

Minireview

Structure–function relationships in the receptor for urokinase-type plasminogen activator

Comparison to other members of the Ly-6 family and snake venom α -neurotoxinsMichael Ploug^{a,*}, Vincent Ellis^{a,b}^a*Finsen Laboratory, Rigshospitalet, Strandboulevarden 49, 2100 Copenhagen Ø, Denmark*^b*Thrombosis Research Institute, Emmanuel Kaye Building, Manresa Road, Chelsea, London, SW3 6LR, UK*

Received 25 May 1994; revised version received 20 June 1994

Abstract

Plasminogen activation is regulated by the interaction between urokinase-type plasminogen activator (uPA) and its specific glycolipid-anchored cell surface receptor (uPAR). uPAR is composed of three homologous domains and is the only multi-domain member of the Ly-6 family of glycolipid-anchored membrane proteins. Recent evidence has highlighted similarities between the individual domains of uPAR and the large family of secreted, single domain snake venom α -neurotoxins, suggesting that uPAR may adopt the same gross folding pattern as these structurally well characterized proteins. Structural aspects of the binding between α -neurotoxins and the acetylcholine receptor may have a major influence on future studies of the interaction between uPA and uPAR.

Key words: uPA; uPAR; Bungarotoxin; Glycosyl-phosphatidylinositol; CD59; MIRL

1. Introduction

Specific binding sites for urokinase-type plasminogen activator (uPA) are present on many cells, including neutrophils, monocytes, migrating keratinocytes and certain cancer cells [1–4]. The cellular binding is of high affinity ($K_d \approx 0.1$ – 1 nM), and involves the interaction between a single, specific membrane protein (uPAR) and the NH_2 -terminal growth factor-like module of uPA [5,6]. uPAR can bind both active two-chain uPA and its single-chain proenzyme (pro-uPA) [7], and receptor-bound uPA can be inactivated by its specific inhibitors, the serpins PAI-1 and PAI-2 [8,9].

The primary function of uPAR is to strictly confine uPA-catalyzed plasminogen activation to the cell surface. This is achieved not only by the high-affinity binding but also by the favourable kinetics of the individual reactions in this reciprocal zymogen activation system which lead to a greatly enhanced generation of cell-associated plasmin [10,11]. These mechanisms are considered to play a major role in mediating the controlled break-

down of extracellular matrix during cell migration and invasion in many conditions, including cancer invasion and metastasis (as reviewed in [12]).

The purpose of this minireview is to summarize recent structural data concerning the multi-domain organisation of uPAR and to relate these to the high-affinity interaction between uPA and uPAR.

2. Primary structure and membrane attachment of uPAR

The specific binding of uPA to cells was first observed with human monocytes and the monocyte-like cell line U937 [1]. The membrane protein responsible for this binding was later purified and characterized from phorbol ester stimulated U937 cells [13,14] and was designated as the uPA receptor (uPAR) [15]. It is a heterogeneously glycosylated, single-chain polypeptide of $M_r \approx 50$ – $60,000$ which decreases to $35,000$ upon deglycosylation. Sequencing of human uPAR cDNA revealed that the protein is encoded as a 335 residue polypeptide of which the first 22 amino acids constitute the signal peptide [16]. The nascent uPAR contains five potential glycosylation sites for N-linked carbohydrate (Asn-Xaa-Thr/Ser). Murine, bovine and rat uPAR cDNAs with similar characteristics have also been sequenced [17–19].

Despite being an integral membrane protein there is no obvious transmembrane segment in the cDNA

*Corresponding author. Fax: (45) 31 38 54 50.

