Biochem. J. (1988) 256, 703–712 (Printed in Great Britain)

The complete amino acid sequence of the human erythrocyte membrane anion-transport protein deduced from the cDNA sequence

Michael J. A. TANNER,* Peter G. MARTIN and Stephen HIGH
Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, U.K.

1. We have isolated cDNA clones corresponding to the red cell membrane anion-transport protein (Band 3). 2. The cDNA clones cover 3475 bases of the mRNA and contain the entire protein-coding region, 150 bases of the 5′ untranslated region and part of the 3′ non-coding region, but do not extend to the 3′ end of the mRNA. 3. The translated protein sequence predicts that the human red cell anion transporter contains 911 amino acids. 4. The availability of the amino acid sequence allows the interpretation of some of the many studies on the chemical and proteolytic modification of the human protein aimed at examining the structure and mechanism of this membrane transport protein.

INTRODUCTION

The exchange transport of anions across the erythrocyte membrane is mediated by the major integral membrane protein (Band 3, reviewed by Jay & Cantley, 1986). This protein has an M_r of about 95000 and two domains. The 40000-M_r *N*-terminal domain is located within the cytoplasm and forms a binding site for the red cell skeleton and other proteins, while the 55000-M_r *C*-terminal domain is membrane-associated and is involved in anion exchange.

The complete sequence of the mouse red cell anion-transport protein has been deduced from the cDNA sequence (Kopito & Lodish, 1985). Portions of the amino acid sequence of the human red cell protein have also been determined. These include about 200 amino acids of the *N*-terminal cytoplasmic domain (Kaul *et al.* 1983) and fragments from the anion-transporting membrane domain (Mawby & Findlay, 1982; Brock *et al.*, 1983). The amino acid sequence of the chicken red cell protein inferred from the cDNA sequence has also very recently been reported (Cox & Lazarides, 1988). In addition, cDNA clones coding for a protein with an amino acid sequence closely homologous with the red cell anion-transport protein in the membrane domain, but clearly distinct from it, have also been isolated from the human K562 cell line (pHKB3, Demuth *et al.*, 1986). The extent of the homology with the transport domain of the red cell protein suggests that this protein also functions as an anion transporter.

There have been many studies on the mechanism of anion transport in human erythrocytes. The locations in the polypeptide chain of the binding sites of different anion-transport inhibitors and the structure and topology of the human protein have been the subject of numerous investigations (reviewed by Jennings, 1985; Jay & Cantley, 1986; Passow, 1986). Studies on the properties of the cytoplasmic domain and its interaction with other proteins have also mainly involved the human protein (reviewed by Low, 1986). Although the availability of the complete amino acid sequence of the mouse protein has been of considerable value in interpreting these results, there are known sequence differences throughout the mouse and human proteins, particularly in the *N*-terminal cytoplasmic domain. Detailed interpretation of the results of the many physico-chemical studies on the human protein requires that the complete amino acid sequence of the human protein be available. We have isolated cDNA clones for the human red cell anion-transport protein which contain the entire protein-coding region and allow us to deduce the complete amino acid sequence of the protein.

MATERIALS AND METHODS

Isolation of cDNA for human K562 band 3

A 24 base synthetic oligonucleotide complementary to bases 1516–1539 of pHKB3 (Demuth *et al.*, 1986) was end-labelled using polynucleotide kinase (Maniatis *et al.*, 1982) and used to screen a human foetal liver cDNA library in λgt11 (Clontech Laboratories, Palo Alto, CA, U.S.A). Hybridization was done in 5 × SSPE at 65 °C (1 × SSPE contained 0.18 M-NaCl, 1 mM-EDTA and 10 mM-sodium phosphate buffer, pH 7.0) and the filters were washed in 5 × SSPE at 68 °C. The insert from the DNA of one of the positive phage [KB3–5 (1.4 kb), see the Results and discussion section for further information] was excised with *Eco*RI, gel purified and subcloned into *Eco*RI-cut Bluescript vector (Stratagene Cloning Systems, San Diego, CA, U.S.A.). The DNA sequence from the ends of the insert was identical with corresponding portions of the pHKB3 cDNA isolated by Demuth *et al.* (1986).

Isolation of cDNA clones for the human red cell anion-transport protein

The KB3-5 insert was nick-translated by using a commercial kit (Amersham International, Amersham,

Abbreviation used: H_2-DIDS, di-isothiocyanatodihydrostilbene disulphonate.

* To whom correspondence should be addressed.

These sequence data have been submitted to the EMBL/Genbank Data Libraries under the accession number X12609.

U.K.) and used to screen a human reticulocyte cDNA library in λgt11 (Conboy *et al.*, 1986) under low-stringency conditions. Hybridization was done at 65 °C in 6 × SSC (1 × SSC contained 0.15 M-NaCl and 15 mM-sodium citrate buffer, pH 7.0) and the filters were washed at 60 °C in 2 × SSC. The inserts in two positive phage that were isolated (HEB3-1 and HEB3-2) were sequenced independently on both strands. HEB3-1 was subcloned into Bluescript and the sequence obtained progressively using synthetic oligonucleotide primers complementary to the sequence of the insert. The two *Eco*RI fragments comprising HEB3-2 were each sequenced completely by using internal oligonucleotide primers, as well as using data obtained from sequencing the fragments obtained after shotgun cloning of a *Pst*I digest of the *Kpn*I fragment of the intact HEB3-2 phage DNA containing the internal *Eco*RI site of the HEB3-2 insert (Fig. 1). The sequence of the overlap across the *Eco*RI site in HEB3-2 was established in both directions as described in the Results and discussion section.

DNA sequencing

Sequencing was done by using the di-deoxy method (Sanger *et al.*, 1977) with templates obtained by single-strand rescue of the inserts cloned in the Bluescript plasmid vectors using helper phage, as recommended by the suppliers of the vector. Sequences from the ends of the inserts were obtained using primers corresponding to the T3 or T7 RNA polymerase promoter sites in the vectors. The sequence reactions were carried out using Sequenase (U.S. Biochemical Corp., Cleveland, OH, U.S.A.), as described by the manufacturer. Separations were done using buffer gradient gels (Biggin *et al.*, 1983). Wedge gels were used as described by Tabor & Richardson (1987) to obtain longer sequence readings. Most reactions were done in parallel using dGTP and dITP to resolve any compressions.

