

# BIOINFORMATICS

コンピュータ・ネットワークを用いて遺伝情報とそれに関連するデータベースを活用し、生物学の疑問に答えること

# 生物の実験法

今まで

生物体を使う実験----- *in vivo*

人工的環境(試験管内)--- *in vitro*

新たに

コンピュータを使う----- *in silico*

(マイクロプロセッサのsilicon chipを使うという意味)

# 生物情報データベース

DNA塩基配列 遺伝子 タンパク質 ゲノム

|             |  |
|-------------|--|
| DNA/RNA塩基配列 | GenBank (NCBI, USA)<br>EMBL (UK)<br>DDBJ (Japan) |
| タンパク質配列     | PIR (USA)<br>Swiss-Prot (Swiss)<br>PRF (Japan)   |
| タンパク質立体構造   | PDB (USA)<br>SCOP (UK)<br>CATH (Swiss)           |
| タンパク質モチーフ   | PROSITE (Swiss)                                  |
| ゲノム         | KEGG (Japan)<br>SGR (USA)                        |

# 生物ゲノム・手法データベース

---

| Database Resources  |  |
|---|--|
|  <a href="#">Saccharomyces Genome Database</a>                                 |  |
|  <a href="#">Stanford Microarray Database</a>                                  |  |
|  <a href="#">PharmGKB</a>  |  |
|  <a href="#">The Arabidopsis Information Resource</a>                          |  |
|  <a href="#">A functional genomic resource for Human, Mouse, and Rat genes</a> |  |
|  <a href="#">Candida Genome Database</a>                                      |  |
|  <a href="#">Tetrahymena Genome Database</a>                                 |  |

酵母

マイクロアレイ

薬剤

植物

ヒト, マウス, ラット

カンジダ

テトラヒメナ

---

LOCUS AY049786 425 bp mRNA linear PRI 23-OCT-2001

DEFINITION [Homo sapiens synuclein alpha](#) (SNCA) mRNA, complete cds. ACCESSION AY049786 VERSION AY049786.1

GI:16356656

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 425) AUTHORS Hu,X., Xu,Y., Peng,X., Yuan,J. and Qiang,B.

TITLE Direct Submission JOURNAL Submitted (30-JUL-2001) Department of Biochemistry, Institute of Basic Medical Science, Chinese Academy of Medical Sciences, 5 Dong Dan San Tiao, Beijing 100005, P.R. China

FEATURES Location/Qualifiers source 1..425 /organism="Homo sapiens" /mol\_type="mRNA" /db\_xref="taxon:9606" [gene](#) 1..425 /gene="SNCA" [CDS](#) 3..425 /gene="SNCA" /note="non-A4 component of amyloid precursor" /codon\_start=1 /product="synuclein alpha" /protein\_id="[AAL15443.1](#)" /db\_xref="GI:16356657"

/translation="MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKT  
KEGVVHGVATVAEKTKEQVTNVGGAVVTGVTAVAQKTVEGAGSIAAATGFVKKD  
QLGK NEEGAPQEGILEDMPVDPDNEAYEMPSEEGYQDYEPEA"

```
ccatggatgt attcatgaaa ggactttcaa aggccaagga gggagttgtg gctgctgctg 61
agaaaaccaa acaggggtgtg gcagaagcag caggaaagac aaaagagggt gttctctatg 121
taggctccaa aaccaaggag ggagtgggtc atgggtgtggc aacagtggct gagaagacca 181
aagagcaagt gacaaatgtt ggaggagcag tggtgacggg tgtgacagca gtagcccaga 241
agacagtgga gggagcaggg agcattgcag cagccactgg ctttgtcaaa aaggaccagt 301
tgggcaagaa tgaagaagga gccccacagg aaggaattct ggaagatatg cctgtggatc 361
ctgacaatga ggcttatgaa atgccttctg aggaagggtg tcaagactac gaacctgaag 421
cctaa //
```

# ゲノム DNA RNA タンパク質 構造

## I. 配列情報の比較: 相同性 (homology)

- ・性質の推定
- ・類縁関係
- ・進化的関連性
- ・生物系統学分類学

## II. タンパク質立体構造 (Structural bioinformatics)

- ・構造予測
- ・構造比較
- ・グラフィックス

## III. 遺伝子タンパク質間の相互関係 (Systems biology)

- ・タンパク質間相互作用
- ・生体内反応の流れ
- ・生体反応の予測

# Structural Bioinformaticsにむけて

- ・タンパク質構造構築の原理
- ・タンパク質構造モチーフ
  - ヘリックスと シート
  - ドメイン構造
  - ドメイン構造
  - / ドメイン構造
- ・タンパク質立体構造の形成過程

宿題: タンパク質立体構造の表示と説明

# タンパク質の機能と構造

## Protein classes

|                            |       |
|----------------------------|-------|
| ATPases                    | 角田・山口 |
| Cyclins                    | 平野    |
| GTP-binding proteins       | 工藤    |
| GTP/GDP exchange factors   | 呉・張   |
| GTPase activating proteins | 中村    |
| Molecular chaperones       | 山吹    |
| Moterproteins              | 斉藤    |
| Polymerases                | 井上    |
| Proteases                  | 倉間    |
| Protein kinases            | 田中・松山 |
| Protein phosphatases       | 村岡    |
| Tubulins                   | 吉田    |

