Volume 257, number 2, 319-323 FEB 07780 November 1989

Amino acid sequences of cytotoxin-like basic proteins derived from cobra venoms

Seiji Inoue, Kohji Okumura, Michiyo Tsujino, Kazuto Ohkura, Kiyoshi Ikeda, Masayuki Takechi*, Yasuo Tanaka* and Kyozo Hayashi°

*Department of Biochemistry, Osaka University of Pharmaceutical Sciences, Kawai, Matsubara, Osaka 580, *Faculty of Pharmacy, Kinki University, Higashiosaka 577* and *°Department of Pharmacology, Gifu Pharmaceutical University, Gifu 502, Japan*

Received 5 August 1989

Amino acid sequences of cytotoxin-like basic proteins (CLBPs), purified from the venoms of Formosan cobra *(Naja naja atra)* and Indian cobra *(Naja naja)*, were reinvestigated. The determined sequences differed from those reported previously by Takechi et al. [(1985) Biochem. Int. 11, 795–802; (1987) Biochem. Int. 14, 145–152]. The sequence of CLBPs at residues 25–30 was found to be hydrophilic as compared with those of cytotoxins. The difference in the hydrophobicity of this region might be responsible for the difference in their cytotoxic activities.

Amino acid sequence; Cytotoxin; Cardiotoxin; Cytotoxin-like basic protein; *(Naja naja, Naja naja atra)*

1. INTRODUCTION

Cobra venom toxins contain three main classes of homologous proteins; long neurotoxins, short neurotoxins, and cytotoxins (cardiotoxins). Cytotoxins are highly basic polypeptides consisting of 60 amino acid residues, and exhibit cytotoxic activities against many kinds of cells. The cobra venoms also contain small amounts of cytotoxin homologues. They have been reported to exhibit quite low cytotoxic activity, though their sequences were quite similar to those of cytotoxins [1]. Two cytotoxin homologues, CLBP and LCBP were purified from the venoms of Formosan cobra (*Naja naja atra*) [2] and Indian cobra (*Naja naja*) [3], respectively, and their amino acid sequences were determined. Previously, we isolated four cytotoxins (CT-I, CT-II, CT-III, and CT-IV) and also another CLBP from the venom of the Thailand cobra (*Naja naja siamensis*), and determined their amino acid sequences [4,5]. From the study on the *N. naja siamensis* CLBP, it was suggested that the reported sequences of CLBPs from *N. naja* and *N. naja atra* might be mistaken. In the present study, we therefore reinvestigated the sequences of CLBPs from the latter two cobra venoms.

Correspondence address: S. Inoue, Department of Biochemistry, Osaka University of Pharmaceutical Sciences, Kawai, Matsubara, Osaka 580, Japan

Abbreviations: CLBP, cytotoxin-like basic protein; LCBP, less-cytotoxic basic polypeptide; HPLC, high-performance liquid chromatography; Cm, *S*-carboxymethyl; Cmc, *S*-carboxymethylcysteine; N, amino; C, carboxyl

2. MATERIALS AND METHODS

2.1. *Isolation of CLBPs*

CLBPs were isolated from the venoms of *N. naja atra* (Sigma) and *N. naja* (Sigma) as described previously [2,3]. Cm derivatives of the CLBPs were prepared by the method of Crestfield et al. [6].

2.2. *Studies on the amino acid sequences of CLBPs*

Cm-CLBPs were digested with staphylococcal protease (Miles Scientific) in 0.1 M NH_4HCO_3 at 37°C for 12 h or with endoproteinase Arg-C (Boehringer Mannheim, DDR) in 0.1 M NH_4HCO_3 at 30°C for 11 h. The peptides thus obtained were separated by reversed-phase HPLC on a Cosmosil 5C18-300 column (Nacalai tesque, Japan) or VYDAK 214TP54 column (The Separation Group, USA) with 0.1% trifluoroacetic acid, containing a linear gradient from 0 to 70% acetonitrile. The manual Edman degradation method was used to determine the N-terminal sequences of Cm-CLBPs and peptides. Phenylthiohydantoin derivatives were identified by HPLC [4]. The C-terminal sequences of Cm-CLBPs were determined by carboxypeptidase A (Sigma) digestion at 40°C for 120 min in 0.1 M Tris-HCl buffer (pH 8.0) [7]. The released amino acids were determined with the amino acid analyzer.

