Biochem. J. (1994) 304, 595–601 (Printed in Great Britain)

Molecular cloning of the rat analogue of human CD59: structural comparison with human CD59 and identification of a putative active site

Neil K. RUSHMERE,* Richard A. HARRISON,† Carmen W. VAN DEN BERG* and B. Paul MORGAN*‡

* Department of Medical Biochemistry, University of Wales College of Medicine, Cardiff CF4 4XX, U.K. and † MIP Unit, MRC Centre, Cambridge CB2 2QH, U.K.

We have previously described the purification and partial characterization of the rat analogue of the human complement regulatory molecule CD59 [Hughes, Piddlesden, Williams, Harrison and Morgan (1992) Biochem. J. **284**, 169–176]. We present here the molecular cloning and full sequence analysis of this molecule. A PCR-based approach utilizing primers designed from the amino-terminal protein sequence was used to isolate a full-length cDNA clone from a rat kidney cDNA library. This clone encoded a 92 bp 5′-flanking sequence, a 66 bp signal peptide and a 315 bp coding region containing putative glycosylation and GPI-anchor signals. The 3′ untranslated flanking region was approximately 1.1 kbp long and included the poly-A tail and a CATA repeating sequence. The coding region was 58 % identical with the human cDNA at the nucleotide level and 44 % identical at the amino acid level. Despite this relatively low overall sequence conservation, several highly conserved stretches were apparent, particularly in the N-terminal portion of the molecule, in the cysteine-rich region immediately preceding the site of glycolipid attachment and in the C-terminal peptide removed during glycolipid attachment. An N-glycosylation site was identified at Asn-16 and a putative glycosylphosphatidylinositol anchor addition site at Asn-79, indicating that the mature processed protein was two residues longer than human CD59. Comparison of the sequences of rat and human CD59, together with consideration of the published three-dimensional structure of human CD59 and functional data, implicates specific regions of the protein in interactions with C-8 and/or C-9.

INTRODUCTION

Complement is tightly controlled at the cell surface by membrane-bound regulatory proteins which act either on the C-3/C-5 convertase enzymes or the membrane attack complex (MAC) (reviewed in Morgan and Meri, 1994). The major MAC-inhibiting protein on human cells is a small (approx. 20 kDa) glycosylphosphatidylinositol (GPI) anchored molecule known as CD59 (Sugita et al., 1988; Okada et al., 1989a; Holguin et al., 1989; Davies et al., 1989). We have sought to purify CD59 analogues from other species in order to examine their cross-species activities and to identify conserved functionally important regions. To date we have isolated CD59 analogues from rat (Hughes et al., 1992), sheep (van den Berg et al., 1993) and pig (van den Berg and Morgan, 1994) erythrocytes. All resemble human CD59 in terms of molecular weight, membrane anchorage and stage of complement inhibition. Although others have reported that human CD59 is species-restricted in its complement-inhibiting activity (Okada et al., 1989a,b; Rollins et al., 1991), we have found that human CD59 and each of the available species analogues inhibit complement from many different sources, suggesting that active sites are conserved between species (van den Berg and Morgan, 1994).

Human CD59 has been cloned and shows no sequence similarities with other complement-inhibiting proteins or with components of the lytic pathway (Davies et al., 1989; Okada et al., 1989b). However, it shows limited sequence identity with murine LY-6 antigens, HSV-15 protein from herpesvirus saimiri, mouse thymocyte B-cell antigen ThB, urokinase plasminogen activator receptor and several neurotoxins (Albrecht et al., 1992; Gumley et al., 1992; Fleming et al., 1993 and refs therein). To date only limited N-terminal sequence information has been available on MAC-inhibitory proteins from species other than humans.

Here we report the isolation of a cDNA clone encoding rat CD59 which consists of 92 bp of 5′-flanking sequence, a 22 amino acid signal peptide, a 104 amino acid coding region containing putative glycosylation and GPI-anchor signals and approx. 1.1 kbp of 3′-flanking region containing the signal for poly-A tail. Comparison of the full sequences of rat and human CD59 and the limited N-terminal sequences of sheep and pig CD59 highlights conserved regions which, together with structural information on the human protein (Fletcher et al., 1994), suggests specific regions likely to be of functional importance in complement inhibition.

MATERIALS AND METHODS

Murine Moloney leukaemia virus (MMLV) reverse transcriptase and corresponding buffer, DNA molecular weight markers and restriction enzymes were from Gibco BRL (Paisley, U.K.). Rat Kidney cDNA Lambda gt11 expression library, ExAssist helper phage and Nutrap purification columns were from Stratagene (Cambridge, U.K.). The dNTPs, RNAase inhibitor RNasin, Magic PCR columns and the fmol DNA sequencing kit were from Promega (Southampton, U.K.). *Taq* DNA polymerase was from Bioline (Finchley, London, U.K.). Denatured, sonicated Salmon Sperm DNA was from ILS Biotechnology (Llandysul, Dyfed, U.K.). RNAzol was from AMS Biotechnology (Witney, Oxford, U.K.). Filter membranes (nylon) were supplied by

Abbreviations used: MAC, membrane attack complex; GPI, glycosylphosphatidylinositol; ORF, open reading frame; MMLV, murine Moloney leukaemia virus; RT, reverse transcriptase.

‡ To whom correspondence should be addressed.

Amersham International U.K. (Little Chalfont, Bucks., U.K.). Oligonucleotide primers were synthesized on an Applied Biosystems 392 RNA/DNA synthesizer (Warrington, Cheshire, U.K.). PCR was carried out using a Perkin-Elmer Cetus DNA thermal cycler (Beaconsfield, Bucks., U.K.). All other materials were from Sigma (Poole, Dorset, U.K.) or Fisons (Loughborough, Leics., U.K.) and were the highest grade available.

