The arrestin superfamily: cone arrestins are a fourth family

Cheryl M. Craft[a,*], Donald H. Whitmore[b]

[a]*Department of Cell and Neurobiology, University of Southern California School of Medicine, Mary D. Allen Laboratories, Doheny Eye Research Institute, Bishop Medical Teaching Bldg. 401, 1333 San Pablo, Los Angeles, CA 90033, USA*
[b]*Department of Biology, Box 19498, The University of Texas at Arlington, Arlington, TX 76019, USA*

Received 15 February 1995

Abstract Arrestins constitute a superfamily of regulatory proteins that down-regulate phosphorylated G-protein membrane receptors, including rod and cone photoreceptors and adrenergic receptors. The potential role of arrestin in color visual processes led us to identify a cDNA encoding a cone-like arrestin in *Xenopus laevis*, the principle amphibian biological model system. Alignment of 18 deduced amino acid sequences of all known arrestins from both invertebrate and vertebrate species reveals five arrestin families. Further analysis identifies 7 variable and 4 conservative arrestin structural motifs that may identify potential functional domains. The adaptive evolutionary relationship of *Xenopus* cone arrestin to the arrestin gene tree suggests high intrafamily homology and early gene duplication events.

Key words: Arrestin; Color vision; Retina; G-protein receptor; *Xenopus laevis*; Adaptive evolution

1. Introduction

Signal transduction of many hormonal, neurotransmitter and sensory stimuli from membrane receptors through G-protein mechanisms involves members of several large gene families. The archetype of such systems is the visual cascade that occurs in vertebrate retinal rod photoreceptors. Rhodopsin is one member of a large family of serpentine protein receptors that have a 7-transmembrane domain motif. Upon excitation rhodopsin interacts with transducin, one of many G proteins. Rhodopsin is subsequently phosphorylated by rhodopsin kinase, one of the receptor-specific kinases. In the process of receptor down-regulation that follows, a protein called S-antigen or arrestin interacts with the activated, phosphorylated rhodopsin to complete its return to the quiescent state. We hypothesize that unique arrestin families exist for each of the receptor families that function in this manner, and, furthermore, arrestins constitute a superfamily of regulatory proteins.

Support for this hypothesis continues to mount as new arrestin families are discovered and described . Four vertebrate arrestin families are now recognized: S-antigens (SAG) involved with retinal rod function; β-arrestins (BAR1 and BAR2) involved with β-adrenergic receptors; and retinal cone arrestins (CAR) associated with retinal cone function. The latter family is the most recently established and has a single described member, human cone arrestin [1,2] .

*Corresponding author. Fax: (1) (213) 342 2709;
E-Mail: CCRAFT@HSC.usc.edu

This manuscript is dedicated to Mary D. Allen for her generous endowment and charitable contributions to vision research and the establishment of the Mary D. Allen Laboratories.

Elucidation of the visual transduction mechanism has depended heavily upon the mammalian rod photoreceptor model. Analogous studies of the high level of visual acuity and perception of color initiated by cone stimulation are fragmentary because the proteins in this pathway are not known in detail yet. The mammalian retina is a duplex structure generally dominated by rods except for the central fovea and its high cone concentration. It has been suggested that early mammalian ancestors were nocturnal and this experience set the stage for retinal evolution different from that of non-mammalian vertebrates. Primates appear to have diverged from this pattern and acquired trichromacy in a recent evolutionary step. While primate color vision is probably most highly developed among mammals, other vertebrate classes appear to have evolved cone dominated retinas and exquisite color vision [3].

Against this evolutionary background, we chose to examine the retina of *Xenopus laevis* to further our understanding of cone arrestin distribution, structure and function. Not only is *Xenopus* the principle amphibian biological model system [4], but it is phylogenetically distant from mammals and possesses color vision. Its retina contains four photoreceptor classes: a green rod, a red rod, double cone and a single cone. The objectives of this study were to determine if cone-like arrestins were used in amphibian visual processes and, if present, to compare the primary structure(s) with the human cone arrestin.

2. Materials and methods

2.1. *Xenopus tissue and cDNA library*

Male *Xenopus laevis* were purchased from Carolina Biological Supply Co. (Burlington, NC). Animals were anesthetized and pithed prior to tissue removal. Surgical removal of tissues was performed in a darkroom under dim red light. All tissues were immediately frozen on dry ice and subsequently stored at −70°C until RNA extraction.

2.2. *RNA preparation*

Total RNA was extracted from liver, heart, brain, retina, spleen and testes using previously published methods [5,6].

2.3. *Library screening*

Xenopus arrestins were identified by screening a λZapII *Xenopus laevis* retinal cDNA library (kindly provided by Drs. K. Green and J. Besharse). Retinal tissue used to construct this library was isolated from animals four hours after dark onset. An [α-^{32}P]dCTP-labeled random primed cDNA probe used to screen the library was the complete coding region (nucleotides +146/−1040) of human cone arrestin (hrCAR5B) [2]. Cross species identification for the screen was determined empirically utilizing lower prehybridization and hybridization stringency conditions of 30% formamide, 37°C and a final wash of 0.5 × SSC, 40°C).

2.4. *Northern blot analysis*

To examine the mRNA expression of *Xenopus* SAG and CAR genes, two arrestin cDNA clones from the *Xenopus* retina library were used as probes for Northern blot analysis. Total RNA was electrophoresed

