

Identification of a member of mouse semaphorin family

S. Inagaki^{a,*}, T. Furuyama^b, Y. Iwahashi^{a,c}

^aDepartment of Laboratory Sciences, School of Allied Health Sciences, Osaka University Faculty of Medicine, 2-2 Yamadaoka, Suita-shi, Osaka 565, Japan

^bSecond Department of Anatomy, ^cDepartment of Ophthalmology, Osaka University Faculty of Medicine, 2-2 Yamadaoka, Suita-shi, Osaka 565, Japan

Received 12 June 1995; revised version received 1 July 1995

Abstract Grasshopper semaphorin I (Sema I) and its related proteins, chick collapsin and mouse Sema III contribute to the axon guidance by their repellent actions [5,9,12]. We have identified a member of semaphorin gene family from the mouse brain and named it M-Sema F. The N-terminal encodes a semaphorin domain that is similar between Sema I–III [6] followed by a single putative immunoglobulin-like domain, a transmembrane domain, and a proline-rich intracellular domain. M-Sema F mRNA is expressed widely in the nervous tissues during development. These suggest that M-Sema F is a transmembrane member of the semaphorin family of the vertebrate which may function in the developing neuronal network.

Key words: Semaphorin; Collapsin; Axon guidance; Cloning; Transmembrane; Mouse

1. Introduction

Neural networks which are very complicated but specific to each neuron are formed during development when growth cones make specific pathway choices and find their correct targets using a variety of guidance molecules in its surroundings. Extension-promoting molecules such as chemoattractants and adhesion molecules have long been considered to control the direction of growth cone extension [4,10]. Recently, evidence has been obtained for repulsive signals which can direct growth cone extension [2,10]. Collapsin purified from chick brain causes the collapse of the growth cones of dorsal root ganglion neurons in vitro [9]. Luo et al. [9] have shown that a domain of about 500 amino acids of collapsin is highly homologous to grasshopper Fasciclin IV (Fas IV), later renamed G-Sema I, which is required for the proper guidance and fasciculation of the T11 growth cones in the limb bud of the grasshopper embryo [5]. The domain is called semaphorin (sema) domain. The semaphorin genes having the sema domain form a family of transmembrane and secreted molecules which appear to function during growth cone guidance [5,6,9,12]. Two members of the semaphorin gene family have been identified from *Drosophila* [6]. *Drosophila* Sema I (D-Sema I) is predicted to be a transmembrane protein and D-Sema II to be a secreted protein. In the vertebrate, human and murine Sema IIIs are likely to be the homologues of chick collapsin and so far, only secreted semaphorin members have been identified [6,9,12].

In the present study, we have used a polymerase chain reaction (PCR) to clone a novel semaphorin gene in mice that is

named mouse-semaphorin F (M-Sema F). This gene has a putative transmembrane region and most closely related to murine semaphorin C (Sem C) of a new subfamily, semaphorin IV (Sema IV) identified recently by Puschel et al. [12].

2. Materials and methods

2.1. Cloning of M-Sema F

PCR [13] was performed using Taq polymerase (Pharmacia), cDNA to poly(A)⁺ RNA from neonatal mouse brain and partially degenerate oligonucleotides corresponding to collapsin amino acids FIGTD and DPYCA. The PCR cycling condition was as follows: 94°C for 3 min followed by 35 cycles of 94°C for 1 min, 52°C for 1 min, and 72°C for 2 min and finally 72°C for 10 min. The reaction products (260 bp) were cloned into the pCRscript (Stratagene) and sequenced by using dyedexy terminator cycle sequencing (Perkin Elmer). A 260 b PCR product from one clone, chosen from several clones, was used to screen 1×10^6 clones from whole body of embryonic day 10. cDNA were subcloned into the plasmid pBluescript KS (Stratagene). Analysis of the largest insert (clone E2), about 2 kb, showed that it did not contain a complete open reading frame. With the same PCR fragment, another round of screening was done from mouse neonatal whole brain cDNA libraries (ICR outbred strain, Stratagene). Of several clones, the largest 3.5 kb insert (clone E51) was analysed. This encoded the entire open reading frame. Nucleotide sequence was determined by dye primer cycle sequencing and dyedexy terminator cycle sequencing (Perkin Elmer).