N-terminal sequence analysis

Rat CD59 was purified and partially sequenced using an Applied Biosystems 477A protein sequencer equipped with 120 analyser and 610A data analysis software as described previously (Hughes et al., 1992).

Rat cDNA synthesis

Total RNA was extracted from rat liver and kidney and from a CD59 positive subline of the human myelomonocytic cell line U937 (van den Berg et al., 1994) using RNAzol. RNA was then used in rapid amplification of cDNA ends derived adapter ($GACTCGAGTCGACATCGAT_{17}$)-primed cDNA synthesis (Frohman, 1990) catalysed by MMLV reverse transcriptase (MMLV-RT) using the following conditions: total RNA (1 μg) was incubated in MMLV-RT buffer, 50 mM Tris/HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$ containing 1 mM dNTPs, 10 mM dithiothreitol, 1 unit RNasin, 200 units MMLV-RT, 45 pmol $oligoT_{17}$-adapter primer in a total volume of 20 μl at 22 °C for 10 min and subsequently at 42 °C for 50 min; the solution was denatured at 95 °C (2 min) and quenched on ice.

PCR amplification of rat CD59 cDNA

A portion (approx. 7%) of the original cDNA reaction was amplified using degenerate primers based on deduced peptide sequence (see Figure 1 for primer selection) using a variation of the touchdown procedure of Don et al. (1991). The cDNA in 10 mM Tris/HCl (pH 9.0)/50 mM KCl/1.5 mM $MgCl_2$/0.1% Triton X-100 containing 0.13 mM dNTPs was mixed with 75 pmol of a 1:1 mixture of the primers $RCD59A_1$ and $RCD59A_2$ and 75 pmol of the primer RCD59B with final addition of 2.5 units *Taq* at the initial denaturation temperature of 95 °C for 3 min. The initial cycling parameters comprised 94 °C for 1 min, 65 °C for 30 s and 72 °C for 1 min. Thereafter, the annealing temperature of the reaction was decreased 2 °C every second cycle from 65 °C to a touchdown at 47 °C, at which temperature 17 cycles were carried out. An aliquot was removed, diluted 1:50, reamplified and analysed by agarose gel electrophoresis. The 108 bp DNA fragment was isolated from a 6% preparative polyacrylamide gel by electroelution (Gobel et al., 1987) and quantified by comparison with DNA standards.

Initial sequencing of rat CD59 and production of labelled probe

The 108 bp fragments derived from PCR reactions of rat liver and kidney cDNAs were sequenced using the fmol sequencing kit according to the manufacturer's protocol employing $RCD59A_1$ end-labelled primer and the following conditions. An initial denaturation cycle of 95 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 42 °C for 30 s and 70 °C for 1 min. Sequence data were used to generate two additional primers RCD59C and RCD59D (see Figure 1). Labelling was achieved by PCR employing the following conditions: 150 ng of 108 bp DNA fragment in 16 mM $(NH_4)_2SO_4$, 67 mM Tris/HCl (pH 8.8) containing 0.01% Tween-20, 1 mM $MgCl_2$, 10 μM dNTPs, 2.5 units *Taq* and 75 pmol each of RCD59A, $RCD59A_2$ and RCD59B. A 5 μl volume of each of [α-^{32}P]dCTP and [α-^{32}P]dTTP (3000 Ci/mmol) in 50 μl was amplified using the following cycling parameters: initial denaturation step 94 °C for 3 min, 10 cycles at 93 °C for 20 s, 57 °C for 20 s, 72 °C for 20 s and finally 72 °C for a further 10 min. Unincorporated dNTP was removed on Nutrap push columns.

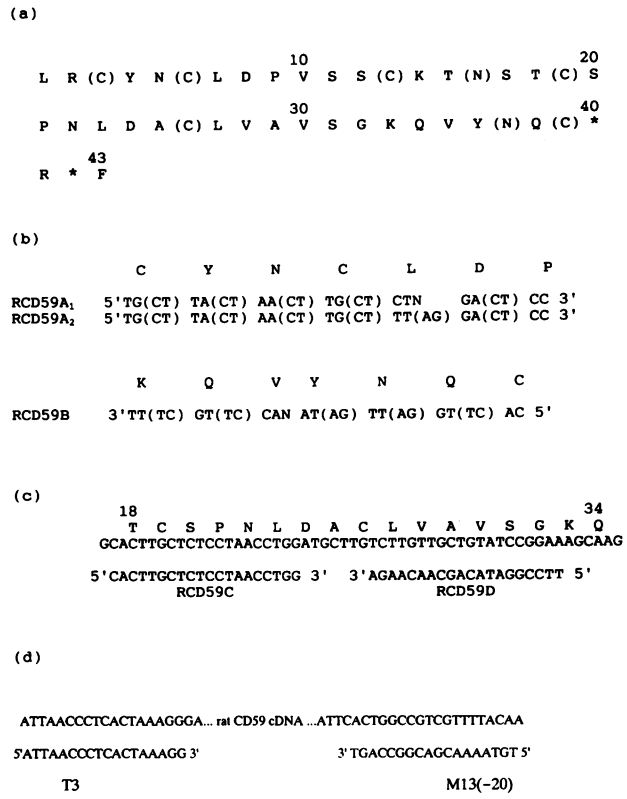

Figure 1 Derivation of primers from N-terminal sequence

(**a**) N-terminal sequence analysis of purified rat CD59. No signal was obtained at residues in brackets (mainly cysteines) and these were tentatively assigned by comparison with the human CD59 sequence. Residues identified with (*) were unassigned. (**b**) Protein sequence allowed the manufacture of degenerate oligonucleotide primers based on the N-terminal residues ($RCD59A_1$, $RCD59A_2$) and C-terminal residues (RCD59B) of the 43 amino acid fragment. The C-terminal primer was synthesized on the assumption that residue 37, which was not assigned in protein sequencing, was identical to human residue 37 i.e. Asn. (**c**) Partial DNA sequence of the RT-PCR generated 108 bp probe. Translation of this sequence is consistent with the N-terminal sequence data residues (18–34) and permitted the manufacture of DNA sequencing primers RCD59C and RCD59D. (**d**) Oligonucleotide primers T3 and M13 (−20) are located 5′ and 3′ to the rat CD59 DNA sequence within the Lambda Zap vector and allowed the amplification and sequence determination of rat CD59 cDNA clones 5A and 5B.

