

①

1999年3月

阜阳师范学院学报(自然科学版)

March 1999

第16卷第1期

Journal of Fuyang Normal College (Natural Science)

Vol. 16 NO. 1

信息生物学技术在人神经元蛋白 17.3 (NP17.3) 结构和功能研究中的应用

崔映宇^① 余龙^② 龚若沐 戴方珍 董长宇

(复旦大学遗传学研究所, 遗传工程国家重点实验室 上海 200433)

Q510.1
Q811.4

摘要 本文利用信息生物学技术, 对新近克隆的人神经元蛋白基因(np17.3)所编码的人神经元蛋白 NP17.3 的一、二级结构特点及其理化特性进行分析, 同时就其空间三维构象进行了模拟, 为研究其在活体内的功能提供了理论指导。

关键词 信息生物学、人神经元蛋白 NP17.3、国际互联网、PCGENE 软件包、123D 软件包

分类号 Q954.52

结构 功能

0 前言

信息生物学(Bioinformatics)是计算机科学与生命科学联姻的结晶。计算机技术的发展、国际互联网的涌现、信息高速公路的开通及其广泛延伸使生物学的研究方法正经历着一场深刻的质的变革。目前, 信息生物学在对蛋白质进行空间构象模拟、晶体结构分析及潜在功能预测方面的强大功能已初露端倪。基于此, 本文作了一些有益的探索, 简述如下。

1 材料和方法

1.1 材料

NP17.3 全长 mRNA、PCGENE 软件包、国际互联网、微型计算机 DTK586、IBM pentium[®] II、COMPAQ586、打印机 Deskjet500、EPSON stylus_{TM}COLOR600。

1.2 方法

1.2.1 一级结构和二级结构分析

利用 PCGENE 软件包进行

具体操作如下: MS-DOS 状态下

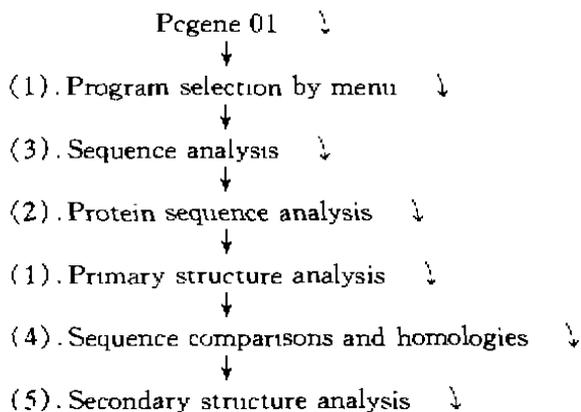
```
C: \>cd pcgene ↓
C: \pcgene>pcgene ↓
```

• 国家自然科学基金项目(39680019)资助课题;

①安徽省阜阳师范学院生物讲师, 硕士研究生, 通讯联系人;

