Genomic Structure of the Human Complement Protein C8γ: Homology to the Lipocalin Gene Family†

Kenneth M. Kaufman and James M. Sodetz*

Department of Chemistry and Biochemistry and the School of Medicine, University of South Carolina, Columbia, South Carolina 29208

Received January 27, 1994; Revised Manuscript Received February 28, 1994®

ABSTRACT: Human C8 is one of five complement components (C5b, C6, C7, C8, C9) that interact to form the cytolytic C5b–9 complex on target cells. It contains three subunits (C8α, C8β, C8γ) which are encoded in separate genes. In relation to other proteins of the complement system, C8γ is unusual in that it is not structurally related to any other component nor does it have an obvious function. Based on weak but significant sequence similarity, it is proposed to be a member of the lipocalin family of widely distributed proteins that bind and transport small hydrophobic ligands. In this study, the human C8γ gene has been characterized and found to contain seven exons spanning ~1.8 kb. S1 nuclease and anchored PCR were used to identify the transcription initiation site. This site is preceded by putative regulatory elements that include two SP1 binding sites, several glucocorticoid response elements, and two SV40 enhancer core consensus sequences. A comparison to genes of other lipocalins reveals a remarkably close correlation in exon number, lengths, and phases. A close correspondence in exon boundaries is also observed and suggests that C8γ contains the same discrete structural elements that define the characteristic β-barrel shape of the lipocalins. These results establish that C8γ is indeed ancestrally related to the lipocalin family and strengthens the likelihood that its role in the complement system is to bind an as yet unidentified ligand.

Human C8 is one of five complement components (C5b, C6, C7, C8 and C9) that interact as a consequence of complement activation to form the cytolytic C5b–9 complex on target cell membranes (Müller-Eberhard, 1988; Sodetz, 1988). It contains three nonidentical subunits designated α (M_r = 64 000), β (M_r = 64 000), and γ (M_r = 22 000). These are arranged asymmetrically as a disulfide-linked α–γ dimer that is noncovalently associated with β. cDNA cloning and sequencing established that each subunit is encoded in a separate gene (Rao et al., 1987; Howard et al., 1987; Ng, et al., 1987; Haefliger et al., 1987). The C8α and C8β loci were found to be closely linked on chromosome 1p (Theriault et al., 1992), whereas C8γ is located on 9q (Kaufman et al., 1989). Several different roles for C8α and C8β in the assembly and function of C5b–9 have been defined; however, the role of C8γ remains to be determined (Sodetz, 1988). A derivative of C8 lacking C8γ was found to be functionally equivalent to native C8, thus indicating it is not essential for lytic activity (Brickner & Sodetz, 1984; Davé & Sodetz, 1990).

Throughout the complement system there are numerous examples of components that exhibit significant sequence similarity amongst themselves either because of a common genetic origin or because of the need to utilize common structural modules or motifs to carry out their functions. Overall sequence similarity indicative of a close ancestral relationship can be found in proteins such as the A, B, and C chains of C1q; C1r and C1s; C2 and factor B; C3, C4, and C5; and the family of terminal components C6, C7, C8α, C8β, and C9 (Müller-Eberhard, 1988). Also found are common structural features such as the collagen-like domains of the three C1q chains, the serine protease domains of C1r, C1s, C2, factors B, D, and I; the multiplicity of short consensus repeat (SCR) sequences found in the regulatory components DAF, MCP, C4bp, and factor H and complement receptors CR1 and CR2; and the thrombospondin repeat, LDL receptor, and epidermal growth factor sequence motifs found in C6, C7, C8α, C8β, and C9 (Reid & Day, 1989). In most cases, these common structural features are related to common functions, i.e., catalysis, ligand-binding, protein–protein interactions, etc.

C8γ is an exception because it is not structurally related to any complement component. Based on weak but significant sequence similarity (~15–25% identity), it is proposed to be a member of the lipocalin family of proteins (Pervaiz & Brew, 1987; Hunt et al., 1987). Approximately 25 members of this expanding family have been identified, primarily on the basis of sequence comparison (Igarashi et al., 1992). Members are widely distributed in both vertebrates and invertebrates and are characterized by their similar size ($M_r \cong$ 20 000), a generally weak pairwise sequence identity averaging ~20%, and the common ability to bind small hydrophobic ligands (Godovac-Zimmerman, 1988; North, 1991). Most are secreted into body fluids where they bind a variety of ligands. Several are retinol-binding proteins such as serum RBP[1] (Newcomer et al., 1984), LCN1 from tears (Redl et al., 1992), and BLG from milk (Dufour & Haertlé, 1991). The serum lipocalin α1AGP binds steroid hormones (Kute & Westphal, 1976), while APO-D binds cholesteryl esters (Drayna et al., 1987). Other examples include the odorant-binding protein of the nasal mucosa (Tirindelli et al., 1989) and the pheromone-binding proteins α2UG and MUP from rodent urine (Böcskei et al., 1992). In insect hemolymph, the lipocalins INSEC (Holden et al., 1987) and BBP (Huber et al., 1987) provide coloration and camouflage by binding biliverdin, while coloration in lobster is provided by α-crustacyanin complexed

† Supported by NIH Grant GM42898.
® Abstract published in *Advance ACS Abstracts,* April 1, 1994.

