Amino acid sequence of band-3 protein from rainbow trout erythrocytes derived from cDNA

Stefan HÜBNER, Frank MICHEL, Victor RUDLOFF and Heribert APPELHANS

Max-Planck-Institut für Biophysik, Heinrich-Hoffman-Strasse 7, 6000 Frankfurt/Main, Germany

In this report we present the first complete band-3 cDNA sequence of a poikilothermic lower vertebrate. The primary structure of the anion-exchange protein band 3 (AE1) from rainbow trout erythrocytes was determined by nucleotide sequencing of cDNA clones. The overlapping clones have a total length of 3827 bp with a 5'-terminal untranslated region of 150 bp, a 2754 bp open reading frame and a 3'-untranslated region of 924 bp. Band-3 protein from trout erythrocytes consists of 918 amino acid residues with a calculated molecular mass of 101 827 Da. Comparison of its amino acid sequence revealed a 60–65 % identity within the transmembrane spanning sequence of band-3 proteins published so far. An additional insertion of 24 amino acid residues within the membrane-associated domain of trout band-3 protein was identified, which until now was thought to be a general feature only of mammalian band-3-related proteins.

INTRODUCTION

Band-3 protein is the most abundant integral membrane protein in erythrocytes. It was first identified in human erythrocytes and has also been detected in other vertebrate erythrocytes. There are about 10^6 molecules per cell [1], comprising about 25 % of the total membrane protein [2–4]. Proteolytic digestion of inside-out vesicles or unsealed human ghost erythrocytes with trypsin generates two fragments: a hydrophilic cytoplasmic fragment of 41 kDa and a hydrophobic 52 kDa integral membrane fragment [5]. This digestion pattern represents the two-domain structure of band-3 protein. Four major functions of band-3 protein can be attributed to the two domains. The *N*-terminal hydrophilic domain protrudes into the cytoplasm and provides a binding site for ankyrin, thereby organizing and stabilizing the erythrocyte cytoskeleton [6]. In addition, binding sites for the glycolytic enzymes aldolase and glyceraldehyde-3-phosphate dehydrogenase [7] as well as a major antigenic determinant have been assigned to this domain [8]. Band-3 protein also binds deoxygenated haemoglobin [9] at its extreme *N*-terminus, inducing a clustering of band-3 protein in the membrane. After clustering, binding sites for autoantibodies are provided, which in turn serve as a signal for phagocytosis by circulating macrophages [10–13]. Thus band-3 protein is also involved in the elimination of senescent or abnormal erythrocytes.

The *C*-terminal membrane-spanning domain of band-3 protein comprises an electroneutral chloride/bicarbonate exchange function (for review see ref. [14]). This exchange facilitates buffering of intracellular bicarbonate and provides the means to dilute intracellular bicarbonate by providing additional extracellular volume of the blood for transport of bicarbonate from peripheral tissue to the lung.

Until now cDNAs encoding erythroid band-3 protein have been cloned and isolated from mouse [15], human [16], chicken [17,18] and rat [19]. Comparison of the amino acid sequences revealed a strong similarity within the transmembrane domains of 85–90 %, whereas the cytoplasmic domain is less conserved.

Erythroid band-3 protein is a member of an anion-exchanger gene family from which at least two other members have been identified in various vertebrates by cloning of the corresponding cDNAs. These band-3-related cDNAs have been isolated from a

variety of non-erythroid tissues such as stomach, brain [20], kidney [21] and certain erythropoietic precursor cells [22] and show about 60 % amino acid identity within their transmembrane domains with their erythroid counterparts. A unique feature of all band-3-related proteins known so far is the existence of a single insertion of 25–29 amino acids within the transmembrane domain, whose involvement in anion transport is not yet known.

Earlier studies by Romano & Passow [23] on the kinetics of band-3-mediated anion transport in trout erythrocytes revealed that the anion transport in these erythrocytes is similar to, but yet differs in certain aspects from, that found for human erythrocyte membranes. Most striking are the differences in the activation enthalpies of sulphate equilibrium exchange. In addition to these findings, Michel & Rudloff [24] demonstrated that treatment of native trout erythrocytes with proteinases results in a digestion pattern of band-3 protein that is somewhat different from that observed for the mammalian protein. A major prerequisite for understanding the structures responsible for the characteristics of trout band-3-mediated anion transport is the availability of its complete amino acid sequence. As a first step towards this goal, we report the cloning and isolation of the full-length cDNA coding for trout erythrocyte band-3 protein.

MATERIALS AND METHODS

Materials

Restriction enzymes, T_4 DNA ligase, T_4 polynucleotide kinase and nick-translation kit were from Bethesda Research Laboratories (Eggenstein, Germany). Oligo(dT)-cellulose and T_7 DNA sequencing kit were from Pharmacia. [α - 32 P]dCTP (3000 Ci/mmol), [γ - 32 P]ATP (6000 Ci/mmol) and [α - 35 S]thio[d]ATP (1300 Ci/mmol) were from Du Pont de Nemours (Bad Homburg, Germany). PCR was performed in a thermal cycler (Perkin-Elmer Cetus, Langen, Germany) using a gene amplification kit from Perkin-Elmer Cetus.