タンパク質の機能の説明

他の生物との相同性

その構造



# バイオインフォマティクス

データベースを活用し、生物科学の疑問に答える

## 疑問

生存に必須の遺伝子(タンパク質)とはどのようなものか？

必須タンパク質と非必須タンパク質には何らかの違いがあるか？

必須タンパク質は細胞のどこにどの程度存在するか？

必須タンパク質のサイズは？

必須タンパク質の量は？

細胞内の局在とタンパク質の性質には関連があるか？

生物(酵母)の特徴を知る試み

必須遺伝子は酵母しかわかっていない。

# データベースの構築

|              |       |
|--------------|-------|
| 401          | 角田    |
| 402          | 山口    |
| 403          | 工藤    |
| 404          | 呉     |
| 405          | 張     |
| 406          | 中村    |
| 407          | 山吹    |
| 408          | 斉藤    |
| 409          | 倉間    |
| 410          | 田中    |
| 410          | 松山    |
| 412          | 村岡    |
| 413とまとめ      | 平野・井上 |
| Nonessential | 吉田    |

リスト 遺伝子名 タンパク質機能

アミノ酸数

等電点

分子数 / 細胞

局在部位

ヒトとの相同性

| PLATE | ROW | COLUMN | OFF | Reference Gene Name | Gene Description | Gene Product                               | Protein Info Length (aa)  | pI   | Localization Resources... subcellular | SD             | Localization resources--GFP DB File | Model organism (BLAST Best Hit to human) |           |
|-------|-----|--------|-----|---------------------|------------------|--|---|------|---------------------------------------|----------------|-------------------------------------|--|-----------|
| 1     | 401 | H      | 10  | YOR048C             | DEP8             | Dead Box Protein                           | RNA helicase  | 492  | 7.48                                  | 14000          |                                     | n  | 1.00E-261 |
| 2     | 401 | A      | 4   | YAL030W             | CDC19            | Cell Division Cycle                        | protein kinase  | 500  | 7.68                                  | 281000         | ±24180                              | c  | 1.00E-261 |
| 3     | 401 | B      | 6   | YLL019W             | DDY7             | none                                       | 50 snRNP protein  | 480  | 10.22                                 | 8620           | ±1820                               | n  | 1.00E-261 |
| 4     | 401 | B      | 7   | YLL190C             | DGC7             | none                                       | aspartyl-tRNA synthetase  | 957  | 8.54                                  | not visualized |                                     | c  | 1.00E-261 |
| 5     | 401 | D      | 1   | YLR068W             | FRS1             | phenylalanyl (F)-tRNA Synthetase           | phenylalanine-tRNA ligase subunit                                       | 595  | 5.61                                  | 5440           |                                     | c  | 1.00E-261 |
| 6     | 401 | E      | 1   | YLR106C             | MDJ1             | MDJ1                                       | nuclease  | 4910 | 4.73                                  | 530            |                                     | c/n                                      | 1.00E-261 |
| 7     | 401 | E      | 4   | YLR117C             | CLF1             | none                                       | pre-mRNA splicing factor  | 607  | 5.57                                  | 2140           | ±264                                | n  | 1.00E-261 |
| 8     | 401 | G      | 2   | YMR229C             | RXN5             | Ribosomal RNA Processing                   | 50 snRNP protein  | 1729 | 6.1                                   | 8060           |                                     | n  | 1.00E-261 |
| 9     | 401 | G      | 11  | YMR280W             | AGM155           | Human Gap Homolog                          | none  | 871  | 7.26                                  | 521            |                                     | c/n                                      | 1.00E-261 |
| 10    | 401 | G      | 12  | YMR280C             | AKS1             | Helicase Associated with Spt1              | RNA-dependent helicase (guanine)  | 905  | 10.13                                 | not visualized |                                     | none                                     | 1.00E-261 |
| 11    | 401 | H      | 6   | YMK280C             | PKS1             | none                                       | karyopherin   | 1089 | 4.48                                  | 15600          |                                     | c/n                                      | 1.00E-261 |
| 12    | 401 | H      | 11  | YOR048C             | KIX1             | none                                       | 3'-5' exonuclease   | 1008 | 8.8                                   | 820            |                                     | n  | 1.00E-261 |
| 13    | 401 | A      | 12  | YOR030W             | MYO3             | MYO3                                       | class V myosin  | 1874 | 8.28                                  | 4380           |                                     | Bad Neck c/n Cell periphery Bad          | 1.00E-261 |
| 14    | 401 | B      | 2   | YOR039C             | AL17             | 5L-Aryl-tRNA synthetase                    | RNA Polymerase I subunit  | 890  | 9.27                                  | 33800          | ±6740                               | c  | 1.00E-261 |
| 15    | 401 | B      | 4   | YOR031W             | RNA190           | RNA Polymerase A                           | RNA polymerase I subunit  | 1804 | 7.08                                  | 2840           |                                     | n  | 1.00E-261 |
| 16    | 401 | C      | 2   | YOL021C             | DOS1             | none                                       | 3'-5' exonuclease complex subunit                                       | 1001 | 8.63                                  | 800            |                                     | c/n                                      | 1.00E-261 |
| 17    | 401 | C      | 12  | YPL294W             | HPS25            | none                                       | peptide kinase (tyrosine)   | 494  | 10.07                                 | 18300          | ±1890                               | Bad Neck c/n Bad                         | 1.00E-261 |
| 18    | 401 | D      | 6   | YPL229W             | RFH2             | RFV-B-like                                 | transcriptional regulator   | 471  | 5.08                                  | 3030           |                                     | n  | 1.00E-261 |
| 19    | 401 | F      | 4   | YPL168W             | CDC60            | Cell Division Cycle                        | leucine-tRNA ligase   | 1090 | 5.68                                  | 158000         |                                     | c  | 1.00E-261 |
| 20    | 401 | F      | 6   | YPL175W             | GPT14            | Suppressor of Ty                           | N-acetylglucosaminyl-phenylalanyl-tRNA synthetase                       | 452  | 7.25                                  | 1480           |                                     | n  | 1.00E-261 |
| 21    | 401 | F      | 10  | YBR196C             | PGI1             | none                                       | glucose-6-phosphate isomerase   | 954  | 8.43                                  | 91000          | ±34000                              | c  | 1.00E-261 |
