Volume 250, number 2, 411–415 FEB 07322 July 1989

Complete primary structure of a calcium-dependent serine proteinase capable of degrading extracellular matrix proteins

Hirohisa Kinoshita[+], Hisako Sakiyama[+], Kastuo Tokunaga*, Shinobu Imajoh-Ohmi°, Yoshio Hamada[†], Kaichi Isono and Shigeru Sakiyama*

[+] *Division of Physiology and Pathology, National Institute of Radiological Sciences, 4-9-1 Anagawa, Chiba 260,* * *Division of Biochemistry, Chiba Cancer Center, 666-2 Nitona, Chiba 280,* ° *The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minatoku, Tokyo 108,* [†] *Laboratory of Tissue and Cell Culture, National Institute for Basic Biology, Okazaki 444* and *2nd Department of Surgery, Chiba University School of Medicine, Chiba 280, Japan*

Received 5 May 1989

A novel calcium-dependent serine proteinase (CASP) secreted from malignant hamster embryo fibroblast Ni 12C2 degrades extracellular matrix proteins. A complementary DNA encoding CASP has been isolated with the use of oligonucleotide probes synthesized based on partial amino acid sequences of CASP. The complete amino acid sequence of CASP revealed that it has a serine active site at the C-terminal side. Glu rich and proEGF homologous sites are found at the N-terminal site suggesting that it is structurally similar to blood coagulation factors such as IX, X and an anti-coagulation factor, protein C.

Ca^{2+}; Serine protease, Ca^{2+}-dependent; cDNA; Protein, extracellular matrix

1. INTRODUCTION

We have isolated a novel calcium-dependent serine proteinase [1] with a trypsin-type substrate specificity from the conditioned medium of hamster embryo fibroblasts Ni 12C2 cells which show high metastatic and phagocytotic activities [2,3]. Ca^{2+}-dependent serine proteinase (CASP) degrades type I and IV collagen and fibronectin in the presence of calcium in the millimolar range. For the hydrolysis of synthetic peptides acetyl Gly-L-Lys-naphthyl ester (AGLNE), however, CASP does not require calcium. The enzyme has been shown to be an 88 kDa heterodimer, consisting of heavy (66 kDa) and light (33 kDa) chains with disulfide bond(s) [1]. In the present study, we have cloned CASP cDNA and determined the complete primary structure of the CASP.

Correspondence address: H. Sakiyama, Division of Physiology and Pathology, National Institute of Radiological Sciences, 4-9-1 Anagawa, Chiba 260, Japan.

2. MATERIALS AND METHODS

2.1. *Cell culture*

Ni 12C2 hamster fibroblasts were cultured in Eagle's minimal essential medium supplemented with 10% newborn calf serum on plastic tissue culture dishes.

2.2. *Protein sequence analysis*

CASP was purified from the conditioned medium of Ni 12C2 hamster fibroblasts as described [1]. The heavy and light chains of CASP were eluted from the SDS-polyacrylamide gel. N-terminal amino acid sequences of both chains have been determined using a model 470A gas phase sequencer (Applied Biosystems). The sequences thus obtained were EPTMHX-EILSPNYPQAYPNEMEXXXXIXV (X denotes undetermined amino acid) and FXXFPAKIQSFPXQVFFEFPXAXXALI for heavy and light chains, respectively.

2.3. *Oligonucleotide synthesis*

Oligonucleotides based on the N-terminal amino acid sequences of both heavy and light chains were synthesized according to the most frequent codon usage in rat α and β actin genes [4]. (A) CTCCATCTCGTTGGGGTAGGCCTGGGGG-TAGTTGGGGGACAGGATCTC for heavy chain and (B) GGGGAACTCGAAGAACACCTGGGCGGGGAAGGACT-GGATCTTGGCGGGGAA for light chain were synthesized with an automatic DNA synthesizer (Applied Biosystems).

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

2.4. *Construction of λgt 10 cDNA library*

A cDNA library of 1.0×10^7 independent clones was prepared from total poly(A)$^+$ RNA of Ni 12C2 with the use of λgt10 vector system as described [5].