Northern-blot analysis

Reticulocyte RNA was prepared using the method described by Clemens (1984). Rabbit bone marrow RNA was prepared by using the guanidine thiocyanate method (Chirgwin *et al.*, 1979) from animals made anaemic with phenylhydrazine (Foxwell & Tanner, 1981). Polyadenylated RNA was prepared (Aviv & Leder, 1972) and blots were prepared using formaldehyde-containing gels (Maniatis *et al.*, 1982) and 5 μg of polyadenylated RNA per track. Total RNA (20 μg) was used in the human reticulocyte RNA track.

RESULTS AND DISCUSSION

Isolation cDNA clones for the human erythrocyte anion transporter

Several attempts were made to isolate cDNA clones for the human erythrocyte anion-transport protein by using degenerate oligonucleotide probes corresponding to different sequences in the regions of the protein where portions of the amino acid sequence had been determined (Kaul *et al.*, 1983; Brock *et al.*, 1983), but these all proved unsuccessful. The successful strategy was based on the close homology between the membrane domains of the erythrocyte protein and a related human protein for which a cDNA (pHKB3) was isolated from the K562 cell line by Demuth *et al.* (1986). A 24-base oligonucleotide was prepared which was complementary to bases 1516–1539 of pHKB3, which corresponds to a region within the membrane domain of pHKB3 which is absent from the red cell protein. The oligonucleotide was used to screen a human foetal liver cDNA library in λgt11 and several positive phage were isolated. One of these (KB3-5) contained a 1.4 kb insert and DNA sequencing from the ends of the insert showed that it corresponded to bases 1492–2759 of the pHKB3 cDNA isolated by Demuth *et al.* (1986).

The KB3-5 insert was used to screen a human reticulocyte cDNA library in λgt11 at low stringency. Two positive phage were isolated, one of which contained an insert of 2.4 kb (HEB3-1, Fig. 1). The other phage (HEB3-2) yielded two fragments of 1.9 kb and 0.8 kb when *Eco*RI was used to excise the insert from the phage DNA.

Sequence of human red cell anion-transport protein cDNA

The DNA sequence of HEB3-1 and the two *Eco*RI fragments of HEB3-2 were determined independently and the relationship between the two cDNAs is shown in Fig. 1. The overlap across the *Eco*RI site in HEB3-2 was established by cutting the intact phage DNA containing HEB3-2 with *Kpn*I and isolating the 1.6 kb *Kpn*I fragment which contained the *Eco*RI site in HEB3-2 (Fig. 1). The

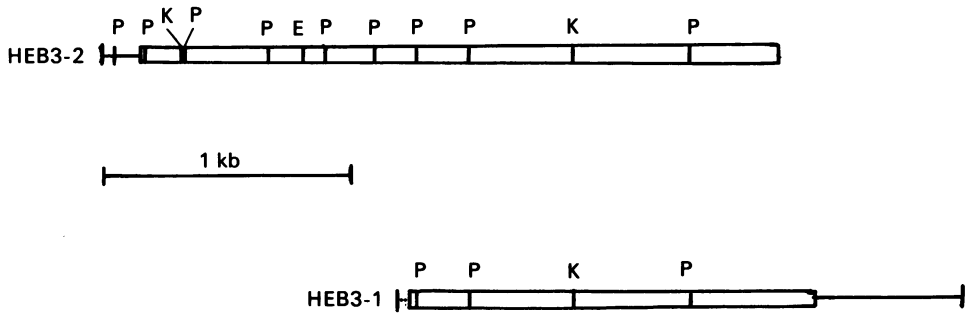

Fig. 1. Relationship between cDNA clones for human red cell anion transporter

The Figure shows the relationship between HEB3-1 and HEB3-2 and a partial restriction map of the two cDNAs. P, *Pst*I; K, *Kpn*I; E, *Eco*RI. The open boxes represent the portions containing the protein-coding region of the mRNA. The dotted portion at the 5′ end of HEB3-1 represents the artefactual sequence discussed in the text.