| 22    | 401 | F      | 12  | YOR020W             | CDC47            | Cell Division Cycle                        | histone H4 acetyltransferase  | 845  | 5                                     | 100            |                                     | c/n                                      | 1.00E-261 |
| 23    | 401 | H      | 6   | YOR062W             | osecure1         | Long-Chain Base                            | serine palmitoyltransferase component                                   | 561  | 8.05                                  | 54500          |                                     | ER                                       | 1.00E-261 |
| 24    | 401 | A      | 4   | YOR091C             | RL1              | Release 1 inhibitor                        | ATP-binding cassette (ABC) superfamily transmembrane protein (putative) | 600  | 8.02                                  | 8200           |                                     | c  | 1.00E-261 |
| 25    | 401 | B      | 10  | YOR094W             | RXN3             | none                                       | none  | 430  | 5.20                                  | not visualized |                                     | n  | 1.00E-261 |
| 26    | 401 | D      | 1   | YEL003W             | WCM3             | Mitochondrion Maintenance                  | none  | 871  | 5.18                                  | 39100          | ±6940                               | none                                     | 1.00E-261 |
| 27    | 401 | D      | 9   | YBR012W             | HPS17            | Pre-rRNA Processing                        | helicase like protein   | 1140 | 7.78                                  | 807            |                                     | c/n                                      | 1.00E-261 |
| 28    | 401 | B      | 1   | YMR028W             | GCD17            | General Control Nonfermentable             | translational inhibition factor eIF-2 gamma subunit                     | 827  | 8.77                                  | 20800          | ±2840                               | c  | 1.00E-261 |
| 29    | 401 | B      | 2   | YMR038C             | AAR1             | ATP-binding cassette protein involved in   | trapping protein, ATP binding cassette protein                          | 810  | 8.48                                  | 17000          |                                     | c  | 1.00E-261 |
| 30    | 401 | B      | 4   | YMR043C             | SMT1             | S-Adenylyl-tRNA synthetase                 | S-adenosyl-L-homocysteine hydrolase (putative)                          | 449  | 8.2                                   | not visualized |                                     | c  | 1.00E-261 |
| 31    | 401 | F      | 2   | YOR189C             | TRJ1             | Chromosome Region Maintenance              | lysine-tRNA ligase  | 394  | 8.78                                  | 2710           | ±305                                | c  | 1.00E-261 |
| 32    | 401 | F      | 10  | YOR010W             | CRM1             | Chromosome Region Maintenance              | none  | 1804 | 8.24                                  | 7000           |                                     | n  | 1.00E-261 |
| 33    | 401 | F      | 11  | YMR019C             | DED1             | Defence Essential Domain                   | asparaginyl-tRNA synthetase   | 954  | 5.68                                  | not visualized |                                     | none                                     | 1.00E-261 |
| 34    | 401 | G      | 5   | YMR065C             | RXN9             | Ribosomal RNA Processing                   | weak RNA-dependent ATPase activity which is not specific for rRNA       | 501  | 9.08                                  | not visualized |                                     | none                                     | 1.00E-261 |
| 35    | 401 | G      | 10  | YMR074W             | GMC7             | glutamine (G)-dependent Rad51 synthetase   | glutamine-dependent NAD synthetase                                      | 714  | 6.5                                   | 172            |                                     | c/n                                      | 1.00E-261 |
| 36    | 401 | H      | 10  | YMR104C             | DNA2             | DNA synthesis defective                    | DNA replication helicase  | 1522 | 8.43                                  | low signal     |                                     | n  | 1.00E-261 |
| 37    | 401 | A      | 1   | YMR166C             | KOG1             | Non-rod Cell Growth                        | none  | 1557 | 7.78                                  | low signal     |                                     | m  | 1.00E-261 |
| 38    | 401 | A      | 9   | YCL017C             | MPS1             | none                                       | none  | 497  | 8.25                                  | 504            |                                     | m  | 1.00E-261 |
| 39    | 401 | B      | 1   | YCL054W             | SER1             | Suppressor of Pab1                         | Aspartyl-tRNA synthetase  | 841  | 8.2                                   | not visualized |                                     | n  | 1.00E-261 |
| 40    | 401 | B      | 3   | YCR012W             | PSI1             | 3-Phosphoglycerate Kinase                  | 3-phosphoglycerate kinase   | 410  | 7.77                                  | 214000         | ±130000                             | n/c                                      | 1.00E-261 |
| 41    | 401 | B      | 6   | YLR128W             | SAP2             | COM1 Interacting Protein                   | 50 snRNP protein  | 843  | 5.92                                  | 8620           |                                     | n/c                                      | 1.00E-261 |
| 42    | 401 | B      | 12  | YLR193C             | MM27             | mitochondrial processing peptidase subunit | mitochondrial processing peptidase subunit                              | 462  | 7.04                                  | 1380           | ±180                                | m  | 1.00E-261 |
| 43    | 401 | C      | 1   | YLR198C             | SFC10            | SFC10                                      | ribocyt complex component   | 871  | 8.20                                  | 2380           |                                     | Bad Neck and cell periphery and to       | 1.00E-261 |
| 44    | 401 | C      | 3   | YLR178W             | CEP5             | Centrosome Binding Factor                  | major low affinity 90 kDa centrosome microtubule binding protein        | 493  | 9.8                                   | 23800          | ±2870                               | n  | 1.00E-261 |
| 45    | 401 | C      | 6   | YLR188W             | PEP3             | Peptide Inhibitor 3/PO                     | none  | 676  | 4.28                                  | 43000          |                                     | n/c                                      | 1.00E-261 |