2.5. *Isolation of CASP cDNA clones*

About 4×10^5 phages were screened for CASP cDNA clone by hybridization analysis using 5'-end labeled oligonucleotide probes A and B. The hybridization was carried out in 40% formamide–5 × SSPE (1 × SSPE is 0.18 M NaCl–0.01 M sodium phosphate–10 mM EDTA–5 × Denhardt's solution–0.1% SDS–100 μg/ml salmon sperm DNA for 16 h at 42°C and the nylon filter (Pall) was washed with 6 × SSC–0.1% SDS at room temperature. Plaques that gave positive signals were purified by a secondary plating and rescreening. Only probe A gave a positive signal under the present condition. A positive clone was isolated and subsequently subcloned into pTZ18R (Pharmacia) (pHCSP).

2.6. *Nucleotide sequence analysis*

The DNA sequence was determined by the dideoxy method using progressive deletion of the plasmid with *exo*III (USB). A search for homology of CASP sequence with known amino acid sequences was carried out using the GenBank/NBRS data library.

3. RESULTS AND DISCUSSION

Using probe A, we have isolated cDNA clone for CASP. From the nucleotide sequence, the primary structure for CASP was determined. Complete sequence analysis revealed the presence of an open reading frame of 695 amino acids and the poly(A) tract (18 base pairs) (fig.1). The potential initiation codon ATG which conforms to Kozak's criteria [6] is at 281 nucleotide followed by a typical signal peptide sequence of 21 amino acids. Experimentally determined N-terminal sequences of the heavy and light chains were completely identical to those derived from the nucleotide sequences (amino acid 22–50 and 446–472 residues, respectively). Heavy and light chains, therefore, are tandemly located in this order. RNA blot hybridization analysis revealed that the size of the mRNA of CASP is 3.0 kb (not shown) indicating that CASP is derived from a single transcript and that the isolated cDNA covers a nearly full-length region of the mRNA. Thus, two chains encoded by a single gene might arise through proteolytic cleavage of the single polypeptide chain. The cleavage site is thought to be between Arg (444) and Ile (445) (fig.1) based on the experimentally determined N-terminal amino acid sequence of the light chain. The same sequence (Arg-Ile) is also cleaved in the case of chymotrypsinogen [7] (fig.2B). Two possible *N*-glycosylation sites are found in the heavy chain but not in the light chain. The heavy chain indeed has *N*-linked oligosaccharide(s) since the treatment of CASP with endoglycosidase F resulted in the reduction of the molecular mass of the heavy chain but not the light chain (Sakiyama et al., in preparation). The overall structure of CASP is schematically shown in fig.2A.

All members of the serine proteases share strongly conserved amino acid sequence blocks that make up the active site of the enzyme. This sequence (amino acid residues 633–640) is also highly conserved in CASP. By comparing the amino acid sequence of CASP with that of chymotrypsinogen with which CASP shares 33.6% homology, we predict the three reactive residues of this enzyme to lie at positions 482(His), 536(Asp) and 638(Ser) (fig.1). The consensus of the amino acids surrounding these active residues is also seen (fig.2B). The light chain of CASP shares a high homology with protein C (34%) whose homologous sites, like chymotrypsinogen, are centralized around the regions of active site residues of serine proteases. The catalytic site of CASP, therefore, is thought to be the light chain. The overall homology of the amino acid sequence between the light chain and serine proteases was between 22% and 40%.

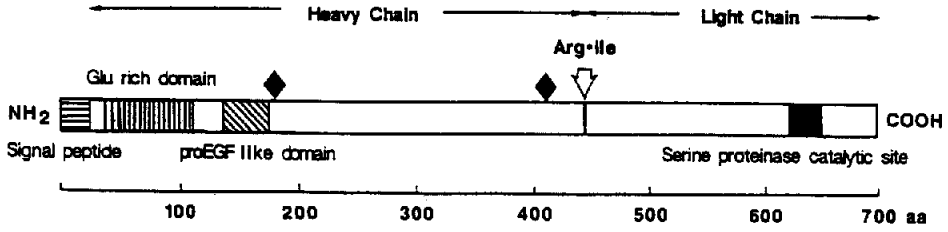
The heavy chain which we consider to be a regulatory subunit contains Glu-rich region (15 of 103 amino acid residues) at the N-terminal part (positions 22–124) followed by a proEGF homologous domain (positions 137–177). An alignment of amino acid residues of the domains of the proteases homologous to proEGF is shown in