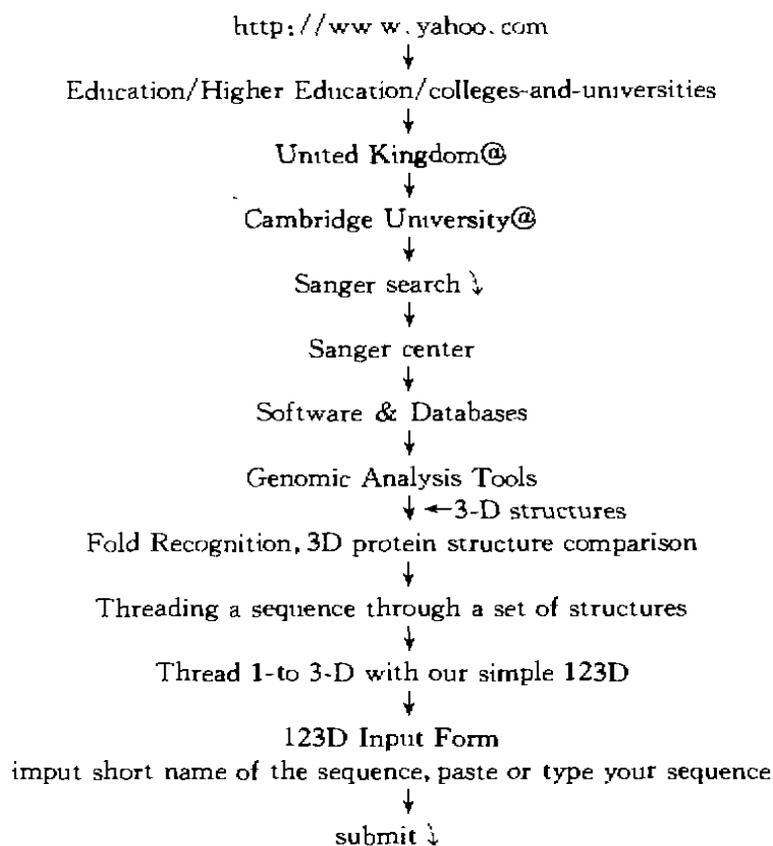
②复旦大学遗传学研究所教授, 博士生导师。

收稿日期: 1998-10-06

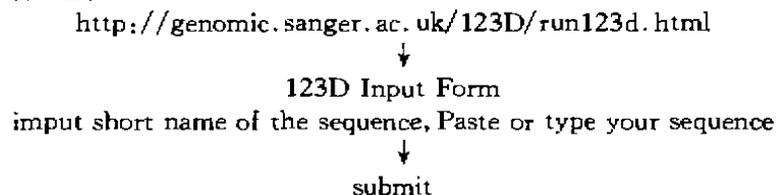


1.2.2 三维空间构象模拟

通过国际互联网,应用 123D 软件包的系列分析软件,进行目标蛋白与公用库存蛋白的同源比较及其空间构象模拟。具体操作如下:



或直接进入下列网址:



1.2.3 蛋白功能预测

据一、二级结构特点,理化特性,信号序列和功能位点的有无及分布,空间结构特点等综合进行。

2 结果

2.1 NP17.3 蛋白氨基酸组成及分子量

```

          5          10          15          20          25          30
I M A A G L F G L S A R R P L A A A A T R G L P A A R V R W E
3I S S F S R T V V A P S A V A G K R P P E P T T P W Q E D P E
6I P E D E N L Y E K N P D S H G Y D K D P V L D V W N M R L V
9I F F F G V S I I L V L G S T F V A Y L P D Y R M K E W S R R
12I E A E R L V K Y R E A N G L P I M E S N C F D P S K I Q L P
15I E D E

```

Number of residues:153.

Molecular weight(MW):17300;Checking number(CN):122915

Amino acid composition

15A	1C	1H	4M	5T
13R	2Q	4I	7F	4W
5N	14E	13L	15P	5Y
9D	8G	6K	11S	11V

图 1 NP17.3 蛋白氨基酸组成

2.2 蛋白等电点分析

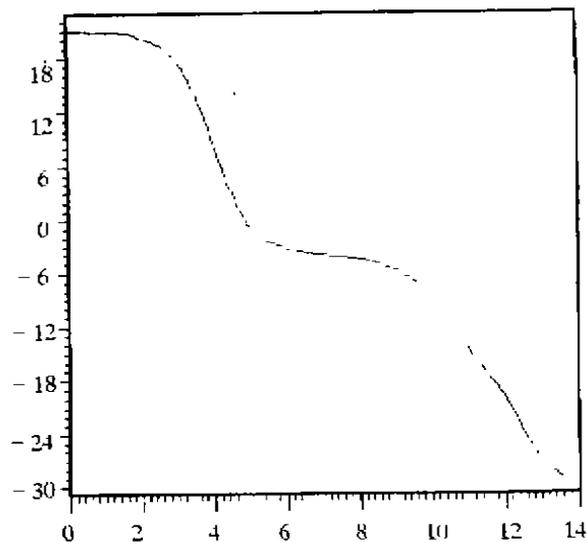
Residues and pK values
taken in account in the
computation.