Screening rat kidney cDNA library

Approximately 3.6×10^5 colony plaques obtained from the rat kidney cDNA library were transferred in duplicate to nylon filters and screened by hybridization with the 108 bp probe generated as described above. Initially, filters were prehybridized in 0.8 M NaCl, 0.02 M Pipes (pH 6.5), 50% formamide, 0.5% SDS and 100 μg/ml denatured sonicated salmon sperm DNA at 42 °C for 1 h, and subsequently hybridized overnight at 42 °C in fresh prehybridization solution containing denatured PCR labelled probe. The filters were washed at room temperature with 2 × SSC (1 × SCC = 150 mM NaCl/15 mM sodium citrate)/0.5% SDS and autoradiographed. Colonies positive on both of the paired filters were eluted in lambda dilution buffer (100 mM NaCl/50 mM Tris/HCl (pH 7.5)/8 mM $MgSO_4$,$7H_2O$/0.01%

gelatin) and screened for the presence of rat CD59 by PCR using RCD59A$_1$/A$_2$ and RCD59B primers.

Sequence of rat CD59 cDNA

Positive colonies were PCR-amplified using T3 and RCD59B primers. T3 primer is the T3 RNA polymerase binding-site located 5′ to the polylinker and rat cDNA insert in the Lambda Zap vector (Figure 1d). An aliquot of eluted phage (10 μl), previously denatured at 95 °C for 3 min, was incubated in 16 mM $(NH_4)_2SO_4$/67 mM Tris/HCl (pH 8.8)/0.01 % Tween-20/1 mM $MgCl_2$/100 μM dNTPs containing 2.5 units *Taq* in 50 μl using the following cycling parameters: initial denaturation step at 94 °C for 3 min, 35 cycles at 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min. The resultant PCR products were purified and sequenced using an end-labelled RCD59D primer, yielding 5′-flanking and N-terminal coding-region sequence. PCR amplification was also performed (conditions as above) with the RCD59A$_1$/A$_2$ primer and the 3′-located M13-20 primer (located 3′ to the polylinker and rat cDNA insert in the vector) (Figure 1d). Sequencing of the purified PCR products using end-labelled RCD59C primer revealed the composition of the C-terminal part of the coding region and a portion of the 3′ untranslated sequence. Colonies were excised as phagemids using ExAssist helper phage and sequenced to confirm the above results. Initial DNA and protein homology analysis was accomplished using the Needleman and Wunsch (1970) alignment method in the Genepro DNA software supplied by Biotech Instruments (Luton, Beds., U.K.), and final refinements were performed 'by eye'.

Structural analysis and comparisons

Structural analysis of human CD59 was performed as previously described (Fletcher et al., 1993; 1994) using coordinates for the lowest structure calculated for human CD59. Structural diagrams and comparisons with rat CD59 were obtained using MAGE software (copyright David C. Richardson, Little River Institute, 5820 Old Stony Way, Durham, NC 27705, U.S.A.).

RESULTS

N-terminal protein sequencing and probe design

Protein sequence analysis extended the previously reported 15 residue N-terminal sequence (Hughes et al., 1992) by a further 28 amino acids. Residues 3, 6, 13, 19, 26 and 39 were unidentified but predicted from the human CD59 sequence to be Cys. As no signal for residue 16 was obtained and residue 18 was identified as Thr, a tentative assignment of 16 as a glycosylated Asn was made. Availability of additional peptide sequence enabled the design of degenerate primers to the N- and C-terminal regions of this 43 residue stretch (Figure 1a and b). Two primers encoding amino acids 3–9 (Cys-Tyr-Asn-Cys-Leu-Asp-Pro) were produced (RCD59A$_1$ and RCD59A$_2$) to eliminate the possibility of TTT (Pro) at the Leu position and exhibited individual degeneracies of 128 and 64 respectively. A primer encoding amino acids 33–39 (Lys-Gln-Val-Tyr-Asn-Gln-Cys) of the sequenced 43 residue stretch, RCD59B, displayed a degeneracy of 256 (Figure 1b). This primer was designed on the assumption that residue 37 was Asn from comparison with the human CD59 sequence.

Sequencing the 5′ segment of rat CD59

PCR from rat liver and kidney cDNAs identified a 108 bp fragment, consistent with the predicted peptide length of 36 amino acid residues. There was no specific signal from the control human (U937) cDNA. Direct PCR sequencing of both rat-kidney and -liver 108 bp fragments using the RCD59A$_1$ primer revealed an internal 55 bp sequence which exhibited 58 %