[1] Abbreviations: APO-D, apolipoprotein-D; α2UG, rat α-2u-globulin; α1AGP, α-1-acid glycoprotein; BBP, *Pieris brassicae* bilin-binding protein; BLG, β-lactoglobulin; EPSI, rat epididymal secretory protein I; INSEC, *Manduca sexta* insecticyanin; LCN1, tear prealbumin; MUP, murine major urinary protein; PDS, brain prostaglandin D_2 synthase; PHC-BK, human protein HC (α_1-microglobulin)-bikunin precursor; PP14, human placental protein 14; RBP, serum retinol-binding protein.

0006-2960/94/0433-5162$04.50/0

with the chromophore astaxanthin (Zagalsky et al., 1991).

Three-dimensional structures have been determined for human RBP (Newcomer et al., 1984), bovine BLG (Monaco et al., 1987), α2UG and MUP (Böcskei et al., 1992), INSEC (Holden et al., 1987), and BBP (Huber et al., 1987). Despite low sequence identity, these proteins have highly conserved folding patterns consisting of a core of eight antiparallel β-strands arranged into two orthogonal sheets to form a β-barrel structure with a cup-shaped ligand-binding site (North, 1991). This site is lined with apolar residues that confer specificity toward hydrophobic ligands. Members of the fatty acid-binding protein family, which also bind small hydrophobic ligands, have a similar β-barrel structure. Thus, it has been suggested that both families form part of a larger "structural superfamily" that might appropriately be named the calycins to reflect the cup-shaped structure of its members (Flower et al., 1993). Interestingly, molecular modeling of C8γ against bovine BLG suggests it also is capable of forming a β-barrel structure with a distinct binding site (Haefliger et al., 1991).

In the present study, we provide direct evidence that C8γ is indeed ancestrally related to the lipocalin family. The entire human C8γ gene has been sequenced, and the exon number, lengths, and phases correlate remarkably well with those found in other lipocalin genes. Furthermore, the correspondence in exon boundaries suggests C8γ contains the same discrete structural elements that define the β-barrel in other lipocalins.

EXPERIMENTAL PROCEDURES

Isolation and Characterization of Genomic Clones. A human genomic DNA library in the Lambda-Dash vector was obtained from Stratagene. Approximately 1×10^6 phage were plated and screened using Colony/Plaque Screen filters (Dupont/New England Nuclear) according to the manufacturer's recommendations. A 750-bp fragment containing the full-length C8γ cDNA was used as the probe (Ng et al., 1987). Phage DNA from a positive clone was purified and partially digested with *Bgl*II to yield a 5.2-kb fragment containing the entire C8γ gene. The fragment was isolated, tailored by partial fill-in with dGTP and dATP, and subcloned into pBluescript II (Stratagene). The resulting plasmid (pBG1) was sonicated and shotgun cloned into M13 (Deininger, 1983). Exon-containing clones were isolated and subjected to dideoxy sequencing using Sequenase (USB). Remaining sequence gaps were analyzed by subcloning three *Pst*I fragments (330, 445, and 1700 bp) from pBG1 into M13 and targeting regions with specific primers. The 3′ end of the gene was characterized by excising a 2.4-kb *Kpn*I fragment from pBG1, religating, and sequencing the resulting plasmid pBG2.

Identification of the Transcription Initiation Site. A human liver tissue sample was acquired with the assistance of Dr. John E. Volanakis from the Tissue Procurement Core Facility of the Comprehensive Cancer Center at the University of Alabama at Birmingham and used to isolate total and poly-(A)$^+$ RNA. S1 nuclease experiments were performed using published procedures (Sharp et al., 1980). The C8γ primer AGCTGCCAGGAGCAGAGTCAA was radiolabeled, hybridized to a M13 single-stranded template containing a 5′ C8γ genomic *Pst*I fragment (1.7 kb), and extended with Klenow. This was then cut with *Apa*I to generate a 414-bp product spanning bases 79–492 of the C8γ gene. Single-stranded probe was isolated using a 1.2% alkaline agarose gel, hybridized to 50 μg total RNA, and digested with 2000 units of S1 nuclease (USB). Products were resolved on a 6% polyacrylamide, 8 M urea sequencing gel in 1XTBE.