SSDI 0014-5793(95)00213-8

```
          10          20          30          40          50
ATG GCA GAA AGT TCT AAA GTA TTT AAG AAG TCC AGT GGT GAT GGG AAG CTG GCC
M   A   E   S   S   K   V   F   K   K   S   S   G   D   G   K   L   A

      60          70          80          90          100
ATC TAT CTT GCC AAG CGA GAC TAT GTG GAC CAC GTG GAC CAT GTT GAA CCT GTA
I   Y   L   A   K   R   D   Y   V   D   H   V   D   H   V   E   P   V

110         120         130         140         150         160
GAT GGA ATG ATC ATA ATT GAC CCT GAA TAT CAA AAG GAC AAG AAA GTC TTT GTG
D   G   M   I   I   I   D   P   E   Y   Q   K   D   K   K   V   F   V

        170         180         190         200         210
ACC CTC GCT TGT ACT TTT CGT TAT GGA CGA GAT GAC CAT GAG CTC ATT GGC TTG
T   L   A   C   T   F   R   Y   G   R   D   D   H   E   L   I   G   L

  220         230         240         250         260         270
AGC TTT AAG AAG GAA CTG TGC TTT CTG CAT TGT CAG GTG TAC CCC CCA CTG CCT
S   F   K   K   E   L   C   F   L   H   C   Q   V   Y   P   P   L   P

          280         290         300         310         320
GAA GAT AAG AAG CCC CTC ACA CCA CTA CAG GAA AAG TTA TCT AAA AAG CTG GGA
E   D   K   K   P   L   T   P   L   Q   E   K   L   S   K   K   L   G

    330         340         350         360         370
GTG AAT GCA TTT CCC TTC TGT TTT AAT ATG ACC ACG GAT TTG CCA TGC TCG GTG
V   N   A   F   P   F   C   F   N   M   T   T   D   L   P   C   S   V

380         390         400         410         420         430
ACA CTT CAG CCA GGA CCA GAG GAT ACT GGG AAG AAA TGT GGA GTT GAT TTT GAG
T   L   Q   P   G   P   E   D   T   G   K   K   C   G   V   D   F   E

        440         450         460         470         480
GTG AAA GGT TTC TGG GCA GAT AAT GTG GAA GAG AAA ATA TCC AGA AAG AAC AGC
V   K   G   F   W   A   D   N   V   E   E   K   I   S   R   K   N   S

  490         500         510         520         530         540
GTT CAG CTC ATA ATC AGG AAA GTG CAG TTT GCT CCA GAA GCC ACA GGA ACT GCC
V   Q   L   I   I   R   K   V   Q   F   A   P   E   A   T   G   T   A

          550         560         570         580         590
TCA TGT GTC CAA ACA ACA CGC CAA TTC ATG ATG TCA GAC AAA CCA CTA CAG GTG
S   C   V   Q   T   T   R   Q   F   M   M   S   D   K   P   L   Q   V

    600         610         620         630         640
GAG GTC TCA CTG GAC AAA GAA GTT TAT TAT CAT GGG GAG CCA GTT GGC ATC AAA
E   V   S   L   D   K   E   V   Y   Y   H   G   E   P   V   G   I   K

650         660         670         680         690         700
TTA AAA ATA AAC AAC AAC ACC AGC AAG ATT GTG AAA AAG ATC AAA ATA ACA GTG
L   K   I   N   N   N   T   S   K   I   V   K   K   I   K   I   T   V

        710         720         730         740         750
GAG CAG TTG ACC GAT GTG GTT CTT TAT TCA CTG GAC AAG TAC ACC AAA ATT GTC
E   Q   L   T   D   V   V   L   Y   S   L   D   K   Y   T   K   I   V

  760         770         780         790         800         810
TGC TGT GAG GAG ATA AAT GAG ACG GTC GCA GCA AAT GCC AAT TTC TCA GGG TCA
C   C   E   E   I   N   E   T   V   A   A   N   A   N   F   S   G   S

          820         830         840         850         860
TAT TCG CTG ACA CCG CTT CTG GCC AAC AAC AAG GAG AAA CGT GGT CTG GCC CTA
Y   S   L   T   P   L   L   A   N   N   K   E   K   R   G   L   A   L

    870         880         890         900         910
GAT GGC AAA CTG AAA CAC GGC GAT ACC AAC CTT GCA TCA TCT ACA ATC CTA CGA
D   G   K   L   K   H   G   D   T   N   L   A   S   S   T   I   L   R

920         930         940         950         960         970
CCA GGG ATG GAT AAA GAG GTG CTA GGA ATG TTG GTG TCT TAT AAA GTT CGA GTC
P   G   M   D   K   E   V   L   G   M   L   V   S   Y   K   V   R   V

        980         990         1000        1010        1020
AGT CTG GTG GTG GCC AGA GGA GGA ATT CTG GGA GAC CTG ACA TCA AGT GAT GTG
S   L   V   V   A   R   G   G   I   L   G   D   L   T   S   S   D   V

  1030        1040        1050        1060        1070        1080
TCG GTG GAG CTG CCA TTT ACC TTG ATG CAT CCC AAA CCA TCA CCG GAC CAG ACA
S   V   E   L   P   F   T   L   M   H   P   K   P   S   P   D   Q   T

          1090        1100        1110        1120        1130
AAC ATC GAG GAT GTG GTG ATA GAG GAA TTC GCC AGG CAA AAG CTA CAG GGA GCA
N   I   E   D   V   V   I   E   E   F   A   R   Q   K   L   Q   G   A

    1140        1150        1160
GAG GGT GAA GAT GAC AAG GAT GAT GCA
E   G   E   D   D   K   D   D   A
```

← Fig. 1. Nucleotide and deduced amino acid sequence of mRNA encoding *Xenopus* cone arrestin. The analysis was performed using the PCGENE program. The numbering is spaced every 10 nucleotides and the predicted amino acid sequence of the open reading frame is indicated. The entire sequence is submitted to EMBL/GenBank Data Libraries under Accession No. L40463.

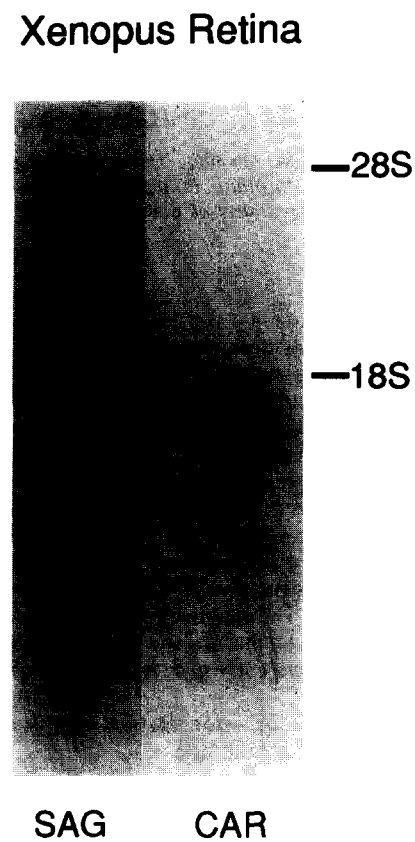


Fig. 2. Northern (RNA) blot analysis of SAG and CAR mRNAs from *Xenopus* retina. Total RNA was extracted from retinal tissue, electrophoretically separated, transferred to membrane and probed with either *Xenopus* CAR or SAG cDNAs. The SAG transcript was 1.9 kb and that of CAR was 1.7 kb. Locations of the 28S and 18S ribosomal RNAs are indicated.

on a 1.5% formaldehyde agarose gel in MOPS buffer (*N*-morpholino)propanesulfonic acid), electrophoretically transferred to Zeta-probe (Bio-Rad) in 1× TAE (40mM Tris-acetate, 2 mM EDTA, pH 8.5) at 4°C. The blot was irradiated with UV for 1 min, prehybridized and hybridized in 50% deionized formamide, 1× Denhardt's, 6× SSC, 100 mM sodium phosphate buffer, 0.1 mg/ml yeast tRNA at 42°C as previously described [5].

