. The 374 amino acid extracellular domain of CD6 is followed by a 26 amino acid hydrophobic membrane spanning domain and a 44 amino acid cytoplasmic domain. There are two serine residues in the cytoplasmic domain, one or both of which could be phosphorylation substrates during T cell activation (7, 11, 12).

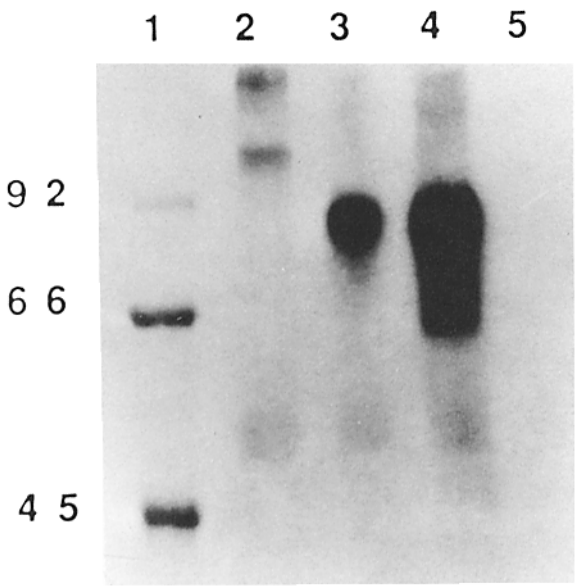

Figure 1. Surface expression of the CD6 antigen. Lane *1*; molecular mass standards. Lanes *2–5*; anti-CD6 mAb immunoprecipitates of HPB-ALL cells, COS cells transfected with CD6-15, COS cells transfected with CD6-13 and mock transfected COS cells.

Sequence analysis of CD6-13 (data not shown) showed that it is missing ~ 1.5 kb of the 3′ sequence relative to CD6-15. This results in a truncated protein, missing 23 amino acids of the cytoplasmic domain. The last amino acid in this truncated protein was encoded by sequences derived from the cloning vector which also provides an in-frame stop codon. This observation, in conjunction with the RNA blot analysis (data not shown), which shows a single 3-kb species in RNA obtained from either T cells or the HPB-ALL cell line, indicate that the shorter clone was probably generated during the cloning procedure and does not represent a naturally occurring CD6 transcript.

The predicted amino acid sequence of CD6-15 was compared with the National Biomedical Research Foundation (NBRF) database using the FASTP algorithm. Significant homology was found with the human and mouse T cell surface protein CD5 (18, 19), the cysteine-rich domain of the bovine and mouse macrophage scavenger receptor type I (13, 20), the extracellular domain of the sea urchin spermatozoa protein responsible for crosslinking the egg peptide speract (21), the human complement factor 1 (22), and the human lung tumor antigen L3 (23) (Fig. 2). Over an ~100 amino acid stretch (a region defined by the cysteine-rich domain of the macrophage scavenger receptor type I, SRCR) (13) six cysteine residues are shared with approximately equal spacing