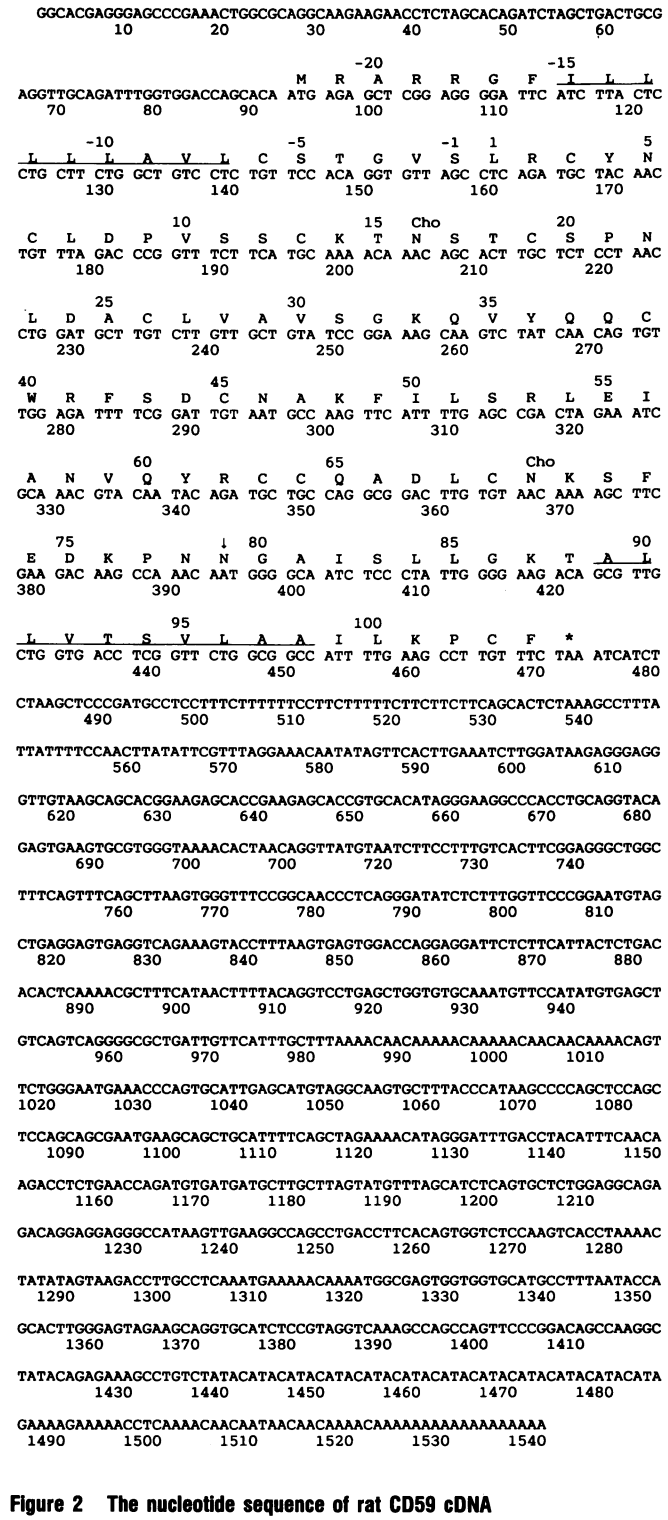

Figure 2 The nucleotide sequence of rat CD59 cDNA

All 1540 bases of the rat cDNA nucleotide sequence are given. The translation of the ORF is given in single letter code with the probable initiation codon (Met-22) and stop codon (*) indicated. The putative GPI anchor addition site (Asn-79) is indicated by (↓). The hydrophobic residues in both the signal peptide (Ile-15 to Leu-7) and the C-terminal region adjacent to the putative GPI anchor site (Ala-89 to Ala-98) are underlined. Putative N-glycosylation signals are also indicated (Cho).

nucleotide identity with the corresponding human CD59 region (Figure 1c). The deduced amino acid sequence of this region was consistent with the experimentally determined protein sequence (residues 10–32) and confirmed that the 108 bp fragment was derived from rat CD59.

Screening rat kidney cDNA library

The 108 bp fragment was used to screen a rat kidney cDNA library and yielded four positive colonies. These colonies were subjected to PCR screening using the degenerate primers, RCD59A$_1$/A$_2$ and RCD59B. Two of the colonies yielded putative full-length cDNA clones (5A and 5B) which were isolated and sequenced as detailed in the Materials and methods section (Fig. 1c).

Sequence analysis of 5A phagemid clone confirmed PCR sequence data and extended the 3′-untranslated region to the poly-A tail. The complete cDNA was 1540 bp long and comprised 92 bp of 5′-flanking sequence, a 381 bp open reading frame (ORF) and a 3′-untranslated region comprising a further 1067 bp, including the poly-A tail (Figure 2). The rat CD59 cDNA was considerably longer than that for human CD59 (1140 bp), primarily due to its longer 3′-untranslated region. Sequence identity between the rat and human CD59 cDNAs was found throughout: 49% in the 5′-flanking region, 47% in the 3′-flanking region and 58% in the coding region. The 3′-flanking region contained a tetranucleotide repetitive sequence (CATA), occurring 11 times. This repeating element is not present in either the human CD59 cDNA or genomic sequences (Davies et al., 1989; Sawada et al., 1990) but is found in other rodent sequences (Watanabe and Ohshima, 1988; Pierce et al., 1992; Kimura et al., 1989; Kawamura et al., 1992).

The deduced ORF amino acid sequence consisted of 126 amino acids (compared with 128 in humans) and contained a 22 amino acid putative signal peptide predominantly composed of hydrophobic residues (25 amino acids in humans). The putative native protein was encoded from amino acids 1 (Leu) to 104 (Phe) (Figure 2). The deduced amino acid sequence for rat CD59 differed from the experimentally determined protein sequence only at residue 43. All of the assigned Cys residues were correct as was assignment of residue 16 as Asn. However residue 37, assigned as Asn and utilized in the design of primer RCD59B, was in fact Gln. This error led to mismatches at positions 13 and 15, TT(AG) instead of GT(TC), in primer RCD59B.