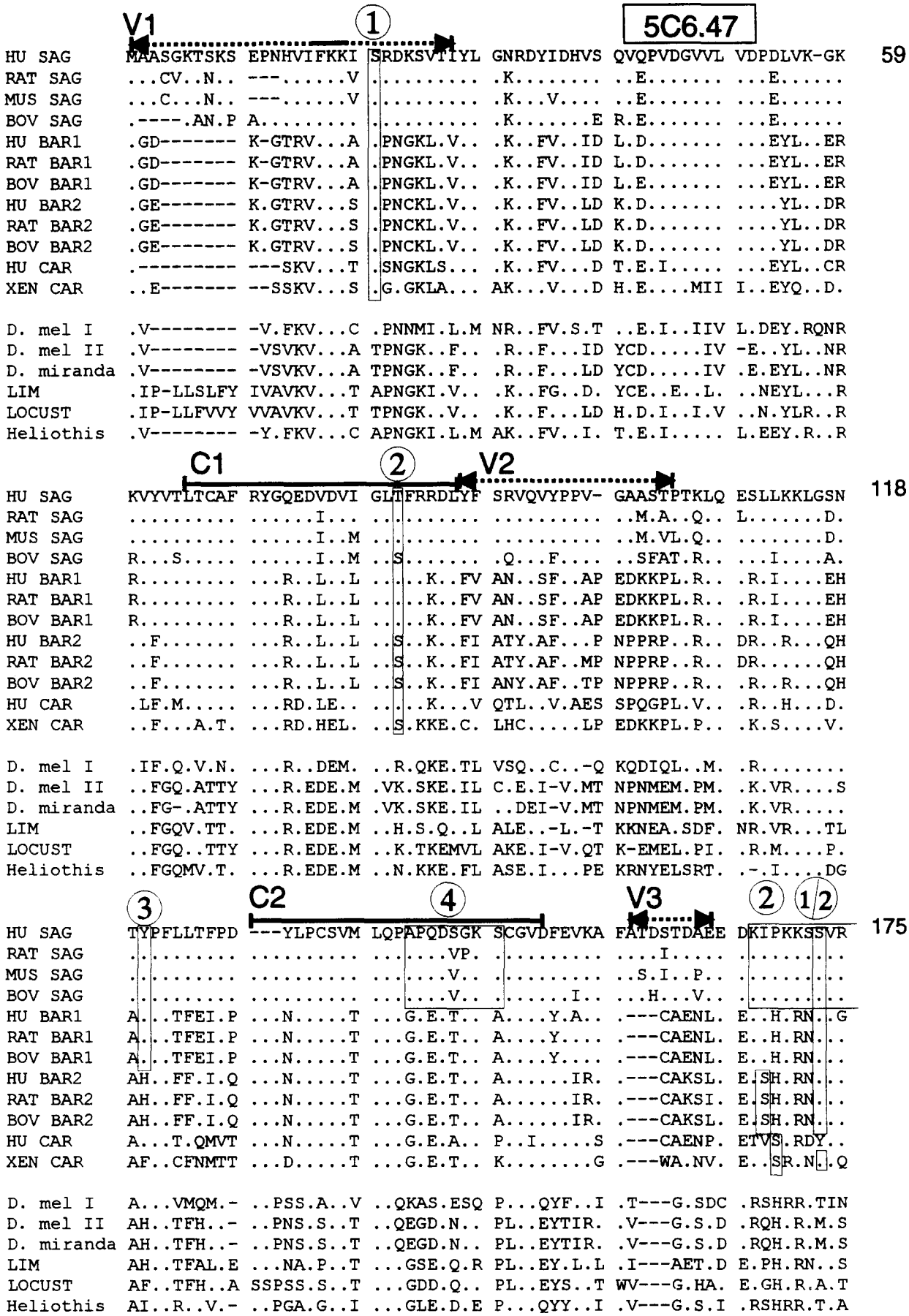
2.5. DNA sequencing

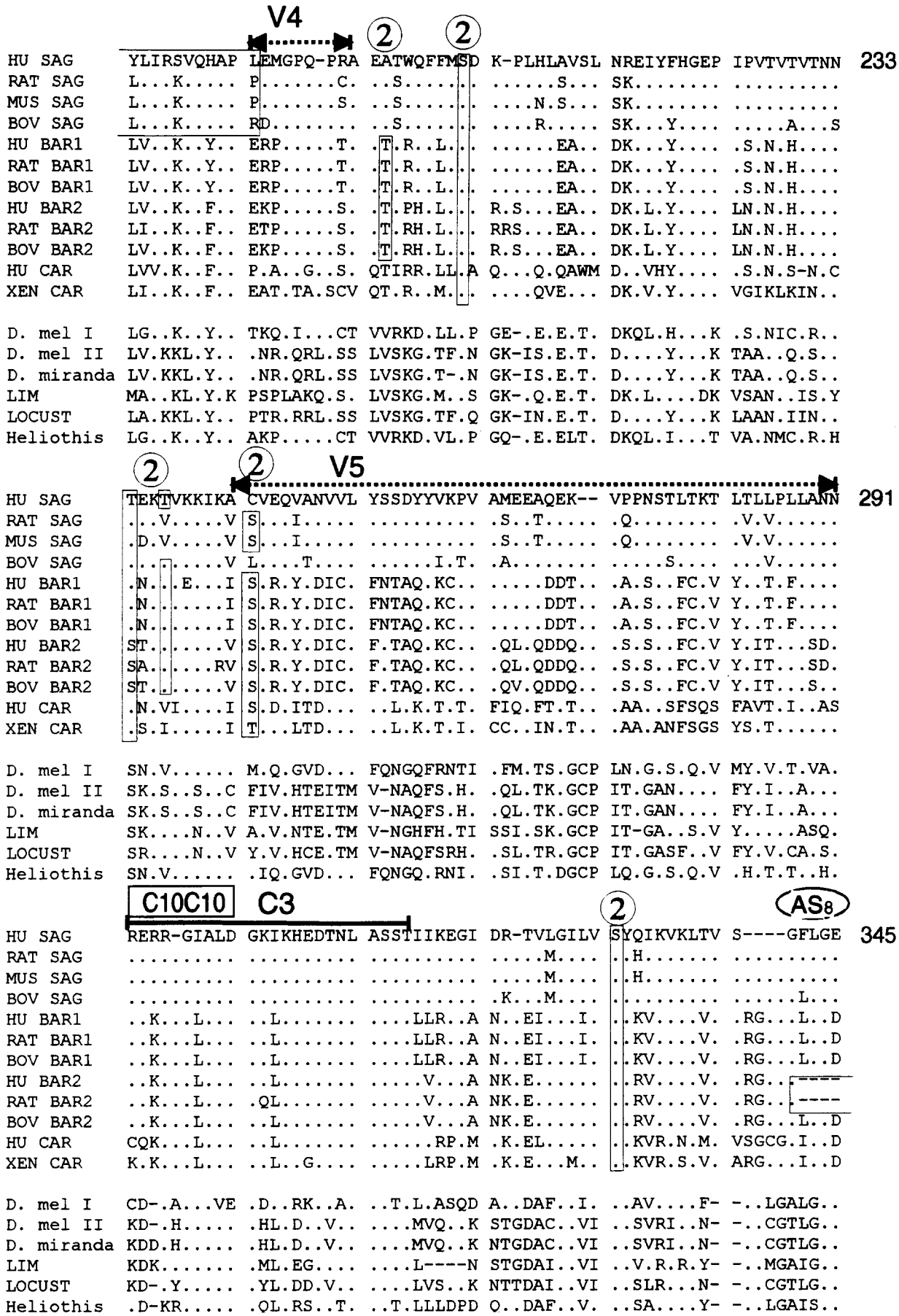
Both strands of the cDNA were sequenced by the dideoxy chain-termination method [7] using Sequenase DNA kit V. 2.0 (USB, Cleveland, OH).

2.6. Data analysis

DNA sequences were analyzed with the computer program PCGENE (IntelliGenetics, Mountain View, CA). Homologous arrestin amino acid sequences were aligned using the CLUSTAL V program [8]. Arrestin gene trees were constructed with subroutines of the MEGA package [9].