2.3 *Amino acid analysis*

The protein and peptide samples were hydrolyzed with 6 N HCl containing 0.2% phenol at 110°C for 24 h in evacuated sealed tubes. The amino acid compositions were determined with an amino acid analyzer, a Hitachi model L-8500.

3. RESULTS

3.1. *Amino acid sequence of* N. naja *CLBP*

The manual Edman degradations of *N. naja* Cm-CLBP revealed its N-terminal sequence up to residue 25. Carboxypeptidase A released asparagine (0.78 mol/mol protein) and carboxymethylcysteine (0.13 mol/mol protein) after a 120 min incubation, suggesting the C-terminal residues of Cm-CLBP being

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

Table 1

Amino acid compositions of *N. naja* CLBP, staphylococcal protease peptides (IS-1 and IS-2), and endoproteinase Arg-C peptides (IR-1 to IR-3)

Amino acid	*N. naja* CLBP	IS-1	IS-2	IR-1	IR-2	IR-3
Cys[a]	7.65 (8)	1.92 (2)	5.71 (6)	3.58 (3)		5.34 (5)
Asp	6.77 (7)	1.16 (1)	6.21 (6)	2.40 (2)		5.11 (5)
Thr	3.74 (4)	1.92 (2)	2.10 (2)	2.41 (3)		0.98 (1)
Ser	1.77 (2)		1.97 (2)			1.56 (2)
Glu	2.00 (2)	1.86 (2)		1.86 (2)		
Pro	5.05 (5)	2.32 (2)	2.70 (3)	1.93 (2)	1.60 (2)	0.70 (1)
Gly	2.08 (2)		2.07 (2)	1.03 (1)		1.04 (1)
Ala	2.87 (3)		3.21 (3)	0.88 (1)		1.99 (2)
Val	1.92 (2)		2.33 (2)		0.96 (1)	0.92 (1)
Ile	0.98 (1)	0.79 (1)		0.81 (1)		
Leu	6.88 (7)	1.87 (2)	5.04 (5)	3.79 (4)	1.08 (1)	1.99 (2)
Tyr	1.85 (2)	0.65 (1)	0.94 (1)	0.98 (1)		1.08 (1)
Phe	4.30 (4)	0.91 (1)	3.14 (3)	2.32 (2)	2.04 (2)	
Lys	9.83 (11)	1.95 (2)	9.15 (9)	5.85 (5)	2.87 (3)	3.32 (3)
His	0.99 (1)	1.08 (1)		0.88 (1)		
Arg	1.00 (1)		1.04 (1)		0.94 (1)	
Total	62	17	45	28	10	24

[a] Cysteine was determined as carboxymethylcysteine
Values in parentheses were taken from the sequence

Cmc-Asn-COOH. The staphylococcal protease digest of Cm-CLBP was subjected to HPLC. The amino acid compositions of the two separated peptides, IS-1 and IS-2, are shown in table 1. The C-terminal fragment IS-2 was subjected to manual Edman degradation. The endoproteinase Arg-C digest of Cm-CLBP was separated by HPLC. The amino acid compositions of three of the four separated fragments were determined (table 1), and the peptides IR-1 and IR-3 were subjected to manual Edman degradation. The summary of sequence studies on the *N. naja* CLBP described above are shown in fig.1.