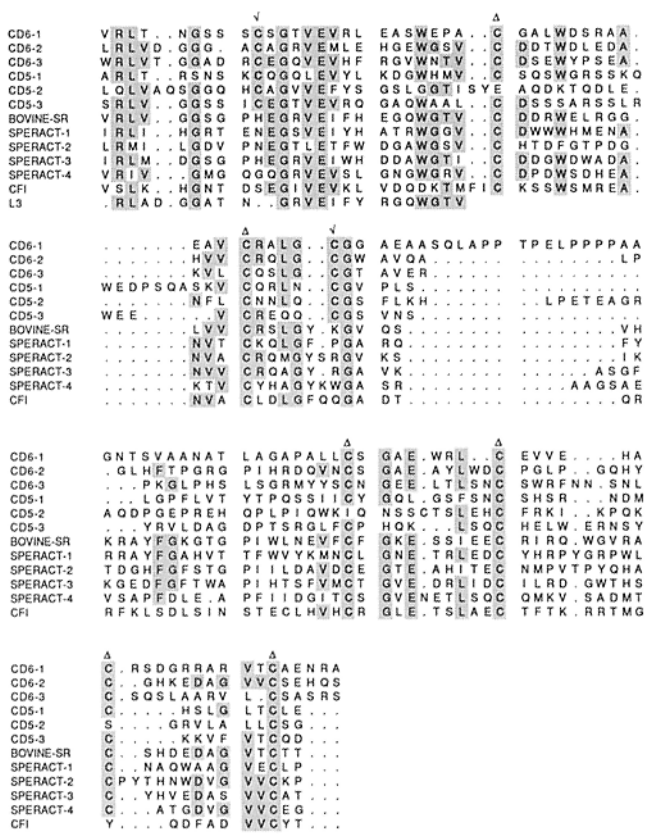

Figure 2. Homology between the CD6 sequence and human CD5, the bovine macrophage scavenger receptor type I (BOVINE-SR), the speract crosslinking protein from sea urchin sperm (SPERACT), the human complement factor I (CFI), and the human lung tumor antigen L3 (L3). The CD6-1, 2, and 3 extracellular domains correspond to residues (21-136), (137-240) and (241-341), respectively. The CD5 extracellular domains 1–3 and the speract receptor extracellular domains 1–4 were previously described (13). Cysteine residues shared by all of the SRCR family members are denoted by a Δ, the two cysteine residues shared only by CD6 and CD5 are denoted with a √

by six of the seven members of the SRCR family. At the moment only a small number of residues of the L3 antigen have been reported based on NH_2-terminal sequence; further work is needed to confirm if the L3 sequence will fully conform to that of the SRCR domain. Within this region we also find 31 highly conserved residues and two additional cysteine residues which are shared by CD6 and CD5 but are not present in the other members of the SRCR family. From this we conclude that CD6 and CD5 are more closely related to each other than to the other members of the SRCR family. The inclusion of CD5 in the immunoglobulin supergene family has been controversial (24–26). This data and a recent report describing the homology between CD5 and the type I scavenger receptor (13) indicate that both CD6 and CD5 are more closely related to the SRCR family than to the immunoglobulin supergene family.

Analysis of the sequence of CD6 with a dot matrix homology program showed that the extracellular domain of CD6 is composed of three SRCR domains (Fig. 2). This repeated domain structure is also found in the other members of the SRCR family of proteins which are integral membrane proteins. The extracellular domain of CD5 is also composed of three SRCR motifs (13) while the extracellular region of the speract crosslinking protein is composed of four SRCR domains (13) (Fig. 2). Although the type I scavenger receptor gene encodes a single cysteine-rich domain, the protein is expressed on the cell surface of macrophages as a trimer thereby bringing together three SRCR domains (20). The human complement factor 1 protein contains a single SRCR domain (13) (Fig. 2) and the small amount of sequence information available for the L3 indicates that L3 contains at least one copy of the SRCR motif (Fig. 2).

Based on the homology between the speract receptor and the other members of the SRCR family of proteins we have entertained the possibility that other membranes of this family are involved in peptide binding. Interestingly, the responses elicited by speract on its target cells are in some cases similar to those elicited by mAbs directed against CD6. For example, like speract, anti-CD6 mAbs can, under the appropriate conditions, cause increased levels of intracellular Ca^{2+} (27). A number of different thymic peptide hormones which play a role in the recruitment of T cell precursors to the thymus and in some steps of intra- and extrathymic T cell differentiation and/or activation have been reported (29–31). Receptors for these thymic peptides have yet to be identified. It is possible that CD6 may be involved in the binding of thymus derived peptides. On the other hand, it is possible that like CD5, CD6 is involved in protein binding (28). CD72, the ligand of CD5 is a member of a large family of surface proteins with an inverted membrane orientation (32). Based on the homology between CD6 and CD5, it is possible that CD6 binds to one of the proteins in the CD72 gene superfamily. The availability of cDNAs encoding CD6 will allow us to investigate these possibilities as well as others in an effort to understand the role of CD6 in the immune response.

We thank C. Morimoto for his generous gift of monoclonal antibody 2H1; J. P. Brown for pointing out the homology between the speract crosslinking protein and the L3 antigen; J. A. Ledbetter, L. J. Dangott, and P. L. Romain for helpful discussions; and M. West for help in the preparation of this manuscript.