RNA isolation

Commercially available rainbow trout (*Salmo gairdnerii*) were anaesthetized with ethyl 3-aminobenzoate methanesulphonate

(MS-222 from Sandoz) at a final concentration of 132 mg/l of fresh water. Over a period of 4 days, 1–2 ml of blood was withdrawn by cardiac puncture. After 14 days, trout spleen was removed and total RNA was isolated after centrifugation through a dense cushion of CsCl as described by Chirgwin *et al.* [25] with a Beckman SW41 rotor (32000 rev./min for 20 h at 18 °C). RNA pellet was dissolved in diethyl pyrocarbonate-treated distilled water. Polyadenylated [poly(A)⁺] RNA was isolated from total RNA by oligo(dT)–cellulose chromatography [26]. The resultant poly(A)⁺ RNA was used for the construction of a trout spleen λ gt10 cDNA library.

For the isolation of total RNA from mature trout erythrocytes, which was only used for PCR experiments, 2 ml of blood was collected in heparinized tubes and pelleted in a Sigma 203K centrifuge (2500 rev./min for 3 min at 4 °C). Serum supernatant and buffy coat were carefully removed, and remaining erythrocytes were washed four times with fish Ringer [145 mM-NaCl, 5 mM-CaCl₂, 1 mM-MgSO₄, 4 mM-KCl, 10 mM-Hepes, 5 mM-glucose, 0.5% (w/v) BSA]. After the final wash, cells were resuspended in 2 ml of fish Ringer. Cells (100 μ l) were mixed with 800 μ l of RNazol B (WAK-Chemie Medical, Bad Homburg, Germany) and homogenized in a glass homogenizer using a Teflon pestle. After addition of 0.1 vol. of chloroform, the solution was mixed, left on ice for 5 min and debris was removed by centrifugation (5000 *g* for 15 min). RNA was precipitated by adding an equal volume of propan-2-ol and left at 4 °C overnight. RNA was recovered by centrifugation (5000 *g* for 45 min) and dissolved in diethyl pyrocarbonate-treated water.

Library construction

Trout spleen λ gt10 cDNA library was constructed as described by Huynh *et al.* [27]. Briefly, oligo(dT)-primed first-strand cDNA was synthesized from 5 μ g of poly(A)⁺ RNA using avian myeloblastosis virus reverse transcriptase (Life Science Inc., St. Petersburg, FL, U.S.A.). Second-strand synthesis was performed by the method of Gubler & Hoffman [28]. Blunt-ended double-stranded cDNA was ligated with *Eco*RI adaptors and cloned into *Eco*RI-cut dephosphorylated λ gt10 arms. Approx. 2×10^6 independent clones with insert lengths of 0.8–3 kb were obtained from 100 ng of cDNA. The library was amplified once and stored at 4 °C.

Library screening

The library was screened with a heterologous probe consisting of a 1686 bp *Bgl*I cDNA fragment encoding the membrane-associated amino acid sequence 312–874 of mouse erythroid band 3 [15]. About 300 000 recombinant clones were plated on *Escherichia coli* C600hflA and allowed to grow for approx. 8 h. Replica filters were prepared according to the manufacturer's instruction for Hybond N (Amersham Buchler, Braunschweig, Germany). Filters were prehybridized in $6 \times$ SSC ($1 \times$ SSC is 0.3 M-NaCl/0.03 M-trisodium acetate, pH 7.0)/ $5 \times$ Denhardt's [29]/0.5% (w/v) SDS at 65 °C for 3 h. Hybridization was performed overnight in the same solution, except that ³²P-labelled cDNA fragment with 0.5×10^6 – 1×10^6 c.p.m./ml was added. Hybridized filters were washed twice in $2 \times$ SSC at 65 °C for 15 min, once in $2 \times$ SSC/0.1% (w/v) SDS for 30 min at 65 °C and finally once for 15 min at 65 °C in $0.1 \times$ SSC. Filters were blotted dry and exposed to Kodak XAR X-ray film between two intensifier screens (Cronex Lightning Plus, Bad Homburg, Germany). Positive plaques were picked, replated and rescreened until pure. Insert lengths of purified λ gt10 cDNA clones were determined by PCR as described below.

Oligonucleotide synthesis and DNA sequence analysis

Oligonucleotides were synthesized in 0.2 μ mol scale on a

Cyclone DNA synthesizer from Milligene (Eschborn, Germany) and purified by reversed-phase h.p.l.c. Both strands of the isolated cDNA clones were sequenced in pUC18 vector utilizing [α -³²S]thio]dATP by the dideoxy chain-termination method with the T₇ DNA sequencing system [31]. The primers used were either pUC18 universal primers or specific 15–20-mer oligonucleotides derived from previous sequence data. Sequence similarities to band-3 and band-3-related DNA and protein sequences were determined by using the Mac vector sequence analysis program version 3.0 (International Biotechnologies Inc., New Haven, CT, U.S.A.).