DISCUSSION

Although several human proteins which inhibit complement have been characterized, little is known about control of complement in other species. We have previously shown that rat cells express a complement inhibitory molecule which is an analogue of human CD59. Rat CD59 is broadly expressed and is an important protective factor in many tissues (Hughes et al., 1993; Piddlesden and Morgan, 1993; Funabashi et al., 1994; Matsuo et al., 1994). While others have reported that human CD59 mediates efficient inhibition only against human MAC (Okada et al., 1989a,b; Rollins et al., 1991), we have found that human CD59 and analogues from other species (rat, sheep and pig) are all effective inhibitors of MACs from many different species (van den Berg and Morgan, 1994). While precise measurements of the affinity of their interactions with MAC complexes from different sources have not been performed, it is clear that CD59 analogues are not homologously restricted in their complement inhibitory activity. This implies that C-8 and/or C-9 binding-sites are, for the most part, conserved between species studied thus far. Molecular characterization of the rat CD59 analogue would thus facilitate understanding of the mechanism of CD59-mediated complement inhibition and might identify functionally important conserved regions. We report here the isolation and characterization of a rat kidney cDNA clone encoding the rat CD59 protein. This, together with the recent determination of a high-resolution structure for human CD59 (Fletcher et al., 1994), permits an analysis of the relationship between the mature rat and human proteins.

At the protein level, the human and rat CD59 molecules showed 45% sequence identity (57 amino acid matches out of 126 aligned; Figure 3a) with a high proportion of 'conservative' replacements. Sequence identity between the mature CD59 proteins was 44% (35 matches out of 79 aligned, including all ten Cys residues) and was not evenly distributed throughout the protein, with residues 43–60 being poorly conserved. The human CD59 GPI-anchor site is unequivocally located at Asn-77 (Sugita et al., 1993). Based on the consensus for the GPI addition signal (Gerber et al., 1992) and the spacing between the hydrophobic region and the anchor addition site in the human protein we predict that the anchor site in rat CD59 is at Asn-79, although Asn-78 is also a possibility. The predicted molecular mass of the mature protein is 8936 Da compared with the unprocessed mass of 13790 Da, a reduction of 35% during protein processing. In contrast to human CD59, which contains only one potential N-linked glycosylation site (Asn-X-Thr/Ser), rat CD59 possesses two at Asn-16 and Asn-70. The Asn-70 site lies in the sole region of gross structural difference between human and rat CD59s. There are a number of reasons for believing that this site is not utilized. First, with the exception of the cysteines, this residue is the most highly conserved in the extended CD59/neurotoxin fold family (Albrecht et al., 1992; Fletcher et al., 1994). In all of the solved structures for this fold, including human CD59, the residue packs to the interior of the protein and is therefore not available for glycosylation (Fletcher et al., 1993, 1994). Secondly, in the folded structure this residue is spatially close to the site of GPI anchor attachment and, even if the side-chain were available, glycosylation would be of a residue lying in a 'hidden' environment between the protein domain and the cell membrane (Figure 4a). Thirdly, it is reported that potential sites for N-linked glycosylation that lie close to the C-terminus of a protein are frequently not used (Gavel and Heijne, 1990). In contrast, the other potential N-linked glycosylation site (Asn-16) is used (no phenylthiohydantoin amino acid is seen at this point in protein sequencing), and lies in close proximity to the glycosylated Asn-18 of human CD59.

The displacement of the carbohydrate by two residues might explain the structural variation between rat and human CD59s in their C-terminal regions. One possible role of the carbohydrate would be to maintain the orientation of the protein relative to the membrane. If this required interactions with residues at the C-terminus of the protein (e.g. Phe-71, which is otherwise solvent-exposed), insertion of an additional two residues near the C-terminus of the rat protein might permit preservation of specific interactions with the laterally displaced carbohydrate. A role of carbohydrate in maintaining protein orientation would also afford an explanation as to why deglycosylated human CD59 is non-functional although still able to bind C-8 and C-9 in ligand-blotting experiments (Ninomiya et al., 1992); exposure of carbohydrate-protected surface hydrophobic residues (Val-17 in humans) could result in a rearrangement of proximal residues and distortion of the active site.

Apart from homology with human CD59 and with the available protein sequence for sheep and pig CD59s (Figure 3b),