This work was supported in part by a Leukemia Society of America Fellowship (to A. Aruffo), NIH grants AI-27849 and DK-43031 (to B. Seed), and by an award to the Massachusetts General Hospital from Hoechst AG.

Address correspondence to Alejandro Aruffo, Bristol-Myers Squibb Pharmaceutical Research Institute, 3005 First Avenue, Seattle, WA 98121.

Received for publication 17 May 1991 and in revised form 17 July 1991.

References

1. Reinherz, E.L., S. Meuer, K.A. Fitzgerald, R.E. Hussey, H. Levine, and S.F. Schlossman. 1982. Antigen recognition by human T lymphocytes is linked to surface expression of the T3 molecular complex. *Cell.* 30:735.
2. Morimoto, C., C.E. Rudd, N.L. Letvin, M. Hagan, and S.F. Schlossman. 1988. 2H1-A novel antigen involved in T lymphocyte triggering. *J. Immunol.* 140:2165.
3. Kamoun, M., M.E. Kadin, P.J. Martin, J. Nettleton, and J.A. Hansen. 1981. A novel human T cell antigen preferentially expressed on mature T cells and shared by both well and poorly differentiated B cell leukemias and lymphomas. *J. Immunol.* 127:987.
4. Bastin, J.M., S. Granger, N. Tidman, G. Janossy, and A.J. McMichael. 1981. Recognition of a human T-lymphocyte differentiation antigen by an IgM monoclonal antibody. *Clin. Exp. Immunol.* 46:597.
5. Gangemi, R.M.R., J.A. Swack, D.M. Gaviria, and P.L. Romain. 1989. Anti-T12, an anti-CD6 monoclonal antibody, can activate human T lymphocytes. *J. Immunol.* 143:2439.
6. Mayer, B., I. Funke, B. Seed, G. Riethmüller, and E. Weiss. 1990. Expression of the CD6 T lymphocyte differentiation antigen in normal human brain. *J. Neuroimmunol.* 29:193.
7. Swack, J.A., R.M.R. Gangemi, C.E. Rudd, C. Marimoto, S.F. Schlossman, and P.L. Romain. 1989. Structural characterization of CD6: properties of two distinct epitopes involved in T cell activation. *Mol. Immunol.* 26:1037.
8. Kirkman, R.L., J.L. Araujo, G.J. Busch, C.B. Carpenter, E.L. Milford, E.L. Reinherz, S.F. Schlossman, T.B. Strom, and N.L. Tilney. 1983. Treatment of acute renal allograft rejection with monoclonal anti-T12 antibody. *Transplantation (Baltimore).* 36:620.
9. Reinherz, E.L., R. Geha, J.M. Rappeport, M. Wilson, A.C. Penta, R.E. Hussey, K.A. Fitzgerald, J.F. Daley, H. Levine, F.S. Rosen, and S.F. Schlossman. 1982. Reconstitution after transplantation with T-lymphocyte-depleted HLA haplotype-mismatched bone marrow for severe combined immunodeficiency. *Proc. Natl. Acad. Sci. USA.* 79:6047.
10. Hafler, D.A., R.J. Fallis, D.M. Dawson, S.F. Schlossman, E.L. Reinherz, and H.L. Weiner. 1986. Immunologic responses of progressive multiple sclerosis patients treated with an anti-T-cell monoclonal antibody, anti-T12. *Neurology.* 36:777.
11. Swack, J.A., J.W. Mier, P.L. Romain, S.R. Hull, and C.E. Rudd. 1991. Biosynthesis and post-translational modification of CD6, a T cell signal-transducing molecule. *J. Biol. Chem.* 266:7137.
12. Cardenas, L., A.C. Carrera, E. Yague, R. Pulido, F. Sanchez-Madrid, and M.O. le Landazuri. 1990. Phosphorylation-dephosphorylation of the CD6 glycoprotein renders two isophorms of 130 and 105 kilodaltons. Effect of serum and protein kinase C activators. *J. Immunol.* 145:1450.
13. Freeman, M., J. Ashkenas, D.J.G. Rees, D.M. Kingsley, N.G. Copeland, N.A. Jenkins, and M. Krieger. 1990. An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning the type I and type II murine macrophage scavenger receptors. *Proc. Natl. Acad. Sci. USA.* 87:8810.