N-ter (+) Met, pK:9.21
C-ter (-) Glu, pK:2.19

Arg (+) 13, pK:12.48
Lys (+) 6, pK:10.79
His (+) 1, pK:6

Asp (-) 9, pK:3.65
Glu (-) 14, pK:4.25
Cys (-) 1, pK:8.35
Tyr (-) 5, pK:10.13

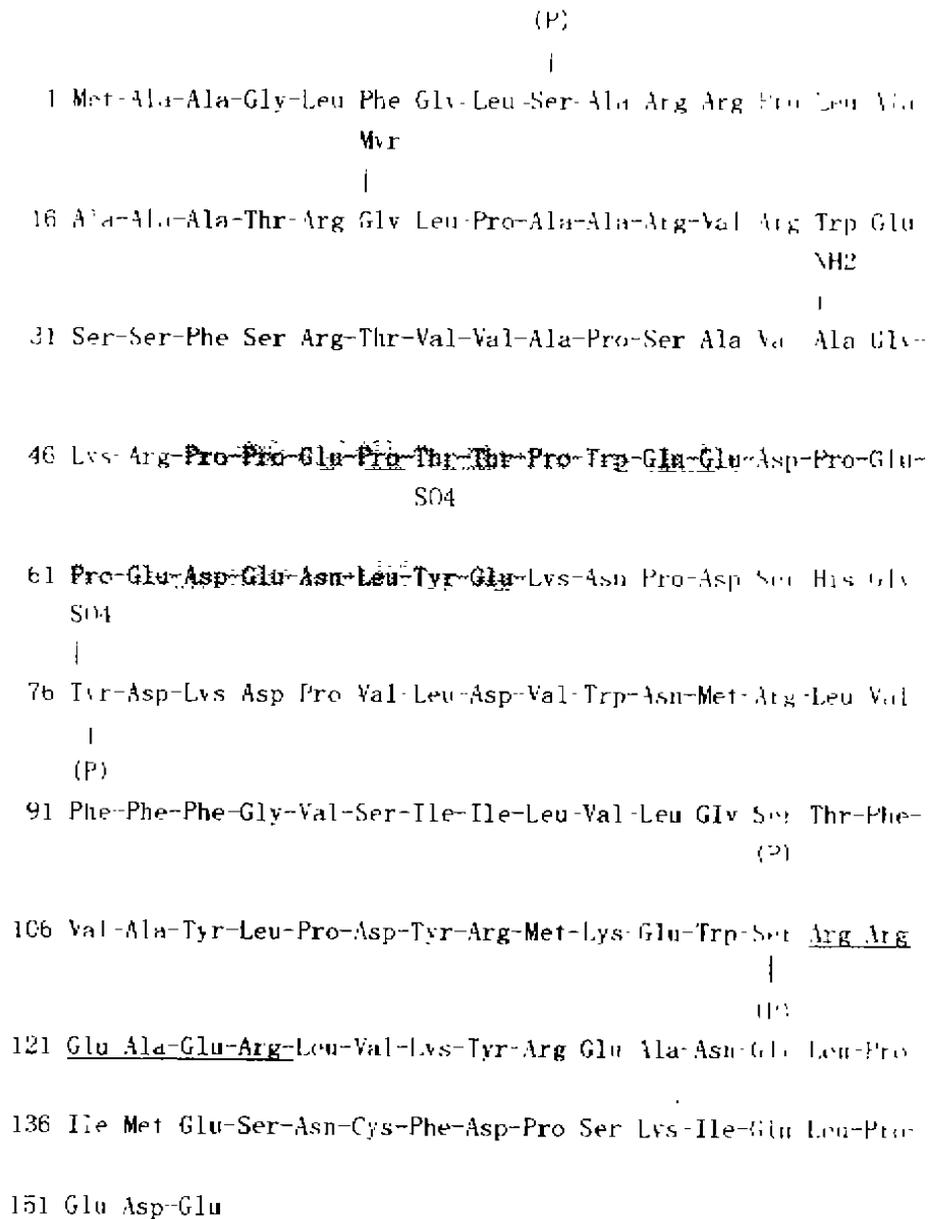
Isoelectric point: 4.89



Curve of the charge of protein NP17.3 as a function of the pH (from 0 to 14)
Calculated on the complete sequence, 153 residues.