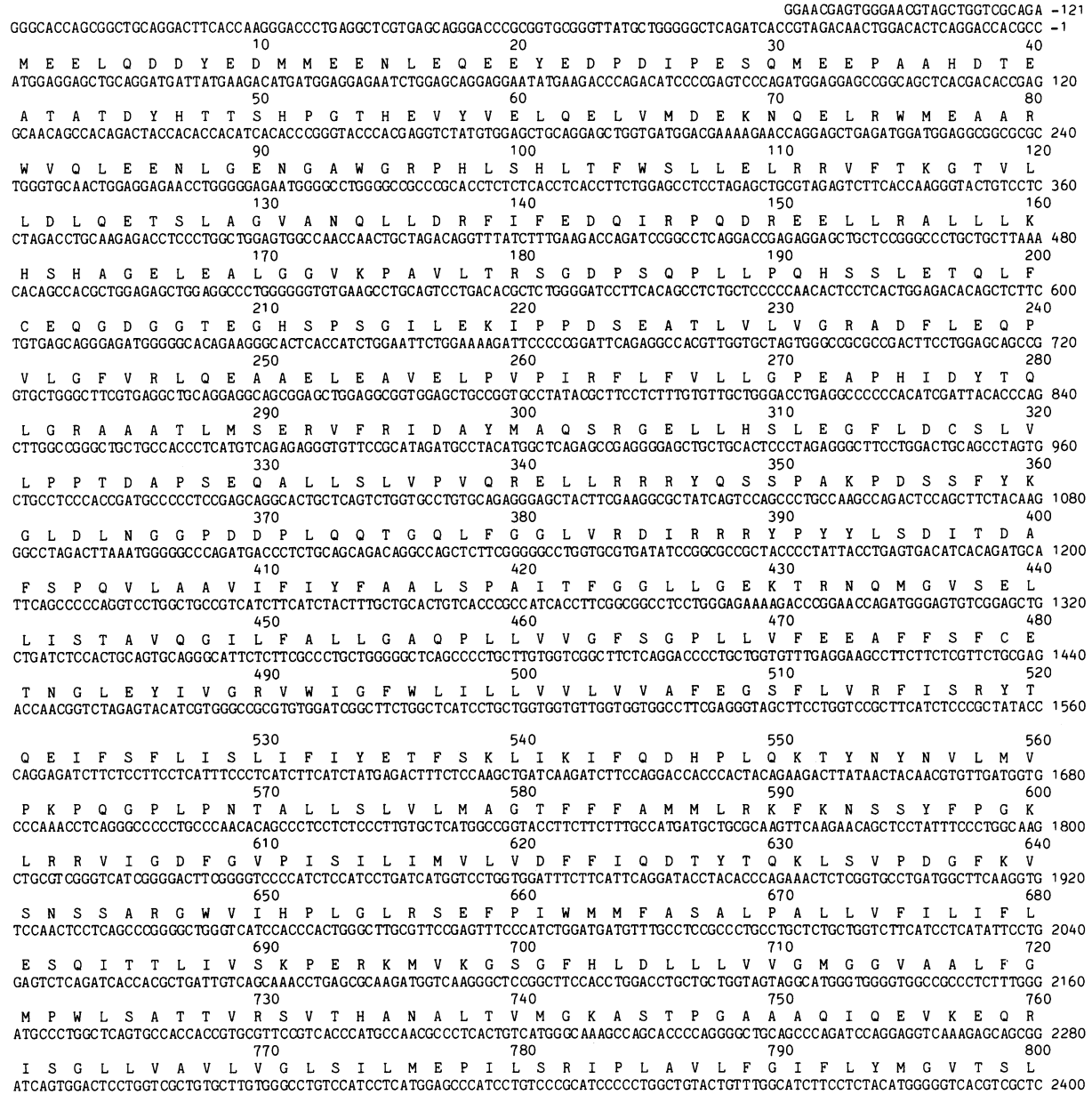

Fig. 2. DNA and inferred protein sequence of human red cell anion-transporter cDNA

The DNA sequence is a composite of that obtained from HEB3-1 and HEB3-2. The sequence is numbered so that base 1 corresponds with the initiation codon for the protein.

*Kpn*I fragment was further digested with *Pst*I and the resulting fragments subcloned in a *Pst*I-cut vector. The *Pst*I fragment containing the *Eco*RI site was sequenced in both directions.

The sequence of the 51 bases at the 5′ end of HEB3-1 was found to be the exact reverse complement of the corresponding portion of HEB3-2 (bases 1028–1079, numbering as in Fig. 2). The translated protein sequence of this region of HEB3-2 showed strong homology with the amino acid sequence of the mouse red cell anion transporter deduced by Kopito & Lodish (1985), whereas that of the corresponding region of HEB3-1 showed no

homology. We conclude that the 51 bases at the 5′ end of HEB3-1 result from a cloning artefact. The remaining overlapping regions of sequence of HEB3-1 and HEB3-2 were identical.

The sequence of the human red cell anion-transport protein cDNA reconstructed from these clones is shown in Fig. 2 and contains 3475 bases. The sequence does not extend to the 3′ end of the mRNA since the cDNAs did not contain a poly(A) tail or polyadenylation signal. The nucleotide sequence around the methionine codon at base 1 (Fig. 2) conforms closely with the consensus sequence CC(A/G)CCAUGG found around initiation codons in eukaryotic mRNAs (Kozak, 1984) and is preceded by two upstream in-frame termination codons at bases −84 and −75. The translated protein sequence from this initiation codon is almost identical with the *N*-terminal sequence of 201 amino acids of the human protein determined by Kaul *et al.* (1983), except that the cDNA predicts methionine instead of aspartic acid at residue 11 and glutamic acid instead of lysine at residue 56. In addition the cDNA does not predict the additional glutamic acid following Asp–67 in the sequence of Kaul *et al.* (1983). The presence of an amino acid at this position would require the insertion of a gap in this region of otherwise strong protein-sequence homology between the human and mouse proteins (see Fig. 4). An open reading frame coding for 911 amino acids extends from the methionine at base 1 and other portions of the predicted amino acid sequence correspond exactly with the amino acid sequences of fragments from the membrane-bound portion of the human protein, which were previously determined [residues 437–473 (Mawby & Findlay, 1982); residue 559–630 (Brock *et al.*, 1983)].

A methionine codon occurs upstream at base −50 and in a different reading frame from the initiator codon and is followed by an in-frame termination codon at base 32. The sequence around this methionine codon is a poor match with the consensus sequence around initiation codons. A methionine occurs in the same position relative to the initiator codon in the 5′ untranslated region of the mouse red cell anion-transport protein (Kopito & Lodish, 1985). In addition to the strong DNA sequence homology in the coding region of the human and mouse cDNAs expected from the strong protein sequence homology (see Fig. 4), homology in the 5′ non-coding regions and the 3′ non-coding regions from the termination codons to base 3230 (of the human cDNA) was also evident (Fig. 3) and may reflect common functional constraints in these regions of the mRNAs for the two proteins. In contrast, no significant homology could be identified in the 5′ and 3′ non-coding regions between the human cDNA and the chicken red cell anion-transport protein cDNA sequence recently reported by Cox & Lazarides (1988) or between the known portions of the 3′ non-coding regions of the human red cell and K562 cell cDNAs. The first nucleotide of the human cDNA corresponds to base −153 of the mouse mRNA, which is one of the transcriptional initiation sites of the mouse mRNA in erythroid tissues (Kopito *et al.*, 1987).

Northern-blot analysis using the 800 bp fragment from the 5′ end of HEB3-2 to the *Eco*RI site detected a 4700 base mRNA in human reticulocyte RNA, while a slightly smaller 4500 base cross-hybridizing mRNA species was detected in rabbit bone marrow and rabbit reticulocyte polyadenylated RNA (results not shown). The cDNA clones we isolated, therefore, lack about 1200 bases of the human red cell anion-transporter mRNA. Comparison with the mouse mRNA, which contains 4400 bases (Kopito & Lodish, 1985), suggests that the missing region comprises mainly the 3′ portion of the non-coding region to the polyadenylation site.

(*a*) 5′ non-coding regions

```
HEB3 -150 GGAACGAGTGGGAACGTAGCTGGTCGCAGAGGGCACCAGCGGC....TGCAC
          || |  ||||| |  |||  ||| ||||||||||| || |||     ||||
MEB3 -153 AGAGCAGTTGGGAGCTCAGCCAGTCACAGAGGGCAC.AGTGGCGGCAGGCAC

HEB3 -100 ACTTCACCAAGGGACCCTGAGGCTCGTGAGCAGGGACCCGCGGTGCGGGT
          ||||| | ||||  ||| ||||||| ||||||||||||| | |||||  |||
MEB3 -100 ACTTCCCGAAGGACCCCCGAGGCTCCTGAGCAGGGACCTGAGGTGCAGGT

HEB3  -50 TATGCTGGGGGCTCAGATCACCGTAGACAACTGGACACTCAGGACCACGC
          ||||||||  ||| ||||   | |   || ||||||||||  |||||  |
MEB3  -50 TATGCTGGCGGCCCAGAGAGCAGAGAACGACTGGACACTGAGGACGGCAT

HEB3   -1 CATGGAGGAGCTGCAGGATGA
          |||||  |||  |||  |||  |
MEB3   -1 CATGGGGGACATGCGGGACCA
```