14. Aruffo, A., and B. Seed. 1987. Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Natl. Acad. Sci. USA.* 84:8573.
15. Seed, B., and A. Aruffo. 1987. Molecular cloning of the CD2 antigen, the T cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl. Acad. Sci. USA.* 84:3365.
16. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
17. von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683.
18. Jones, N.H., M.L. Clabby, D.P. Dialynas, H.J. Huang, L.A. Herzenberg, and J.L. Strominger. 1986. Isolation of complementary DNA clones encoding the human lymphocyte glycoprotein T1/Leu-1. *Nature (Lond.).* 323:346.
19. Huang, H.J., N.H. Jones, J.L. Strominger, and L.A. Herzenberg. 1987. Molecular cloning of Ly-1, a membrane glycoprotein of mouse T lymphocytes and a subset of B cells: molecular homology to its human counterpart Leu-1/T1 (CD5). *Proc. Natl. Acad. Sci. USA.* 84:204.
20. Kodama, T., M. Freeman, L. Rohrer, J. Zabrecky, P. Matsudaira, and M. Krieger. 1990. Type I macrophage scavenger receptor contains α-helical and collagen-like coiled coils. *Nature (Lond.).* 343:531.
21. Dangott, L.J., J.E. Jordan, R.A. Bellet, and D.L. Garbers. 1989. Cloning of the mRNA for the protein that crosslinks to the egg peptide speract. *Proc. Natl. Acad. Sci. USA.* 86:2128.
22. Goldberger, G., G.A. Bruns, M. Rits, M.D. Edge, and D.J. Kwiatkowski. 1987. Human complement factor I: analysis of cDNA-derived primary structure and assignment of its gene to chromosome 4. *J. Biol. Chem.* 262:10065.
23. Linsley, P.S., D. Horn, H. Marquardt, J.P. Brown, I. Hellström, K.E. Hellström, V. Ochs, and E. Tolentino. 1986. Identification of a novel serum protein secreted by lung carcinoma cells. *Biochem.* 25:2978.
24. Bazan, J.F. 1990. Structural design and molecular evolution of a cytokine receptor superfamily. *Proc. Natl. Acad. Sci. USA.* 87:6934.
25. Williams, A.F., and A.N. Barclay. 1988. The immunoglobulin superfamily-domains for cell surface recognition. *Annu. Rev. Immunol.* 6:381.
26. Holmgren, A., and C.I. Bränden. 1989. Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. *Nature (Lond.).* 342:248.
27. Ledbetter, J.A., C.H. June, L.S. Grosmaire, and P.S. Ravinovitch. 1987. Crosslinking of surface antigens causes mobilization of intracellular ionized calcium in T lymphocytes. *Proc. Natl. Acad. Sci. USA.* 84:1384.
28. Van de Velde, V., I. von Hoegen, W. Luo, J.R. Parnes, and K. Thielemans. 1991. The B-cell surface protein CD72/Lyb-2 is the ligand for CD5. *Nature (Lond.).* 351:662.
29. Bach, J.F., M. Dardenne, J.M. Pleau, and J. Rosa. 1977. Biochemical characterization of a serum thymic factor. *Nature (Lond.).* 266:55.
30. Goldstein, A.L., T.L.K. Low, M. McAdoo, J. McClure, G.B. Thurman, J. Rossio, C. Lai, D. Chang, S.S. Wang, C. Harvey, A.H. Ramel, and J. Meienhofer. 1977. Thymosin α1: isolation and sequence analysis of an immunologically active thymic polypeptide. *Proc. Natl. Acad. Sci. USA.* 74:725.
31. Schlesinger, D.H., and G. Goldstein. 1975. The amino acid sequence of thymopoietin II. *Cell.* 5:361.
32. Nakayama, E., I. von Hoegen, and J.R. Parnes. 1989. Sequence of the Lyb-2 B-cell differentiation antigen defines a gene superfamily of receptors with inverted membrane orientation. *Proc. Natl. Acad. Sci. USA.* 86:1352.