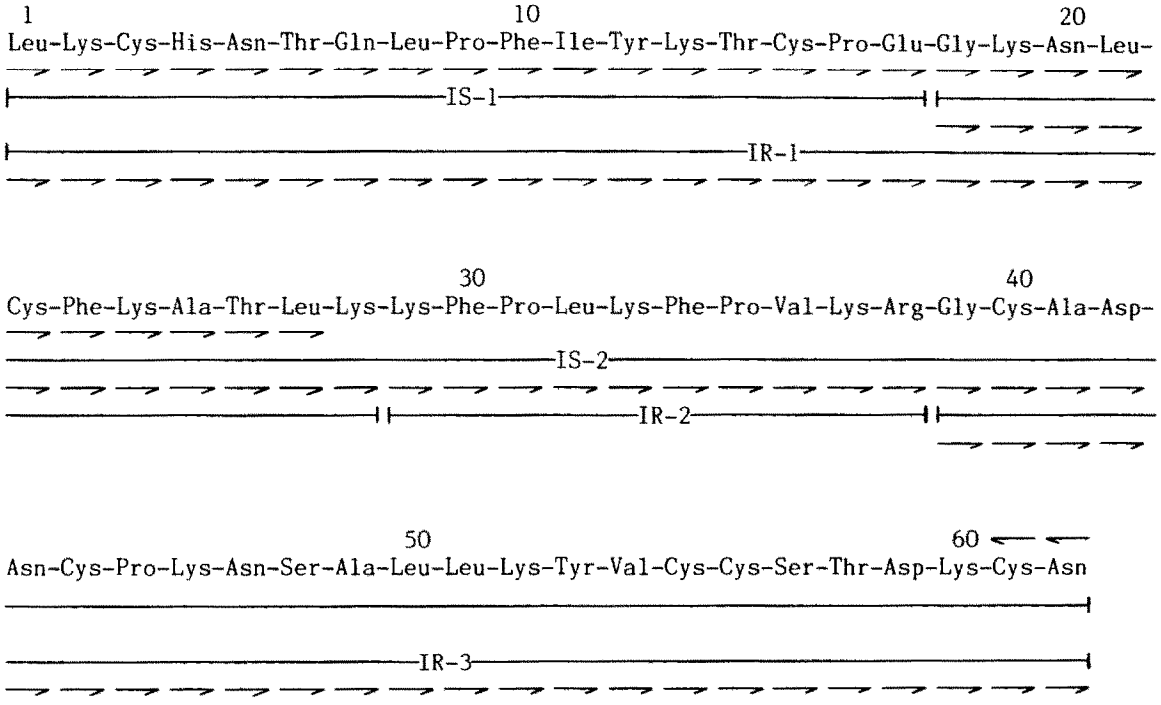

Fig.1. Summary of the sequence studies on *N. naja* CLBP. IS-and IR- refer to the staphylococcal protease and endoproteinase Arg-C peptides, respectively. The amino acid sequence was determined by the manual Edman degradation (→) and by the carboxypeptidase digestion (←) methods.