图 2 NP17.3 蛋白氨基酸电离常数与等电点

2.3 蛋白位点分析(包括"PEST"位点)

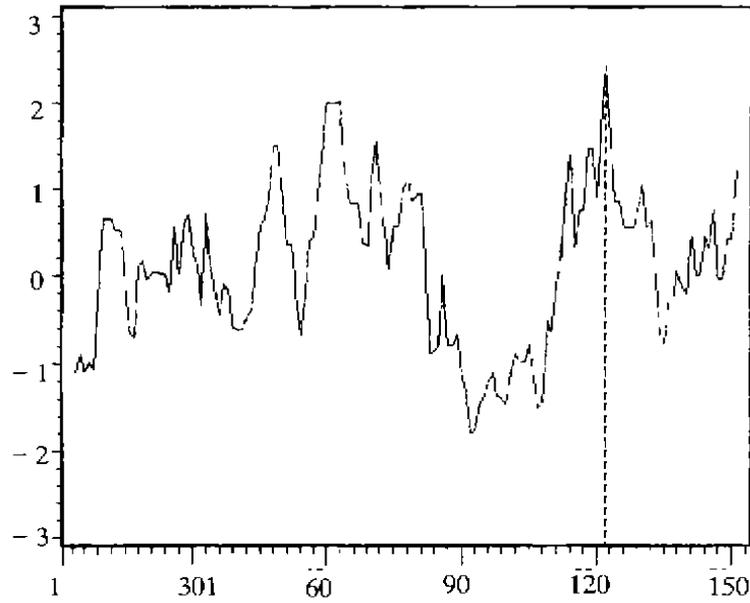


Myr: stands for a myristyl group(十四烷基化基团); NH₂: stands for an amide group(酰胺化基团);

(P): stands for a phosphate group(磷酸化基团); (SO₄): stands for a sulfate group(硫化基团)
阴影区示氨基酸残基"PEST"区域, 下划线处"_____"示潜在抗原决定簇区。