Anchored PCR was performed as described elsewhere (Loh et al., 1989). Human C8γ cDNA was synthesized in a 25-μL reaction containing either 1 μg of total liver RNA or 100 ng of poly(A)$^+$ RNA, 10 mM dithiothreitol, 75 units of RNase inhibitor, 200 units of MLUV reverse transcriptase (Gibco-BRL), and 1 fmol of C8γ-specific primer JS-55 (5′AAACG-GCCAGCGGAGCCCACAGC3′). Total cDNAs were also synthesized using 100 pmol of random hexamers (Perkin-Elmer Cetus) in place of JS-55. Reactions were incubated for 1 h at 40 °C and then incubated for 5 min at 70 °C, phenol extracted, and ethanol precipitated. cDNAs were resuspended in 5 μL of terminal deoxytransferase buffer, 0.1 mM dGTP, and 30 units of terminal deoxytransferase (Gibco-BRL) in order to generate 5′-poly(G) extensions. Reactions were incubated at 37 °C for 30 min, heat inactivated for 2 min at 65 °C, and ethanol precipitated. The cDNA was then resolubilized in 10 μL of H_2O.

PCR was used to amplify C8γ-specific cDNA in a 100-μL reaction containing 1 μL of poly(G)-tailed cDNA, 100 pmol of primer JS-55, 90 pmol of poly(C)$_{18}$ primer (New England Biolabs), 2.5 units of *Taq*I DNA polymerase (Perkin-Elmer Cetus), 0.2 mM of each dNTP, 1.75 mM $MgCl_2$, 50 mM KCl, and 10 mM Tris-HCl, pH 8.3. To facilitate cloning of PCR products, 10 pmol of a primer containing the M13 mp8 polylinker with a 3′-poly(C)$_{14}$ extension [AAGCTTGGCT-GCAGGTCGACGGATCCCCGGAATT(C)$_{14}$] was also included in the reaction. Five cycles of amplification were carried out at 94 °C for 1 min, 35 °C for 2 min, and 72 °C for 3 min, followed by 35 cycles at 94 °C for 1 min, 40 °C for 2 min, and 72 °C for 3 min. One microliter of the reaction mixture was reamplified using 100 pmol of a second C8γ primer (AGGCTTCTGGCCCAGCGAGCCAGC) and 100 pmol of mp8 poly(C)$_{14}$ primer for 35 cycles at 94 °C for 1 min, 55 °C for 2 min, and 72 °C for 3 min. PCR products were blunt-ended by incubating with 5 units of Klenow for 30 min at 30 °C. Products were digested with *Bam*HI and cloned into *Bam*HI/*Eco*RV-digested pBluescript II. cDNAs generated with random hexamers were likewise amplified using JS-55 in the first PCR reaction followed by reamplification with a second C8γ primer (AAGGATCCAGGTCCCTG-CAAACTGC), which contains a 5′ *Bam*HI site. PCR products generated with this primer were digested with *Eco*RI and *Bam*HI and cloned into *Eco*RI/*Bam*HI-digested pBluescript II.

RESULTS AND DISCUSSION

Screening of a genomic library with a C8γ cDNA probe yielded several clones that produced two fragments (2.5 and 2.7 kb) after *Bgl*II digestion. These same fragments were observed after Southern blot analysis of *Bgl*II-digested human genomic DNA, thus indicating the C8γ gene is small and contained within these two fragments. Partial digests of a selected clone yielded a 5.2-kb *Bgl*II fragment that was subcloned and characterized as described in Figure 1. Sequence analysis confirmed that it contains the entire C8γ gene, which is composed of seven exons spanning approximately 1.8 kb. Complete correlation between restriction digests of the genomic clone and Southern blot analysis of genomic DNA (data not shown) indicates C8γ is a single-copy gene.

The complete genomic sequence is shown in Figure 2. The first exon contains untranslated sequence, the putative leader sequence, and a portion of mature C8γ sequence. All exon/intron junctions follow the GT/AG rule (Breathnach & Chambon, 1981). Overall, the nucleotide sequence agrees with the published C8γ cDNA sequence with only three exceptions. Two differences are in the 5′ untranslated region, and one is a silent substitution in the coding region. The transcription initiation site was assigned on the basis of results

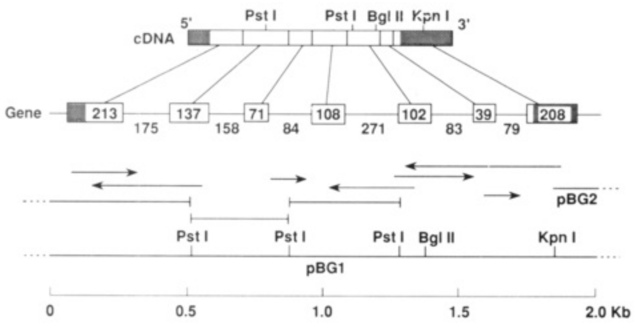

FIGURE 1: C8γ genomic map and sequencing strategy. Shown is the relationship between the C8γ gene and cDNA. Exon and intron lengths are in base pairs. Untranslated regions are shaded. Also shown is a partial restriction map of plasmid pBG1 that contains the entire gene. Arrows identify regions sequenced by shotgun cloning. Brackets identify *Pst*I fragments subcloned into M13 and sequenced. Plasmid pBG2 was constructed by excision of a 3′*Kpn*I fragment and religation.