Abbreviations: uPA, urokinase-type plasminogen activator; uPAR, urokinase-type plasminogen activator receptor PAI, plasminogen activator inhibitor.

derived sequence of uPAR. The reason for this apparent discrepancy is that uPAR is anchored to the plasma membrane by a glycolipid (glycosyl-phosphatidylinositol) moiety which is added concurrently to the posttranslational removal of a COOH-terminal signal sequence [20]. Amino acid composition [20] and mutational [21] analyses indicate that this processing is most likely to occur at Gly²⁸³ of human uPAR, generating a mature protein composed of residues 1–283. As a direct consequence of its glycolipid anchorage uPAR is absent from the cell surface of peripheral blood leukocytes affected by the haemopoietic stem cell disorder paroxysmal nocturnal haemoglobinuria [2], the protein being secreted from these cells in a truncated form [22]. The possible biological significance of the glycolipid-anchorage of uPAR has been reviewed elsewhere [23].

3. Internal sequence homologies in uPAR

The entire sequence of mature uPAR is composed of three repeats of approximately 90 residues. The consensus sequence derived from these repeats is primarily defined by a conserved pattern of cysteine residues, which constitute 10% of the protein, but also includes specified gap regions [24]. These repeats appear to be rather distantly related as their sequence identity is less than 20%

within human, murine or bovine uPAR, whereas the interspecies conservation of the individual repeats is greater than 60% [18]. This primary sequence data led to the assumption that these repeats are autonomous structural entities and a three domain model for uPAR was proposed [23,24]. The validity of this model has been supported by several independent observations. Firstly, limited proteolysis of uPAR preferentially leads to cleavage of the polypeptide connecting the first and second repeats (Fig. 1), a property typical of an interdomain linker region. Secondly, studies on the gene structure show that the protein repeats of uPAR are encoded by symmetrical exon sets flanked by intron/exon boundaries of the same phase i.e. type-1 [25]. Thirdly, as discussed in detail below, single domain proteins homologous to the individual repeats of uPAR, i.e. conforming to the consensus sequence that defines them, have been identified.

4. Relationship between uPAR and the Ly-6 gene family

Protein database searching using the consensus sequence of the putative domains of uPAR revealed homology to a diverse group of single domain glycoproteins [24]; see Fig. 2, upper panel. This group includes a gene family of murine leukocyte antigens collectively