2.2. Northern analysis of the expression of M-Sema F

Mouse embryos were obtained from timed pregnant mouse (the presence of sperm in the vagina equivalent to E0.5). Total RNA was isolated from the head and body of embryonic day 14 (E14) mice, brain, liver and kidney of postnatal day 1 (P1) mice, and adult brain following to the procedures [1]. 30 µg of total RNA from each tissue was fractionated on a 1% agarose-formaldehyde gel and transferred to nylon membrane (Hybond N, Amersham). The membrane was hybridized overnight at 42°C in 50% formamide, 5 × SSPE, 5 × Denhardt's solution, 20 µg/ml salmon sperm DNA and 5×10^5 cpm/ml probe. 2 kb cDNA from clone E2 was labeled with [α -³²P]dCTP (NEN) by random priming. Following hybridization, the filter was washed to a final stringency of 0.1 × SSPE and 0.1% SDS at 50°C. The washed filter was exposed to X-ray film (Kodak XAR-5) with an intensifier for 48 h.

2.3. In situ hybridization

The whole body of E15.5 and E16.5, the head of P1, and the brain of adult mice were dissected under pentobarbital anesthesia (50 mg/kg). Tissues were frozen on dry-ice powder, sectioned on a cryostat at −15°C (15 µm thickness) and processed to in situ hybridization procedure with anti-sense and sense RNA probes labeled with [³⁵S]UTP corresponding to the 600 bp fragment of 5'-terminal of M-sema F from clone E51 by T3 and T7 RNA polymerase. Tissue sections were fixed in 10% formalin in phosphate buffered saline (PBS) for 25 min, rinsed twice in PBS, and rinsed in 50 mM Tris-HCl and 5 mM EDTA. Sections were digested with 1 µg/ml Proteinase K for 20 min at room temperature, washed in PBS, refixed in 10% formalin in PBS, and washed in PBS. The sections were treated with 0.1 M triethanolamine and acetic anhydride for 10 min, washed in PBS, and dehydrated in ethanol. Then the sections were incubated with 50% formamide containing hybridization buffer with 1×10^6 cpm/section of either sense or antisense riboprobe, incubated overnight at 55°C in a humid chamber, washed for 15 min in 5 × SSC and 1% 2-mercaptoethanol (ME) at 55°C, for 30 min

*Corresponding author. Fax: (81) (6) 879-2619.

Abbreviations: Sema I–III, semaphorin I–III; Fas IV, fasciclin IV.

in 2 × SSC, 50% formamide and 5% ME at 65°C, and three times for 10 min each in the RNase buffer containing 10 mM Tris HCl, 1 mM EDTA and 0.5 M NaCl at 37°C, and then treated for 30 min with

RNase A (1 mg/ml) at 37°C. Sections were washed for 30 min in 2 × SSC, 50% formamide and 5% ME at 65°C, and for 10 min each in 2 × SSC and 0.1 × SSC at room temperature. Slides were dehydrated in ethanol, exposed to X-ray film, and processed for microautoradiogram using the emulsion (Ilford K-5) as described previously [3].