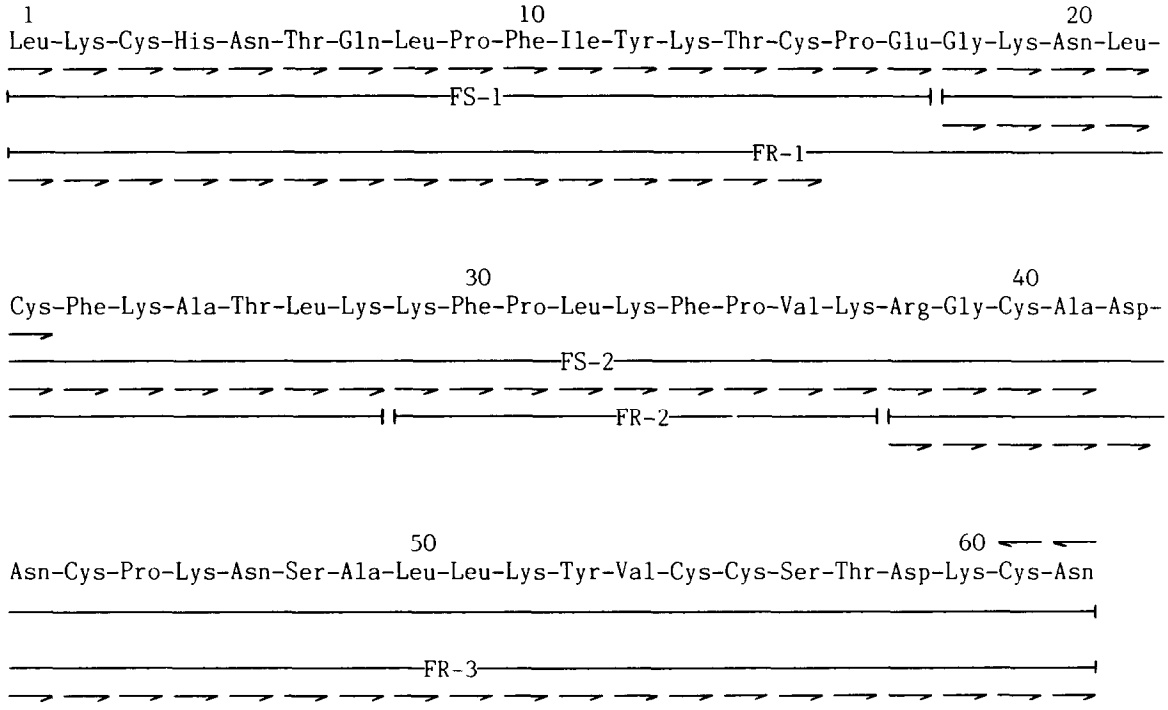

Fig.2. Summary of the sequence studies on *N. naja atra* CLBP. FS-and FR- refer to the staphylococcal protease and endoproteinase Arg-C peptides, respectively. The amino acid sequence was determined by the manual Edman degradation (⟶) and by the carboxypeptidase digestion (⟵) methods.

3.2. *Amino acid sequence of* N. naja atra *CLBP*

Summary of the sequence study on *N. naja atra* CLBP is given in fig.2. The N-terminal sequence of Cm-CLBP was determined by manual Edman degradation up to residue 22. The C-terminal sequence, -Cmc-Asn-COOH, determined by carboxypeptidase A digestion (data not shown), was the same for that of the *N. naja* CLBP. The amino acid compositions of staphylococcal protease peptides (FS-1 and FS-2) and endoproteinase Arg-C peptides (FR-1 to FR-3) of the *N. naja atra* Cm-CLBP are given in table 2. The determined sequence of the *N. naja atra* CLBP was found to be identical to that

Table 2

Amino acid compositions of *N. naja atra* CLBP, staphylococcal protease peptides (FS-1 and FS-2), and endoproteinase Arg-C peptides (FR-1 to FR-3)

Amino acid	*N. naja atra* CLBP	FS-1	FS-2	FR-1	FR-2	FR-3
Cys[a]	8.17 (8)	2.12 (2)	6.15 (6)	2.91 (3)		4.69 (5)
Asp	6.79 (7)	1.21 (1)	5.65 (6)	2.30 (2)		4.85 (5)
Thr	3.56 (4)	1.79 (2)	1.90 (2)	2.80 (3)		0.98 (1)
Ser	1.77 (2)		1.65 (2)			1.65 (2)
Glu	2.06 (2)	1.89 (2)		2.48 (2)		
Pro	5.19 (5)	1.94 (2)	2.98 (3)	2.06 (2)	1.83 (2)	1.39 (1)
Gly	2.13 (2)		1.77 (2)	1.44 (1)		1.12 (1)
Ala	2.92 (3)		2.71 (3)	1.07 (1)		1.92 (2)
Val	1.83 (2)		1.92 (2)		0.99 (1)	1.15 (1)
Ile	0.99 (1)	0.89 (1)		1.02 (1)		
Leu	6.80 (7)	1.99 (2)	4.76 (5)	4.13 (4)	1.03 (1)	2.20 (2)
Tyr	1.94 (2)	0.95 (1)	1.04 (1)	1.03 (1)		0.96 (1)
Phe	4.06 (4)	0.94 (1)	2.83 (3)	2.22 (2)	2.00 (2)	
Lys	10.0 (11)	2.08 (2)	8.88 (9)	4.98 (5)	2.90 (3)	3.50 (3)
His	1.05 (1)	0.87 (1)		1.10 (1)		
Arg	1.02 (1)		1.00 (1)			0.98 (1)
Total	62	17	45	28	9	25

[a] Cysteine was determined as carboxymethylcysteine
Values in parentheses were taken from the sequence

```
1.  IRCF--ITPDITSKDCPNC-HVCYTKTWCDAFCSIRGKRVDLGCAATCPTVKTGVDIQCCSTDNCNPFPTRKRP
2.  LECHNQQSSQTPTTTGCSGGETNCYKKRWRD-H---RGYRTERGC--GCPSVKNGIEINCCITTDRCNN-------
3.  LKCHN--TQLPFIYKTCPEGKNLCFKATLK-FP--LKFPVKRGCADNCPKNSALLKYVCCSTDKCN--------
4.  LKCHN--TQLPFIYKTCPEGKNLCFKATLK-FP--LKIPIKRGCADNCPKNSALLKYVCCSTDKCN--------
5.  LKC-N--KLIPIASKTCPAGKNLCYKMFMMS-D---LTIPVKRGCIDVCPKNSLLVKYVCCNTDRCN--------
6.  LKC-N--KLVPLFYKTCPAGKNLCYKMFMVA-T---PKVPVKRGCIDVCPKSSLLVKYVCCNTDRCN--------
7.  LKC-N--KLIPLAYKTCPAGKNLCYKMFMVA-A---PKVPVKRGCIDACPKNSLLVKYVCCNTDRCN--------
8.  LKC-N--KLIPLAYKTCPAGKNLCYKMFMVS-N---KTVPVKRGCIDACPKNSLLVKYVCCNTDRCN--------
9.  LKC-N--KLVPLFYKTCPAGKNLCYKMFMVS-N---LTVPVKRGCIDVCPKNSALVKYVCCNTDRCN--------
10. RKC-N--KLVPLFYKTCPAGKNLCYKMFMVS-N---LTVPVKRGCIDVCPKNSALVKYVCCNTDRCN--------
```