Fig.1. Nucleotide sequence of CASP cDNA and translated amino acid sequence of CASP. (A) Restriction map of CASP cDNA. The solid region is a open reading frame. (B) The 2085-nucleotide open reading frame is flanked by 3'-untranslated sequences of 364 nucleotides containing a polyadenylation signal, AATAAA (overlined) and a poly(A) tail, and by the 5'-untranslated region of 280 nucleotides. ▦, signal peptide; ----, EGF precursor-like region; : ▽, a possible cleavage site between heavy and light chains; •, sites of possible asparagine-linked glycosylation; ■, His, Asp and Ser which are the active sites of serine protease are indicated. N-terminal amino acid sequence of heavy chain determined using a gas phase sequencer is underlined, and that of light chain is wavy-underlined.

B.

```
ATGTGAAGCAGGGGGGTAGGGAATCTTAGGAGCCGTCAGGACAGGGAGGTCGGGCGGAGGTTCCTGCAGAGGCTGACTCCTCTGCAGTGGCAGGGTCTTCCCAGCATGACCAGTGGCGGG
                                                                                                                        120

AGAGGAGACTGACCCGCTTGTTCCCATCAGCTCCTGAAGGTGACACTGAGCCCTGGGTGGCCCCTCACCACCAAAACAGGTGCCTGTATTTGTTGGATAAAGACAACCAGGCCAGCCGCC
                                                                                                                        240

                                                  M  G  K  S  S  E  A  W  C  I  V  L  F  S  V  F  A  S  F  S  A  E  P  T  M  H  G
CTTTAGTTCCAAGGTCAGAGGTCCTGAGAGCCAGAGAGCCATGGGCAAATCATCAGAGGCATGGTGCATTGTCTTGTTTTCTGTTTTCGCATCATTTTCTGCTGAGCCTACCATGCATGG
                                                                                                                        360

                                                  40
E  I  L  S  P  N  Y  P  Q  A  Y  P  N  E  M  E  K  T  W  D  I  E  V  P  E  G  F  G  V  R  L  Y  F  T  H  L  D  M  E  L
GGAGATCCTGTCCCCTAACTATCCTCAGGCATACCCCAATGAGATGGAGAAAACGTGGGACATAGAAGTCCCAGAAGGGTTTGGGGTTCGCCTCTACTTCACCCATCTGGACATGGAGCT
                                                                                                                        480

                                                  80
S  E  N  C  E  Y  D  S  V  Q  I  I  S  G  G  V  E  E  G  R  L  C  G  Q  R  T  S  K  N  A  N  S  P  I  V  E  E  F  Q  I
GTCAGAGAACTGCGAATATGACTCAGTGCAGATAATCTCAGGAGGCGTCGAGGAAGGGAGACTCTGTGGGCAGAGGACCAGCAAGAATGCCAACTCCCCCATTGTGGAAGAGTTTCAAAT
                                                                                                                        600

                                                  120
F  Y  N  K  L  Q  V  I  F  R  S  D  F  S  N  E  E  R  F  T  G  F  A  A  Y  Y  A  A  I  D  V  N  E  C  T  D  F  T  D  V
CCCATACAATAAACTCCAGGTCATCTTTAGGTCAGACTTCTCCAACGAAGAGCGGTTTACTGGCTTTGCTGCATATTACGCTGCCATTGATGTAAATGAATGCACAGATTTTACAGATGT
                                                                                                                        720

                                                  160
P  C  S  H  F  C  N  N  F  I  G  G  Y  F  C  S  C  P  P  E  Y  F  L  H  D  D  M  R  N  C  G  V  N  C  S  G  N  V  F  T
CCCTTGCAGCCACTTCTGCAACAATTTCATTGGTGGTTATTTCTGTTCCTGCCCCCCAGAATACTTCCTCCATGATGATATGAGGAATTGCGGAGTCAATTGCAGTGGGAATGTATTCAC
                                                                                                                        840

                                                  200
A  L  I  G  E  I  S  S  P  N  Y  P  N  P  Y  P  E  N  S  R  C  E  Y  Q  I  L  L  E  E  G  F  Q  V  V  V  T  I  Q  R  E
TGCCCTGATTGGGGAGATTTCAAGCCCCAATTACCCCAATCCATACCCCGAGAACTCAAGGTGTGAGTACCAGATTTTGTTGGAGGAGGGGTTCCAAGTGGTGGTGACCATCCAGAGAGA
                                                                                                                        960

                                                  240
D  F  D  V  E  P  A  D  S  Q  G  N  C  Q  D  S  L  L  F  A  A  K  N  R  Q  F  G  P  F  C  G  N  G  F  P  G  P  L  T  I
AGATTTTGATGTGGAACCAGCTGATTCGCAGGGGAACTGCCAGGACAGCTTACTGTTTGCTGCAAAAAATCGACAATTTGGTCCTTTCTGTGGCAATGGATTCCCTGGGCCACTAACTAT
                                                                                                                        1080

                                                  280
E  T  H  S  N  T  L  D  I  V  F  Q  T  D  L  T  E  Q  K  K  G  W  K  L  R  Y  H  G  D  P  I  P  C  P  K  E  I  T  A  N
CGAAACCCACAGTAACACTCTTGATATTGTCTTTCAAACCGACCTAACAGAGCAAAAAAAGGGCTGGAAGCTTCGTTACCATGGAGACCCAATCCCTTGTCCCAAGGAAATCACCGCCAA
                                                                                                                        1200

                                                  320