Hybridization analysis

For Southern-blot analysis, DNA was electrophoresed on a 0.8% (w/v) agarose gel, vacuum-blotted on a nylon membrane and prehybridized in a solution containing $6 \times$ SSC/ $5 \times$ Denhardt's [29]/0.5% (w/v) SDS at 65 °C for 2 h. Hybridization was performed in the same solution with ³²P-labelled DNA at 1×10^6 c.p.m./ml. Filters were washed at 60 °C once in $2 \times$ SSC for 60 min, then in $2 \times$ SSC/0.1% (w/v) SDS for 30 min and finally in $0.1 \times$ SSC for 15 min.

PCR amplification

Oligo(dT)-primed single-stranded cDNA was synthesized from 5 μ g of total erythrocyte RNA. cDNA was desalted and purified from excess oligo(dT) by centrifugation in Centricon 100 units (Amicon, Witten, Germany). One-twentieth of the cDNA was PCR-amplified in a 100 μ l reaction volume containing 10 mM-Tris/HCl, pH 8.3, 1.5 mM-MgCl₂, each deoxynucleotide triphosphate at 0.2 mM, 22 pmol of each trout band-3-specific 5'- and 3'-primer and 2.5 units of *Taq* polymerase. Forty cycles were run, each cycle consisting of 40 s of denaturation at 94 °C, annealing at 47 °C for 30 s and extension for 120 s at 72 °C. One-tenth of the amplified product was analysed by agarose-gel electrophoresis and subsequent Southern-blot hybridization. Insert lengths of the recombinant phages were determined by PCR using 2 μ l of a 100 μ l pure phage suspension in the presence of two λ gt10-specific primer oligonucleotides (λ -left-arm primer 5'-GAGGATCCGCAAGTTCAGCCTGGTTAAG-3', λ -right-arm primer 5'-CGCAAGCTTATGAGTATTTCTTCCAGGG-3'). The amplification reaction was initiated by a denaturation step for 8 min at 94 °C, followed by 40 cycles consisting of 30 s at 94 °C, 30 s at 60 °C, 60 s at 72 °C and a final extension step for 5 min at 72 °C. To identify overlapping cDNA inserts, single-stranded DNA was synthesized by asymmetric PCR [30] using 2 μ l of the pure phage solution. PCR amplification was as described above for insert length determination, except that the molar ratios of the λ gt10-specific primer oligonucleotides were 1:50 and 50:1 respectively. Amplified products were purified by Centricon 100 filter units according to the manufacturer's instructions and sequenced by the dideoxy chain-termination method [31] using that primer oligonucleotide which was limited in the previous asymmetric PCR.

Rapid amplification of the 5'-cDNA end was carried out as described by Frohman *et al.* [32] with the modification that two successive amplifications were performed using two nested primer pairs. Briefly, using 5 μ g of total erythrocyte RNA, cDNA synthesis was primed with the band-3-specific oligonucleotide SF23. After enzymic addition of a poly(A) sequence at the 3' end of the cDNA, the reaction mixture was diluted to 500 μ l in TE buffer (10 mM-Tris/HCl, pH 7.6, 0.01 mM-EDTA), and 5 μ l portions were used for second-strand synthesis and the first amplification reaction. Double-stranded cDNA synthesis was performed in the presence of three primer oligonucleotides [2 pmol of (dT)₁₇, adaptor primer 5'-AAGGATCCGTCGAC-TGCAGAATTCAAGCTTC(T)₁₇-3', 25 pmol of first-adaptor

primer 5'-AAGGATCCGTCGACTGCA-3' and 25 pmol of band-3-specific primer SF28] starting with the denaturation of the RNA-cDNA hybrid for 7 min at 94 °C. *Taq* polymerase (2.5 units) was added and incubation proceeded for 45 s at 94 °C, 2 min at 45 °C and 40 min at 72 °C. For amplification, 30 cycles were subsequently carried out by using a step programme of 45 s at 94 °C, 45 s at 52 °C, 3 min at 72 °C followed by a 15 min final extension at 72 °C. To increase the specificity of the amplified product further, 0.005 vol. of the amplification mixture was used in a second round of amplification in the presence of 25 pmol of each second-adaptor primer (5'-CGACTGCAGAATTCAG-CT-3') and band-3-specific primer SF31. PCR was initiated by denaturation for 5 min at 94 °C, followed by 30 cycles consisting of 45 s at 94 °C, 45 s at 50 °C, 3 min at 72 °C and a final extension at 72 °C for 15 min. Amplified products were separated by gel electrophoresis, eluted from the gel matrix and subcloned in pUC18 for DNA sequence analysis. All DNA sequences obtained from PCR-amplified cDNA fragments were verified by sequencing recombinant plasmids originating from eight independent PCR experiments.