The extent of variation at each residue position of the aligned verte-


V1
1
5C6.47
C1
2
V2
3
C2
4
V3
2
1/2


V4
2
2
HU SAG YLIRSVQHAP LEMGPQ-PRA EATWQFFMSD K-PLHLAVSL NREIYFHGEP IPVTVTVTNN 233
RAT SAG L...K..... P.......C. ..S.....S. ......S... SK..... .....
MUS SAG L...K..... P.......S. ..S.....S. ....N.S... SK..... .....
BOV SAG L...K..... RD..... ..S.....S. ....R..... SK...Y.... .....A...S
HU BAR1 LV..K..Y.. ERP.....T. .T.R..L.S. ......EA.. DK...Y.... .S.N.H....
RAT BAR1 LV..K..Y.. ERP.....T. .T.R..L.S. ......EA.. DK...Y.... .S.N.H....
BOV BAR1 LV..K..Y.. ERP.....T. .T.R..L.S. ......EA.. DK...Y.... .S.N.H....
HU BAR2 LV..K..F.. EKP.....S. .T.PH.L.S. R.S...EA.. DK.L.Y.... LN.N.H....
RAT BAR2 LI..K..F.. ETP.....S. .T.RH.L.S. RRS...EA.. DK.L.Y.... LN.N.H....
BOV BAR2 LV..K..F.. EKP.....S. .T.RH.L.S. R.S...EA.. DK.L.Y.... LN.N.H....
HU CAR LVV.K..F.. P.A..G..S. QTIRR.LL.A Q...Q.QAWM D..VHY.... .S.N.S-N.C
XEN CAR LI..K..F.. EAT.TA.SCV QT.R..M.S. ....QVE... DK.V.Y.... VGIKLKIN..
D. mel I LG..K..Y.. TKQ.I...CT VVRKD.LL.P GE-.E.E.T. DKQL.H...K .S.NIC.R..
D. mel II LV.KKL.Y.. .NR.QRL.SS LVSKG.TF.N GK-IS.E.T. D....Y...K TAA..Q.S..
D. miranda LV.KKL.Y.. .NR.QRL.SS LVSKG.T-.N GK-IS.E.T. D....Y...K TAA..Q.S..
LIM MA..KL.Y.K PSPLAKQ.S. LVSKG.M..S GK-.Q.E.T. DK.L....DK VSAN..IS.Y
LOCUST LA.KKL.Y.. PTR.RRL.SS LVSKG.TF.Q GK-IN.E.T. D....Y...K LAAN.IIN..
Heliothis LG..K..Y.. AKP.....CT VVRKD.VL.P GQ-.E.ELT. DKQL.I...T VA.NMC.R.H
2
2
V5
HU SAG TEKTVKKIKA CVEQVANVVL YSSDYYVKPV AMEEAQEK-- VPPNSTLTKT LTLLPLLANN 291
RAT SAG ...V.....V S...I..... .....S..T..... .Q..... .V.V.....
MUS SAG .D.V.....V S...I..... .....S..T..... .Q..... .V.V.....
BOV SAG .....V L....T.... .....I.T. .A..... .....S.... ...V.....
HU BAR1 .N...E...I S.R.Y.DIC. FNTAQ.KC.. .....DDT.. .A.S..FC.V Y..T.F....
RAT BAR1 .N.....I S.R.Y.DIC. FNTAQ.KC.. .....DDT.. .A.S..FC.V Y..T.F....
BOV BAR1 .N.....I S.R.Y.DIC. FNTAQ.KC.. .....DDT.. .A.S..FC.V Y..T.F....
HU BAR2 ST.....V S.R.Y.DIC. F.TAQ.KC.. .QL.QDDQ.. .S.S..FC.V Y.IT...SD.
RAT BAR2 SA.....RV S.R.Y.DIC. F.TAQ.KC.. .QL.QDDQ.. .S.S..FC.V Y.IT...SD.
BOV BAR2 ST.....V S.R.Y.DIC. F.TAQ.KC.. .QV.QDDQ.. .S.S..FC.V Y.IT...S..
HU CAR .N.VI....I S.D.ITD... ..L.K.T.T. FIQ.FT.T.. .AA..SFSQS FAVT.I..AS
XEN CAR .S.I.....I T...LTD... ..L.K.T.I. CC..IN.T.. .AA.ANFSGS YS.T.....
D. mel I SN.V..... M.Q.GVD... FQNGQFRNTI .FM.TS.GCP LN.G.S.Q.V MY.V.T.VA.
D. mel II SK.S..S..C FIV.HTEITM V-NAQFS.H. .QL.TK.GCP IT.GAN.... FY.I..A...
D. miranda SK.S..S..C FIV.HTEITM V-NAQFS.H. .QL.TK.GCP IT.GAN.... FY.I..A...
LIM SK....N..V A.V.NTE.TM V-NGHFH.TI SSI.SK.GCP IT-GA..S.V Y.....ASQ.
LOCUST SR....N..V Y.V.HCE.TM V-NAQFSRH. .SL.TR.GCP IT.GASF..V FY.V.CA.S.
Heliothis SN.V..... .IQ.GVD... FQNGQ.RNI. .SI.T.DGCP LQ.G.S.Q.V .H.T.T..H.
C10C10
C3
2
AS8
HU SAG RERR-GIALD GKIKHEDTNL ASSTIIKEGI DR-TVLGILV SYQIKVKLTV S----GFLGE 345
RAT SAG ....M.... .H..... .....
MUS SAG ....M.... .H..... .....
BOV SAG .... .K...M.... ....L...
HU BAR1 ..K...L... ..L..... ....LLR..A N..EI...I. .KV....V. .RG...L..D
RAT BAR1 ..K...L... ..L..... ....LLR..A N..EI...I. .KV....V. .RG...L..D
BOV BAR1 ..K...L... ..L..... ....LLR..A N..EI...I. .KV....V. .RG...L..D
HU BAR2 ..K...L... ..L..... .....V...A NK.E..... .RV....V. .RG.....
RAT BAR2 ..K...L... .QL..... .....V...A NK.E..... .RV....V. .RG.....
BOV BAR2 ..K...L... ..L..... .....V...A NK.E..... .RV....V. .RG...L..D
HU CAR CQK...L... ..L..... ......RP.M .K.EL..... .KVR.N.M. VSGCG.I..D
XEN CAR K.K...L... ..L..G.... .....LRP.M .K.E...M.. .KVR.S.V. ARG...I..D
D. mel I CD-.A...VE .D..RK..A. ..T.L.ASQD A..DAF..I. ..AV....F- -..LGALG..
D. mel II KD-.H..... .HL.D..V.. ....MVQ..K STGDAC..VI ..SVRI..N- -..CGTLG..
D. miranda KDD.H..... .HL.D..V.. ....MVQ..K NTGDAC..VI ..SVRI..N- -..CGTLG..
LIM KDK..... .ML.EG.... ....L----N STGDAI..VI ..V.R.R.Y- -..MGAIG..
LOCUST KD-.Y..... .YL.DD.V.. ....LVS..K NTTDAI..VI ..SLR...N- -..CGTLG..
Heliothis .D-KR..... .QL.RS..T. ..T.LLLDPD Q..DAF..V. ..SA....Y- -..LGAIS..