from S1 nuclease mapping and anchored PCR. S1 nuclease analysis of liver total RNA using the probe described above yielded products of 121 and 123 bp in length, suggesting the initiation site was at nucleotide 370 or 372. To distinguish between these sites as well as confirm the absence of additional 5′ untranslated exons, anchored PCR was performed on total and poly(A)$^+$ RNA. Both RNA sources yielded a single product ~250 bp in length. Amplification products from both sources were subcloned into pBluescript II, after which a total of 14 clones were isolated and sequenced. One contained a truncated product beginning 5 bp downstream of the 5′ end of the C8γ cDNA sequence, while the remaining 13 contained sequence extending 15 bp upstream of the cDNA to position 370 in the genomic sequence. These results indicate that C8γ does not contain an additional 5′ untranslated exon and that position 370 identified by S1 nuclease mapping is the transcription initiation site.

Analysis of the nucleotide sequence in the promoter region does not reveal any obvious TATA boxes. There are two possible SP1 binding sites: one agrees with the consensus sequence and the other is 70% identical (Maniatis et al., 1987). These sites are often found in genes lacking TATA boxes. Sequences similar to the glucocorticoid response element or GRE occur four times in the 5′ region. Two agree with the consensus sequence, and two are 83% identical (Yamamoto, 1985). Putative GREs at positions 49–54 and 381–386 are in a reverse orientation. The other two are at positions 201–206 and 247–252 in a forward orientation. Two sequences similar to the SV-40 enhancer core consensus sequence were also identified (Wasylyk, 1988). At the present time, it is not known if any of these elements are operational in transcription and/or regulation of the C8γ gene.