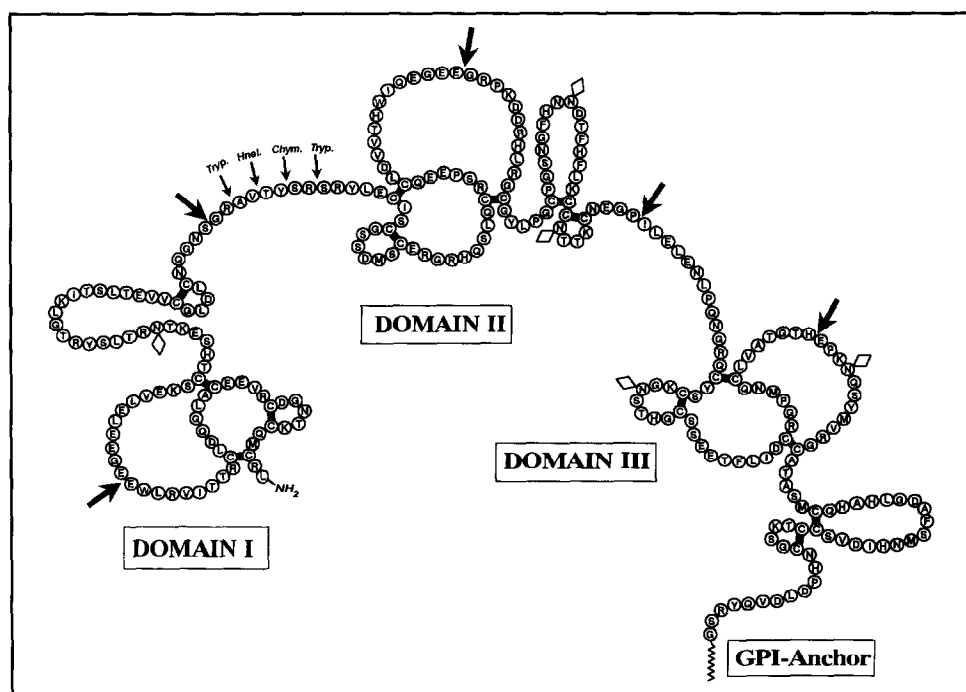


Fig. 1. Primary structure of uPAR. The primary sequence of human uPAR is shown as encircled amino acids in the single letter code. Circles joined by a black bar represent disulphide bonded cysteine residues (note that only the disulphide bonds of the NH₂-terminal domain have been determined experimentally [30]; the remaining disulphide bonds have been located according to the consensus shown in Fig. 2). Diamonds represent potential attachment sites for N-linked carbohydrate. The large arrows indicate positions corresponding to the presence of introns in the uPAR gene [25]. The smaller arrows identify peptide bonds that are extremely susceptible to proteolysis in the native non-denatured protein; - Chym., chymotrypsin [24]; Hnel., human neutrophil elastase; and Tryp., trypsin (cleavage sites were identified by laser desorption mass spectrometry; Rahbek-Nielsen and Ploug, unpublished data).

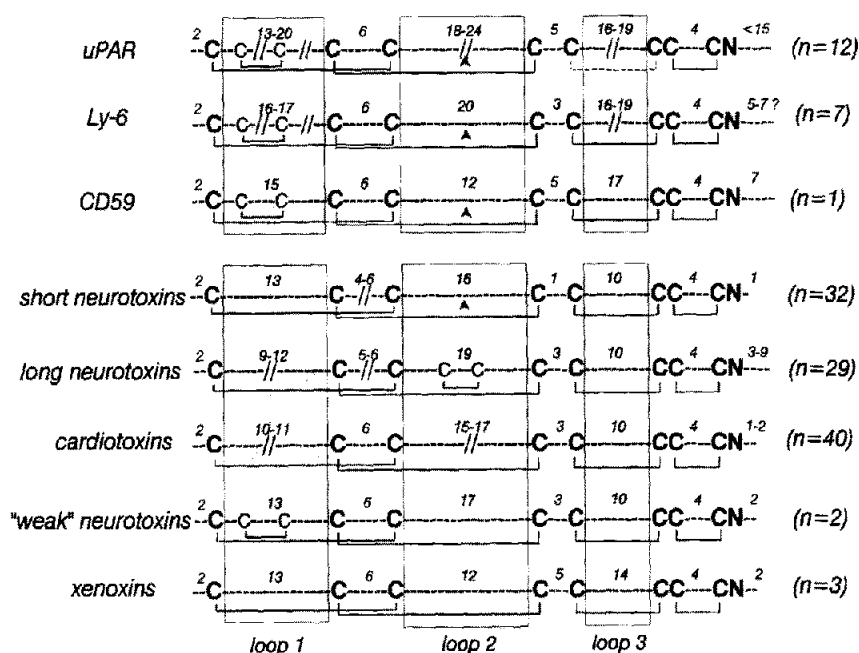


Fig. 2. Sequence alignment of the uPAR/Ly-6 family and the snake venom neurotoxins. The upper panel shows an alignment of the individual domains from the uPAR/Ly-6 family of glycolipid-anchored membrane proteins. The lower panel shows an equivalent alignment for the secreted single domain proteins belonging to the various subgroups and homologs of the snake neurotoxin family. Distances between the conserved cysteines are shown, highlighting that the variability both between and within the individual groups is essentially confined to the loop regions (covered by the shaded boxes). The number of sequences defining each group is shown in brackets (the uPAR domains are from four different species). Additional, non-conserved cysteine pairs found in all members of the uPAR/Ly-6 family, two 'weak' neurotoxins and the long neurotoxins are situated outside of the globular core, within the long loops in the snake neurotoxin model. It should be noted that the NH₂-terminal domain of uPAR is uniquely lacking a cysteine pair conserved in all other proteins represented in this alignment (shown as a dashed line). The arrowheads denote the positions of the single intradomain intron dividing the exons encoding the consensus sequence where known.

known as Ly-6 and the regulatory complement component CD59 (MIRL, membrane inhibitor of reactive lysis). Like uPAR these are all glycolipid-anchored membrane proteins and have gene organizations comparable to those segments of the uPAR gene encoding the individual domains [26,27]. Of these proteins only CD59 has been studied at the structural level, an incomplete disulphide structure consistent with that of the NH₂-terminal domain of uPAR [28] and partial ¹H-NMR assignments [29] having recently been reported.