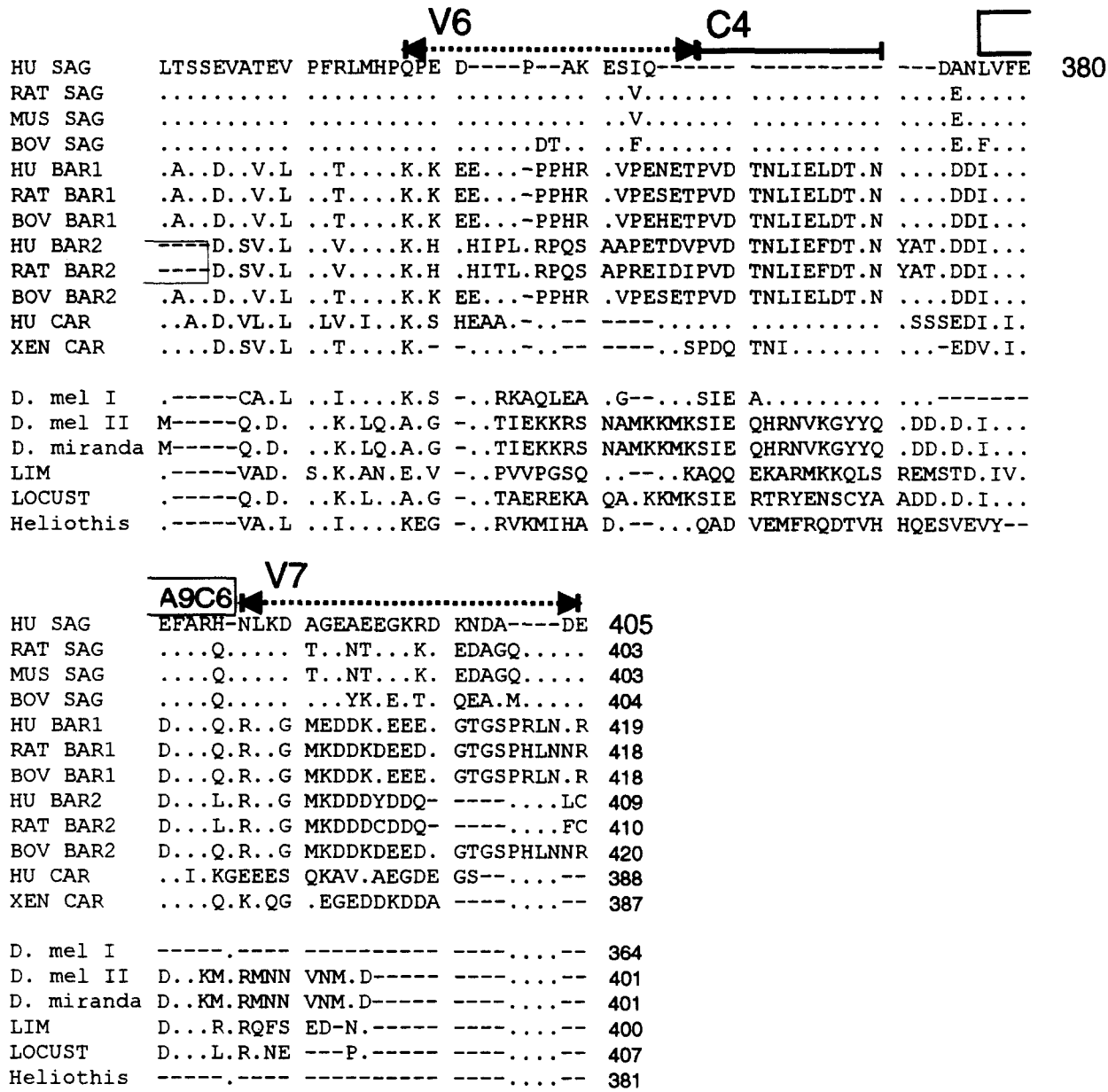


Fig. 3. Deduced amino acid sequences of 18 arrestin proteins aligned by CLUSTAL V. Human S-antigen (HU SAG) [13], rat S-antigen (RAT SAG) [14,5], mouse S-antigen (MUS SAG) [15], bovine S-antigen (BOV SAG) [16], human β-arrestin1 (HU BAR1) [17], rat β-arrestin1 (RAT BAR1) [17,5], bovine β-arrestin1 (BOV BAR1) [19], human β- arrestin2 (HU BAR2) [20], rat β-arrestin2 (RAT BAR2) [18,2], bovine β-arrestin2 (BOV BAR2) [21], human cone arrestin (HU CAR) [1,2], *Drosophila melanogaster* arrestin I (D. mel I) [22,23], *Drosophila melanogaster* arrestin II (D. mel II) [24,25], *Drosophila miranda* arrestin (D. miranda) [26], *Limulus* arrestin (LIM) [27], *Locusta* arrestin (LOCUST) [28], and *Heliothis* arrestin (Heliothis) [28]. Gaps (−) were introduced to obtain maximum similarity. Dots (.) indicate residue identity when compared to human SAG. Single-letter abbreviations for amino acids are used: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile;, K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Glu; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. Primary structure analysis was performed by the PROSITE subroutine in PCGENE. Dashed arrows with accompanying Vs are regions of pronounced variability. Solid bars indicate highly conserved regions designated as 'arrestin signatures'. Boxes with MAB 5C6.47, C10C10, and A9C6 mark antigenic sites for each of these antibodies [29,30]. The designation AS8 marks a position of alternative splicing in the BAR gene [31]. Circled numbers are possible motifs highlighted by PROSITE: 1, cGMP-dependent protein kinase phosphorylation site; 2, protein kinase C phosphorylation site; 3, tyrosine kinase phosphorylation site; 4, ATP-GTP binding motif (P 1 loop).

brate arrestins was estimated by moving a window 10 × 12 amino acids in dimensions sequentially down the aligned arrestins. This produced 45 blocks of information. Within each block, substitutions at each site were counted as 1.0 unless the replacement was with an amino acid of the same group (groups: KHR, C, DENQ, GP, AST, ILMV, and FWY) [10] in which case a value of 0.5 was recorded for these conservative substitutions. Each block possessed 120 positions unless gaps were present. The divergence index value for each block was calculated by dividing the sum of the substitution values by 120 minus the number of gaps. Divergence index values ranged from 0 to 1: 1 would indicate non-conservative substitutions at every site in the block, and 0 would result if no substitutions were observed within the block.

3. Results and discussion

3.1. Isolation of Xenopus arrestin cDNAs

The cDNA recombinants represent two different types of retinal arrestin. Ten cDNAs were identified on low stringency hybridization from an initial screen of 3×10^5 recombinants, plaque purified and excised into pBluescript after the tertiary screen. One set of clones did not hybridize with the human cone arrestin (HCAR5B) cDNA probe at increased wash stringency. One of these (XrSAG9b12, 613 nucleotides) has been sequenced and, based on translated amino acid, is a partial cDNA encoding an SAG-like arrestin (data not shown). The second set of cDNA clones continued to hybridize with the HCAR probe through high stringency wash conditions (0.1 × SSC, 60°C). Both strands of one complete cDNA (XrCAR7b31, 1356 nucleotides) were subsequently sequenced and found to translate into a 387 amino acid cone-like arrestin (Fig. 1).

3.2. Arrestin mRNAs are different sizes and enriched in retina

The retinal mRNA encoding *Xenopus* SAG arrestin was approximately 1.9 kb while that encoding the *Xenopus* CAR arrestin was about 1.7 kb (Fig. 2). Each mRNA was identified using a verified cDNA of XrSAG9b12 or XrCAR7b31, respectively, as a radiolabeled probe for the Northern blot. Both probes identified a distinct, single mRNA in retina. Additionally, XrSAG9b12 mRNA was also observed in brain and testes (data not shown). The mRNA size differences may be due to the smaller size of CAR proteins compared to SAGs, but verification awaits the completion of sequencing a full-length *Xenopus* SAG cDNA.