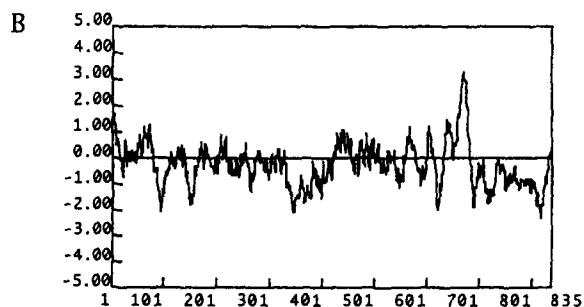
3. Results

3.1. Cloning and structural features of M-Sema F

The 3.5 kb insert contained a single open reading frame starting with an ATG start codon and ending with a TGA stop codon at nucleotide 2504 (Fig. 1). The open reading frame predicts a protein of 834 amino acids. The start codon is a strong translation initiator by sequence context criteria [7]. The stop codon is followed by a 860 nt untranslated sequence ending with a poly(A) tail that is preceded by a consensus polyadenylation signal AATAAA. Hydrophobicity analysis of the predicted amino acid sequence [8] revealed two stretches of hydrophobic sequence, one is located at the N-terminus of M-Sema F that could function as a signal peptide [14], and the other is located at 663–687 residues that could serve as a potential transmembrane-spanning domain. Thus, the deduced protein has an extracellular domain of 641 amino acids, a transmembrane domain, and cytoplasmic domain of 147 amino acids. The deduced protein sequence contains five potential sites for N-glycosylation in the extracellular domain and 13 cysteine residues which are conserved between semaphorin domains of all vertebrate Sema IIIs, and invertebrate Sema I and II (Fig. 1). The sema domain (amino acid residues 45–551) is 45% identical with Sem C, 39% identical with Sem B, 37% identical with H-Sema III and 29% identical with G-Sema I (Fas IV). A single putative immunoglobulin-like domain of about 60 amino acids [15] is conserved between all vertebrate semaphorins and precedes to a putative transmembrane domain. A putative intracellular domain contains a proline-rich sequence.

3.2. Northern analysis of the expression of M-Sema F

Northern blot hybridization of total RNA prepared from E14 and P1 mouse tissues detected a single band of about 5 kb just beneath the 28S rRNA band (Fig. 2). M-Sema F mRNA was expressed in P1 brain and kidney, and in E14 head and



← Fig. 1. The nucleotide and predicted amino acid sequences of M-Sema F. (A) The positions of nucleotide and amino acid residues are given at the right. Thirteen conserved cysteine residues are marked by asterisks. Putative signal peptide and transmembrane domains are underlined. A dotted line indicates a single immunoglobulin-like domain. (B) A hydrophobicity plot for the predicted M-Sema F protein was generated by Kyte and Doolittle [8]. The values of hydrophobicity are plotted against the positions of the increasing hydrophobicity.