图3 人神经元蛋白 NP17.3 蛋白位点分析

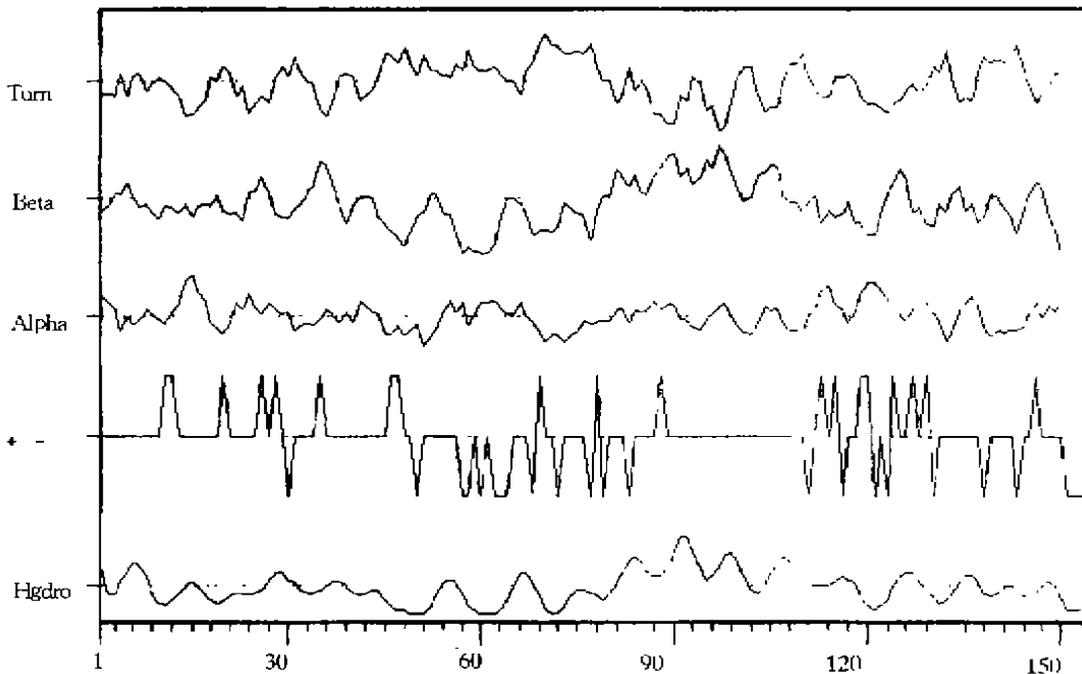
2.4 NP17.3 的氨基酸亲水性分布图



NP17.3 蛋白序列的亲水性特征,以平均每组 6 个氨基酸的组群进行计算机分析结果。图中显示,整个蛋白序列中有三个高亲水性区:60-65, 68-73, 119-124, 其中 119-124 为一潜在抗原决定簇区。尽管该蛋白亲水性占优势,但也包含一相对疏水性区 89-109,其每侧各有一高度亲水性的侧翼区。

图 4 人神经元蛋白 NP17.3 氨基酸亲水性分布图

2.5 二级结构的 NOVOTNY 法分析结果



plot of secondary structure curves for sequence NP17.3 Protein From position 1 to 153.

Turn...reverse turn propensity(转角倾向);Beta...Beta sheet propensity(β -折叠倾向);

Alpha...Alpha helix propensity(α -螺旋倾向);+ - ...charged residues profile(氨基酸残基所带电荷特征);

Hydro...hydrophobicity profile(疏水性特征).

图 5 人神经元 NP17.3 二级结构 NOVOTNY 法分析结果

图中显示, NP17.3 蛋白的 87-109 位氨基酸残基区具高疏水性, 整个蛋白被此疏水区分隔成两个亲水区域。48-68 位氨基酸残基高亲水区为快速衰变蛋白特征性的“PEST”区 (PPEPTTPWQEDPEPEDENLYE) (参见图 3), 高亲水性的 119-124 位氨基酸残基区 (RREAER) 为一潜在抗原决定族区, 与鼠 NP15.6 蛋白的抗原决定族区 69-74 位氨基酸残基区 (RREAER) 序列完全一致。从氨基酸残基所带电荷性质来看, NP17.3 为一酸性质白 (PI 4.89)。

2.6 NP17.3 的三维空间构象图

应用 123D 软件包的系列分析软件, 进行蛋白质的同源比较及其空间构象模拟, 据以推测其功能。

http://www.yahoo.com 上找到 Sanger center 打开 123D face, PDB header 下进行模拟, 结果如图 6:

3 讨论

人神经元蛋白 NP17.3, 是一个由 153 个氨基酸残基组成的蛋白质。计算机分析结果显示其分子量为 17.3KDa (故暂命名其为 NP17.3) (见图 1), 等电点为 4.89 (见图 2), 因而为一酸性蛋白。该蛋白 48~68 (PPEPTTPWQEDPEPEDENLYE) 位氨基酸残基区为一“PEST 域” (Pro-Glu-Ser-Thr) (见图 3)。1986 年, Rogers, Wells 和 Rechsteiner 首次提出“PEST 区域”假说^[1], 认为胞间半衰期短于 2 小时的真核蛋白氨基酸序列中包含一个或多个富含脯氨酸(P)、谷氨酸(E)、丝氨酸(S)和苏氨酸(T)的区域, 此区域两侧通常为带正电荷的氨基酸残基, 故谓“PEST 区”; 而半衰期长于 20 小时的真核蛋白氨基酸序列中则很少有此区域。1987 年, 他们又在一篇综述中列举了 23 种含“PEST 区域”的快速衰变蛋白^[2]。PCGENE 分析“PEST 区域”程序中, 将 PEST 分值等于或大于 5 者定义为标准的“PEST 区”, 而 NP17.3 的 PEST 分值高达 20.54, 强烈提示它是一种含“PEST”的蛋白质。兹将其与部分 PEST 蛋白显示于表 1。