5. Similarity to snake venom α -neurotoxins – a possible template for the gross folding of uPAR

Knowledge of the disulphide structure of the NH₂-terminal domain of uPAR [30] prompted us to re-search the protein data base for homologies using a shorter consensus sequence or motif. The sequence C C X X X X C N was chosen as (a) the asparagine is the only invariant non-cysteine residue in the uPAR/Ly-6 consensus, and (b) the motif forms a short loop structure as its second and third cysteines are disulphide bonded. The search identified the large family of snake venom α -neurotoxins, exemplified by the acetylcholine receptor

antagonist α -bungarotoxin [31–33]. These secreted proteins are of similar size to the uPAR/Ly-6 domains, and most importantly have eight cysteine residues that are comparably spaced and identically paired to those of uPAR (Fig. 2). A similar relationship between Ly-6 and the snake neurotoxins has been independently proposed by others based on linear alignment of sequences [34]. It should be pointed out that this relationship was not identified in the searches which originally revealed the uPAR/Ly-6 relationship as the neurotoxins lack a specific disulphide pairing which is conserved in all members of the Ly-6 family (Fig. 2). Further support for a relationship between these two groups of proteins again comes from the similarity in the structure of the uPAR, Ly-6 and CD59 genes, and that of the only snake toxin analyzed at this level, erabutoxin-c [35]. These similarities may be taken to suggest that these proteins share a similar three-dimensional folding motif.

Fortuitously the snake neurotoxin family has been extremely well studied at the structural level, with over 15 known X-ray and NMR structures. Their overall folding consists of three adjacent loops participating in a flat triple-stranded antiparallel β -sheet, emerging from a small globular core stabilized by the four conserved disulphide bonds. Consistent with this folding the CD spec-

trum of recombinantly expressed uPAR (Casas-Finet and Ploug, unpublished data) is comparable to those of the α -neurotoxins which show essentially only β -structure [32,33]. Recently a gross folding topology of the Ly-6 family member CD59 has been proposed based on almost complete ^1H -NMR assignments. This was found to be compatible with an α -neurotoxin-like structure, particularly in terms of β -sheet formation [29]**.

Fig. 3 shows the primary sequence of the NH_2 -terminal domain of uPAR superimposed onto an α -neurotoxin model structure which has been achieved with only minor modifications. It is apparent from the linear alignment in Fig. 2 and the model in Fig. 3 that the modifications that are necessary occur outside of the cysteine-rich globular core. The distal parts of each of the three major loops coincide with gap regions originally defined in the uPAR consensus sequence and in general such gap regions are predicted to occur in exposed surface loops [36]. The additional non-conserved pair of cysteines present in all uPAR domains and the Ly-6 family fall within loop 1 of the model. A comparable non-conserved cysteine pair also occurs in loop 2 of the long neurotoxins as well as in loop 1 of two 'weak' neurotoxins (Fig. 2).