A

```

GGCAGGAGTGGCGGAGTCAAAACGGAGGGGAGCGCCAGGGATTGGAGTGCACGAAGAGGGCTGCTGGAGT  -24
AAGTTAGACCTGGGTGTCTGCCATGCCGCCCACTGGGTGTCTGGCTGGCGAGCAGGGCTGTGGGGCTG  51
      N A P H M A V W L L A A G L W G L  17
GGCAGCGGGCTGAGATGGTGGCACTGTGCCCGGAAGACAGTATCTCTGGGAGCTGGTCAACAGTGT  126
G I G A E M M W N L V P R K T V S S G E L V T V V  42
AGCGGGTCTCCCGACAGGATCCAGGACTCTGACACTGACCTGACAGACATCTGGCCTTTATATGTG  201
R R F S Q T G I Q D F L T L I L T E H S G L L V V  67
GGGCGCCGAGAGGGCGCTGTTCCTTCAGTGTAGAGCTCTGGAGCTGCAAGGAGCGATCTCTGGGAGGCTCCA  276
G A R E A L F A F S V E A L E L Q G A I S W E A P  92
GCTGAGAAGAAATGAATGATCCACGAAGGGAAGACCAACAGCAGTCTGCTCAACTCTACCGCTCTCT  351
A E K K I E C T Q K G K S N C G A E L F N F I R F L  117
CAGCATAACAATCTCCATCTGATCTCTGCGGACCTCTTCCAGGCTGACCTGACCTACCTACCATG  426
Q P Y N S H L Y V C G T Y A F Q P K C T Y I N M  142
CTCAGCTTACCTTGGACCTGCAGATTTGGAGTGGGAAGGTAAATGCCATATGACCCAGTAAAGGTGAC  501
L T F T L D R A E F E D G K G K C P Y D P A K G H  167
ACCGGACTCTTGTGACGGTGTACTGACGACATCAATAACTTCTGGGACAGAGCCGGTATCTCT  576
T G L L V D G E L Y S A T L M N F L G T E P V I L  192
CGATACATGGGACCCACCTCCATCAAGACAGATGATCTGGCTTTTGGTGAATGAACCCACTTTGTAGG  651
R Y M G T H H S I K T E Y F W L N E P H F V G  217
TCTGCTTTTCTCCCTGAGATGTGGAGCTTCCAGGAGACGATGACAGATCTACTCTCTTCAGTGGCGG  726
S A F Y P E S V G S F T G D D D K I Y F F F S E R  242
GCAGTGGAGTGTACTCTTCCAGGAGTGGTGGCTGTGGGAGAGCTGTAAAGGTGACATCGGGGGA  801
A V E Y D C Y S E Q V V A R V A R V C K G D M G G  267
GCACGGACCTGCAGAGAAATGGACGCTTCTGAAGGCTGGTGGTGTCTGACGCGCTGACTGAGAGTGC  876
A R T L Q K K W T T F L K A R L V C S A P D M K V  292
TACTTCAACAGCTGAAGGGGTGCAACCTCTGGGGGCTCTTGGGCAACACACCTCTCTCGGGGTTTT  951
Y F N Q L K A V Y C T G G G G A S W N N T T F G V F  317
CAAGCGCATGGGCGATGACCTGTCTGCACTTTGTGATACAGTGGACAGATCCAGCAAGTGTGTGAG  1026
Q A R W G D M D L S A V C E Y Q L E Q I Q Q V F E  342
GGTCTCAAGAGGTACAGTGAAGAGCCAGAGTGGGCGCTATCTGACCGGTACCCAGCCCTCGGCT  1101
G P Y K E Y S E Q A Q K W A R Y T D P V P S P R P  367
GGTCTGTATCAACACTGGCAGGAGCAATGGCTACACAGTCTCTGGAAGTCCGGGACCAACACCTCAAC  1176
G S C I N N M H R Y T S L E L P D N T L N  392
TTCTCAAGAGCAACCTGATGAGGAGCAGGTGAAGCTCGTGGGCGCCGCTCACTTGTGAAGAGAAC  1251
F I K K H P L M D Q V K P R L R P L L V K K N  417
ACTAATTCACACAGCTGGTGGCCGACAGGTGCCAGGCTGTGAGTGGTGGTGGTGGTGGTGGTGGT  1326
T N F T H V V A D R V P G L D G A T Y T V L F I G  442
ACAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  1401
T G D G W L L K A V S L G P M I H M V E E L Q V F  467
GATCAGGAGCTGAGGAGTCTGGTGTCTGACAGCAAGAGTGGTCTTGTGGTGGTGGTGGTGGTGGTGG  1476
D Q E P V E S L V L S Q S K K V L F A G S R S Q L  492
GTTCACTGTCTCTGGGAGCTGCACAAAGTACCTGTTCTGTAGAGTGTCTGCTGGCAGGAGCCCTTACT  1551
V Q L S L A D C T K Y R F C V D C V L A R D P Y C  517
GCTGGAAATGTCAACACAGCGCTGTGGGACCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG  1626
A M N V N T S R C V A T T S G R S G S F L V Q H V  542
CGAATCTGGACACTCAAGATGTGATCAGTATGGATTAAGAAAGTCAAGTCTTCCCAAGAACATCACC  1701
A N L D T S K M E N Y Q G I K Y V R S I P K H I T  567
GTTGTGACGACAGCTGGTCTCACTCCCTGACCTCTCTGCTCAATTTGGCCATGGCCACTGGAGCTTGG  1776
V S G T D L V L P C H L S N L A H A H W T F G  592
ACCGAGGACTGCTGGCAGAGCACTGGCTCTTTTATGACAGGAGTCCAGGCGCTGGTGGTATGGCC  1851
S Q D L P A E Q P G S F L Y D T G L Q A L V V M  617
GCACAGTCCCTGCTCTGAGCTCTGTTGATGATGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG  1926
A Q S R H S G P Y R C Y S E E Q G T R L A A E S Y  642
CTGTGCTGTGGTGGGCGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG  2001
L V A V V A G S V T L E A R A P L E N L G L V M  667
CTGCTGTGGTGGGCGCTGGGCGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG  2076
L A V A L G A V L L L L L L L L L L L L L L L L  692
GAGAGCTAGAAAAGGTGCCAAGGCTGTGAGGAGACCTGGTGTACCCCTTGGAACTGCCAAGGAGCTGCC  2151
E L E K G A K A S E R T L V Y P L E L P K E P A  717
AGTCCCGCTCTGCTCTGGGCGGCAAACTGATGAGAACTTGGGATCTGTGGGTACTACTATTGGATGGC  2226
S P P P R P P G P E T D E K L M D P V G Y Y S D G  742
TCTCTCAAGATTGTGGCTGACGCGCGGTGGCAGCTGGGGTGGGCGCCCTTCCCACTCTCTGGCACT  2301
S L K I V P G H A R C Q P G G G P S P P P G  767
GGCGAGCTCTGCTCTCAACTGGCTCCACTAGGAGGTGGTGGAACTCAATGCAATGGTATGTGGCTG  2376
G C Q L P S P T R L H L G G G R N S N A N G Y V R  792
TTACAGTGGGCGGAGGAGGACGAGGAGGATCTGGGACCCACTGCTGAGCTCGGGATGAATTACGACGAA  2451
L Q L G G E D R G G S G H P L P E L A D E L R R K  817
CTACAAAGCGGAGCGCTGCTGACTCAACCAAGAGGAGTCTCAAGTGAAGGAGACCCCACTCATG  2526
L Q Q R Q P L P D S N P E E S S V *  835
GGGGGGGGGGTCTCATGGGAGTGACCTTAACCTTTGCACAGGACAGCTACCTCAGGAGATGGCAGGGG  2601
CACTTGTCTGCTGGGACAGACACTGCCCATCTTTGCCCGGCGTGGAGAGCTGCTGAGCATGGGCACTG  2676
CTTGGTGTGCTCAGCAGGACTTCAGCTCAGAGGACACACCTCTCTGGAATTTGAGACATGTGGGACCC  2751
CAGCAGCAAACTTTGCAAGGAGGCTTTCAAGAGTGGGCGTGTGGTCAATATGTGTTGGTATGATGT  2826
GAAGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  2901
TCTGCTGTGAGCTTGGAGCTATGAAGGGAGAGGGGCTATGAGCACTTGTCTCTCAGCCCACTGCTCC  2976
AGCTTTGGGCGAGTGTATACATACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG  3051
CGGGGGTGGGCTGAGCCCTGCTTCAAGGCTGTGAATTTTTCAGGCGGGGGTGGGGGTGGAGATGGA  3126
CTGCTTTCAGGGGGAGGGTGGGCGAGGCTCCCACTTGGCTGGGCTGGGTTGGTGGTATTTTATTTGCT  3201
CTGACAGGCTGGGAGGGTGTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG  3276
CCCTGCTTGGGAGAGGCTCAAGTGTAACTTATTTGTCCCACTATTTATTTGTGTAATTTTGAGTA  3351
TTTTTATTTGACAAATAAATGGAGAAATGAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAA  3404