(*b*) 3′ non-coding regions

```
HEB3 2710 TACGACGAAGTGGCCATGCCTGTGTGAGGGGCGGGCCCAGGCCCTAGACC
          ||  |||||||||  |||||||||||||||||  |||  ||||||||
MEB3 2764 TATGACGAAGTGCCCATGCCTGTGTGAGCGGCAGGCCCAGG.....

HEB3 2760 CTCCCCCACCATTCCACATCCCCACCTTCCAAGGAAAAGCAGAAGTTCAT
            ||||  ||||  |||||  |  ||||||  |||  |||||  | ||     ||||
MEB3 2815 ..CCCCAACCACTCCACTTTCCCACCCTCCCAGGAAGACCA....CTCAT

HEB3 2810 GGGCACCTCATGGACTCCAGGATCCTCCTGGAGCAGCAGCTGAGGCCCCA
          ||||   | ||
MEB3 2849 GGGCGTGTTAT.....

HEB3 2860 GGGCTGTGGGTGGGGAAGGAAGGCGTGTCCAGGAGA.CCTTCCACAAAGG
          ||||||||||||||||||  ||    ||  ||||  | ||  ||  |   | |
MEB3 2860 .GGCTGTGGGTGGGGAAGGAGGGGTGTTCTAGGAAACCCCTCTATCTAAG

HEB3 2909 GTAGCCTGGCTTTTCTGGCTGGGGATGGCCGATGGGGCCCACATTAGGGG
            |||  ||||  ||||||||||||  ||  |  ||    ||    |  |  ||  ||||
MEB3 2909 GTACACTGGATTTTCTGGCTGGAGACGACC...GGAAGCTATATCAGGGA

HEB3 2959 GTTTGTTGCACAGTCCCTCCTGTTGCCACACTTTCACTGGGGATCCCGTG
          ||  |  |||||||  ||                |  |  ||  |||  |
MEB3 2956 ATTGGGGTCACAGTCTCT.....TGCAACTGTATGGGCCCCAGG

HEB3 3009 CTGGAAGACTTAGATCTGAGCCCTCCCTCTTCCCAGCACAGGCAGGGGTA
          ||||||  |  |||||||||  |  |    ||||||    |||  ||  |    ||
MEB3 2995 CTGGAAAAGTTAGATCTGTGTCTCTTCTCTTCATGGCATAGTTATAAATA

HEB3 3059 GAAGCAAAGGCAGGAGGTGGGTGAGCGGGTGG..GGTGCTTGCTGTGTGA
          |  |     |  ||  ||  ||          |    ||||  ||||||||
MEB3 3045 GCAAAGGGGAGAGACGGAGGACACTTTTCATGTTCCTGCTGGCTGTGTGG

HEB3 3107 CCTTGGGCAAGTCCCTTGACCTTTCCAGCCTATATTTCCTCTTCTGTAAA
          ||||  ||  |||||||||||||||  ||  ||||  |  |||||  |||  ||||
MEB3 3095 CCTTAGGTAAGTCCCTTGACCTCTCTGGCCTCTGCTTCCTACTCTATAAA

HEB3 3157 ATGGGTATATTGATGATAATACCCACATTACAGGATGGTTACTGAGGACC
          |||  |  ||||  |  |  |  ||||  |  ||||   ||||   |||||||||
MEB3 3145 ATGTATGCATTGGTAAGAGTACCTATGTTACCATATGGCAGCTGAGGACC

HEB3 3207 AAAGATACATGTAAAATAGGGCTT 3230
          ||   | ||||  ||  |   ||
MEB3 3195 AAGAGCAAATGTCCAAGAACCATT 3218
```

Fig. 3. Sequence homologies in the 5′ and 3′ non-coding regions of human and mouse red cell anion-transporter cDNAs

The sequences of human red cell Band 3 cDNA (HEB3) and mouse red cell Band 3 cDNA (MEB3, Kopito & Lodish, 1985) were compared using the NUCALN computer program (Wilbur & Lipman, 1983). (*a*) 5′ non-coding regions. The protein initiator codons are underlined. (*b*) A portion of the 3′ non-coding regions. The termination codons for the proteins are underlined.

Amino acid sequence of the human red cell anion transport protein

The predicted amino acid sequence of the human red cell protein contains 911 amino acids, slightly fewer than the 929 amino acids predicted for the mouse protein (Kopito & Lodish, 1985). The sequence conservation within the *N*-terminal cytoplasmic domain of the tryptophan-rich region (residues 81–105, numbering based on the human red cell protein sequence; Fig. 4) and the proline-rich hinge regions (residues 175–190) of the mouse and human proteins has already been noted (Kopito & Lodish, 1985), as has the conservation between the human, mouse and chicken proteins of the putative ankyrin-binding domain between residues 106 and 146 (Cox & Lazarides, 1988). Four other regions in the *N*-terminal domain are strongly conserved between the human and mouse proteins (Fig. 4) and are located at residues 217–247, 267–301, 314–347 and 379–400. Residues 217–247 show a less marked conservation in the chicken and human K562 proteins. The conservation between all four known anion-transport protein sequences between residues 379 and 400 is striking. This region is on the *C*-terminal side of a proline-rich region, which contains Tyr-358 and Lys-359 at which the native protein is readily cleaved by intracellular chymotrypsin and trypsin respectively (Mawby & Findlay, 1982). These cleavages occur some 50 amino acids away from the point at which the polypeptide probably enters the membrane (see below), and cleavage closer to the membrane only occurs after denaturation of the protein with alkali (by chymotrypsin; Ramjeesingh *et al.*, 1983) or acid (by pepsin; Brock & Tanner, 1986), suggesting that this region is highly structured. The retention of this portion with the membrane domain after proteolysis of the native protein suggests that it is associated with the membrane domain. This, together with the high degree of conservation of residues 379–400 may reflect a functional role of this highly charged region in the anion-transport process.