Figure 3 compares the organization of the C8γ gene to other lipocalin genes that have been characterized. C8γ is the smallest at 1.8 kb, and PHC-BK is the largest at 16.5 kb. The exon organization of C8γ is most similar to BLG, MUP, PP14, and PDS, with all five genes containing seven exons. With the exception of exons 1 and 7, which contain variable-sized 5′ and 3′ untranslated regions, all corresponding exons are nearly identical in length. In addition, all corresponding exon/intron junctions have the same phase with respect to the open reading frame, the only exception being the junction between exons 6 and 7 in C8γ and PP14. When compared to the remaining lipocalins, C8γ differs from α1AGP and EPSI only at the 3′ end, where the former has one less and the latter one more intron. Both RBP and APO-D contain

```
  ┌ SP1 ┐     ┌ SP1 ┐                     ┌ GRE ┐
agcgggcggcggtcgtgggcggggttgcaggcgaggctcaacgaacgctggtctgaccgtcggcgctccctgttgccgggccctgagcaagtggcttcatgaaccccgtgacgttggcca  120
                              ┌ ECC ┐            ┌ ECC ┐     ┌ GRE ┐
tggagataagaccactgggtgatggtttaaggaagataacgtgtaaagggctaaggactgtcggtggaaatcaggggtgcaggagaaatggataaacagccagaggtcaactcggacttt  240
      ┌ GRE ┐
gtacataggacatggtgccaggccctgccaggaagtgcagatcgaagctaggctcacgaggaggctggaggtgggggtggggaggcaacggatggacatggacttcctgggctgggctc  360
                                                                                                  -20                 -10
        ┌exon 1→  ┌ GRE ┐                                                                          M  L  P  P  G  T  A  T  L  L  T  L
tgtgacagcAGAGTAGACTCTGTCCTGGGACTTGGTGGTGCTACCCTTGGCCTCCCACAGTCCTGCCACCCTGCTGCCGCCACCATGCTGCCCCCTGGGACTGCGACCCTCTTGACTCTG  480
                            +1                    10                   20
L  L  A  A  G  S  L  G  Q  K  P  Q  R  P  R  R  P  A  S  P  I  S  T  I  Q  P  K  A  N  F  D  A  Q  Q
CTCCTGGCAGCTGGCTCGCTGGGCCAGAAGCCTCAGAGGCCACGCCGGCCCGCATCCCCCATCAGCACCATCCAGCCCAAGGCCAATTTTGATGCGCAGCAGgtagaagttgggggggt  600
agagggaggcaggtagaagttgtgggaggggtagagggagacaggtagaagttgttgcgggggagagggaagcaggtgaagttgtggggggtgtagagggaagcaggtgaggggccctcc  720
                                        ┌exon 2→ 30                            40                             50
                                         F  A  G  T  W  L  L  V  A  V  G  S  A  C  R  F  L  Q  E  Q  G  H  R  A  E  A  T  T
cacagtgccctcgagttctcccatggtctgcccccagTTTGCAGGGACCTGGCTCCTTGTGGCTGTGGGCTCCGCTTGCCGTTTCCTGCAGGAGCAGGGCCACCGGGCCGAGGCCACCAC  840
                  60                            70
 L  H  V  A  P  Q  G  T  A  M  A  V  S  T  F  R  K  L
ACTGCATGTGGCTCCCCAGGGCACAGCCATGGCTGTCAGTACCTTCCGAAAGCTgtgagtcccagagcagccctgcaccctaaccccaaccctcctctcagccccggacttcagccctg  960
                                                                                            ┌exon 3→          80
                                                                                             D  G  I  C  W  Q  V  R  Q
ctctggcccctgaccccaccccggctgtggcctggactaggattcctggttggggtctcccagcctgtggtgcctcctccccgccccccagGGATGGGATCTGCTGGCAGGTGCGCCAG  1080
                           90
L  Y  G  D  T  G  V  L  G  R  F  L  L  Q  A
CTCTATGGAGACACAGGGGTCCTCGGCCGCTTCCTGCTTCAAGgtgaggcaggggctgcaggtcatgtgggtggggatgacgcagccactgtggctctctgacatggctactgtggctc  1200
          ┌exon 4→   100                           110                            120                           130
            R  G  A  R  G  A  V  H  V  V  V  A  E  T  D  Y  Q  S  F  A  V  L  Y  L  E  R  A  G  Q  L  S  V  K  L  Y  A
tgcccagCCCGAGGCGCCCGAGGGGCTGTGCACGTGGTTGTCGCTGAGACCGACTACCAGAGTTTCGCTGTCCTGTACCTGGAGCGGGCGGGGCAGCTGTCAGTGAAGCTCTACGgtatg  1320

tggggccagcctctgtgaccaggcaggcgctcaagctctgcacactcactgggccaccccgaggggctgggtgagccatggggacacacttcctttctcccatcctgatcctcctgcta  1440
agcaggggcccagggagtagtgacagacaggcctggtgtgggagcagggaggagggccccgagggggcaggggacacacagaccccgttcccagagccctccacgccgcctggtgccagga  1560
                             ┌exon 5→                    140                            150                            160
                               R  S  L  P  V  S  D  S  V  L  S  G  F  E  Q  R  V  Q  E  A  H  L  T  E  D  Q  I  F  Y  F  P
ccccaggaaccctgtctgccctgcagCCCGCTCGCTCCCTGTGAGCGACTCGGTCCTGAGTGGGTTTGAGCAGCGGGTCCAGGAGGCCCACCTGACTGAGGACCAGATCTTCTACTTCCC  1680
                                                                                           ┌exon 6→  170
 K  Y  G                                                                                     F  C  E  A  A  D  Q  F  H
CAAGTACGgtgagtgtccccagcaggtccccagctcagccacccccactctctggctgatgtccagcctgacccctgccttggcgccccagGCTTCTGCGAGGCTGCAGACCAGTTCCAC  1800
                                                                                          ┌exon 7→
V  L  D  G                                                                                  V  R  R  *
GTCCTGGACGgtgagtgcacagcgggggcaagcatggcggcgtggtgaggggggccactcgcaccggctgagtctcgtctctgctgcagAAGTGAGGAGGTGAGGCCGGCACACAGCTCC  1920
AGTGCTGAGAAGTCAGTGCCCCGAGAGACGACCCCACCAGTGGGGTGCCCGCTGCCTGTCCTCCGTGAAACCAGCCTCAGATCAGGGCCCTGCCACCCAGGGCAGGGGATCTTCTGCCGG  2040
CTGCCCCAGAGGACAGTGGGTGGAGTGGTACCTACTTATTAAATGTCTCAGACCCCTctctgactcttctgtccactctggaccggcgccagtaccaccaaggccctctctgcccccacc  2160
ccgcctctttaaaagcccggcgctccctgttggctggagtccacgcagggtcactgggccgatttcggctcttgggatttgggaggggagatcctctctggcatatgccatcttgtgccc  2280
tgctggacctgggggcgtccacgtcactccaaggctgctcttgcctgggccatgcctgcagccc                                                          2344
```

FIGURE 2: Sequence of the C8γ gene. Exon 1 begins with the transcription initiation site determined by S1 nuclease analysis and anchored PCR. The start of the cDNA sequence is at position 385. The end of exon 7 corresponds to the start of the poly(A) extension in the cDNA. Nucleotides that differ from the published cDNA sequence are underlined. The polyadenylation signal is double underlined. Putative regulatory elements located in the 5′ region are identified as SP1 binding site (SP1), SV40 enhancer core consensus sequence (ECC), and glucocorticoid response element (GRE). The sequence is available in GenBank (Accession No. U08198).