S  V  W  A  P  E  K  A  K  Y  V  F  K  D  V  V  K  I  S  C  V  D  G  F  E  A  V  E  G  N  V  G  S  T  F  F  Y  S  T  C
CTCTGTTTGGGCGCCTGAAAAGGCAAAATACGTGTTTAAAGATGTGGTGAAGATATCCTGTGTGGACGGGTTTGAGGCTGTAGAGGGAAATGTTGGCTCAACATTCTTCTATTCTACTTG
                                                                                                                        1320

                                                  360
Q  S  N  G  Q  W  S  N  S  R  L  R  C  Q  P  V  D  C  G  I  P  E  P  I  Q  N  G  K  V  D  D  P  E  N  T  L  F  G  S  V
TCAAAGCAATGGACAGTGGAGCAATTCCAGGCTACGGTGTCAGCCTGTGGACTGTGGCATTCCGGAACCCATTCAGAACGGTAAAGTTGATGATCCAGAAAATACTCTGTTTGGCTCTGT
                                                                                                                        1440

                                                  400
I  H  Y  S  C  E  E  P  Y  Y  Y  N  E  H  A  E  H  G  G  E  Y  R  C  A  A  N  G  S  W  V  N  D  E  L  G  I  E  L  P  K
CATTCACTACTCGTGTGAGGAGCCGTATTACTACATGGAACATGCAGAACACGGCGGGGAGTATCGCTGTGCTGCTAATGGGAGCTGGGTGAATGACGAGCTGGGCATAGAGCTCCCAAA
                                                                                                                        1560

                                                  440
C  V  P  V  C  G  V  P  T  E  P  F  R  I  Q  Q  R▽I  F  G  G  F  P  A  K  I  Q  S  F  P  W  Q  V  F  F  E  F  P  R  A
ATGTGTTCCAGTCTGTGGGGTACCCACTGAGCCCTTTAGAATACAGCAGAGGATATTCGGAGGATTTCCTGCGAAGATCCAGAGTTTTCCTTGGCAAGTCTTCTTTGAGTTCCCGAGGGC
                                                                                                                        1680

                                                  480
G  G  A  L  I  G  E  H  W  V  L  T  A  A  H  V  V  E  G  N  S  D  P  S  M  Y  V  G  S  T  S  V  R  M  E  N  L  A  N  V
CGGTGGGGCTCTTATTGGCGAGCACTGGGTGCTGACGGCAGCTCACGTTGTGGAAGGGAACTCCGACCCATCGATGTATGTGGGGTCCACGTCTGTGAGAATGGAAAACCTGGCAAACGT
                                                                                                                        1800

                                                  520
Q  K  L  T  T  D  R  V  I  I  H  P  G  W  K  P  G  D  D  L  S  T  R  T  N  F  D  N  D  I  A  L  V  R  L  K  D  P  V  K
TCAGAAGCTCACCACCGACCGTGTGATCATCCATCCGGGCTGGAAACCAGGGGATGACCTAAGCACACGGACAAATTTTGACAATGACATCGCACTGGTGCGGCTGAAAGACCCTGTGAA
                                                                                                                        1920

                                                  560
M  G  P  T  V  S  P  I  C  L  P  G  T  S  S  E  Y  E  P  S  E  G  D  L  G  L  I  S  G  W  G  R  T  E  R  N  N  I  V  I
AATGGGGCCCACTGTTTCCCCCATCTGCCTGCCAGGCACCTCCTCAGAGTACGAACCCTCAGAGGGTGATCTGGGACTGATCTCAGGGTGGGGCCGAACAGAGAGGAGAAATATTGTTAT
                                                                                                                        2040

                                                  600
Q  L  R  G  A  K  L  P  V  T  S  L  E  K  C  R  Q  V  K  E  E  N  P  K  A  R  A  D  D  Y  V  F  T  S  N  M  I  C  A  G
CCAGCTCAGAGGGGCAAAGTTACCCGTGACCTCTTTAGAGAAGTGCCGACAGGTGAAAGAGGAGAACCCCAAAGCGAGGGCGGATGACTACGTTTTCACCAGCAACATGATCTGTGCTGG
                                                                                                                        2160

                                                  640
E  K  G  V  D  S  C  Q  G  D  S  G  G  A  F  A  L  P  V  P  N  V  R  D  P  K  F  Y  V  A  G  L  V  S  W  G  K  K  C  G
AGAGAAGGGCGTTGACAGCTGTCAGGGGGACAGCGGTGGGGCTTTCGCTTTGCCGGTCCCCAATGTAAGGGACCCCAAATTCTACGTGGCAGGCCTAGTGTCCTGGGGGAAAAAGTGTGG
                                                                                                                        2280

                                                  680
T  Y  G  I  Y  T  K  V  K  N  Y  K  D  W  I  L  Q  T  M  Q  E  N  S  V  P  S  Q  D
GACCTATGGAATCTACACGAAGGTAAAGAACTACAAGGACTGGATACTGCAAACTATGCAGGAGAATAGTGTCCCCAGTCAGGACTAACCCATTCATACCGCCAGCCCCTCCCCCCAGGA
                                                                                                                        2400

CACAGCAACGCATTGCTCTCAGACCCTGTGGTAGTCCCATTATCTCGCAGTGATGGAGAGAAGATGCAGGAGTGCCGTTAACACTGAACTTGATTGTCAAGACACCTGAATGGATGGAGG
                                                                                                                        2520

GTCTGATCACAGGCCTGTGTTGATTACTCAGGTGTGAATTCCCTCCTGTTATTGGTAACAACATGGCAGAGACTCCTCTGTCTTGCACCATTCTGCAGGGTATCTTAACCTCCGTTTCTT
                                                                                                                        2640

TCACCTGCTTGATTCCATTTATAATCCTTGGAGTCCTTTATTTACTTACAATAAAGCATATTTCTAACCTGAAAAAAAAAAAAAAAAAA
```