Amino acid sequence of the membrane-bound domain

The availability of the amino acid sequence allows the assignment of some of the sites of labelling and chemical and proteolytic cleavage which have been extensively studied by using the human red cell protein. A membrane-bound pepsin fragment (Brock & Tanner, 1986) has been shown to have the amino acid sequence starting at Ile-397 and this is close to the point where the polypeptide on the *N*-terminal side of the membrane domain is likely to enter the membrane from the cytoplasmic surface. The only lysine residue in the *N*-terminal CNBr peptide of the membrane domain is Lys-430 and this has been shown to be susceptible to extracellular labelling (Jennings & Nicknish, 1984). Cys-479, the only cysteine residue in the *N*-terminal portion of the membrane domain, is labelled by extracellular eosin-maleimide (Macara *et al.*, 1983), as well as external *p*-chloromercuribenzene sulphonate (Solomon *et al.*, 1983). An extracellular tyrosine residue in this region, which can be radioiodinated using lactoperoxidase (labelled site 2, Tanner *et al.*, 1979), is likely to be Tyr-486, which is close to this extracellular cysteine. The covalent binding site for the anion-transport inhibitor di-isothiocyanato dihydrostilbene disulphonate (H_2-DIDS) can be localized to Lys-539 (Cox & Lazarides, 1988).

A variety of proteinases cleave the anion-transport protein in the extracellular region with high bend-forming potential between residues 548 and 568. A substantial insertion occurs here in the K562 protein. Papain cleaves the human protein at Gln-550 and Gln-564 while chymotrypsin cleaves at Tyr-553 (Jennings & Adams, 1981; Jennings *et al.*, 1984). Thermolysin cleaves both at this tyrosine and at Tyr-555, while pepsin cleavage occurs at the latter tyrosine residue, as well Leu-558 (Tanner *et al.*, 1979; Brock *et al.*, 1983). However, only one of these tyrosines (at residue 553) is radioiodinated by extracellular lactoperoxidase.

Two potential *N*-glycosylation sites occur in the sequence at Asn-593 and Asn-642. Glycosylation does not occur at the former site and it is likely to have an intracellular location (Brock *et al.*, 1983). However, Asn-642 is extracellular since it is in a polar region adjacent to Tyr-628 (which can be radioiodinated by extracellular lactoperoxidase, Brock *et al.*, 1983) and Gln-630, a site of cleavage by extracellular papain (Jennings *et al.*, 1984). The glycosylation site has been mapped to a position 280 residues from the *C*-terminus of the protein by using an end-labelling method (Jay, 1986), which closely corresponds with Asn-642 located 269 residues from the *C*-terminus.

The sites of proteolytic cleavage in the *C*-terminal portion of the membrane domain are relatively poorly characterized except for a site of intracellular cleavage by trypsin at Lys-743 (Jennings *et al.*, 1986). Rothstein and co-workers (Ramjeesingh *et al.*, 1983) have shown that exhaustive digestion with chymotrypsin yields a membrane-bound fragment of M_r 8000–9000, which contains two cysteine residues. This fragment was assigned a location on the *N*-terminal side of the glycosylation site. It is clear from the amino acid sequence that this fragment originates from the extreme *C*-terminus of the protein and probably results from chymotrypsin cleavage in the polar loop between residues 801 and 834, perhaps at one of the tyrosine residues at positions 818 or 824. Cleavage at these sites would give rise to fragments containing either 93 or 87 amino acids, which would have a size similar to that observed for the chymotryptic fragment.