3.3. Arrestin motifs

Gene duplication leading to the foundation of families of genes whose members have distinct but related functions is a major tenant of the concept of adaptive evolution [11,12]. One approach to the assessment of adaptive evolution is to examine many protein sequences from divergent species in order to examine the evolution of the protein family as it maintains functional similarity. The arrestin multigene family consists of four recognized families of arrestins with evidence of the existence of at least one more. The established families are best characterized by the vertebrate arrestins and are (1) rhodopsin arrestins or S-antigens, (2) β-arrestins 1, (3) β-arrestins 2 and (4) cone arrestins. The insect retinal arrestins probably represent a fifth family, the olfactory or chemosensory arrestins. Confirmation of the existence of this family awaits the discovery of vertebrate homologs.

The sequenced *Xenopus* cDNA was translated and the deduced amino acid sequence was aligned with 17 other deduced arrestin amino acid sequences from both invertebrate and vertebrate species representing four families of the arrestin superfamily (Fig. 3). A divergence index value was calculated for each of 45 blocks of amino acids. The invertebrate arrestins were so divergent not only from the vertebrate arrestins but from one another they contributed little to pattern recognition and were eliminated from this analysis. The sequential display of the calculated divergence indices revealed 7 variable (V1–V7) and 4 conservative (C1–C4) regions (Fig. 4). Most of the variable regions highlighted by this analysis bear a striking characteristic: the sequences are generally conserved among arrestin family members, but highly divergent between the families. For example, V5 (between aa 249–300) is the most extensive region of divergence in the superfamily. Within this segment there are many amino acid differences between the four arrestin families resulting in high divergence indices, but upon close examination of this region, one is struck by the marked intrafamily conservation. This combination of characteristics is often viewed as indicative of members of a multigene family that have diverged, but conserved regions that are specific to the function of each of the families.

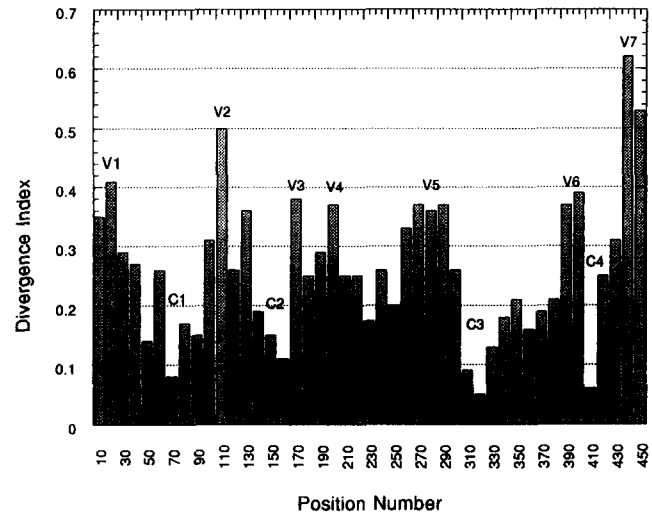


Fig. 4. Estimate of sequence divergence. A divergence estimate was calculated for each block of 10 amino acids in the aligned sequence arrangement of Fig. 3. Twelve vertebrate arrestins were used in the analysis which counted conservative amino acid substitutions as 0.5 and non-conservative as 1.0. The total substitution value of each block was divided by the number of possible substitution sites within the block, 120 minus the number of gaps. V1–V7 are the most variable domains and C1–C4, the most conservative.

Arrestin families diverge at both the N- and C-terminus. Much discussion has focused on the C-terminal domain regarding its role in receptor specificity. The truncated *Drosophila* arrestin I and *Heliothis* arrestin have challenged this concept. In addition, a splice variant of bovine SAG, named P^{44}, with a truncated C-terminus was recently identified and localized to the rod outer segment during retinal dark-adaptation. This splice variant was shown to be a much more potent inactivator of phototransduction [31,32]. Perhaps the C-terminus does not directly participate in the receptor interaction, but rather governs arrestin tertiary structure or configuration changes that occur during receptor interaction. Alternatively, increased