RESULTS AND DISCUSSION

Isolation of band-3-encoding cDNA from rainbow trout erythrocytes

Trout were made anaemic by daily removal of 1–2 ml of blood over a period of 4 days to increase the number of proerythroblasts which are thought to be enriched in band-3 mRNA. Then 14 days after taking the last blood sample, spleen, head kidney and peripheral blood were analysed for the presence of erythrocyte precursor cells. Microscopic analysis of tissue and blood samples revealed that predominantly the proerythroblasts had accumulated in the spleen. Therefore total RNA was extracted from spleen tissue and a cDNA library was constructed in λ gt10 from the poly(A)⁺ fraction. The library was screened with a cDNA fragment of mouse erythroid band-3 encoding the amino acids 312–874 of the membrane-associated domain. This fragment was chosen because it encodes a region of band 3 that is highly conserved in erythroid and non-erythroid band-3 proteins from various species isolated so far. Therefore we thought that this cDNA fragment would be suitable for detecting specific band-3-encoding cDNA clones under highly stringent conditions. Using this hybridization probe, 34 independent clones were extracted from the trout spleen cDNA library. Ten clones were plaque-purified. cDNA-insert lengths of these clones were determined by PCR as described in the Methods section using two λ gt10-specific primer oligonucleotides flanking the *Eco*RI-cloning sites. Analysis of the amplification products revealed insert lengths ranging from 800 bp to 2200 bp. To identify those clones containing different or partially overlapping cDNA inserts, single-stranded DNA was synthesized by asymmetric PCR and subjected to DNA sequence analysis. Two clones with insert lengths of 2200 bp (pFo1) and 1150 bp (pFo2) were obtained which shared an overlap of 111 bp (Fig. 1). Both inserts were subcloned in pUC18 for DNA sequence analysis. The DNA sequence of clone pFo1 contained a short open reading frame with a TGA stop codon and an untranslated region of 924 nucleotides upstream of a poly(dA) stretch. The nucleotide sequence of the 1150 bp insert of pFo2 revealed one open reading frame.

Combining the sequence data of clones pFo1 and pFo2, a polypeptide consisting of 773 amino acid residues could be deduced which had a molecular mass of 85680 Da. Within the cDNA-derived amino acid sequence, a stretch of 19 amino acid residues was found to be identical with one of several protein sequences (pep4) previously determined by microsequencing proteolytic band-3 fragments [33]. This finding indicated that the

isolated cDNA clones pFo1 and pFo2 really encoded band-3 protein. Comparison of the partial amino acid sequence with that of mouse band-3 protein showed an overall identity of about 60% and indicated that the cDNA clones obtained encoded the complete membrane-spanning domain and part of the cytoplasmic domain of trout band-3 protein. However, comparison of the calculated molecular mass of 85680 Da with that of 116000 Da as determined by SDS/PAGE [24] clearly indicated that the major part of the coding region for the cytoplasmic domain was still missing.

To complete the band-3 cDNA sequence, two different PCR methods were performed. First, from the *N*-terminal amino acid sequence which was derived from microsequencing of isolated band-3 protein (pep1) a mixed oligonucleotide SF30 (Fig. 2) was synthesized. In combination with a second primer oligonucleotide SF28 (Fig. 2) which was derived from the sequence of clone pFo2, we amplified further 5'-terminal sequences using single-stranded cDNA as template. As a result of this PCR reaction, a 252 bp fragment was amplified and found to extend the band-3 protein sequence by another 71 amino acid residues (clone pFo3). The cDNA-derived amino acid sequence of clone pFo3

Fig. 1. Schematic representation of the different cDNAs encoding trout band-3 protein and the sequencing strategy

(a) Relative position of the four cDNAs encoding band-3 protein from rainbow trout erythrocytes. PCR indicates the region of clone pFo3 and pFo4 amplified by PCR to produce the *N*-terminal coding sequence of band-3 protein including the 5'-untranslated region. (b) The combined full-length cDNA: (▨), 5'-untranslated region; (■), band-3 coding sequence; (▩), 3'-untranslated sequence. (c) The complete sequence was determined twice using trout band-3-specific primer oligonucleotides the relative positions of which are given below the diagram.

SF 23	5'-GCAGTCGACGAATTCGATCTCACCCTTCGTCGG-3'	+486	+469
SF 28	5'-GGGAATTCAGCTTATCCACCACCTTCTCAGC-3'	+462	+445
SF 30	5'-GG(ACGT)GA(TC)GC(ACGT)CA(GA)GC(ACGT)TA-3'	+223	+239
SF 31	5'-ATTCTCCTCATAGCCAC-3'	+309	+292

Fig. 2. Nucleotide sequence of the oligonucleotides used for PCR experiments

Sequences encoding convenient restriction sites for subsequent subcloning are underlined. For the oligonucleotide SF30 the mixed sequence in the third position is given in parentheses. The positions within the coding region of the cDNA according to Fig. 3 are shown above each oligonucleotide.

The numbering of the nucleotide sequence is given on the right-hand side of each lane and starts with +1 at the first nucleotide of the translation initiation codon ATG. Stop codons in all three reading frames in the 5'-untranslated region are underlined. The putative polyadenylation signal in the 3'-untranslated region is printed in bold letters. The numbering of the amino acid sequence is also given at the right-hand side of each lane and in addition in the middle of the sequence lane under the respective amino acid residue. Negatively charged amino acid residues in the extreme *N*-terminus are printed in bold letters. The two residues found to be *N*-glycosylated are encircled. The sequence of the four tryptic peptides (pep1 to pep4) is underlined and marked by an arrow. The sequence of the 24-amino acid Z-loop structure is underlined with asterisks (*).

proved to be identical with the sequence of two additional peptides (pep2 and pep3) whose primary structures were known from previous microsequencing [33].