Jennings *et al.* (1986) have shown that one of the sites of cross-linking of the anion-transport protein by extracellular H_2-DIDS is a lysine located between the trypsin cleavage site at Lys-743 and an *S*-cyanylation cleavage site 5000–8000 Da from the *C*-terminus of the protein. This is consistent with the latter cleavage occurring at Cys-843. Inspection of the sequence shows that the H_2-DIDS cross-linking site is probably one of the four lysine residues which are located in clusters between residues 814 and 829 (Lys-757 is located close to the intracellular trypsin cleavage site and is unlikely to be reactive with extracellular H_2-DIDS). Since the *C*-terminus of the protein is intracellular (Lieberman *et al.*, 1987; S. D. Wainwright & M. J. A. Tanner, unpublished results), the polypeptide must cross the membrane an odd number of times between this extracellular site and the *C*-terminus. This is consistent with the membrane association of the *C*-terminal chymotryptic fragment which contains this region and confirms the conclusion of Jennings *et al.* (1986), which were based on extrapolation from the mouse protein sequence. Two short, but closely adjacent, hydrophobic segments are present in this portion of the sequence at residues 835–850 and 854–869. It should

```
                1    6          7                                    40
MUR RBC  B3      gdmr heev leipdr s e el niig ia  r lt  vte  qd e lp
HUM RBC  B3     MEELQD---- ------DYED MMEENLEQEE YEDPDIPESQ MEEPAAHDTE
                41   46    47                                        86
CHK RBC  B3     .......... .......... .......... .........    g  lh
MUR RBC  B3     q     vpss ts p  ssgq         m   qr    q v    h ig
HUM RBC  B3     ATATDY---- HTTSHPGTHE VYVELQELVM DEKNQELRWM EAARWVQLEE
                87                                                   136
HUM K562 B3     .......... .......... ..    tlah  a     dqq t p   - vv
CHK RBC  B3     smepg    - s  pl  yh      h a a   v    vaa     a  hv
MUR RBC  B3       r d v        y          qk  s    f  g a          h
HUM RBC  B3     NLGENGAWGR PHLSHLTFWS LLELRRVFTK GTVLLDLQET SLAGVANQLL
                137                                  174
HUM K562 B3     eqmvis   k ae  anv           psd ekdfsfprni sagslgscwg
CHK RBC  B3      ql   g lk   h ddv       r k ps  a svwtl--- ----------
MUR RBC  B3      c                        r   ed  gn e   -- ----------
HUM RBC  B3     DRFIFEDQIR PQDREELLRA LLLKHSHAGE LEALGGVK-- ----------
                 175                                   216
HUM K562 B3     itm rglrvt  ts s swev f r r ev r erdvpppap  a  trskskh
CHK RBC  B3     -  aqlqcs  ge kdae ra l re rav m relhga q    -----raql
MUR RBC  B3     -        g a e    hqp       y g  ae  s  p t   tl------
HUM RBC  B3s    -PAVLTRSGD PSQPLLPQHS SLETQLFCEQ GDGGTEGHSP SGI-------
                    217                                    256   261
HUM K562 B3     elkl     e na   v     cve  sr tm a    r  v   d  l v
CHK RBC  B3     gpqlhqql e  t       a c a     l  al   aglvr -d  lav l
MUR RBC  B3     ------         t        n   k         k  vp   dlv-  e
HUM RBC  B3     ----LEKIPP DSEATLVLVG RADFLEQPVL GFVRLQEAAE LEAV-ELPVP
                262                                                  311
HUM K562 B3     v    l     ssanm  hei   sis    d kq hea  l  de ed  tai
CHK RBC  B3     v  vltv    rq apqlprd r     v ad     r   lc gg a   gg
MUR RBC  B3     vg  l                          t       t s   hn e   r
HUM RBC  B3     IRFLFVLLGP EAPHIDYTQL GRAAATLMSE RVFRIDAYMA QSRGELLHSL
                312                                    354
HUM K562 B3     na     v     sevqg el  r vahf  q m kk eeqgr ll tgaglep
CHK RBC  B3     q   ea i     qev    h  aa i l  h av     hpd tvrtpggp--
MUR RBC  B3      s                k    n     k        lp      -------
HUM RBC  B3     EGFLDCSLVL PPTDAPSEQA LLSLVPVQRE LLRRRYQSSP AKP-------
                    355          367    368                        394
HUM K562 B3     ksaq kallq mverq qlkm ipsa ----- - aa    i   v     h
CHK RBC  B3     ---------t apkdt dkgq apqd    lr  rrp             k
MUR RBC  B3     ---- pnl n t      kgg pgde    rr   ri    i
HUM RBC  B3     ----DSSFYK GLDLNGG--- ---PDDPLQQ TGQLFGGLVR DIRRRYPYYL
                395                                                  444
HUM K562 B3       fr  ld   c                            qdl i     im
CHK RBC  B3        r  ln   c                             gm        l
MUR RBC  B3           l                      v            l
HUM RBC  B3     SDITDAFSPQ VLAAVIFIYF AALSPAITFG GLLGEKTRNQ MGVSELLIST
                445                                                  494
HUM K562 B3      l   v c          i                 ss h    l
CHK RBC  B3     s  cl  s   s                      r   dh
MUR RBC  B3                    l                    s n        a
HUM RBC  B3     AVQGILFALL GAQPLLVVGF SGPLLVFEEA FFSFCETNGL EYIVGRVWIG
                495                                                  544
HUM K562 B3        vf al m   l         v  f r    a             y  v
CHK RBC  B3            l     c   v    yl                      a  vt l
MUR RBC  B3            m           q y
HUM RBC  B3     FWLILLVVLV VAFEGSFLVR FISRYTQEIF SFLISLIFIY ETFSKLIKIF
```

```
                545                       565                              567
HUM K562 B3      e   hgcsa s sse dgge nmtwagarpt lgpgnrslag qsgqgkprgq
CHK RBC  B3      a    qs d td ste sv- ---------- ---------- -------- k
MUR RBC  B3       y   q  a -p v k     --------- ---------- -------- v
HUM RBC  B3     QDHPLQKTYN YNVLMVPKPQ G--------- ---------- --------PL
                568                                                    617
HUM K562 B3         p           i ff        rf    ri           a
CHK RBC  B3                   l  lf   q    vfl    v  l          fv
MUR RBC  B3          f        ll  t        t
HUM RBC  B3     PNTALLSLVL MAGTFFFAMM LRKFKNSSYF PGKLRRVIGD FGVPISILIM
                618                                                    667
HUM K562 B3         ys e          s   s tapek      n   ek  p  v   v
CHK RBC  B3     a a    k       k  r  le t gt     f   m sat p
MUR RBC  B3         s  kg           l                  yr l  t
HUM RBC  B3     VLVDFFIQDT YTQKLSVPDG FKVSNSSARG WVIHPLGLRS EFPIWMMFAS
                668                                                    717
HUM K562 B3     l   i        m t       i  k  mlq            i a   ic
CHK RBC  B3     pv             t             l              i a   l
MUR RBC  B3     v
HUM RBC  B3     ALPALLVFIL IFLESQITTL IVSKPERKMV KGSGFHLDLL LVVGMGGVAA
                718                                                    767
HUM K562 B3       l   a a              s  va   dkpk         vt
CHK RBC  B3                  t         v  sav   er h v       l
MUR RBC  B3                              g                          s
HUM RBC  B3     LFGMPWLSAT TVRSVTHANA LTVMGKASTP GAAAQIQEVK EQRISGLLVA
                768                                                    817
HUM K562 B3     l      vig dl rq                   n   fye  lh  lm
CHK RBC  B3       i v         ky                 w f           lm
MUR RBC  B3
HUM RBC  B3     VLVGLSILME PILSRIPLAV LFGIFLYMGV TSLSGIQLFD RILLLFKPPK
                818                                                    867
HUM K562 B3     h    t   k  r l       al ll   l   a m  a     f  i
CHK RBC  B3        ke   t        inn lt  lvv l   g  vs     rc vf v
MUR RBC  B3           f
HUM RBC  B3     YHPDVPYVKR VKTWRMHLFT GIQIICLAVL WVVKSTPASL ALPFVLILTV
                868                                                911
HUM K562 B3        m v tr   tdr mk     ne epv    r  v   n m p
CHK RBC  B3         l   r   sei  k    t   vv  e  a  q v n   q  s
MUR RBC  B3         li       el       g    v      n l       p
HUM RBC  B3     PLRRVLLPLI FRNVELQCLD ADDAKATFDE EEGRDEYDEV AMPV
```

Fig. 4. Alignment of known anion-transport protein sequences

The sequences of the human K562 cell band 3 (HUM K562 B3; Demuth *et al.*, 1986), chicken red cell Band 3 (CHK RBC B3; Cox & Lazarides, 1988), mouse red cell Band 3 (MUR RBC B3; Kopito & Lodish, 1985) and human red cell Band 3 (HUM RBC B3) were aligned using the GAP program (University of Wisconsin Genetics Computer Group). The numbering is based on the human red cell Band 3 sequence which is shown in capital letters. Differences in the other proteins from this sequence are shown in lower case letters. Positions at which the other proteins have sequences identical with that of the human red cell Band 3 are shown with a blank space. Gaps in the protein sequences, which were required to obtain the alignments, are shown by (−). The *N*-terminal sequence of the chicken red cell Band 3 starts at position 76 and the absence of a corresponding sequence for this protein before position 76 is indicated by full stops. The known sequence of human K562 Band 3 starts at position 109 and does not contain the *N*-terminus of the protein (Demuth *et al.*, 1986). The lack of a corresponding sequence for this protein before position 109 is also indicated by full stops.

however be noted that the presence of arginine at residue 858 in the chicken protein makes it unlikely that this position is located within the membrane. The region between the H_2-DIDS cross-linking site and the *C*-terminus probably only traverses the membrane once since it is unlikely that the hydrophobic regions are of sufficient length to cross the membrane three times.