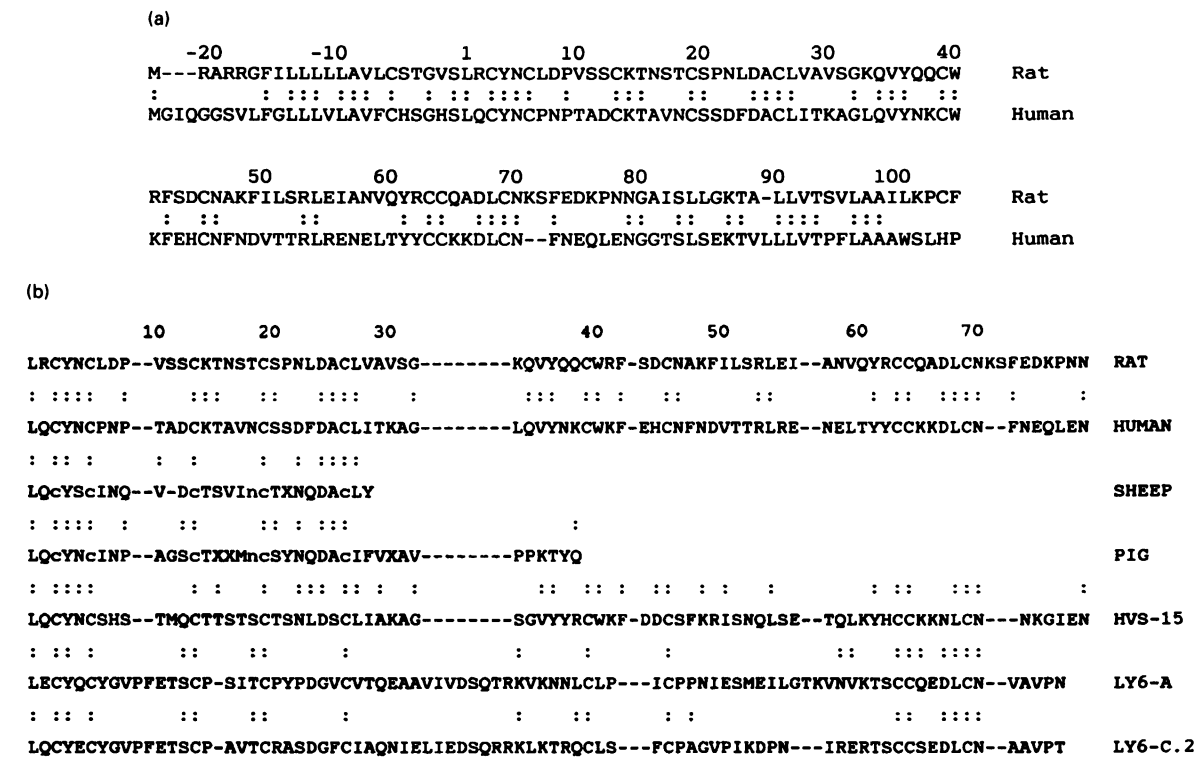

Figure 3 Alignment of rat CD59 with related proteins

(**a**) Amino acid alignment of rat (top) and human (bottom) CD59. Numbering is from the first residue of the mature rat protein. Residues conserved in both sequences are indicated by (:). Gaps inserted to optimize alignment are indicated by (—). (**b**) Comparisons of the amino acid sequences of rat, human, sheep and pig CD59 and the homologous proteins HVS-15 and the murine LY-6 antigens A and C.2. Gaps inserted to optimize alignment are indicated by (—). Numbering refers to the rat CD59 sequence and residues conserved with rat CD59 are indicated by (:). For sheep and pig CD59 only partial N-terminal sequences are available; here 'X' indicates a residue which could not be identified during protein sequencing and for which no confident prediction could be made. Probable cysteine residues, not determined by sequencing, are indicated by (c). For HVS-15, the N- and C-terminal residues of the mature protein are not known; here they are based on the predicted leader peptide and comparison with human CD59.

databank searches revealed significant sequence identities with the HVS-15 protein from herpesvirus saimiri (34%) (Albrecht et al., 1992) and with the murine LY-6 antigens (LY-6A, 21%; LY-6C, 16%) (Khan et al., 1990; Albrecht et al., 1992; Fleming et al., 1993 and references therein) (Figure 3b). The LY-6 antigens have been postulated to be murine analogues of CD59 (Forsberg et al., 1993). However, detailed analysis of the chromosomal location of the human CD59 gene and comparison with the locus for the LY-6 genes does not support this hypothesis (Bickmore et al., 1993). Our demonstration that homology between rat and human CD59 (44%) is much stronger than that between rat CD59 and the LY-6 antigens (21% or less) also makes it unlikely that LY-6 is murine CD59.

Finally, comparative structural analysis of rat (and other mammalian species) and human CD59s gives significant insight into the likely location and nature of the C-8 and/or C-9 binding site of the molecule. Peptide binding studies have shown that a fragment of human CD59 comprising residues 1–14, 15–30 and 39–41, maintained in native disulphide linkage, binds nascent C5b-8, whereas the major peptide derived from the C-terminal region (44–53, 55–65/66, 66/67–77) does not (Ferriani et al., 1993). Structural features likely to be retained in the first peptide fragment are the anti-parallel 2-stranded β-sheet with its linking turn structure, and the 20–24 loop maintained by the 3–26 and 19–39 disulphide bonds (Figure 4a). Examination of the human CD59 structure shows that in the native protein the solvent-exposed residues of this peptide are separated by N-linked carbohydrate (Figure 4c). Furthermore, of the fully-exposed sidechains in the 7–13 loop, Pro-7, Asn-8 and Asp-12, none are conserved in rat CD59 (Leu, Asp and Ser respectively), only one is conserved in pig (Asn-8), and two are conserved in sheep (Asn-8, Asp-12) CD59s, and the partially-exposed side-chain of 15-Thr is found only in rat and human CD59s. In contrast, solvent-exposed residues of the 20–24 loop are more highly conserved. In particular, Ser-20, Asp-24 and the partially exposed N-terminal Leu residue are fully conserved between rat, pig and human CD59s (Ser-20 is replaced by the similar hydroxyl-containing Thr residues in sheep CD59) and are clustered on the surface of human CD59 (Figure 4d). As rat, pig and sheep CD59s are effective inhibitors of human complement (van den Berg and Morgan, 1994), these data strongly implicate this loop, and in particular, these core residues, to comprise the C-8 and/or C-9 binding site. In human CD59 this cluster is flanked by Ser-21, Asp-22 and Lys-41, also contained in the active peptide fragment, by Phe-42 at the start of the crossover-loop connecting strands two and three of the central β-sheet and by Lys-65 from the C-terminal peptide region. In rat CD59 Phe-42 is conserved and other substitutions are 'conservative' in nature: Ser-21 is replaced by Pro, Asp-22 by Arg (conserved in sheep and pig), Lys-41 by Arg and Lys-65 by Gln. As the functional groups on each of these side-chains, particularly that of Ser-21 which points away from the cluster, lie on the periphery of the proposed