酵母遺伝子数 約6000個

必須遺伝子 1142

非必須遺伝子 94を比較

ヒトとの相同性は必須遺伝子と  
非必須遺伝子で違うか同じか？

# 必須遺伝子

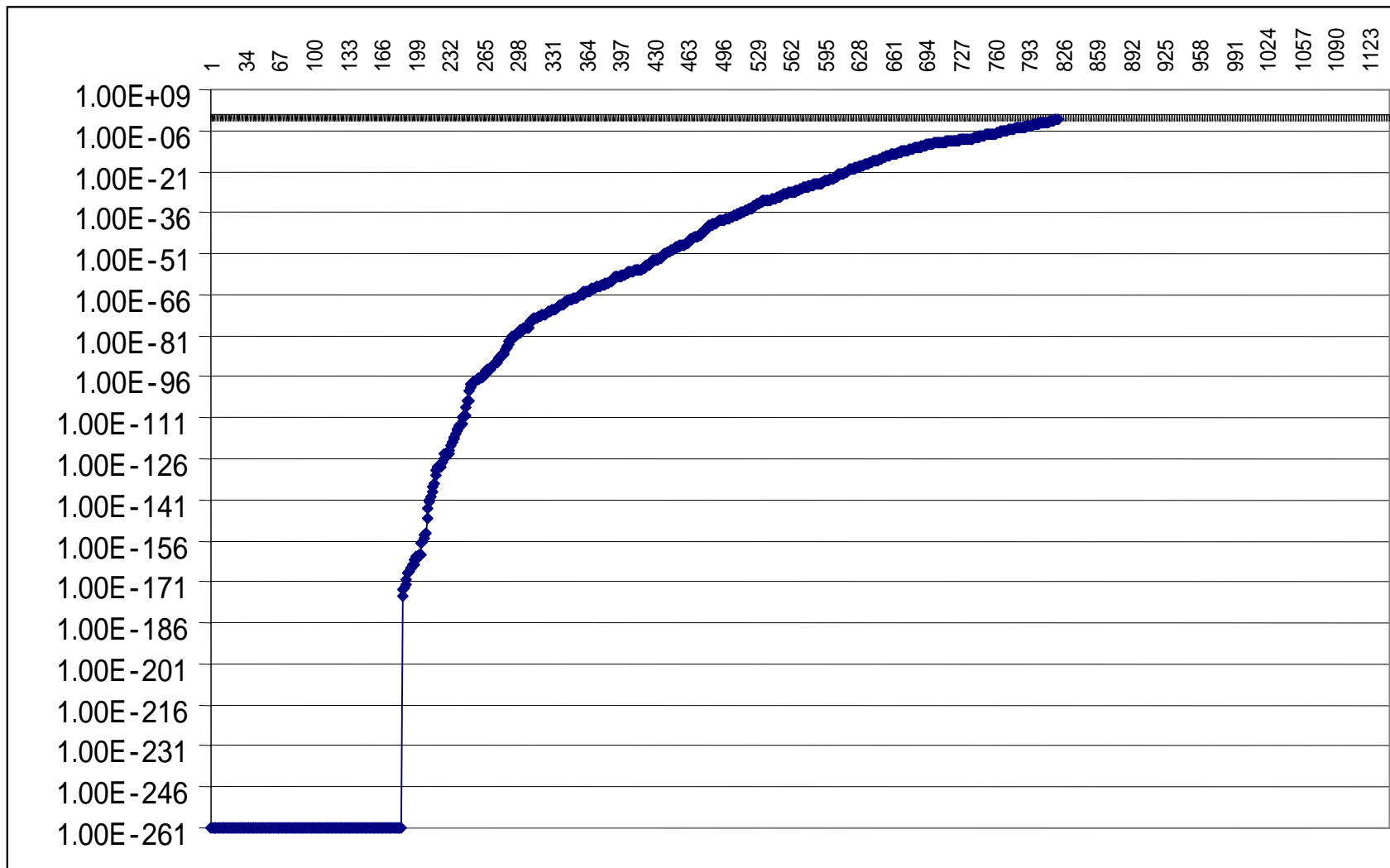
低い



相同性



高い