Second, the cDNA sequence encoding the complete *N*-terminus of band-3 protein and the 5'-untranslated region was completed by combining the procedure of rapid amplification of cDNA ends (5'-RACE) [34] with that of nested PCR [35]. A highly specific amplification was achieved by first amplifying 5'-sequences with a band-3-specific oligonucleotide (SF28 in Fig. 2) and an 18-mer adaptor primer (first-adaptor primer) followed by an amplification with a nested primer pair representing a second band-3-specific oligonucleotide (SF31 in Fig. 2) and a second-adaptor primer, which was shifted downstream by 10 nucleotides. A cDNA fragment of 485 bp was obtained giving rise to clone pFo4. The complete nucleotide sequence and the deduced amino acid sequence obtained from the four overlapping clones pFo1-pFo4 are presented in Fig. 3. The total length of the combined cDNAs is 3827 nucleotides. The translation of band-3 mRNA probably starts at the methionine codon at the position 151-153 of the cDNA sequence (Fig. 3) and is followed by an open reading frame of 2754 nucleotides. One other putative ATG start codon encoded by base number -57 to -55 is followed by an in-frame stop codon at position -42 to -40. In addition to this termination codon, two other stop codons in the two other possible reading frames are encoded by base numbers -34 to -32 and -12 to -10. The 3'-untranslated region of 924 nucleotides is probably complete and terminates in an A-rich region representing the poly(A) tail. The consensus polyadenylation signal AATAAA is missing. However, a sequence AATTAAA, possibly triggering the polyadenylation, is located 22 nucleotides upstream of the poly(A) tail (Fig. 3).

Amino acid sequence of trout erythrocyte band-3 protein

The primary translation product encoded by band-3 mRNA is 918 amino acid residues long and has a calculated molecular mass of 101 827 Da. The amino acid sequences of four proteolytic fragments of band-3 protein obtained after microsequencing were identified within the cDNA-derived amino acid sequence. Three peptides (pep1-pep3) were found to be located in the cytoplasmic domain, two of them overlapping in four amino acid residues. Peptide 4 comes from the transmembrane domain (Fig. 3). In all cases the sequences obtained by microsequencing and cDNA cloning were identical, with one or two exceptions in the last detectable steps of the Edman degradation. This result clearly shows that the isolated band-3 cDNA is a reverse transcript of an mRNA expressed in trout erythrocytes.

The trout protein is shorter than the mouse (929 amino acid residues) and the chicken (922 amino acid residues) band-3 protein, but, compared with human band-3 protein, trout band-3 protein is seven amino acid residues longer. The extreme *N*-terminus is highly enriched in acidic amino acids: within the first 19 residues, five glutamic acid and three aspartic acid residues can be found. This cluster of negatively charged residues might be the binding site for deoxygenated haemoglobin [35].

A proline-rich hinge region has been noted between positions 189 and 204 in mouse [15] and 175 and 190 in human [16] band-3 protein. In the primary structure of trout band-3 protein, a comparable sequence can be detected at position 23-31 immediately after the negatively charged residues at the *N*-terminus. These five proline residues cluster within a stretch of nine amino acid residues at positions 23, 25, 28, 30 and 31 of the trout band-3 protein sequence.

At positions 14 and 37 two tyrosine residues are located in the *N*-terminal sequence which are good candidates for binding the glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, aldolase and phosphofructokinase. For human band-3 protein it

has been proved that phosphorylation of either Tyr-8 or -21 quantitatively blocks the association of the glycolytic enzymes with band-3 protein [35]. This binding function could be ascribed best to the tyrosine at position 14 in trout band-3 protein because this residue is located in front of the proline-rich hinge region, as is also the case for mouse, human and chicken band-3 protein.

At positions 527 and 530 two lysine residues are located which are homologous to lysine residues 558 and 561 in mouse band-3 protein. Previous studies by Bartel *et al.* [36] on mouse band-3 protein identified residue 558 as the 4,4'-di-isothiocyanostilbene-2,2'-disulphonate (H_2 DIDS)-binding site (Lys a). This lysine residue is highly conserved in all band-3 and band-3-related proteins cloned so far. A second amino acid residue which is highly conserved in erythroid band-3 proteins, but not in band-3-related proteins (AE2), is Lys-860 in trout band-3 protein. This residue is homologous to the lysine residue at position 869 in mouse band 3. By site-directed mutagenesis of mouse band-3 cDNA, this residue has been shown to be involved in the covalent binding of the band-3 inhibitors pyridoxal phosphate and H_2 DIDS [37].