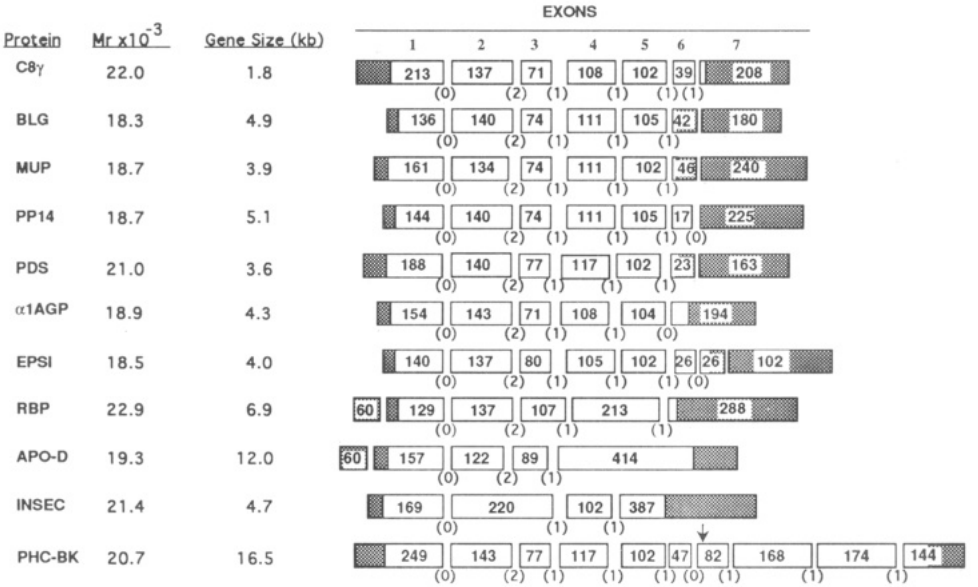

FIGURE 3: Comparison of the C8γ and lipocalin genes. Lengths of exons are in base pairs. Untranslated regions are shaded. Exon–intron phases are shown in parentheses. Numbering of exons is based on the C8γ genomic structure. Genes represented are ovine BLG (Ali & Clark, 1988), MUP (Clark et al., 1985), PP14 (Vaisse et al., 1990), human PDS (White et al., 1992), human α1AGP (Merritt & Board, 1988), EPSI (Girotti et al., 1992), rat RBP (Laurent et al., 1985), human APO-D (Drayna et al., 1987), INSEC (Li & Riddiford, 1992), and PHC-BK (Vetr & Gebhard, 1990). The arrow above the PHC-BK gene indicates the end of the protein HC coding region and the start of the bikunin coding region.

an additional 5′ untranslated exon. Also noteworthy is the increased length of the fifth exon in both genes. These differences suggest introns may have been lost or gained during evolution of the lipocalins. The fifth exon in RBP is identical in length or differs by only 1 or 2 codon multiplicities from the combined lengths of exons 4 and 5 in C8γ, BLG, MUP, PP14, PDS, α1AGP, or EPSI. The size of the fifth exon in APO-D is likewise similar to the combined lengths of those exons corresponding to exons 4–7 in C8γ. In INSEC, the second and fourth exons appear to correspond to C8γ exons 2 and 3 and 5–7, respectively. The PHC-BK gene has the most unusual structure amongst the lipocalins. It encodes a single precursor protein that is posttranslationally cleaved into two different proteins. One is the lipocalin protein HC (or α1-microglobulin), and the other is bikunin (BK), one of the Kunitz family of protease inhibitors (Vetr & Gebhard, 1990). The first six exons and part of the seventh encode protein HC, and all are similar in length and phase to corresponding exons in the other lipocalin genes.

The close correspondence of exon number, length, and phases illustrated in Figure 3 is convincing evidence that C8γ is ancestrally related to the lipocalin family. Although suggested from amino acid sequence comparisons, membership in this family was questionable because sequence similarity is generally weak and there is no obvious rationale for a lipocalin-like role for C8γ in the complement system. Close correspondence amongst the genes suggests this family may have evolved by gene duplication followed by divergence of sequence and function. A survey of the Human Genome Database revealed that several lipocalin loci are clustered on chromosome 9q, including C8γ at 9q22.3-q32, α1AGP at 9q33, and PHC-BK, PP14, PDS, and LCN1 at 9q34. Clustering of genes for structurally and functionally related proteins is not unusual, and duplication in this region may have resulted from a need to produce lipocalin-like proteins with varying ligand specificities.

Lipocalins whose three-dimensional structures are known all contain a series of eight β-strands followed by an α-helix and an additional β-strand. The nine β-strands are designated A–I, with the first eight folding in an antiparallel arrangement