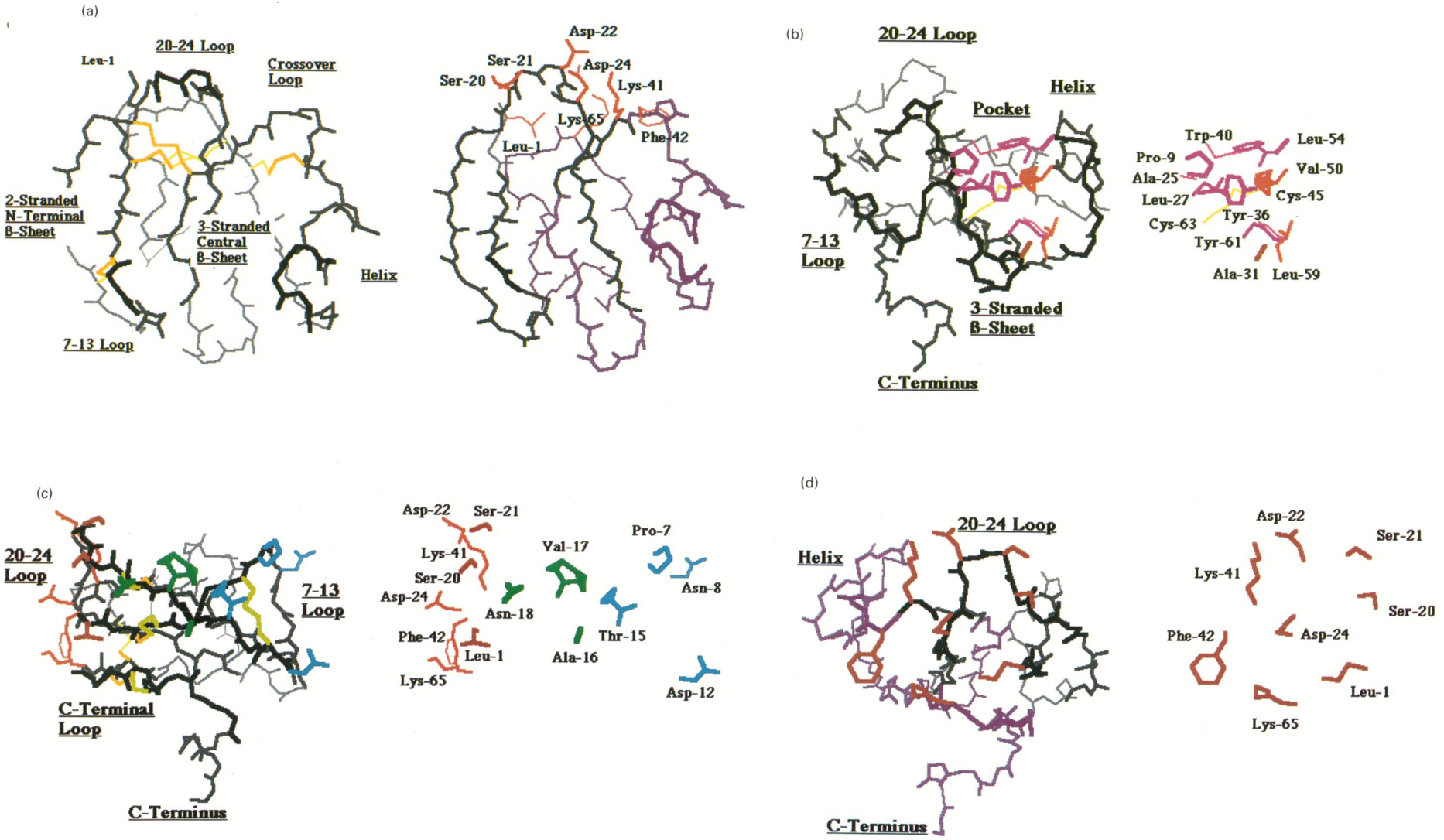

Figure 4 Identification of a putative active site in CD59

(**a**) Folding of human CD59. In this orientation ('top') the C-terminal tail lies beneath the central β-sheet. On the left, disulphide bonds are coloured in yellow. On the right, the location of residues of the proposed binding site are shown (orange). All except Phe-42 and Lys-65 are contained in the major peptide fragment (residues 1–14, 15–30, 39–41) that retains C5b-8 binding activity. The peptide backbone of this fragment is coloured black, and that of the remainder of the molecule purple. (**b**) Side-chain conservations in the hydrophobic pocket lying between the helix and the central sheet. In this projection ('side') the 3-stranded sheet is seen 'edge-on', and the C-terminal peptide lies at the base of the structure. Residues conserved in human and rat CD59 are coloured in magenta and those that are replaced in the rat are coloured orange. It is particularly noteworthy that Val-50, which lies at the centre of this pocket bounded by Trp-40, Tyr-36 and Tyr-61, is replaced by a similar bulky hydrophobic residue Ile-50 in the rat. The conserved disulphide bond between Cys-45 and Cys-60 that lies across the base of the pocket is coloured in yellow. (**c**) The location of the N-glycosylation site of human CD59. 'Side' projection, rotated to illustrate the relative locations of the carbohydrate, the 7–13 and the 20–24 loops. Asn-18 (green) which is glycosylated in human CD59, lies between the exposed residues of the 7–13 (blue) and the 20–24 (orange) loops. In human CD59, both Val-17 and Ala-16 (also coloured green) contact the N-linked carbohydrate. The glycosylation site of rat CD59 is at residue 16 and Val-17 is replaced by a hydrophilic residue, Ser. Disulphide bonds are again coloured yellow. (**d**) The spatial organization of side-chains of the proposed binding site. This is again a 'side' projection further rotated to view the proposed binding site 'face-on'. Components of the site are coloured in brown. Backbone colouring is as described in (**a**).

binding site, they are likely to influence only the affinity of the interaction and not the ability of the molecule to function as a cross-species MAC restriction factor. Proof of the importance of this putative binding site awaits the outcome of ongoing mutagenesis studies.