Two potential *N*-glycosylation sites could be detected at positions 545 and 568 (NDS and NKT). From proteolytic cleavage studies of trout band-3 protein, we know that both sites are indeed glycosylated (B. Stirnberg & V. Rudloff, unpublished work). Treatment of native trout erythrocytes with trypsin or chymotrypsin and subsequent gel-electrophoretic separation of the fragments from stripped membranes results in two diffuse band-3 fragments of apparent molecular masses about 62 000 Da and 54 000 Da. This post-translational modification may explain the difference between the molecular mass of 101 827 Da, as calculated from the cDNA-derived amino acid sequence, and molecular mass of about 116 000 Da, as determined by the relative mobility in SDS/PAGE.

In mouse and trout erythroid band-3 protein, not only the number of glycosylation sites but also locations of the identical sites are different. In trout band-3 protein, these sites are shifted in the *N*-terminal direction and are located in the next extracellular loop. They are found in positions identical to the *N*-glycosylation sites of mouse band-3-related protein AE2.

Comparison of the primary structures of trout and mouse band-3 protein

Fig. 4 shows the comparison of the amino acid sequences of trout and mouse band-3 protein. A detailed alignment including all band-3 and band-3-related proteins cloned so far is presented elsewhere [38]. The putative transmembrane domain starts with Pro-392 in trout corresponding to residue 422 in mouse band-3 protein. Comparing only identical amino acid residues and no conservative exchanges, the overall identity comprising cytoplasmic and transmembrane domain between both proteins is 50%. This comparison also considers regions in which single amino acid sites are matched with a gap in either trout or the mouse sequence. If, however, the two domains are considered separately, the cytoplasmic domains share 33% identical amino acid residues, whereas the transmembrane domains exhibit 62%. As noted for the alignment of the membrane-spanning part of mouse and human band-3 protein [16], identical amino acid residues in the transmembrane domain of trout band-3 protein are also clustered in those regions that most probably span the lipid bilayer and make up the transmembrane helices, whereas extra- and intra-cellular loops exhibit a higher degree of variability.

It should be mentioned that several gaps had to be introduced in either trout or mouse band-3 protein to obtain an optimal alignment (see Fig. 4 and [39]). Four will be discussed in detail. The first gap to be introduced in mouse band-3 protein is located

MENDL	SFGED	VM---	-SYEE	ESDSA	FPSPY	PRTPP	GRSGM	YDLAQ	SQJKE	46	TROUT
*GEMR	DME*V	LEIPD	POS**	*LEMT	-----	-----	IQGIA	*MDLT	IPVT*	40	MOUSE
DENQA	IQSIV	VHTDP	EAYLN	LYTNA	NT--R	GDAGA	TYELN	EL-M-	-GH--	89	TROUT
--MQ	DPEAL	PTREQ	--TDY	VPSST	STPMP	SSGVV	****Q	**MDQ	QR*QE	86	MOUSE
-SQKE	TGSGV	GYEEN	FHDGY	GRONG	SHVSY	LYVKS	LILQA	KIDST	GATIL	138	TROUT
LQ*V*	AAH*Y	*L***	LKEDG	V--R*	P*LL*	****M	*LE*Q	*VVF*	*TFL*	135	MOUSE
DLQAS	SLSAV	AEKVY	DELAT	NGEIR	AARDQ	GLIRA	LLOQR	SQS--	-----	181	TROUT
G*MET	*AG*	*NELL	*CFIY	EDQ**	PQ**E	E****	*LK*	*E-AB	DIGNL	185	MOUSE
-----	-----	-----	-----	EGAVA	QPLG-	GDDEM	QTFVS	TQRD		205	TROUT
EGVGP	AVLTR	SGGAS	EPILP	ROPFL	*TQLY	CGQME	*GS*G	PST*G	*L-KI	233	MOUSE
TTDSV	EASIV	LSGVN	DSLEK	PAVAF	VRLED	SVVIE	GALAE	FVSVR	FUVVL	254	TROUT
PP**	*TFL*	*V*RA	NF***	*VLG*	**NKE	A*PL*	D-VL	*E**G	*LL**	281	MOUSE
VGFSG	GGVDY	HESGR	AMJAL	MADWV	FSLEA	YLAFT	NKELT	NAJAD	FMDCS	304	TROUT
L*EA	PH***	TQL**	*A*T*	*TER*	*RIT*	SH*EN	RE*LL	RSLES	*L***	331	MOUSE
IVIPP	TRI--	--QDE	GHQAP	I-IDF	QOQOL	KDRIR	PSDTR	IIFGC	GAKAD	349	TROUT
L*L**	*DAPS	ERALL	N-*Y*	VQKEL	LKRVY	LPSPA	KP*YN	LXVTL	DLNGG	380	MOUSE
KADES	PRDOP	LARTG	IPFGG	MURON	KRRYR	NYISD	FTDAL	DPQVL	AAVIF	399	TROUT
KGGPG	DED**	*R***	RI***	L*R*Y	R**P*	Y*LL*	I****	S****	*****	430	MOUSE
ITPAA	LSPAI	TFGGL	LADRT	EBBGG	VSELM	ISTCV	QGLIF	AFYIA	QPTLV	449	TROUT
*****	*****	*****	*GE**	RJL**	*****	***A*	***L*	*LLG*	*L**	480	MOUSE
IGFSG	PLLVF	ELAAV	AFCKS	QRIEY	IVGRI	WUGLN	LVIIV	VVIVA	VEGSE	499	TROUT
L****	*****	*****	S**E*	RJL**	*****	*Y*Y*	*ILL*	MLV**	F****	530	MOUSE
LVKFI	SNFTQ	KIFSI	LISLI	FIYET	FSKLG	KIFPA	HFVLV	NYEHL	MDSL-	548	TROUT
QY	*Y**	*Y**	*****	*****	*****	*Y*Q*	T*A--	PVVKR		578	MOUSE
DNEFT	PVVKR	HIEYH	EDGKE	TVHEV	IBERA	YFHTA	LISMC	LAFGC	FFIAY	598	TROUT
PK*QG	*****	-----	-----	-----	-----	*****	*F*LV	*A*AT	*LL*H	604	MOUSE
FLAQF	KNGHF	LPGFI	RUMIG	DFGVF	LAIFT	MLAVD	ITIED	AYTKK	LVPVK	648	TROUT
T**K*	*ASTY	F**KL	***V*	*****	*S*LI	*VL**	S*FKG	T****	*S**D	654	MOUSE
GLAVS	MFNAR	GWFIN	PIGEX	KFFPA	WMGGA	CCVPA	LIVFI	LIFLE	SQITT	698	TROUT
*K**	*SS**	*V*H*	***LY	RL**T	*Y*Y*	SVL**	*****	*****	*****	704	MOUSE
LIVSK	PERON	VKSGG	FHLDL	LILVT	MGGIA	SIFGV	PWLSA	AYVRS	VTHAN	748	TROUT
*****	*****	I****	*****	*LVGG	*V*V*	A**H*	*****	T****	*****	754	MOUSE
ALTVM	SRGKP	P----	EIERV	LEQRI	SGMLV	AMVVG	VSILL	EPILK	MIPMT	794	TROUT
*****	*C*ASG	*GAAA	*Q*Q*	K****	*L**	SVL**	L***M	*****	R**LA	804	MOUSE
ALFGI	FLMDG	ITSLS	GIQNM	DRMLL	LIVPR	KYTPA	DAAQA	RYTIM	KSELF	844	TROUT
V****	*****	V****	***LF	*Y**	*FK*P	*H*H*	VFFVK	*K*W*	R****	854	MOUSE
TLIQM	VCLGA	LAMVK	MSAFS	LALPF	VLIIT	IFLPM	AITGT	LFYDK	EMRCL	894	TROUT
*G*YI	I**AV	*V**	STPA*	*****	V***R	L*LPL	I*REL	*LQ**		904	MOUSE
DASDG	KVRYE	KEPGE	DMY-E	SPLP							
*GD*A	*T*TD	*M*AL	*E*H*	V**V*							