This model, whilst providing compelling evidence for a structural relationship between these proteins, reveals some striking differences between uPAR and all of the other proteins, which may be related to the unique multi-domain structure of uPAR. The model shows that the consensus motif used to detect the homology to the snake neurotoxins forms a short loop within the globular core of the protein. In the neurotoxins residues within this loop are involved in interactions considered important for the structural maintenance of the protein, the side chain of the invariant asparagine being engaged in several hydrogen bonds that anchor the loop to the globular core [37,38] and the highly conserved third position of the loop motif C C X X X C N (X being either Asp (85%) or Asn/Glu (15%)) being able to form a salt-bridge (or hydrogen bond) to the terminal α -amino group [31,32]. This position is also conserved in the NH_2 -terminal domain of uPAR and all the single domain Ly-6 family proteins. However due to its multi-domain structure the potential formation of a similar salt-bridge is precluded in the second and third domains of uPAR. It is therefore perhaps not surprising, in light of the proposed model structure, that the 'carboxyl terminal' sequences of these two domains of uPAR do not conform to the archetypal C C X X D X C N consensus motif, but have the aspartic acid replaced by the non-conservative substitutions, threonine and serine.

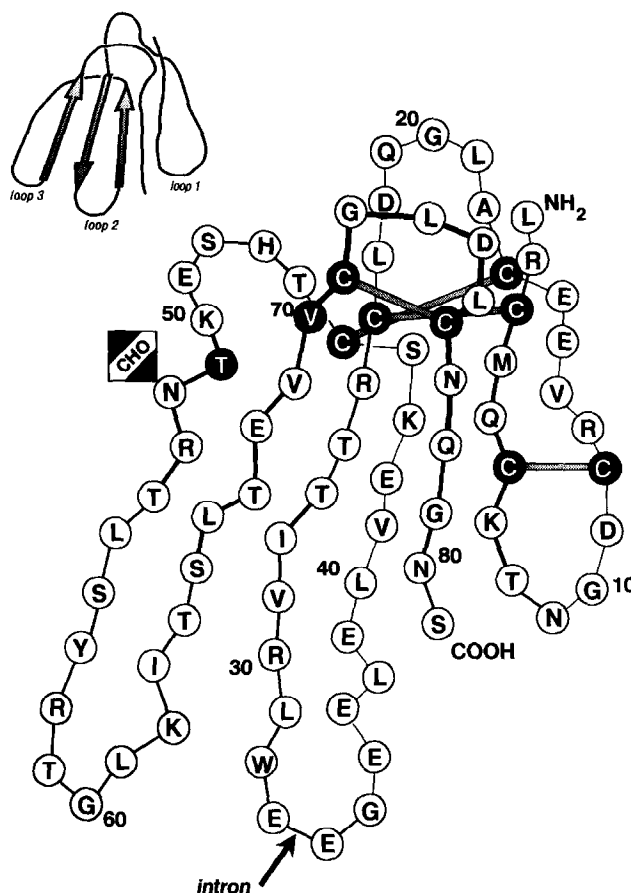


Fig. 3. The NH_2 -terminal sequence of uPAR superimposed onto a simplified folding model of the snake neurotoxins. The amino acid sequence of the NH_2 -terminal domain of uPAR (as encoded by the first exon-set in the gene) is 'threaded' onto a representative α -neurotoxin backbone [31]. Cysteine residues are highlighted, as are Thr⁵¹ and Val⁷⁰ the predicted positions of the cysteine pairing uniquely lacking in this domain of uPAR. The arrow denotes the position of the single intradomain intron. The inset shows the major β -sheet structure in a simplified snake neurotoxin model.

Another major, and more surprising, difference between uPAR and the other single domain proteins is that the NH_2 -terminal domain of uPAR lacks one of the cysteine pairs that is strictly conserved in all of the Ly-6 family, the snake neurotoxins and the second and third domains of uPAR. The significance of this is unclear but it may be related to the unique multi-domain structure of uPAR which, as discussed in the following section, may be of critical importance for its function.