Fig.3. Comparison of the amino acid sequences of long- and short-neurotoxins, cytotoxins, and CLBPs. 1, *N. naja siamensis* α-cobratoxin [12]; 2, *N. naja atra* cobrotoxin [11]; 3, *N. naja* and *N. naja atra* CLBPs (present study); 4, *N. naja siamensis* CLBP [5]; 5, *N. naja atra* and *N. naja siamensis* cytotoxins-I [4,8]; 6, *N. naja atra* cytotoxin-III [10] and *N. naja siamensis* cytotoxin-IV [4]; 7, *N. naja siamensis* cytotoxin-II [4]; 8, *N. naja siamensis* cytotoxin-III [4]; 9, *N. naja atra* cytotoxin-II [9]; 10, *N. naja atra* cytotoxin-IV [9]. The boxed residues indicate the amino acids which differ from those of *N. naja atra* cytotoxin-II.

of the *N. naja* CLBP described above. Thus, both CLBPs were found to be composed of 62 amino acid residues as shown in figs 1 and 2, giving a molecular mass of 7007 Da.

4. DISCUSSION

The determined sequences of the *N. naja* and *N. naja atra* CLBPs were different from those already reported [2,3] in having Ala-Thr-Leu-Lys-Lys-Phe-Pro-Leu-Lys at residues 25–33, in place of Leu-Phe-Pro-Lys-Ala-Thr-Leu-Lys. These differences have probably arisen from misalignments of the tryptic peptides in the previous studies.

As shown in fig.3, the corrected sequences of *N. naja* and *N. naja atra* CLBPs were compared with the reported sequences of *N. naja siamensis* CLBP [5], *N. naja atra* and *N. naja siamensis* cytotoxins [4,8–10], *N. naja atra* short-neurotoxin (cobrotoxin) [11], and *N. naja siamensis* long-neurotoxin (α-cobratoxin) [12]. When the sequences of the *N. naja* and *N. naja atra* CLBPs were compared with that of *N. naja siamensis* CLBP [5], only two amino acid replacements were observed. This conservativeness of the sequence of CLBP among these cobra venoms might suggest that CLBP has some important physiological roles.

On the basis of the results of X-ray crystallographic studies [13], cytotoxins and neurotoxins share a similarity in structure, with three loops forming a three-finger shape. Since the amino acid sequences of CLBPs are similar to those of cytotoxins and neurotoxins, the three-dimensional structures of CLBPs were thought to be similar to those of cytotoxins. When the sequences of CLBPs were compared with those of cytotoxins, remarkable amino acid replacements were observed at

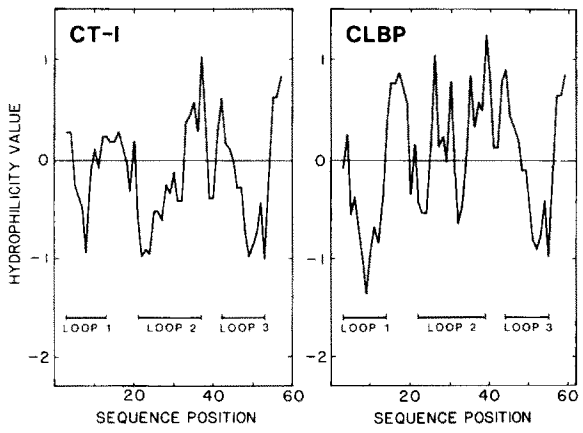

Fig.4. Hydrophilicity profiles of CLBP and cytotoxin-I from *N. naja atra*. The three loops in their three-dimensional structures are indicated.