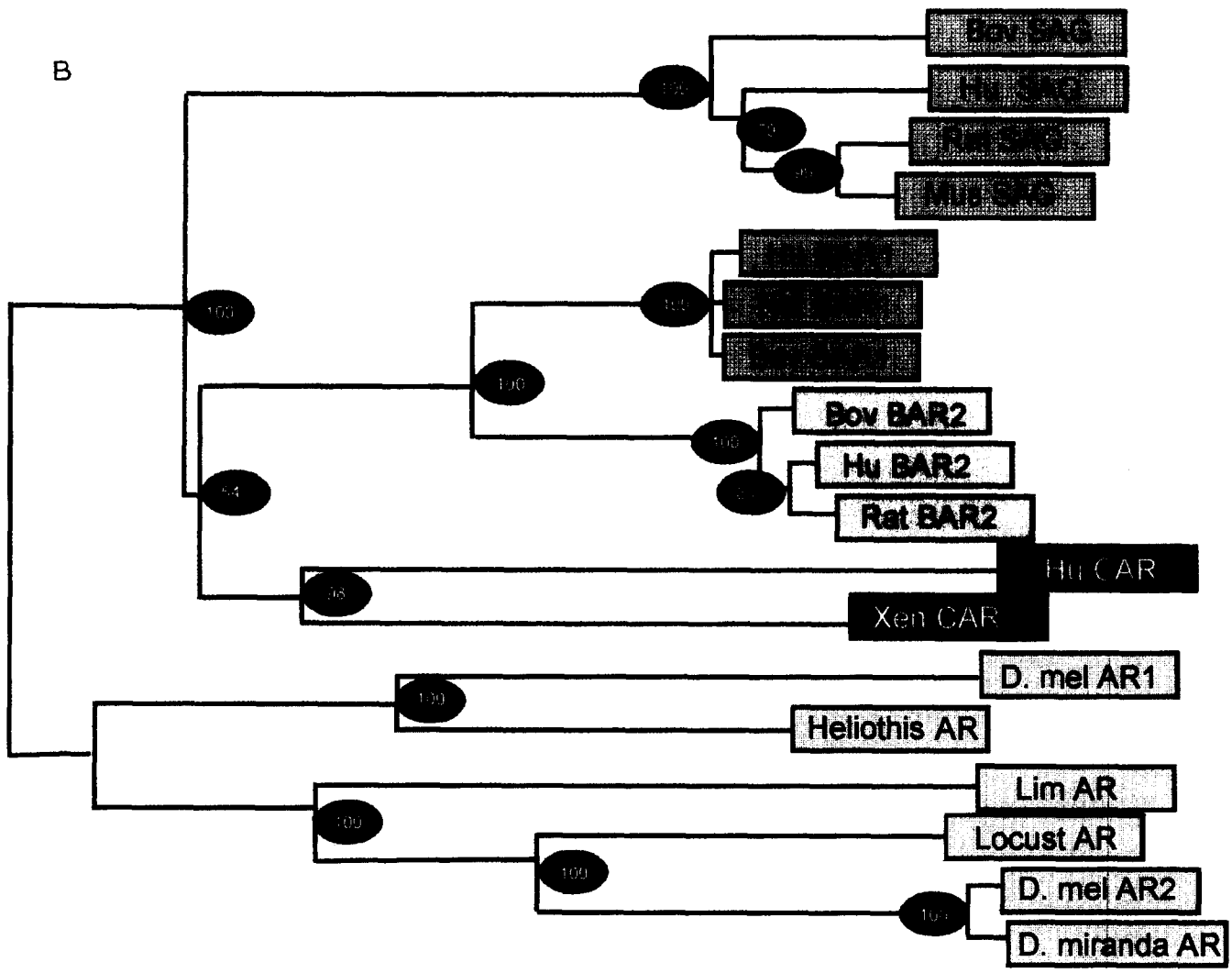
→

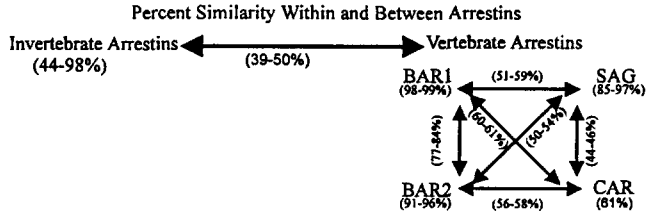
Fig. 5. (A) Pair-wise comparisons of deduced arrestin amino acid sequences. Values below the diagonal are the number of different amino acids and those above the diagonal are the percentage of identical amino acids between each pair of arrestins. (B) An arrestin gene tree was constructed based on calculations of differences in amino acids between pairs of arrestins. Tree topology and branch lengths were obtained by the neighbor-joining algorithm [36]. Circled numbers resulting from 1000 bootstrap replications, indicate the percentage of times the branches were included in a bootstrap tree with a topology identical to the original NJ tree.

A

	Human SAG	Rat SAG	Mouse SAG	Cow SAG	Human BAR1	Rat BAR1	Cow BAR1	Human BAR2	Rat BAR2	Cow BAR2	Human CAR1	Xenopus CAR1	D. melanogaster I	D. melanogaster II	D. miranda I	Limulus	Locusta	Heliothis
Human SAG	_	89	88	85	57	58	58	56	55	56	50	53	44	42	42	40	42	44
Rat SAG	46	_	97	86	57	57	58	56	55	56	51	54	45	41	42	42	43	46
Mouse SAG	50	12	_	86	56	58	59	56	55	56	52	55	46	42	43	42	44	47
Cow SAG	60	56	57	_	58	58	59	55	55	56	51	54	43	41	42	41	43	44
Human BAR1	168	168	165	166	_	98	99	78	77	84	59	60	46	43	43	45	47	49
Rat BAR1	167	167	164	166	9	_	99	79	78	86	61	61	47	43	44	45	48	48
Cow BAR1	165	164	161	163	6	6	_	78	78	84	60	64	47	43	43	45	47	50
Human BAR2	170	168	167	171	89	86	87	_	96	93	56	57	43	45	45	44	46	46
Rat BAR2	172	170	169	173	92	89	90	16	_	91	56	57	43	44	44	43	47	46
Cow BAR2	175	174	172	173	68	60	66	28	35	_	58	58	45	45	45	45	47	47
Human CAR1	195	187	183	190	166	161	161	174	178	173	_	61	45	38	39	43	43	44
Xenopus CAR1	177	172	170	172	154	151	152	163	162	162	154	_	42	39	39	42	42	44
D. melanogaster I	195	189	188	197	189	187	186	198	200	195	194	199	_	46	46	44	49	66
D. melanogaster II	215	216	213	216	218	217	218	212	215	212	230	222	194	_	98	54	72	45
D. miranda I	214	213	210	213	217	216	217	211	214	211	229	221	193	7	_	54	72	45
Limulus	222	213	212	219	208	207	206	211	215	207	211	211	199	177	179	_	57	49
Locusta	218	213	208	214	203	201	202	206	206	205	208	211	184	112	112	170	_	49
Heliothis	198	189	185	196	185	183	182	197	198	192	200	197	123	206	208	190	194	_

B





Scheme 1.

functional interaction or specific binding with other proteins may occur without the C-terminus.

Domains C1–C3 are conserved across the superfamily. One of these, C1 spanning aa 79–97 is cited in PCGENE's subprogram PROSITE as the 'arrestin signature'. The third domain (C3) occurs between aa 304 and 324 and is more conservative than C1. C1 and C3 are fundamental domains of all arrestins and, therefore, are worthy of the designation as 'arrestin signatures'.

Each arrestin was compared to a catalog of known structural motifs in PROSITE to identify possible functional domains. Four different putative functional domains were identified by this analysis and are designated in Fig. 3 as 1, possible cGMP-dependent protein kinase phosphorylation sites; 2, possible protein kinase C phosphorylation sites; 3, possible tyrosine kinase phosphorylation site; and 4, possible ATP-GTP binding motif (also called a P loop). Invertebrate arrestins were viewed as too divergent to be included with the vertebrate arrestin PROSITE analysis.

Downstream from the C2 domain between 170–201 of the aligned sequences in Fig. 3, a boxed area rich in the basic amino acids, arginine and lysine is located. It has been repeatedly suggested this region of SAG is a putative binding site for phosphorylated rhodopsin as well as heparin [34,35]. It is evident in comparisons across the arrestin families that these basic amino acids are conserved in all the families. It is tempting to speculate that a fundamental aspect of all arrestin families is to interact with phosphorylated receptors and that part of this interaction takes place at this domain.

3.4. Relationship of Xenopus CAR to the arrestin gene tree

From deduced and aligned arrestin amino acid sequences a data matrix of amino acid similarities and differences was computed by MEGA (Molecular Genetics Analysis) [9] for each pair of arrestins. The matrix data (Fig. 5A) are arranged to show the numbers of amino acid differences between each arrestin pair (below the diagonal line) and the percent similarity between each pair (above the diagonal line). Gaps in the aligned sequences were handled using the pairwise-deletion option that computes distances for each pair, ignoring only those gaps that are involved in the comparison. An arrestin gene tree was constructed using the distance method neighbor-joining [36] in MEGA which calculates a pairwise evolutionary distance for all arrestins (Fig. 5B). Reliability of the gene tree structure was evaluated by 1000 bootstrap replications. The percentage of times each branch node appeared in a tree is noted. A nearly identical gene tree was produced by the parsimony method in PAUP [37](data not shown).

Except for the near identity of *D. melanogaster* arrestin II and *D. miranda* arrestin, the invertebrate arrestins are surprisingly divergent. In the following diagram, arrestin families are compared based on amino acid similarities. These data reveal that the known invertebrate arrestins are surprisingly divergent, except for the *D. melanogaster* arrestin II and *D. miranda* arrestin, and are in concordance with the more detail relationships depicted by the gene tree.

Part of invertebrate arrestin divergence might be explained by substantial phylogenetic distances between some of the species as well as the liklihood of a mixture of two arrestin families in this category. Two of the invertebrate arrestins (*Heliothis* and *Locusta*) were isolated from antennae and perhaps represent the first members of a fifth arrestin family yet to be described, the olfactory or chemosensory arrestins. The remaining arrestins were all isolated from eyes. The two most closely related invertebrates in this analysis, *D. melanogaster* and *D. miranda* are thought to have diverged about 45 million years ago [26] and since that time the arrestin gene of *D. miranda* has become X-linked while those of *D. melanogaster* remained autosomal. Although mammalian SAG is autosomal, human cone arrestin maps to the X chromosome and is a candidate gene for cone dystrophy [2].