6. Structural aspects of the uPA/uPAR interaction

The uPAR binding-site of uPA has been localized to the growth factor-like module of this mosaic protein [6] which also contains a kringle module and a serine protease domain. The growth factor-like module of uPA (residues 4–43) retains the high affinity binding to uPAR

**Subsequent to submission of this review two almost identical solution structures of CD59 determined by NMR have been independently reported [49,50]. Both structures conform well to those of the α -neurotoxins, the major difference being the presence of a short helical element in the outer part of loop 3 in CD59.

when isolated from the rest of the molecule [39]. The recently solved solution structure of the entire NH₂-terminal fragment of uPA (i.e. the growth factor-like module and the kringle module) provides an explanation for this as these two domains were shown to be structurally independent with no interdomain interactions [40]. In contrast to this the multi-domain structure of uPAR appears to be essential for its high-affinity interaction with uPA. Ligand binding involves the NH₂-terminal domain of uPAR [24], but proteolytic liberation of this domain has recently been shown to cause a dramatic decrease in its binding affinity both for uPA and the low molecular weight fluorophore 8-anilino-1-naphthalene sulphonate (ANS) [41]. This has been interpreted to suggest that interdomain interactions within uPAR play a crucial role in ligand binding, either by stabilizing a discrete conformation of the NH₂-terminal domain or being directly involved in the ligand binding interaction. Two interesting observations appertain to this. Firstly, intermolecular domain associations have been found in the crystal structures of α -cobratoxin [37] and cardiotoxin V₄^{II} [38], and also in the solution structure of κ -bungarotoxin [42]. These toxins dimerize by intermolecular hydrogen bonding of the β -strand in the loop 3 regions, generating 6-stranded antiparallel β -sheets. In κ -bungarotoxin this dimerization is thought to generate the physiologically active species of the toxin [43]. Secondly, as emphasized previously, the NH₂-terminal domain of uPAR uniquely lacks one of the cysteine pairs that is strictly conserved in the other proteins. This pair of cysteines are situated within the globular core at a position corresponding to the end of the β -strand in loop 3 in the neurotoxin structure, where they may be expected to be intimately involved in the maintenance of this structure. The lack of these cysteines may be expected to lead to a conformational instability in the isolated NH₂-terminal domain of uPAR, which may be restricted by interdomain interactions in the intact protein.

7. Perspectives

The structural similarity between uPAR and the snake venom α -neurotoxins addressed in this review may prove to be useful not only in elucidating the structure of intact three domain uPAR, but may also give important leads as to structural aspects of the interaction between uPA and uPAR. Detailed structure-function analyses of α -neurotoxins by chemical modification, site-directed mutagenesis, X-ray crystallography and NMR [44–47] suggest a multipoint interaction between these toxins and the acetylcholine receptor, involving several locally separated structural elements within the neurotoxin. This has been verified by a recent solution structure solved for a complex of α -bungarotoxin and a synthetic peptide derived from the acetylcholine receptor [48]. A similar mul-

tipoint attachment is likely to occur between the growth factor-like module of uPA and the NH₂-terminal domain of uPAR.

Acknowledgements: We would like to thank Dr. Jose Casas-Finet (Structural Biochemistry, NCI-Fredricks, MD) for the CD spectrum of uPAR, Dr. Henrik Rahbek-Nielsen (Molecular Biology, Odense University, Denmark) for the matrix-assisted laser desorption ionization mass spectrometry analyses, and Dr. Charles Pyke (Finsen Laboratory, Denmark) for originally suggesting the relationship between uPAR and the snake neurotoxins. This work was supported financially by the Danish Cancer Society.