Chemical modifications of both basic and acidic residues in the protein result in the inhibition of anion transport (Weith *et al.*, 1982). Inhibition of an extra-

cellular arginine residue by phenylglyoxal results in labelling of the *C*-terminal 8000–9000-M_r chymotryptic fragment (Bjerrum, 1983). The site of modification is probably Arg-827 or Arg-832, since the remaining arginine residues in this fragment are located in the cytoplasmic domain at the extreme *C*-terminus of the molecule. Modification of an intracellular arginine residue results in partial inhibition of anion transport and labelling of a 10000-M_r peptic fragment (Bjerrum, 1983). This fragment contains residues 549–630 of the protein (Brock *et al.*, 1983) suggesting that one of the arginine residues at positions 589, 602 or 603 (all of which are located in an intracellular cluster of basic residues) is modified in this case. Although modifications of Lys-590 by phenyl isothiocyanate results in the inhibition of anion transport (Brock *et al.*, 1983), this amino acid may not itself be involved in transport since Cox & Lazarides (1988) point out that it is replaced by a glutamine residue in the chicken protein. Inhibition of transport by the extracellular application of an impermeant carbodiimide is also associated with labelling of this region (Bjerrum, 1983). Jay & Cantley (1986) suggest that the modified residue is likely to be Asp-621 since Asp-626 is not conserved in the mouse protein. In contrast, Woodward's reagent K labels an extracellular glutamic acid on the *C*-terminal side of this region and also causes inhibition of transport (Jennings & Anderson, 1987). The glutamate residues at positions 755 and 758 appear to form part of the intracellular loop which contains the trypsin cleavage site at Lys-743 and those from Glu-882

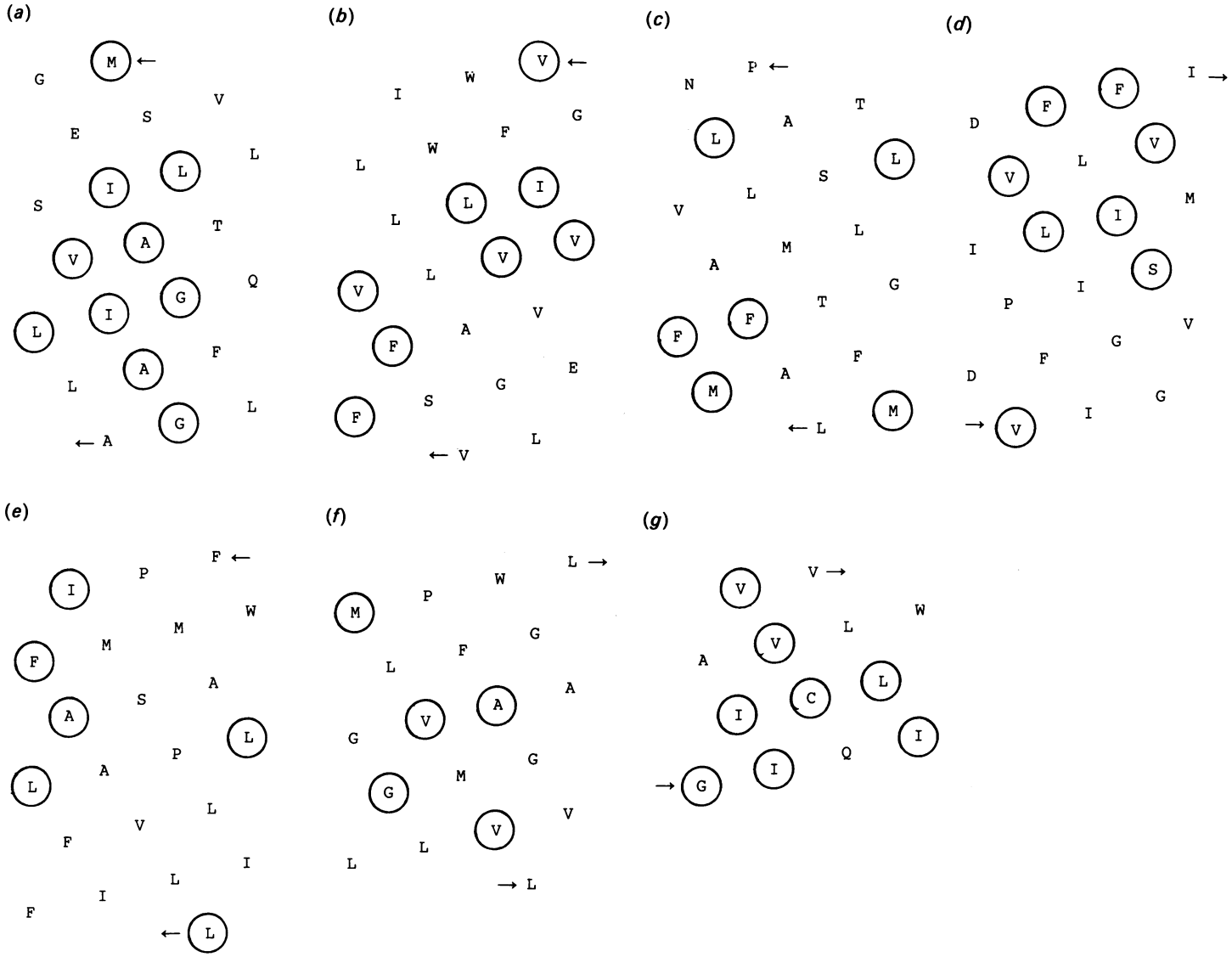

Fig. 5. Locations of amino acid substitutions in some of the hydrophobic segments of the anion-transport protein family

The amino acid sequence of some of the hydrophobic segments of the human anion transport protein are shown drawn on an α-helical net. The *N* to *C*-terminal directions of the sequence are shown by arrows. The circled residues are at the positions where amino acid substitutions are found in the mouse and chicken red cell proteins and in the K562 protein. (*a*) residues 435–457; (*b*) residues 491–513; (*c*) residues 568–588; (*d*) residues 604–624; (*e*) residues 659–681; (*f*) residues 706–724; (*g*) residues 838–850.

to the *C*-terminus are also located in the cytoplasm. Glu-658 is extracellular but not present in the other anion-transport proteins and is thus unlikely to be involved in the transport process. Glu-682 and Glu-694 are likely to be intracellular since the opposite (*N*-terminal) end of the adjacent membrane-traversing segment is extracellular. The most likely candidate for the reactive glutamic acid is Glu-777. This residue could be extracellular since the *N*-terminal end of the adjacent hydrophobic domain is located in the cytoplasm. Although this residue is not conserved in the K562 protein, in this case the adjoining residue is replaced by aspartic acid and there is a corresponding shift in the opposite direction of the adjacent arginine residue so that the general characteristics of this portion of the sequence are retained. An intracellular histidine residue has also been suggested to be involved in anion transport (Matsuyama *et al.*, 1986). All the histidine residues in the membrane domain, apart from His-703 and His-734, are located in regions which are likely to be extracellular, suggesting that one of these residues may be involved in anion transport.