This work was supported by The Wellcome Trust. B.P.M. is a Wellcome Senior Fellow. We thank C. M. Fletcher and D. Neuhaus for permission to use the coordinates for human CD59 in constructing Figure 4.

REFERENCES

Albrecht, J.-C., Nicholas, J., Cameron, K. R., Fleckenstein, B. and Noness, R. W. (1992) Virology **190**, 527–530

Bickmore, W. A., Longbotton, D., Oghene, K., Fletcher, J. M. and van Heyningen, V. (1993) Genomics **17**, 129–135

Davies, A., Simmons, D. L., Hale, G., Harrison, R. A., Tighe, H., Lachmann, P. J. and Waldmann, H. (1989) J. Exper. Med. **170**, 637–654

Don, R. H., Cox, P. T., Wainwright, P. J., Baker, K. and Mattick, J. S. (1991) Nucleic Acids Res. **19**, 4008–4010

Ferriani, V. P., Harrison, R. A. and Lachmann, P. J. (1993) Mol. Immunol. **30**, 10 (abstr.)

Fleming, T. J., O'Huigin, C. and Malek, T. R. (1993) J. Immunol. **150**, 5379–5390

Fletcher, C. M., Harrison, R. A., Lachmann, P. J. and Neuhaus, D. (1993) Protein Sci. **2**, 2015–2027

Fletcher, C. M., Harrison, R. A., Lachmann, P. J. and Neuhaus, D. (1994) Structure **2**, 185–199

Forsberg, U. H., Bazil, V., Stefanova, I. and Schroder, J. (1993) Immunogenetics **30**, 188–193

Frohman, M. A. (1990) in PCR Protocols: A Guide to Methods and Applications (Innis, M. A., Gelfand, D. H., Sninsky, J. J. and White, T. J., eds.), pp. 28–38, Academic Press, London

Funabashi, K., Okada, N., Matsuo, S., Yamamoto, T., Morgan, B. P. and Okada, H. (1994) Immunology **81**, 444–451

Received 3 May 1994/5 July 1994; accepted 14 July 1994

Gavel, Y. and Heijne, G. (1990) Protein Eng. **3**, 433–442

Gerber, L. D., Kodukula, K. and Udenfriend, S. (1992) J. Biol. Chem. **267**, 12168–12173

Gobel, U., Maas, R. and Clad, A. (1987) J. Biochem. Biophys. Methods **14**, 245–260

Gumley, T. P., McKenzie, I. F. C., Kozac, C. A. and Sandrin, M. S. (1992) J. Immunol. **149**, 2615–2618

Holguin, M. H., Fredrick, L. R., Brensham, N. J., Rosse, W. F. and Parker, C. J. (1989) J. Clin. Invest. **84**, 7–17

Hughes, T. R., Piddlesden, S. J., Williams, J. D., Harrison, R. A. and Morgan, B. P. (1992) Biochem. J. **284**, 169–176

Hughes, T. R., Meri, S., Davies, M., Williams, J. D. and Morgan, B. P. (1993) Immunology **80**, 439–444

Khan, K. D., Lindwall, G., Maher, S. E. and Bothwell, A. L. M. (1990) Mol. Cell. Biol. **10**, 5150–5159

Kawamura, N., Singer, L., Wetsel, R. A. and Colten, H. R. (1992) Biochem. J. **283**, 705–712

Kimura, S., Hanioka, N., Matsunaga, E. and Gonzalez, F. J. (1989) DNA **8**, 503–516

Matsuo, S., Nishikage, H., Yoshida, F., Nomura, A., Piddlesden, S. J. and Morgan, B. P. (1994) Kidney Int., **46**, 191–200

Morgan, B. P. and Meri, S. (1994) Springer Semin. Immunopathol. **15**, 369–396

Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. **48**, 443–453

Ninomiya, H., Stewart, B. H., Rollins, S. A., Zhao, J., Bothwell, A. L. M. and Sims, P. J. (1992) J. Biol. Chem. **267**, 8404–8410

Okada, N., Harada, R., Fujita, T. and Okada, H. (1989a) Int. Immunol. **1**, 205–208

Okada, N., Harada, R., Fujita, T. and Okada, H. (1989b) J. Immunol. **143**, 2262–2266

Piddlesden, S. J. and Morgan, B. P. (1993) J. Neuroimmunol. **48**, 169–176

Pierce, R. A., Alatawi, A., Deak, S. B. and Boyd, C. D. (1992) Genomics **12**, 651–658

Rollins, S. A., Zhao, J., Ninomiya, H. and Sims, P. J. (1991) J. Immunol. **146**, 2345–2351

Sawada, R., Ohashi, K., Anaguchi, H., Okazaki, H., Hattori, M., Minato, N. and Naruto, M. (1990) DNA Cell Biol. **9**, 213–220

Sugita, Y., Nakano, Y. and Tomita, M. (1988) J. Biochem. (Tokyo) **104**, 633–637

Sugita, Y., Nakano, Y., Oda, E., Noda, K., Tobe, T., Miura, N.-H. and Tomita, M. (1993) J. Biochem. **114**, 473–477

van den Berg, C. W. and Morgan, B. P. (1994) J. Immunol. **152**, 4095–4101

van den Berg, C. W., Harrison, R. A. and Morgan, B. P. (1993) Immunology **78**, 349–357

van den Berg, C. W., Williams, O. M. and Morgan, B. P. (1994) Immunology **81**, 637–642

Watanabe, N. and Ohshima, Y. (1988) Eur. J. Biochem. **174**, 125–132