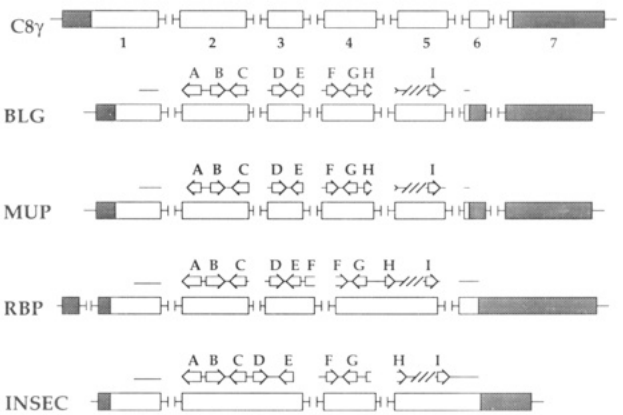

FIGURE 4: Alignment of β-barrel structural elements with exons in the lipocalins. Location of β-strands A–I in bovine BLG, MUP, human RBP, and INSEC were determined from three-dimensional structures cited in the text. The antiparallel nature of these strands is indicated by opposing arrows. The single helix is designated by diagonal lines, while connecting loops and random coils are designated by horizontal lines. Exons 1–7 of human C8γ are shown for comparison. Genes used for the alignment are described in Figure 3. Elements of bovine BLG structure are aligned with the ovine gene because sequence for the former is not available. At the protein level, the two BLG species differ in only 5 out of 162 residues. Elements of human RBP structure are aligned with the rat gene. Available sequence of the partially characterized human RBP gene (D'Onofrio et al., 1985) indicates the third, fourth, and fifth exons encoding the elements are identical in length and phases between the two species. α2UG is not included in the comparison because it is the rat counterpart of MUP and has essentially the same elements and gene structure.

to form the core β-barrel structure (North, 1991). Small hydrophobic ligands insert into the pocket formed by the barrel. Ligand specificity is likely determined by sequence differences in the interior of the barrel, which may explain the low level of sequence similarity throughout the family. Despite these differences, structural elements that define the folding pattern are conserved. The comparative analysis in Figure 4 suggests such elements are also present in C8γ. A comparative alignment of structural elements and exons in lipocalins whose three-dimensional and gene structures are known reveals a close correlation. This was originally observed for RBP and

BLG, which have 21% sequence identity (Ali & Clark, 1988), but the correlation is now also apparent for BLG and MUP (20% identity) and extends to RBP and INSEC if intron insertions and deletions are considered (Figure 3). Importantly, alignment of C8γ with BLG and MUP (20% and 22% identity, respectively) implies that exons 2–4 contain the equivalent of β-strands A–H that form the β-barrel core structure. This correlation is consistent with results from molecular modeling and further strengthens the likelihood that C8γ can assume a β-barrel three-dimensional structure (Haefliger et al., 1991).

Similarity in genomic organization to the other lipocalin genes suggests the functional role of C8γ is to bind an as yet unidentified ligand. The properties of such a ligand are not apparent since no distinctive chromophore copurifies with C8. One study suggested that retinol may be a ligand and that C8γ may function as a vitamin A transporter (Haefliger et al., 1991). However, experiments using purified C8γ failed to detect binding of retinoids, possibly because the internal disulfide bond was cleaved during separation from C8α. In contrast, binding to intact C8α-γ dimer could be detected but only at high ionic strength (2 M NaCl). It was suggested that these conditions were necessary to disrupt strong ionic interactions between C8α and C8γ and thereby expose a retinol-binding site. This explanation is questionable, however, since it has been shown that C8α and C8γ readily dissociate at low ionic strength after limited reduction and cleavage of the single interchain disulfide bond (Brickner & Sodetz, 1984). Thus, the significance of retinoid-binding to C8α–γ under conditions that differ considerably from physiological ionic strength is unclear.

An alternative explanation for the lipocalin-like structure of C8γ is that it may facilitate binding to a narrowly defined hydrophobic domain on C8α prior to intracellular formation of the disulfide-linked C8α–γ dimer. Since C8α has the intrinsic ability to interact with target membranes (Sodetz, 1988), association with C8γ may protect it from premature membrane interactions during normal biosynthetic processing. Such a role is consistent with results from preliminary experiments in which the human C8 subunits were expressed in the baculovirus–insect cell system (Kaufman et al., 1994). Cells infected with C8α baculovirus constructs failed to secrete C8α, whereas cells co-infected with C8α and C8γ secrete functional C8α–γ. Details regarding the ligand-binding properties of C8γ and the existence of a possible recognition site on C8α should be revealed from studies using the recombinant proteins, which are currently in progress.

REFERENCES

Ali, S., & Clark, A. J. (1988) *J. Mol. Biol. 199*, 415–426.

Böcskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E. V., Cavaggioni, A., Findlay, J. B. C., & North, A. C. T. (1992) *Nature 360*, 186–188.

Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349–383.

Brickner, A., & Sodetz, J. M. (1984) *Biochemistry 23*, 832–837.

Clark, A. J., Ghazal, P., Bingham, R. W., Barrett, D., & Bishop, J. O. (1985) *EMBO J. 4*, 3159–3165.

D'Onofrio, C., Colantuoni, V., & Cortese, R. (1985) *EMBO J. 4*, 1981–1989.

Davé, S. J., & Sodetz, J. M. (1990) *J. Immunol. 144*, 3087–3090.

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

Drayna, D. T., McLean, J. W., Wion, K. L., Trent, J. M., Drabkin, H. A., & Lawn, R. M. (1987) *DNA 6*, 199–204.

Dufour, E., & Haertlé, T. (1991) *Biochim. Biophys. Acta 1079*, 316–320.

Flower, D. R., North, A. C. T., & Attwood, T. K. (1993) *Protein Sci. 2*, 753–761.