```

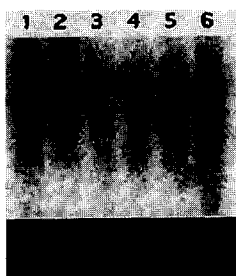


Fig. 2. Expression of M-Sema F by Northern blot analysis. Upper panel: Total RNA was separated on 1% agarose-formaldehyde gels and subjected to the blot hybridization with probes specific for M-Sema F. Lane 1, adult brain; 2, P1 brain; 3, P1 liver; 4, P1 kidney; 5, E14 head; 6, E14 body. Lower panel: 28S rRNA from the same total RNA samples stained with ethidium bromide.

body, while no or very low expression was detected in P1 liver and adult brain.

3.3. *In situ* hybridization

By using *in situ* hybridization histochemistry, the signals were expressed throughout the brain and spinal cord of embryos and P1 animals. In the central nervous system, the signals were very strong in the primordia of the neocortex, hippocampus, thalamus, hypothalamus, tectum, pontine nuclei, spinal cord and retina. The strong signals were also expressed in the primordia of various tissues such as the olfactory epithelium, epithelium of the vomeronasal organ, enamel epithelium of teeth, anterior and intermediate lobes of the pituitary, epithelium of the inner ear, sensory ganglia including trigeminal and dorsal root ganglia. Moreover, the lung and kidney expressed M-Sema F. The expression markedly decreased by adulthood, and found very weakly in several restricted regions of the brain including the hippocampus (the data are not shown).