A.



B.

```
CASP(Top line)
Chymotrypsinogen A-Bovine(bottom line)

(432-490) CGVPT-EP-FRIQQRIFGGFPAKIQSFPWQVFF-E---FPRAGGALIGEHWVLTAAHVVEGNSDP
          :::    :      ::  :  :   : ::::      :   :: :: : :: ::::   :
(  1- 63) CGVPAIQPVLSGLSRIVNGEEAVPGSWPWQVSLQDKTGFHFCGGSLINENWVVTAAHC--GVTTS

(491-555) SMYVGSTSVRMENLANVQKLTTDRVIIHPGWKPGDDLSTRTNFDNDIALVRLKDPVKMGPTVSPI
            :              :::    :      :   : :     ::: :  :         :::
( 64-121) DVVVAGEFDQGSSSEKIQKLKIAKVFKNS--KY-NSL-T---INNDITLLKLSTAASFSQTVSAV

(556-620) CLPGTSSEYEPSEGDLGLISGWGRTERRNIVIQLRGAKLPVTSLEKCRQVKEENPKARADDYVFT
          :::  :       :       ::: : :           :   ::          :
(122-177) CLPSASDDF-AA-GTTCVTTGWGLT-R-YTNANTPD-RLQQASLPL--LSNTNCKKYWGTK-I-K

(621-685) SNMICAGEKGVDSCQGDSGGAFALPVPNVRDPKFYVAGLVSWGKKCGTYGIYTKVKNYKDWILQT
            :::::  :: :: :::::           :       : : :  :     :  ::
(178-241) DAMICAGASGVSSCMGDSGGPLVCKKNGAWTL-VGIVSWGSSTCSTSTPGVYARVTALVNWVQQT
```