已知的含 PEST 域蛋白包括有, 原癌基因 (myc, fos, myb) 产物: C-Myc, C-Fos, C-Myb; 酶: 鸟氨酸脱羧酶 ODC, 羟甲基戊二酸单酰辅酶 A (HMG-CoA) 还原酶 (NADPH), 酪氨酸氨基转移酶 (TAT) 和信号传导途径的组分: 激酶、甾体化合物受体、压阀钙离子通道蛋白 α -2/ δ 亚基、初粘素-X 等。NP17.3 蛋白中 PEST 区域的存在暗示该神经元蛋白可能为神经信号传导途径的组分、且半衰期短于两小时, 再结合人 NP17.3 与鼠 NP15.6 两蛋白结构的保守性,



图 6 NP17.3 蛋白天然构象模拟图

表 1 部分快速衰变蛋白的氨基酸顺序

蛋白质	PEST 分值	氨基酸残基号	氨基酸顺序	半衰期 (小时)
C-Myc	8.6	206-241	KSCASQDSSAFSPSSDLSLSTESSPQ	0.5
	25.4	241-269	GSPEPLVLH HEETPPTTSSDSEEEQEEDEEIDVVSVEK	
腺病毒 早期蛋白 (E1A)	11.6	125-149	HEAGFPSPDDEDEEGEEFVLDYVEH	0.5
	11.8	177-202	RTCGMFVYSPVSEPEPEPEPEPAR	
	13.3	223-244	RECHSSTDSCDSGPSNTPPEIH	
C-Fos	10.1	128-139	KVEQLSPEEEEK	0.5
	5.7		205-250	
鸟氨酸 脱羧酶 (ODC)	5.2	423-449	HGFPPEVEEQD DGTLPMSCAQESGMDR	0.5
HMG-CoA 还原酶	5.5	429-442	RTQELEIELPSEPR	1.5~3
α -酪蛋白	6.6	58-79	KEMEAESISSSEEIVPNSVQEK	2~5
β -酪蛋白	6.6	1-25	RELEELNVPGEIVESLSSSEESTTR	2~5
(压阀)钙离 子通道 α -2 / δ 亚基前体*	5.32	719-761	ARFVVTDGGITRVYPKEAGENWQENPE TYEDSFYKRSLDNDNY	待定
韧带素-X*	28.33	1441-1463	AEEEPSPTEPSTEAPPEPTEEPL	待定
人神经元 蛋白 (NP17.3)	20.54	48-68	PPEPTTPWQEDPEPEDENLYE	待定

PEST 区氨基酸顺序以单字母代码表示

A—丙氨酸; C—半胱氨酸; D—天冬氨酸; E—谷氨酸; F—苯丙氨酸; G—甘氨酸; H—组氨酸; I—异亮氨酸; K—赖氨酸; L—亮氨酸; M—蛋氨酸; N—天冬酰胺; P—脯氨酸; Q—谷氨酰胺; R—精氨酸; S—丝氨酸; T—苏氨酸; V—缬氨酸; W—色氨酸; Y—酪氨酸;

*—示 NP17.3 的同源蛋白。

我们推测该蛋白可能在人神经发育和神经信号传导中起重要作用^[3]。

此蛋白还有如下特征:它包含有 14 个谷氨酸残基,主要分布在离 N 端 30 个氨基酸残基始的近 40 个残基区内和 C 端,正是这两个区域决定了该蛋白的酸性特征(PI4.89),所含的 15 个脯氨酸残基大部分也分布在这两个区域;亲水性分布图(图 4)揭示,整个蛋白序列中,有三个高亲水性区:60~65(Glu-Pro-Glu-Asp-Glu-Asn)、68~73(Glu-Lys-Asn-Pro-Asp-Ser)、119~124(Arg-Arg-Glu-Ala-Glu-Arg),其中 119~124 为一潜在抗原决定簇区,尽管该蛋白亲水性占优势,但也包含一相对疏水性区 89~109(LVFFFGVAILVVGSTFVAYL),其每侧各有一高度亲水性的侧翼区, NP17.3 蛋白似乎被疏水性区域分隔成两个亲水域,阐述这些亲水域的功能是十分有趣的事情;以 GGBSM 法推测其二级结构含三个螺旋区、一