Fig. 4. Comparison of the amino acid sequence of trout and mouse band-3 protein

The numbering of each protein is given at the right-hand side. Identical amino acid residues in the mouse are indicated by asterisks (*). Gaps in either protein are dashed.

between residues 25 and 26 corresponding to amino acids 22–31 in trout band-3 protein. It is exactly that position which represents a proline-rich region in trout band-3 protein as discussed above. The opposite situation is found for the second gap, which had to be introduced in trout band-3 protein between amino acid residues 181 and 182. It is matched in mouse band-3 protein, with residues 179–210 representing the proline-rich hinge region in the mouse protein. The third minor gap of four amino acids is located in the membrane-spanning domain of the trout band-3 sequence between residues 760 and 761. In the corresponding region of the mouse protein there are one glycine (position 766) and three alanine residues (positions 767–769) which lead to a somewhat larger loop structure in mouse band-3 protein. The fourth gap is the most evident one, located within the mouse band-3 protein and flanked by residues 585 and 586. In trout band-3 protein, this region corresponds to the numbering 557–580 encoding 24 additional amino acid residues. This amino acid insertion is located in an extracellular loop structure between the fourth and fifth hydrophobic region. Alignment of the trout transmembrane sequence with that of band-3-related proteins identifies this sequence as the so-called Z-loop structure [22,39],

which is a general feature of all non-erythroid band-3 proteins AE2 and AE3. The existence of this structure in erythroid band-3 protein of trout is verified by the fact that the amino acid sequence of proteolytic fragment 4 (pep4) covers 11 amino acid residues from the C-terminus of the Z-loop structure (Fig. 3) and the adjacent PNTALL sequence which represents one of the most conserved regions in all band-3 proteins known so far. The existence of the Z-loop in trout band-3 protein suggests that this structural feature may be part of the erythroid form of band-3 proteins in lower vertebrates. This assumption is further supported by the observation that in several positions of the cytoplasmic domain the amino acid residues are identical for trout band-3 and non-erythroid band-3 proteins of human, mouse and rat, whereas other residues in trout band-3 protein are identical with those of the erythroid forms of other species. On the basis of these findings it seems conceivable that, during evolution, the band-3 gene has been duplicated and mutated, giving rise to a mammalian band-3 gene family in which the erythroid member has mutated at single amino acid sites and eventually lost the Z-loop structure.