# 非必須遺伝子

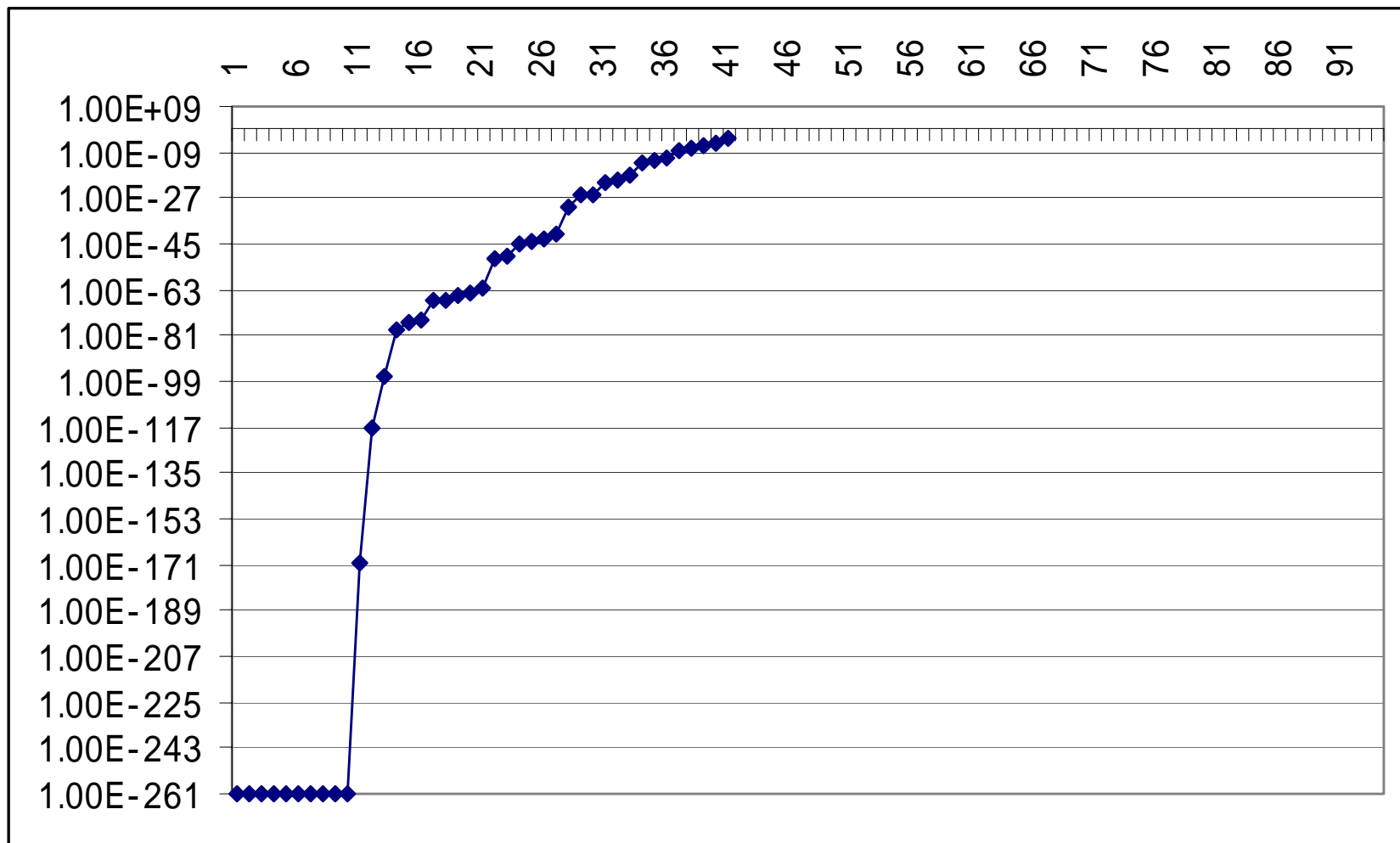
低い



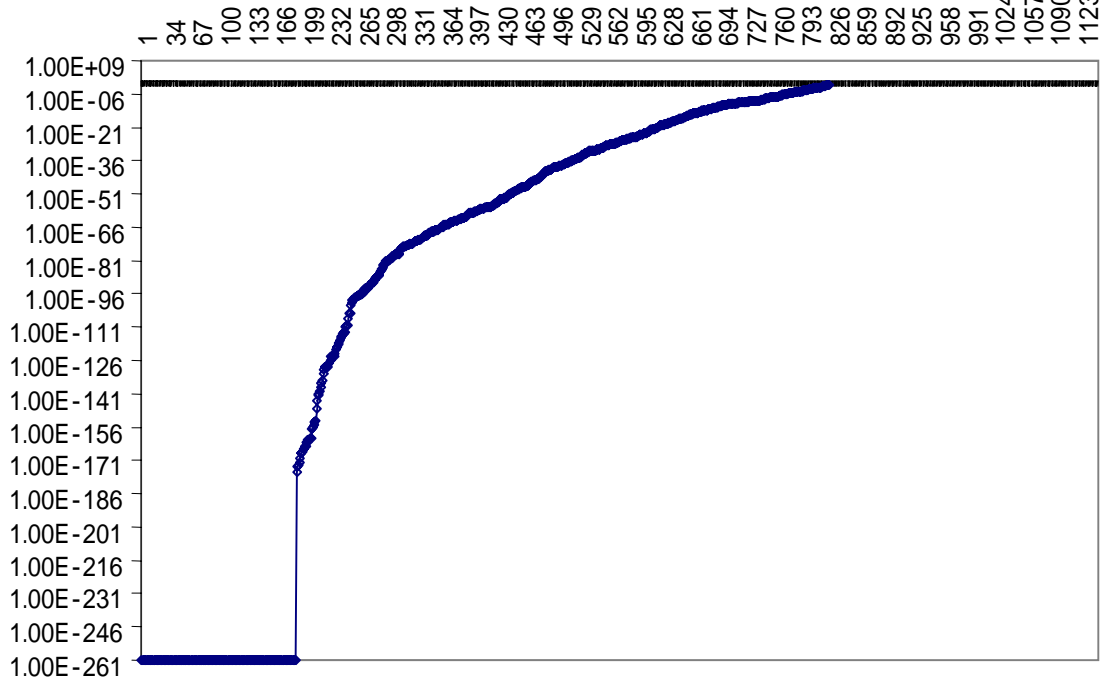
相同性



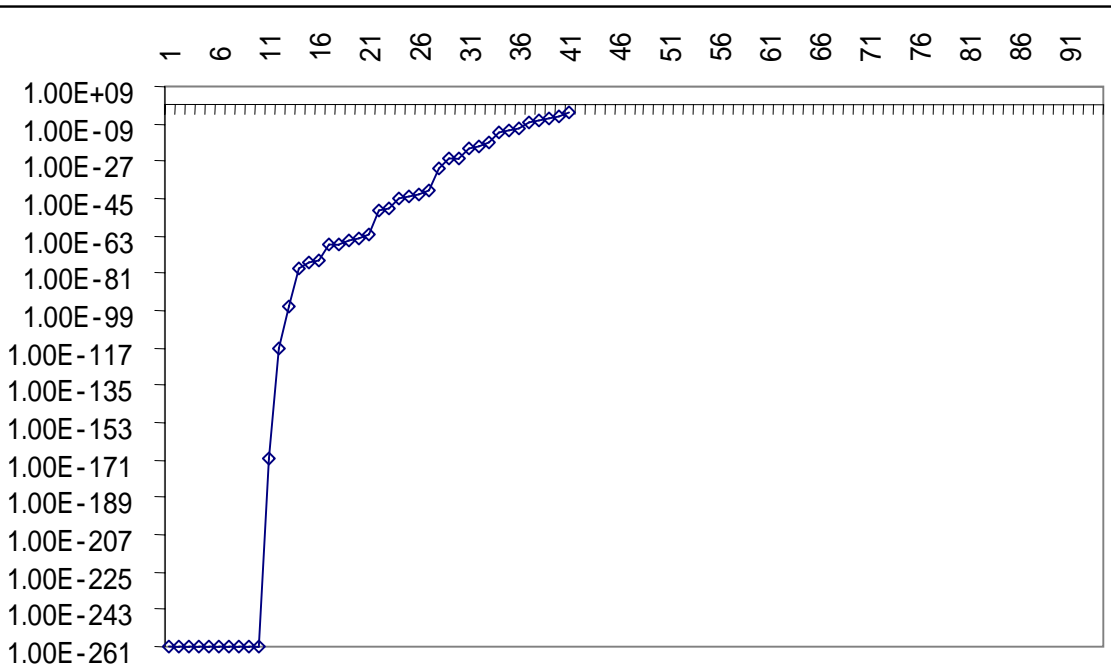