个折叠片和两个无规卷曲(39~80、130~153),其中 87~107(MRLVFFFGVAAIL-VLGSTFVA)具高疏水性,为一跨膜螺旋区,95(Val)~106(Phe)形成跨膜多体;而且 PCGENE 的 PROSITE 程序分析表明该蛋白还是蛋白激酶 C、酪蛋白激酶 II、酪氨酸激酶的底物,含蛋白激酶 C 磷酸化位点和酪氨酸硫酸化位点各两个,酪蛋白激酶的底物,含蛋白激酶 C 磷酸化位点和酪氨酸硫酸化位点各两个,酪蛋白激酶 II 磷酸化位点、酪氨酸激酶磷酸化位点、N-十四烷基化位点、酰胺化位点各一个(图 3),这些位点的存在说明该蛋白易为多种激酶激活,烷基化和酰胺化位点的存在增强了该蛋白的不稳定性。这从另一个方面论证了 NP17.3 这一 PEST 蛋白的功能特征。

此外,用 PCGENE 软件包分析人 NP17.3 的二级结构发现,其 32.6%的氨基酸参与形成 α 螺旋、15.0%的氨基酸参与形成 β 折叠片;相比之下,鼠 NP15.6 的 25.5%和 19.5%的氨基酸分别参与形成二级结构的 α 螺旋和 β 折叠片(Rogister, B. M. et al. 1996)。有趣的是, α 螺旋结构几乎靠近该蛋白序列的 -COOH 端,从 87 到 107 的 21 个氨基酸残基中有 16 个、将近 80%的氨基酸残基参与 α 螺旋的构成。

综上所述,基于人 NP17.3 还有分子量较小的特征,我们推测 NP17.3 是一种寿命极短的参与神经信号传导的蛋白质,它在神经元及神经元效应器皆可表达,但在神经元效应器表达之后,迅即通过突触后膜、间隙、前膜而进入神经轴突,作为逆向轴流的一种成份参与神经的调节功能,可能在神经元间神经信号传导中发挥中介作用。我们将用重组表达、转基因及定点突变、酵母双杂交系统等技术对其确切的功能作进一步的研究。

参考文献

- 1 Rogers, S Wells, R Rechsteiner, M. Amino Acid sequences common to Rapidly Degraded Proteins; The PEST Hypothesis. Science 1986 Oct;234(17):364-368;
- 2 Rechsteiner, M.; Rogers, S; Rote, K. Protein structure and intracellular stability. Trends in Biochem Sci 1987 Oct;12:390-394
- 3 Rogister, B; Mazy-Servais, C, et al.; Molecular cloning of mouse Np15.6 a novel neuronal protein whose expression is developmentally regulated Proc-Natl-Acad-Sci USA. (private letter)(1996. Oct.);

The Application of Bioinformatical Techniques in the Study of Structure and Function of Human Neuronal Protein NP17.3

Cui Yingyu Yu Long Gong Ruomu Dai Fangzhen Dong Changyu

(State key laboratory of Genetics Engineering, Institute of Genetics, Fudan University Shanghai, 200433)

Abstract: In this article, taking advantage of bioinformatical techniques, we have analyzed the primary and secondary structure of human neuronal protein NP17.3 and its physical & chemical characters. In the meantime, we have also simulated its three dimensional conformation. This work has provided solid theoretical foundations for its function research *in vivo*.

Key Words: Bioinformatics, human neuronal protein 17.3, Internet, PCGENE software, 123D software.