References

- [1] Vassalli, J.D., Baccini, D. and Belin, D. (1985) *J. Cell Biol.* 100, 86–92.
- [2] Ploug, M., Plesner, T., Rønne, E., Ellis, V., Høyer-Hansen, G. and Danø, K. (1992) *Blood* 79, 1447–1455.
- [3] Rømer, J., Lund, L.R., Eriksen, J., Pyke, C., Kristensen, P. and Danø, K. (1994) *J. Invest. Dermatol.* 102, 519–522.
- [4] Pyke, C., Kristensen, P., Ralfkiær, E., Grøndahl-Hansen, J., Eriksen, J., Blasi, F. and Danø, K. (1991) *Am. J. Pathol.* 138, 1059–1067.
- [5] Stoppelli, M.P., Corti, A., Soffientini, A., Cassani, G., Blasi, F. and Associan, R.K. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4939–4943.
- [6] Appella, E., Robinson, E.A., Ulrich, S.J., Stoppelli, M.P., Corti, A., Cassani, G. and Blasi, F. (1987) *J. Biol. Chem.* 262, 4437–4440.
- [7] Cubellis, M.V., Noll, M.L., Cassani, G. and Blasi, F. (1986) *J. Biol. Chem.* 261, 15819–15822.
- [8] Cubellis, M.V., Andreasen, P.A., Ragno, P., Mayer, M., Danø, K. and Blasi, F. (1989) *Proc. Natl. Acad. Sci. USA* 86, 4828–4834.
- [9] Ellis, V., Wun, T.-C., Behrendt, N., Rønne, E. and Danø, K. (1990) *J. Biol. Chem.* 265, 9904–9908.
- [10] Ellis, V., Scully, M.F. and Kakkar, V.V. (1989) *J. Biol. Chem.* 264, 2185–2188.
- [11] Ellis, V., Behrendt, N. and Danø, K. (1991) *J. Biol. Chem.* 266, 12752–12758.
- [12] Ellis, V., Ploug, M., Plesner, T. and Danø, K. (1994) in: *Fibrinolysis in Disease*, CRC Press, Boca Raton, FL, in press.
- [13] Nielsen, L.S., Kellerman, G., Behrendt, N., Picone, R., Danø, K. and Blasi, F. (1988) *J. Biol. Chem.* 263, 2358–2363.
- [14] Behrendt, N., Rønne, E., Ploug, M., Petri, T., Løber, D., Nielsen, L.S., Schleuning, W.-D., Blasi, F., Appella, E. and Danø, K. (1990) *J. Biol. Chem.* 265, 6453–6460.
- [15] Blasi, F., Vassalli, J.-D. and Danø, K. (1987) *J. Cell Biol.* 104, 801–804.
- [16] Roldan, A.L., Cubellis, M.V., Masucci, M.T., Behrendt, N., Lund, L., Danø, K., Appella, E. and Blasi, F. (1990) *EMBO J.* 9, 467–474.
- [17] Kristensen, P., Eriksen, J., Blasi, F. and Danø, K. (1991) *J. Cell Biol.* 115, 1763–1771.
- [18] Krättschmar, J., Haendler, B., Kojima, S., Rifkin, D.B. and Schleuning, W.-D. (1993) *Gene* 125, 177–183.
- [19] Rabbani, S.A., Rajwans, N., Achbarou, A., Murthy, K.K. and Goltzman, G. (1994) *FEBS Lett.* 338, 69–74.
- [20] Ploug, M., Rønne, E., Behrendt, N., Jensen, A.L., Blasi, F. and Danø, K. (1991) *J. Biol. Chem.* 266, 1926–1933.
- [21] Møller, L.B., Ploug, M. and Blasi, F. (1992) *Eur. J. Biochem.* 208, 493–500.
- [22] Ploug, M., Eriksen, J., Plesner, T., Hansen, N.E. and Danø, K. (1992) *Eur. J. Biochem.* 208, 397–404.
- [23] Ploug, M., Behrendt, N., Løber, D. and Danø, K. (1991) *Semin. Thromb. Hemostasis* 17, 183–193.