高い



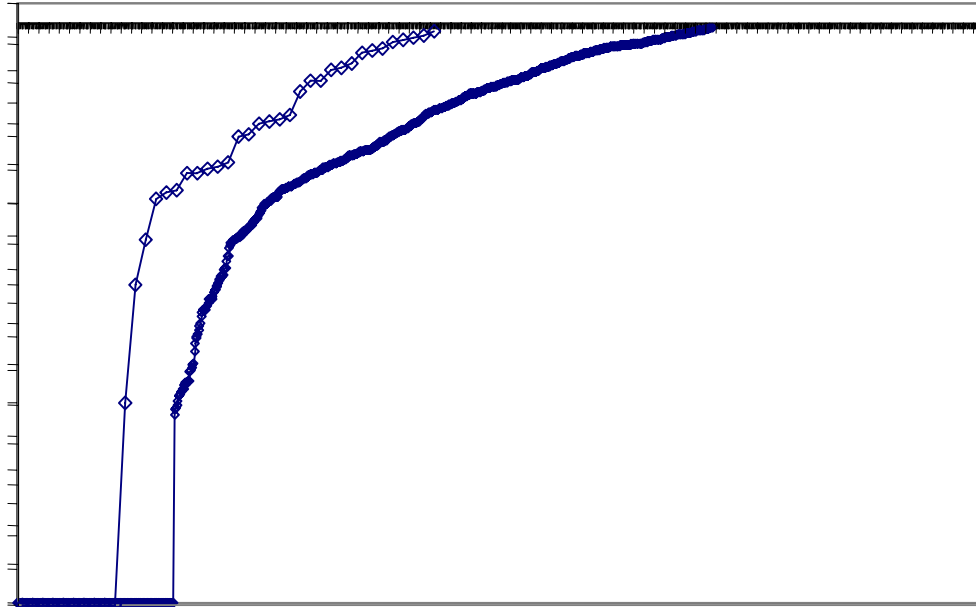
# 必須遺伝子



# 非必須遺伝子



必須遺伝子の方が相同性の高いタンパク質の割合が多い

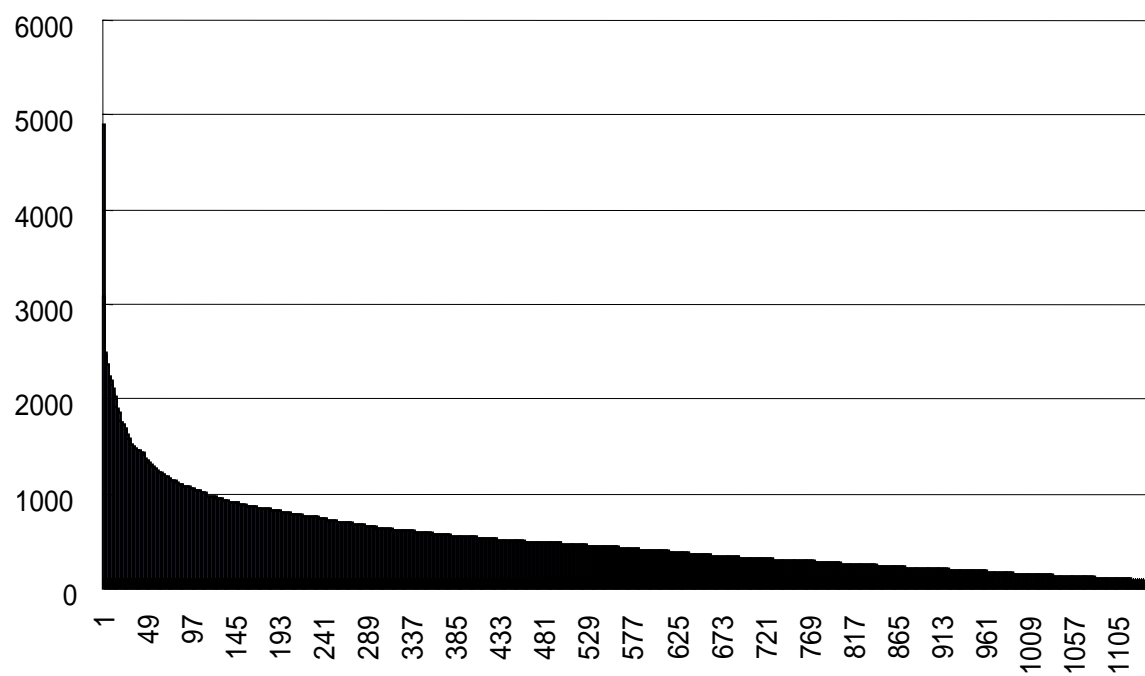