C.

```
EGF precursor-Mouse      (362-401)  DVNECAT-QNHGCTLGCENTPGSYHCTCPTGFVLLPDGKQC
                                    :::::      :   : :  : : : ::    :  :   :
CASP                     (137-177)  DVNECTDFTDVPCSHFCNNFIGGYFCSCPPEYFLHDDMRNC
                                    :       : : :    :   ::: : : : ::   :
Protein C precursor-Bovine(133-172) RFSNCSA-ENGGCAHYCMEEEGRRHCSCAPGYRLEDDHQLC
```

Fig.2. (A) The overall structure of CASP precursor protein. Sites of possible asparagine-linked glycosylation are indicated by rhombuses. (B) Comparison with junctional and catalytic regions of chymotrypsinogen. (C) Comparison with epidermal growth factor-like region and protein C precursor. :, identical residues; △, cleavage site ; ■, His, Asp and Ser active site residues of serine proteases; -, gaps inserted to maximize sequence identity; numbers in parentheses are the region of amino acid sequences shown.

fig.2C. The sequence of CASP, as protein C, has the arrangement of cysteine residues characteristic of proEGF domain. This domian comprises Asp whose β-hydroxylation can be correlated with Ca^{2+} binding [8]. Although it is still under investigation, Glu in the N-terminal region may be γ-carboxylated and function as a Ca^{2+} binding site [9]. If that is the case, CASP binds with Ca^{2+} through γ-carboxyglutamic acid (Gla) and β-hydroxyaspartate via a mechanism similar to that of blood coagulation factors such as IX, X and protein C [10,11]. When the Ca^{2+} binding regulatory domains of coagulation factors are removed, they are activated as serine proteinases. CASP, however, is secreted as an active form since it degrades (AGLNE) without any cleavage of the enzyme. The addition of Ca^{2+} induces autodigestion. It is possible that the self-digestion of the regulatory domain of the enzyme is required to potentiate the enzyme to degrade native proteins.

Acknowledgements: We wish to thank Dr T. Hiwasa for many helpful discussions and Mr H. Kageyama and Dr T. Shirasawa for help in computer work. We also thank Mrs K. Kuriiwa and Ms Y. Nakamura for technical assistance. This work was supported in part by Special Coordination Funds for Promoting Science and Technology from the Science and Technology Agency of Japan and by a Grant-in-Aid from the Ministry of Health and Welfare for Comprehensive 10-year Strategy for Cancer Control, Japan.

REFERENCES

[1] Sakiyama, H., Nishino, Y., Tanaka, T., Tomosawa, T., Kinoshita, H., Nagata, K., Chiba, K. and Sakiyama, S. (1989) Biochim. Biophys. Acta 990, 156–161.
[2] Sakiyama, H. and Robbins, P.W. (1973) Fed. Proc. 32, 86–90.
[3] Sakiyama, H., Nishino, Y., Nishimura, K. and Otsu, H. (1984) Cancer Res. 44, 2023–2032.
[4] Nudel, U., Zakut, R., Shani, M., Neuman, S. and Yaffe, D. (1983) Nucleic Acid Res. 11, 1759–1771.
[5] Huynh, T.V., Young, R.A. and Davis, R.W. (1985) in: DNA Cloning (Glover, D.M. ed.) Vol.1, p.49, IRL Press, Oxford.
[6] Kozak, M. (1987). Nucleic Acid Res. 15, 8125–8132.
[7] Stroud, R.M. (1974). Sci. Am. 231, 74–88.
[8] Rees, D.J.G., Jones, I.M., Handford, P.A., Walter, S.J., Esnouf, M.P., Smith, K.J. and Brownlee, G.G. (1988) EMBO J. 7, 2053–2061.
[9] Price, P.A., Fraser, J.D. and Metz-Virca, G. (1987) Proc. Natl. Acad. Sci, USA 84, 8335–8339.
[10] Furie, B. and Furie, B.C. (1988) Cell 53, 505–518.
[11] Thompson, A.R. (1986) Blood 67, 565–572.