- [24] Behrendt, N., Ploug, M., Patthy, L., Houen, G., Blasi, F. and Danø, K. (1991) *J. Biol. Chem.* 266, 7842–7847.
- [25] Casey, J.R., Petranka, J.G., Kottra, J., Fleenor, D.E. and Rosse, W.F. (1994) *Blood*, in press.
- [26] Petranka, J.G., Fleenor, D.E., Sykes, K., Kaufman, R.E. and Rosse, W.F. (1992) *Proc. Natl. Acad. Sci. USA* 89, 7876–7879.
- [27] Bothwell, A., Pace, P.E. and LeClair, K.P. (1988) *J. Immunol.* 140, 2815–2820.
- [28] Sugita, Y., Nakano, Y., Oda, E., Noda, K., Tobe, T., Miura, N.-H. and Tomita, M. (1993) *J. Biochem.* 114, 473–477.
- [29] Fletcher, C.M., Harrison, R.A., Lachmann, P.J. and Neuhaus, D. (1993) *Protein Science* 2, 2015–2027.
- [30] Ploug, M., Kjalke, M., Rønne, E., Weidle, U., Høyer-Hansen, G. and Danø, K. (1993) *J. Biol. Chem.* 268, 17539–17546.
- [31] Dufton, M.J. and Hider, R.C. (1983) *CRC Crit. Rev. Biochem.* 14, 113–171.
- [32] Endo, T. and Tamiya, N. (1991) in: *Snake Toxins* (Harvey, A.L., Ed.) pp. 259–302, Pergamon Press, New York.
- [33] Rees, B. and Bilwes, A. (1993) *Chem. Res. Toxicol.* 6, 385–406.
- [34] Fleming, T.J., O'hUigin, C. and Malek, T.R. (1993) *J. Immunol.* 150, 5379–5390.
- [35] Fuse, N., Tsuchiya, T., Nonomura, Y., Menez, A. and Tamiya, T. (1990) *Eur. J. Biochem.* 193, 629–633.
- [36] Patthy, L. (1989) *Acta Biochim. Biophys. Hung.* 42, 3–13.
- [37] Betzel, C., Lange, G., Pal, G.P., Wilson, K., Maelicke, A. and Saenger, W. (1991) *J. Biol. Chem.* 266, 21530–21536.
- [38] Rees, B., Bilwes, A., Saman, J.P. and Moras, D. (1990) *J. Mol. Biol.* 241, 281–297.
- [39] Mazar, A.P., Buko, A., Petros, A.M., Barnathan, E.S. and Henkin, J. (1992) *Fibrinolysis* 6, Suppl 1, 49–55.
- [40] Hansen, A.P., Petros, A.M., Meadows, R.P., Nettesheim, D.G., Mazar, A.P., Olejniczak, E.T., Xu, R.X., Pederson, T.M., Henkin, J. and Fesik, S.W. (1994) *Biochemistry* 33, 4847–4864.
- [41] Ploug, M., Ellis, V. and Danø, K. (1994) *Biochemistry*, in press.
- [42] Oswald, R.E., Sutcliffe, M.J., Bamberger, M., Loring, R.H., Braswell, E. and Dobson, C.M. (1991) *Biochemistry* 30, 4901–4909.
- [43] Chiappinelli, V.A. (1991) in: *Snake Toxins* (Harvey, A.L., Ed.) pp. 223–258, Pergamon Press, New York.
- [44] Pillet, L., Trémeau, O., Ducancel, F., Prevet, P., Zinn-Justin, S., Pinkasfeld, S., Boulain, J.-C. and Menez, A. (1993) *J. Biol. Chem.* 268, 909–916.
- [45] Mori, N. and Tu, A.T. (1991) *J. Pharm. Pharmacol.* 43, 785–789.
- [46] Zinn-Justin, S., Roumestrand, C., Prevet, P., Menez, A. and Toma, F. (1993) *Biochemistry* 32, 6884–6891.
- [47] Martin, B.M., Chipper, B.A. and Maelicke, A. (1983) *J. Biol. Chem.* 258, 8714–8722.
- [48] Basus, V.J., Song, G. and Hawrot, E. (1993) *Biochemistry* 32, 12290–12298.
- [49] Kieffer, B., Driscoll, P.C., Campbell, I.D., Willis, A.C., van der Merwe, P.A. and Davies, S.J. (1994) *Biochemistry* 33, 4471–4482.
- [50] Fletcher, C.M., Harrison, R.A., Lachmann, P.J. and Neuhaus, D. (1994) *Structure* 2, 185–199.