In sharp distinction to the invertebrate arrestins, the known vertebrate arrestins exhibit high intrafamily homology, with the exception of the two cone arrestins. The extent of divergence seen between the CARs is undoubtedly exaggerated by the phylogenetic distances between their sources. Amphibians are thought to have arisen about 345 million years ago, while mammals first appeared about 80 million years ago. The only known members of the other three arrestin families all come from mammals. The CAR family is more closely related to the two BAR families than to the SAG family. Tree data suggest an early gene duplication event that gave rise to the SAG branch and a second branch which experienced another gene duplication event that gave rise to the two BAR families and a CAR family.

3.5. Evolutionary perspectives

All vertebrate classes have members with color vision, but it appears to be most highly developed in birds. Unlike other vertebrate classes, mammals are thought to have arisen from nocturnal ancestors whose influence on visual processes is still strongly evident: cones are usually few, often of two types, while hyoids and oil droplets seen in the visual systems of other vertebrates are lacking. Primates are the exception to this mammalian heritage, in that primates have trichromatic vision. Rhodopsin may have evolved from cone pigments [38], which implies that daylight vision, not twilight vision is the more ancestral state. If this assessment is correct, cone arrestins may be more ancient than rhodopsin arrestins (SAGs). Further analysis of arrestins from other vertebrate classes will add perspective to our emerging view that arrestins have distinct and ancient histories that contribute to the genealogy of the arrestin superfamily.

Acknowledgements: We would like to thank Drs. Green and Besharse for providing the *Xenopus* retina library. CMC is the Mary D. Allen Professor and Chair. Research support in part provided by grants NIH R29 NS28126, EY00395, and EY03040.

References

[1] Murakami, A., Yajima, T., Sakuma, H., McLaren, M.J. and Inana, G. (1993) FEBS Lett. 334, 203–209.

[2] Craft, C.M., Whitmore, D.H. and Wiechmann, A.F. (1994) J. Biol. Chem. 269, 4613–4619.
[3] Goldsmith, T.H. (1990) Quart. Rev. Biol. 65, 281–322. [4] Cannatella, D.C. and DeSa, R.O. (1993) Syst. Biol. 42, 476–507.
[5] Craft, C.M., Whitmore, D.H. and Donoso, L.A. (1990) J. Neurochem. 55, 1461–1473.
[6] Chomczynski, P. and Sacchi, N. (1987) Ann. Biochem. 162, 156–159.
[7] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[8] Higgins, D.G. and Sharp, P.M. (1988) Gene 73, 237–244.
[9] Kumar, S., Tamura, K. and Nei. M. (1993) The Pennsylvania State University, University Park, PA.
[10] French, S. and Robson, B. (1983) J. Mol. Evol. 19, 171–175
[11] Ohno, S, Wolf, U. and Alkin, N.B. (1968) Hereditas 59, 169, 187.
[12] Li, W.-H. and Gojobori, T. (1983) Mol. Biol. Evol. 1, 94–108.
[13] Yamaki, K., Tsuda, M. and Shinohara, T. (1988) FEBS Lett. 234, 39–43.
[14] Abe, T., Yamaki, K., Tsuda, M., Singh, V.K., Suzuki, S., McKinnon, R., Klein, D.C., Donoso, L.A. and Shinohara, T. (1989) FEBS Lett. 247, 307–311.
[15] Tsuda, M., Kikuchi, T., Yamaki, K. and Shinohara, T. (1991) Eur. J. Biochem. 200, 95–101.
[16] Shinohara, T., Dietzschold, B., Craft, C.M., Wistow, G., Early, J.J., Donoso, L.A., Horwitz, J. and Tao, R. (1987) Proc. Natl. Acad. Sci. USA 84, 6975–6979.
[17] Parruti, G., Peracchia, F., Sallese, M., Ambrosini, G., Masini, M., Rotilio, D. and De Blasi, A. (1993) J. Biol. Chem. 268, 9753–9761.
[18] Attramadal, H., Arriza, J.L., Aoki, C., Dawson, T,M., Codina, J., Kwatra, M.M., Snyder, S.H., Caron, M.G. and Lefkowitz, R.J. (1992) J. Biol. Chem. 267, 17882–17890.
[19] Loshe, M.J., Benovic, J.L., Codina, J., Caron, M.G. and Lefkowitz, R.J. (1990) Science 248, 1547–1550.
[20] Rapoport, B., Kaufman, K.D. and Chazenbalk, G.D. (1992) Mol. Cell Endocrinol. 84, R39-R43.
[21] Sterne-Marr, R., Gurevich, V.V., Goldsmith, P., Bodine, R.C., Sanders, C., Donoso, L.A. and Benovic, J.L. (1993) J. Biol. Chem. 268, 15640–15648.
[22] Smith, D.P., Shieh, B.-H. and Zuker, C.S. (1990) Proc. Natl. Acad. Sci. USA 87, 1003–1007.
[23] Hyde, D.R., Mecklenburg, K.L., Pollock, J.A., Vihtelic, T.S. and Benzer, S. (1990) Proc. Natl. Acad. Sci. USA 87, 1008–1012.
[24] Yamada, T., Takeuchi, Y., Komori, N., Kobayashi, H. Sakai, Y., Hotta, Y. and Matsumoto, H. (1990) Science 248, 483–486.
[25] LeVine, H. III., Smith, D.P., Whitney,M., Malicki, D.M., Dolph, P. Smith, G.F.H., Burkhart, W. and Zuker, C.S. (1991) Mechan. Dev. 33, 19–26.
[26] Krishnan, R. and Ganguly, R. (1990) Nucleic Acids Res. 18, 5894.
[27] Smith, W.C., Calman, B.G., Hutchinson, L. Greenber, R.M., Donoso, L.A. and Batelle, B.-A. (1993) Invest. Opthalmol. Vis. Sci. 34, 807.
[28] Raming, K., Freitag, J., Krieger, J. and Breer, H. (1993) Cell. Signal. 5, 69–80.
[29] Donoso, L.A., Yamaki, A.K., Merryman, C.F., Shinohara, T., Yue, S. and Sery, T.W. (1988) Curr. Eye Res. 7, 1077-
[30] Merryman, C.F., Donoso, L.A., Zhang, X., Heber-Katz, E. and Gregerson, D.S. (1991) J. Immunol. 146, 75–80.
[31] Yamaki, K., Tsuda, M., Kikuchi, T., Chen, K.H., Huang, K.P. and Shinohara, T. (1990) J. Biol. Chem. 265, 20757–20762.
[32] Palczewski, K., Buczylko, J., Ohguro, H., Annan, R.S., Carr, S.A., Crabb, J.W., Kaplan, M.W., Johnson, R.S., and Walsh, K.A. (1994) Protein Sci. 3, 319–329.
[33] Smith, W.C., A.H. Milam, D. Dugger, A. Arendt, P.A. Hargrave, and K. Palczewski. (1994) J. Biol. Chem. 269, 15407–15410.
[34] Palczewski, K., Buczylko, J., Imami, N.R., McDowell, J.H. and Hargrave, P.A. (1991) J. Biol. Cehm. 266, 15334–15339.
[35] Gurevich, V.V. and Benovic, J.L. (1993) J. Biol. Chem. 266, 11628–11638.
[36] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[37] Swofford, D.L. (1990). PAUP: Phylogeny Analysis Using Parsimony V 3.0. Illinois Natural History Survey, Chamaign, IL.
[38] Yoshizawa, T. (1994) Biophys. Chem. 50, 17–24.