非必須遺伝子にはヒトまで保存されたタンパクが少ない



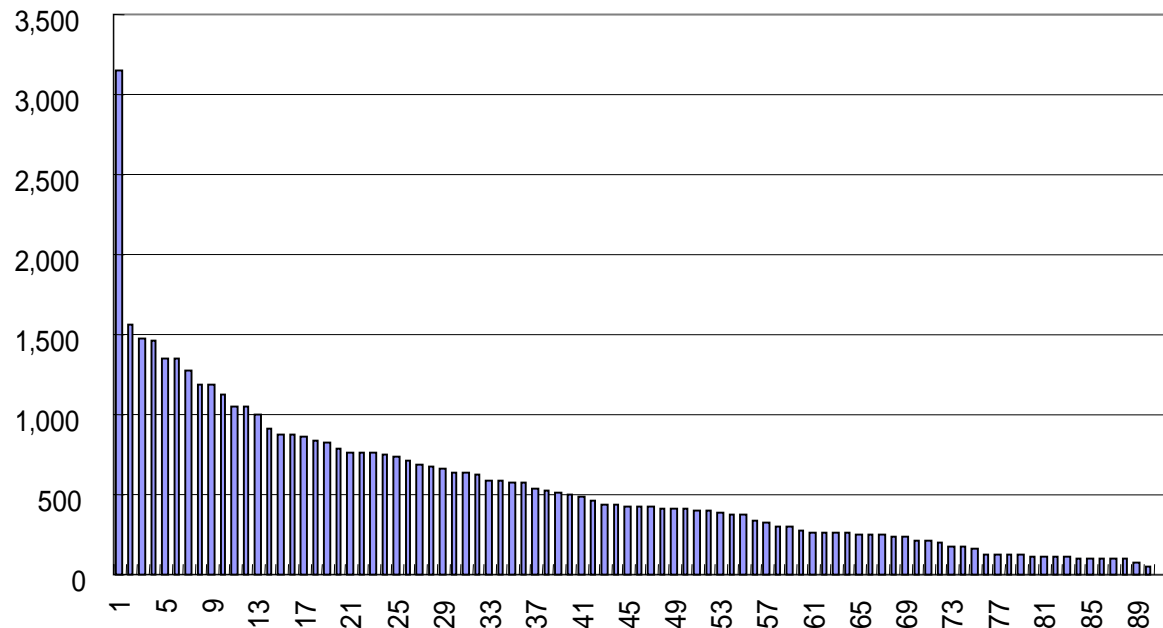
# タンパク質 サイズ

アミノ  
酸数



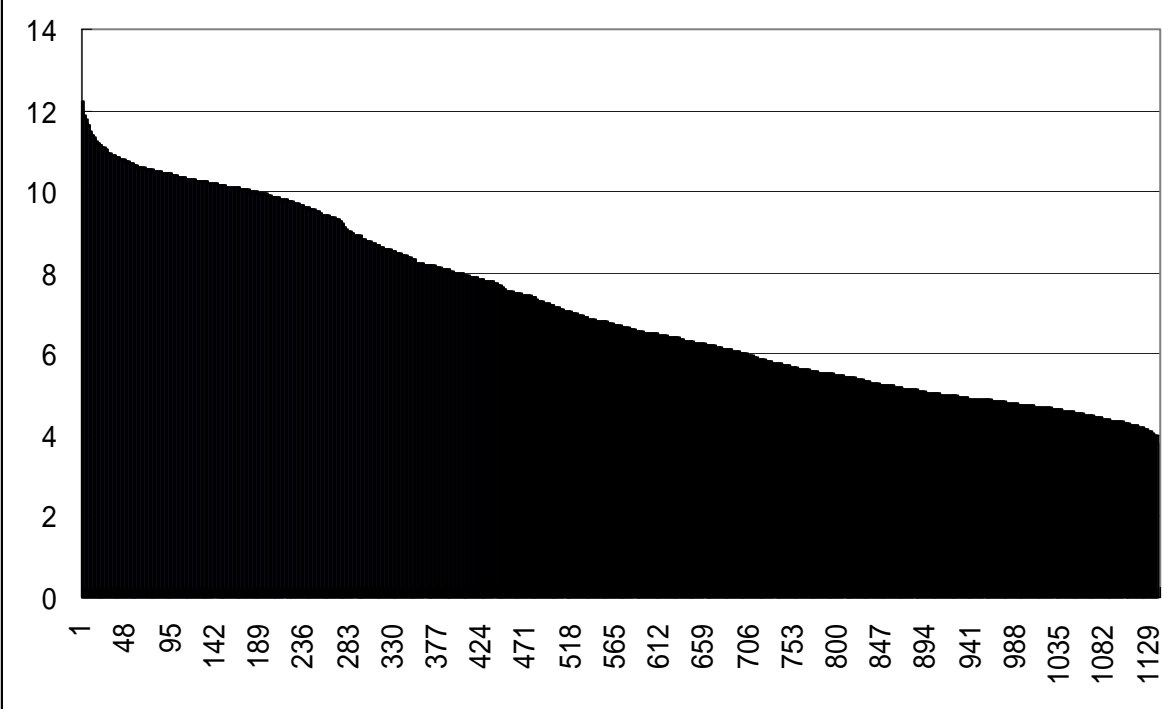
必須  