Inspection of the sequence alignments in Fig. 4 shows that the hydrophobic segments of the four proteins show a high degree of homology and are usually bounded by polar regions in which the majority of the amino acid differences between the proteins are clustered. This is consistent with the proteins having a common core structure within the membrane, as would be expected for a group of proteins with similar transport functions. The packing constraints of the core structure within the membrane probably limit amino acid variation in the hydrophobic segments, while the polar loops at the membrane surface which link the membrane-crossing strands would generally be subject to less constraint. A comparable situation is found in the structures of other families of homologous proteins. It is interesting that the hydrophobic segments themselves vary considerably in their extent of amino acid substitution. Some are almost totally conserved (such as residues 406–428, 459–476 and 738–798), while others contain many conservative substitutions, mainly of bulky hydrophobic residues (for example, residues 435–457, 491–513, 568–588, 604–624, 659–681, 706–724 and 838–850). When the amino acid substitutions are mapped on an α-helical net they appear not to be randomly distributed but tend to cluster in patches on the surface of the helices (Fig. 5). A reasonable structural interpretation of these observations is that the highly conserved segments are tightly packed within the interior of the protein and contact only other portions of polypeptide chain. Since the human red cell protein, at least, forms a tightly associated dimer in the membrane (reviewed by Jay & Cantley, 1986), these protein contacts are likely not only to be intramolecular but also at the sites of association of the monomers. The variable regions in the other hydrophobic segments may comprise the interface within the membrane between the exterior of the protein and the hydrophobic lipid domain. The higher frequency of substitutions at this surface may reflect either less stringent constraints in the interaction of the protein with the membrane lipid or differences in the lipid compositions of the plasma membranes of the cells from which the proteins are derived.

This work was supported by a grant from the Medical Research Council. We are grateful to Dr. N. Mohandas for making available the human reticuloctyte cDNA library and Professor L. Luzzatto and Dr. T. Cox for supplying human reticulocytes.

REFERENCES

Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. **68**, 1408–1412

Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3963–3965

Bjerrum, P. J. (1983) in Structure and Function of Membrane Proteins (Quagliarello, E. & Palmieri, F., eds.) pp. 107–115, Elsevier, Amsterdam

Brock, C. J. & Tanner, M. J. A. (1986) Biochem. J. **235**, 899–901

Brock, C. J., Tanner, M. J. A. & Kempf, C. (1983) Biochem. J. **213**, 577–586

Chirgwin, J. M., Przblya, A. E., Macdonald, R. J. & Rutter, W. J. (1979) Biochemistry **18**, 5294–5299

Clemens, M. J. (1984) in Transcription and Translation: A Practical Approach (Hames, B. D. & Higgins, S. J., eds.), pp. 211–230, IRL Press, Oxford

Conboy, J., Kan, Y. W., Shohet, S. B. & Mohandas, N. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 9512–9516

Cox, J. V. & Lazarides, E. (1988) Mol. Cell. Biol. **8**, 1327–1335

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

Foxwell, B. M. J. & Tanner, M. J. A. (1981) Biochem. J. **195**, 129–137

Jay, D. (1986) Biochemistry **25**, 554–556

Jay, D. & Cantley, L. (1986) Annu. Rev. Biochem. **55**, 511–538

Jennings, M. L. (1985) Annu. Rev. Physiol. **47**, 519–533

Jennings, M. L. & Adams, M. F. (1981) Biochemistry **20**, 7118–7122

Jennings, M. L. & Anderson, M. P. (1987) J. Biol. Chem. **262**, 1691–1697

Jennings, M. L. & Nicknish, J. S. (1984) Biochemistry **23**, 6432–6436

Jennings, M. L., Adams-Lackey, M. & Denney, G. H. (1984) J. Biol. Chem. **259**, 4652–4660

Jennings, M. L., Anderson, M. P. & Monaghan, R. (1986) J. Biol. Chem. **261**, 9002–9010

Kaul, R. K., Murthy, P. S. N., Reddy, A. G., Steck, T. L. & Kohler, H. (1983) J. Biol. Chem. **258**, 7981–7990

Kopito, R. R. & Lodish, H. F. (1985) Nature (London) **316**, 234–238

Kopito, R. R., Andersson, M. A. & Lodish, H. F. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 7149–7153

Kozak, M. (1984) Nucleic Acids Res. **12**, 857–872

Lieberman, D. M., Nattriss, M. & Reithmeier, R. A. F. (1987) Biochim. Biophys. Acta **903**, 37–47

Low, P. S. (1986) Biochim. Biophys. Acta **864**, 145–167

Macara, I. G., Kuo, S. & Cantley, L. C. (1983) J. Biol. Chem. **258**, 1785–1792

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, pp. 202–203, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Matsuyama, H., Kawano, Y. & Hamasaki, N. (1986) J. Biochem. (Tokyo) **99**, 495–501

Mawby, W. J. & Findlay, J. B. C. (1982) Biochem. J. **205**, 465–475

Passow, H. (1986) Rev. Physiol. Biochem. Pharmacol. **103**, 62–203

Ramjeesingh, M., Gaarn, A. & Rothstein, A. (1983) Biochim. Biophys. Acta **729**, 150–160

Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467

Solomon, A. K., Chasan, B., Dix, J. A., Lukacovic, M. F., Toon, M. R. & Verkman, A. S. (1983) Ann. N.Y. Acad. Sci. **414**, 97–124

Tabor, S. & Richardson, C. C. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 4767–4771

Tanner, M. J. A., Williams, D. G. & Kyle, D. (1979) Biochem. J. **183**, 417–427

Weith, J. O., Andersen, O. S., Brahm, J., Bjerrum, P. J. & Borders, C. L. (1982) Phil. Trans. R. Soc. Lond. B. **299**, 383–399

Wilbur, W. J. & Lipman, D. J. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 726–730

Received 23 June 1988; accepted 1 August 1988