We are grateful to Professor H. Passow for his continued advice and encouragement. We thank C. Weindel for excellent technical assistance and for the synthesis of oligonucleotides. We also thank Professor H. Passow, Dr. H. Reiländer and Dr. P. Wood for their comments and critical reading of the manuscript.

REFERENCES

1. Lepke, S., Fasold, H., Pring, M. & Passow, H. (1976) *J. Membr. Biol.* **29**, 147–177
2. Fairbanks, G., Steck, T. L. & Wallach, D. F. H. (1971) *Biochemistry* **10**, 2606–2617
3. Steck, T. L. (1974) *J. Cell Biol.* **62**, 1–19
4. Steck, T. L. (1978) *J. Supramol. Struct.* **8**, 311–324
5. Steck, T. L., Koziarz, J. J., Singh, M. K., Reddy, G. & Köhler, H. (1978) *Biochemistry* **17**, 1216–1222
6. Bennet, V. & Stenbruck, P. J. (1980) *J. Biol. Chem.* **255**, 6424–6432
7. Jay, D. & Cantley, L. (1986) *Annu. Rev. Biochem.* **55**, 511–538
8. Bennett, V. & Stenbruck, P. J. (1980) *J. Biol. Chem.* **255**, 6424–6432
9. Sayare, M. & Fickiet, M. (1981) *J. Biol. Chem.* **256**, 13152–13158
10. Kay, M. M. B. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3521–3525
11. Schlüter, K. & Drenckhahn, D. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6137–6141
12. Shaw, G. M., Aminoff, D., Balcerak, S. P. & LoBulgio, A. F. (1980) *J. Immunol.* **125**, 501–507
13. Waugh, R. F. & Low, P. S. (1985) *Biochemistry* **24**, 34–39
14. Passow, H. (1986) *Rev. Physiol. Biochem. Pharmacol.* **103**, 61–203
15. Kopito, R. R. & Lodish, H. F. (1985) *Nature (London)* **316**, 234–238
16. Tanner, M. J., Martin, P. G. & High, S. (1988) *Biochem. J.* **256**, 703–712
17. Cox, J. V. & Lazarides, E. (1988) *Mol. Cell Biol.* **8**, 1327–1335
18. Kim, H. R. C., Yew, N. S., Ansoorge, W., Voss, H., Schwager, C., Venström, B., Zenke, M. & Engel, J. D. (1988) *Mol. Cell Biol.* **8**, 4416–4424
19. Kudrycki, K. E. & Shull, G. E. (1989) *J. Biol. Chem.* **264**, 8185–8192
20. Kudrycki, K. E., Newmann, P. R. & Shull, G. E. (1990) *J. Biol. Chem.* **265**, 462–471
21. Alper, S. L., Kopito, R. R., Libresco, S. M. & Lodish, H. F. (1988) *J. Biol. Chem.* **263**, 17092–17099
22. Demuth, D. R., Showe, L. C., Ballantine, M., Palumbo, A., Fraser, P. J., Cioe, L., Rovera, G. & Curtis, P. J. (1986) *EMBO J.* **5**, 1205–1214
23. Romano, L. & Passow, H. (1984) *Am. J. Physiol.* **246**, C330–C338
24. Michel, F. & Rudloff, V. (1989) *Eur. J. Biochem.* **181**, 181–187
25. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
26. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408–1412
27. Huynh, T. V., Young, R. A. & Davies, R. (1985) in *DNA Cloning* (Glover, D. M., ed.), pp. 49–78, IRL Press, Oxford and Washington DC

28. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269
29. Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun.* **23**, 641–646
30. McCabe, P. C. (1990) in *PCR Protocols* (Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J., eds.), pp. 76–83, Academic Press, San Diego, New York, Berkley, Boston, London, Sydney, Tokyo and Toronto
31. Korneluk, R. G., Quan, F. & Gravel, R. A. (1985) *Gene* **40**, 317–323
32. Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998–9002
33. Michel, F. (1989) Ph.D. Thesis, University of Frankfurt/Main
34. Frohman, M. A. & Martin, G. R. (1989) *Techniques* **1**, 165–170
35. Low, P. S., Willardson, B. M. & Thevenin, B. (1989) in *Anion Transport Protein of Red Blood Cell Membrane* (Hamasaki, N. & Jennings, M. L., eds.), pp. 103–118, Elsevier, Amsterdam, New York and Oxford
36. Bartel, D., Lepke, S., Layh-Schmitt, G., Legrum, B. & Passow, H. (1989) *EMBO J.* **8** (12), 3601–3609
37. Wood, P. G., Müller, H., Sovak, M. & Passow, H. (1992) *J. Membr. Biol.*, in the press
38. Wood, P. G. (1992) in *Progress in Cell Research* (Passow, H. & Bamberg, E., eds.), Elsevier, Amsterdam, New York and Oxford, in the press
39. Hübner, S., Weindel, C., Michel, F., Rudloff, V. & Appelhans, H. (1991) *Adv. Mol. Genet.* **4**, 219–227

Received 28 August 1991/27 November 1991; accepted 5 December 1991