Girotti, M., Jones, R., Emery, D. C., Chia, W., & Hall, L. (1992) *Biochem. J. 281*, 203–210.

Godovac-Zimmerman, J. (1988) *Trends Biochem. Sci. 13*, 64–66.

Haefliger, J., Jenne, D., Stanley, K. K., & Tschopp, J. (1987) *Biochem. Biophys. Res. Commun. 149*, 750–754.

Haefliger, J. A., Peitsch, M. C., Jenne, D. E., & Tschopp, J. (1991) *Mol. Immunol. 28*, 123–131.

Holden, H. M., Rypniewski, W. R., Law, J. H., & Rayment, I. (1987) *EMBO J. 6*, 1565–1570.

Howard, O. M. Z., Rao, A. G., & Sodetz, J. M. (1987) *Biochemistry 26*, 3565–3570.

Huber, R., Schneider, M., Mayr, I., Müller, R., & Deutzmann, R. (1987) *J. Mol. Biol. 198*, 499–513.

Hunt, L. T., Elzanowski, A., & Barker, W. C. (1987) *Biochem. Biophys. Res. Commun. 149*, 282–288.

Igarashi, M., Nagata, A., Toh, H., Urade, Y., & Hayaishi, O. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5376–5380.

Kaufman, K. M., Snider, J. V., Spurr, N. K., Schwartz, C. E., & Sodetz, J. M. (1989) *Genomics 5*, 475–480.

Kaufman, K. M., Letson, C. S., Michelotti, G. A., Platteborze, P. L., & Sodetz, J. M. (1994) *FASEB J. 8*, A474.

Kute, T., & Westphal, U. (1976) *Biochim. Biophys. Acta 420*, 195–213.

Laurent, B. C., Nilsson, M. H. L., Båvik, C. O., Jones, T. A., Sundelin, J., & Peterson, P. A. (1985) *J. Biol. Chem. 260*, 11476–11480.

Li, W., & Riddiford, L. M. (1992) *Eur. J. Biochem. 205*, 491–499.

Loh, E. Y., Elliott, J. F., Cwirla, S., Lanier, L. L., & Davis, M. M. (1989) *Science 243*, 217–220.

Maniatis, T., Goodbourn, S., & Fisher, J. A. (1987) *Science 236*, 1237–1245.

Merritt, C. M., & Board, P. G. (1988) *Gene 66*, 97–106.

Monaco, H. L., Zanotti, G., & Spadon, P. (1987) *J. Mol. Biol. 197*, 695–706.

Müller-Eberhard, H. J. (1988) *Annu. Rev. Biochem. 57*, 321–47.

Newcomer, M. E., Jones, T. A., Aqvist, J., Sundelin, J., Eriksson, U., Rask, L., & Peterson, P. A. (1984) *EMBO J. 3*, 1451–1454.

Ng, S. C., Rao, A. G., Howard, O. M. Z., & Sodetz, J. M. (1987) *Biochemistry 26*, 5229–5233.

North, A. C. T. (1991) *Biochem. Soc. Symp. 57*, 35–48.

Pervaiz, S., & Brew, K. (1987) *FASEB J. 1*, 209–214.

Rao, A. G., Howard, O. M. Z., Ng, S. C., Whitehead, A. S., Colten, H. R., & Sodetz, J. M. (1987) *Biochemistry 26*, 3556–3564.

Redl, B., Holzfeind, P., & Lottspeich, F. (1992) *J. Biol. Chem. 267*, 20282–20287.

Reid, K. B. M., & Day, A. J. (1989) *Immunol. Today 10*, 177–180.

Sharp, P. A., Berk, A. J., & Berget, J. M. (1980) *Methods Enzymol. 65*, 750–768.

Sodetz, J. M. (1988) *Curr. Top. Microbiol. Immunol. 140*, 19–31.

Theriault, A., Boyd, E., Whaley, K., Sodetz, J. M., & Connor, J. M. (1992) *Hum. Genet. 88*, 703–704.

Tirindelli, R., Keen, J. N., Cavaggioni, A., Eliopoulos, E. E., & Findlay, B. C. (1989) *Eur. J. Biochem. 185*, 569–572.

Vaisse, C., Atger, M., Potier, B., & Milgrom, E. (1990) *DNA Cell Biol. 6*, 401–413.

Vetr, H., & Gebhard, W. (1990) *Biol. Chem. Hoppe-Seyler 371*, 1185–1196.

Wasylyk, B. (1988) *CRC Crit. Rev. Biochem. 23*, 77–120.

White, D. M., Mikol, D. D., Espinosa, R., Weimer, B., Le Beau, M. M., & Stefansson, K. (1992) *J. Biol. Chem. 267*, 23202–23208.

Yamamoto, K. R. (1985) *Annu. Rev. Genet. 19*, 209–252.

Zagalsky, P. F., Eliopoulos, E. E., & Findlay, J. B. C. (1991) *Biochem. J. 274*, 79–83.