4. Discussion

Since M-Sema F contains a semaphorin domain that is conserved among semaphorin domains of invertebrate Sema I and II, and vertebrate Sema III [6], it is strongly suggested that M-Sema F belongs to a member of the semaphorin family. Moreover, there are two interesting features in the M-Sema F sequence. M-Sema F contains a potential transmembrane-spanning region following to a putative immunoglobulin-like domain, suggesting that this may be a transmembrane type of the semaphorin family in the vertebrates. A large family of semaphorin genes have been shown very recently in vertebrate and invertebrate, however, only secreted members have been identified from vertebrates [6,9,12]. M-Sema F is most closely related to mouse sem B and C of subfamily IV which is very recently identified by Puschel et al. [12]. Hydrophobicity analysis suggests that sem B and C appear to encode a hydrophobic domain that could be a putative transmembrane domain [12], although they have described that sem B and C are likely to be secreted proteins. In the intracellular region of M-Sema F, there is a proline-rich sequence, suggesting that this may interact with some cytoskeletal proteins [11].

M-sema F mRNA is widely expressed in the brain and spinal cord of embryonic and neonatal animals followed by a marked decrease in the adult. A predominant appearance of the mRNA



Fig. 3. Expression of M-Sema F in head and body of E16.5 mouse. A semisagittal section. The strong expression is found in the brain, particularly in the primordia of the olfactory bulb (O), neocortex (C), thalamus (T), hypothalamus (H), pontine nuclei (P), tectum. The strong expression is also detected outside the central nervous system such as the olfactory epithelium (E), lung (L) and kidney (K). Scale = 1 mm.

in the developing brain and spinal cord suggests that this semaphorin member plays an important role in forming neural network during development, although it is not known whether M-Sema F functions in the same fashion in the nervous system as Sema III functions as repulsive or inhibitory guidance cues to the sensory nerve.

Acknowledgements: This work was supported by grants-in-Aid for Scientific Research (07308053 to S.I. and 07780679 to T.F.). We thank Drs. M. Tohyama, A. Wanaka, T. Takagi, and Y. Imai for encouraging us and advising for experiments, and Dr. Y. Takahashi and Ms. K. Tokutomi for helping with a part of our experiments.

References

- [1] Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* 162, 156–159.
- [2] Dodd, J. and Schuchardt, A. (1995) *Cell* 81, 471–474.
- [3] Furuyama, T., Inagaki, S. and Takagi, H. (1993) *Mol. Brain Res.* 20, 335–344.
- [4] Kennedy, T.E. and Tessier-Lavigne, M. (1995) *Curr. Opin. Neurobiol.* 5, 83–90.
- [5] Kolodkin, A.L., Matthes, D.J., O'Connor, T.P., Patel, N.H., Admon, A., Bentley, D. and Goodman, C.S. (1992) *Neuron* 9, 831–845.
- [6] Kolodkin, A.L., Matthes, D.J. and Goodman, C.S. (1993) *Cell* 75, 1389–1399.
- [7] Kozak, M. (1989) *J. Cell. Biol.* 108, 229–241.
- [8] Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105–132.

- [9] Luo, Y., Raible, D. and Raper, J.A. (1993) *Cell* 75, 217–227.
- [10] Luo, Y. and Raper, J.A. (1994) *Curr. Opin. Neurobiol.* 4, 648–654.
- [11] Pawsom, T. (1995) *Nature* 373, 573–580.
- [12] Puschel, A.W., Adams, R.H. and Betz, H. (1995) *Neuron* 14, 941–948.
- [13] Saiki, R.K., Gelfand, D.H., Dtoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Ehrlich, H.A. (1988) *Science* 239, 487–494.
- [14] von Heijne, G. (1986) *Nucl. Acids Res.* 14, 4683–4690.
- [15] Williams, A.F. and Barclay, A.N. (1988) *Annu. Rev. Immunol.* 6, 381–405.