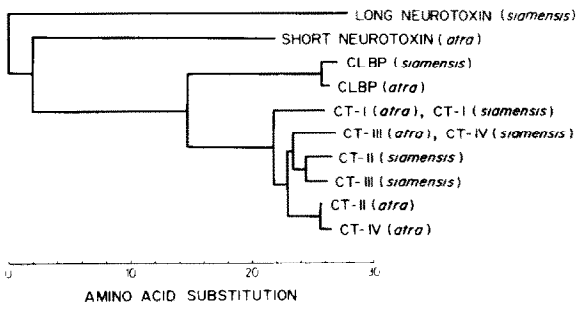

Fig.5. A phylogenetic tree of short-and long-neurotoxins, cytotoxins (CT) and CLBPs. Branch lengths were calculated on the basis of the amino acid difference matrices.

residues 25–30, where hydrophobic amino acid residues in cytotoxins were replaced by hydrophilic residues in CLBP. The hydrophilicity profile of CLBP was compared with that of *N. naja atra* cytotoxin I, according to the methods of Hopp and Woods [14] (fig.4). It is obvious that the hydrophobic region of cytotoxin at residues 25–30 changed into hydrophilic in the CLBP molecule. This region is involved in the second loop of the cytotoxin molecule [13]. As cytotoxins express their cytotoxic activities by interacting with membrane phospholipids, the low cytotoxicities of CLBPs could be ascribable to the replacement of the amino acid residues at the second loop by hydrophilic residues, which may lead to the low binding ability to the membrane. It seems likely that the integrity of the hydrophobic amino acid residues in the second loop of cytotoxins is essential for the cytotoxicity, suggesting that the three-dimensional structure of cytotoxins is very important for their biological functions.

Fig.5 shows the most probable molecular phylogenic tree. The tree was constructed according to Fitch and Margoliash [15] on the basis of the matrix showing amino acid difference among the proteins shown in fig.3. It was suggested that the cytotoxin and CLBP genes were derived by the gene duplication of the short-neurotoxin gene and that the following gene duplication derived the cytotoxin and CLBP genes.

REFERENCES

[1] Dufton, M.J. and Hider, R.C. (1983) CRC Crit. Rev. Biochem. 14, 113–171.
[2] Takechi, M., Tanaka, Y. and Hayashi, K. (1985) Biochem. Int. 11, 795–802.
[3] Takechi, M., Tanaka, Y. and Hayashi, K. (1987) Biochem. Int. 14, 145–152.
[4] Ohkura, K., Inoue, S., Ikeda, K. and Hayashi, K. (1988) Biochim. Biophys. Acta 954, 148–153.
[5] Inoue, S., Ohkura, K., Ikeda, K. and Hayashi, K. (1987) FEBS Lett. 218, 17–21.
[6] Crestfield, A.M., Moore, S. and Stein, W.H. (1963) J. Biol. Chem. 238, 622–627.
[7] Ambler, R.P. (1972) Methods Enzymol. 25, 262–272
[8] Hayashi, K., Takechi, M., Kaneda, N., and Lee, C.Y. (1975) Biochem. Biophys. Res. Commun. 64, 360–366.
[9] Kaneda, N., Sasaki, T. and Hayashi, K. (1977) Biochim. Biophys. Acta 491, 53–66.
[10] Hayashi, K., Takechi, M., Kaneda, N. and Sasaki, T. (1976) FEBS Lett. 66, 210–214.
[11] Yang, C.C., Yang, H.J. and Huang, J.S. (1969) Biochim. Biophys. Acta, 188, 65–77.
[12] Arnberg, H., Eaker, D. and Karlsson, E., unpublished results, cited by Dayhoff, M.O. (1976) in: Atlas of Protein Sequence and Structure, vol. 5, suppl. 2 (Dayhoff, M.O. ed.) National Biomedical Research Foundation, Washington, DC.
[13] Rees, B., Samama, J.P., Thierry, J.C., Gilibert, M., Fischer, J., Schweitz, H., Lazdunski, M. and Moras, D. (1987) Proc. Natl. Acad. Sci. USA 84, 3132–3136.
[14] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3828.
[15] Fitch, W.N. and Margoliash, E. (1967) Science 155, 279–284.