517.6

アミノ  
酸数

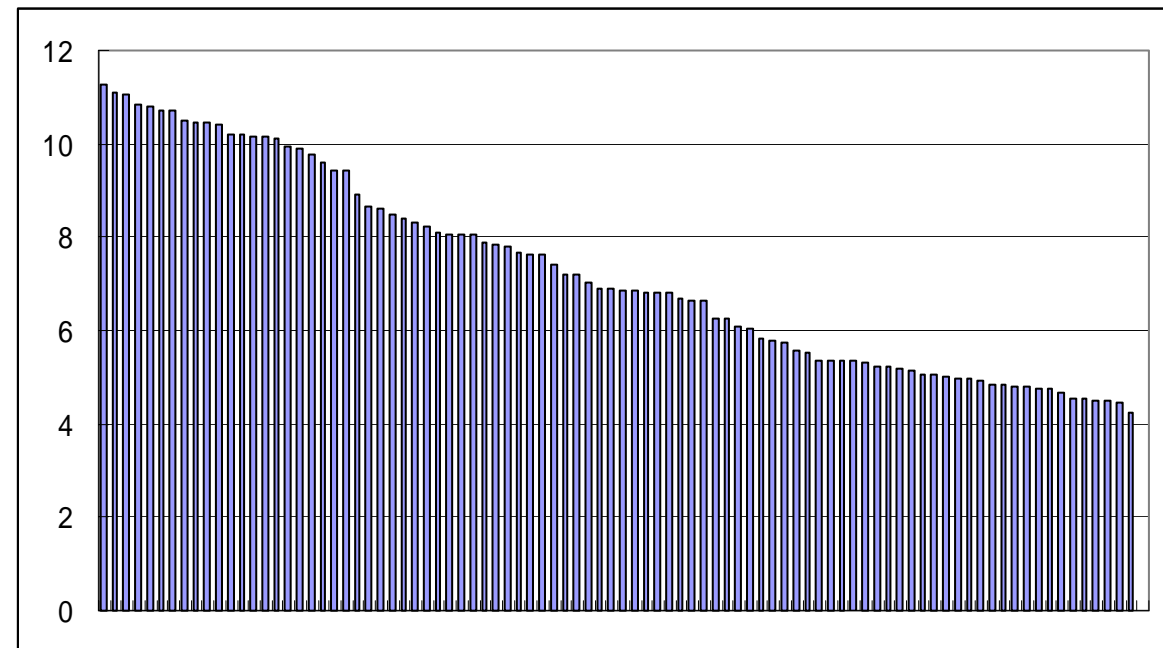


非必須  
553.9

# タンパク質 等電点

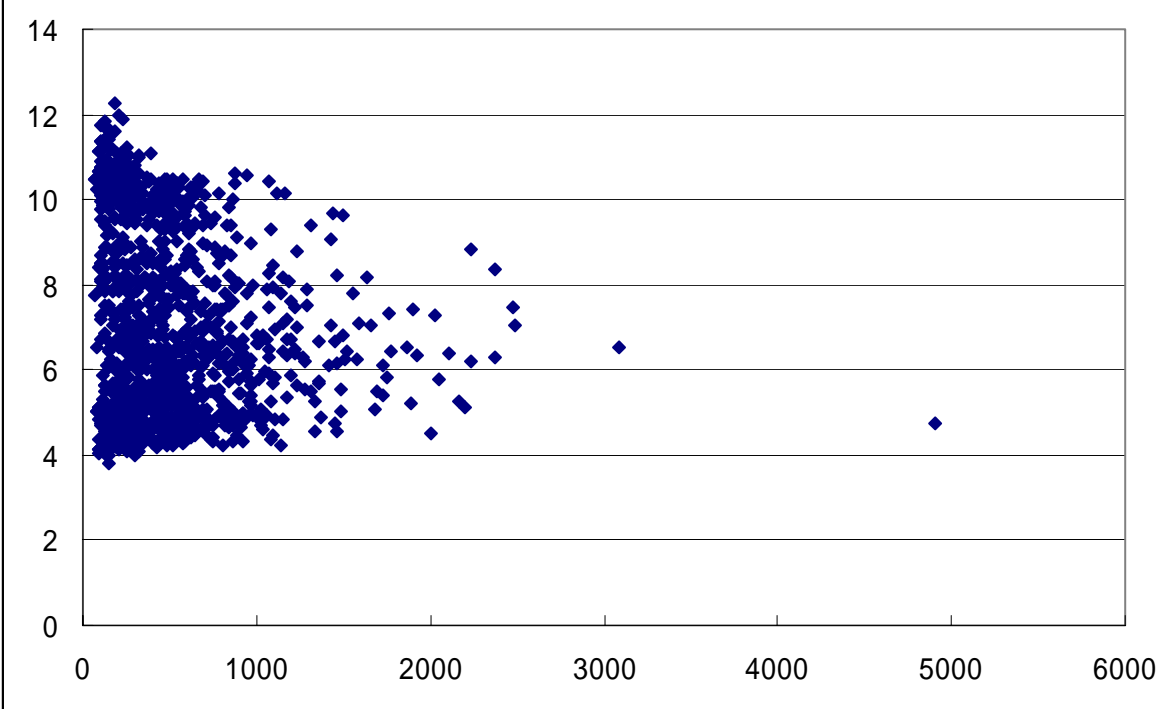


必須  
7.15



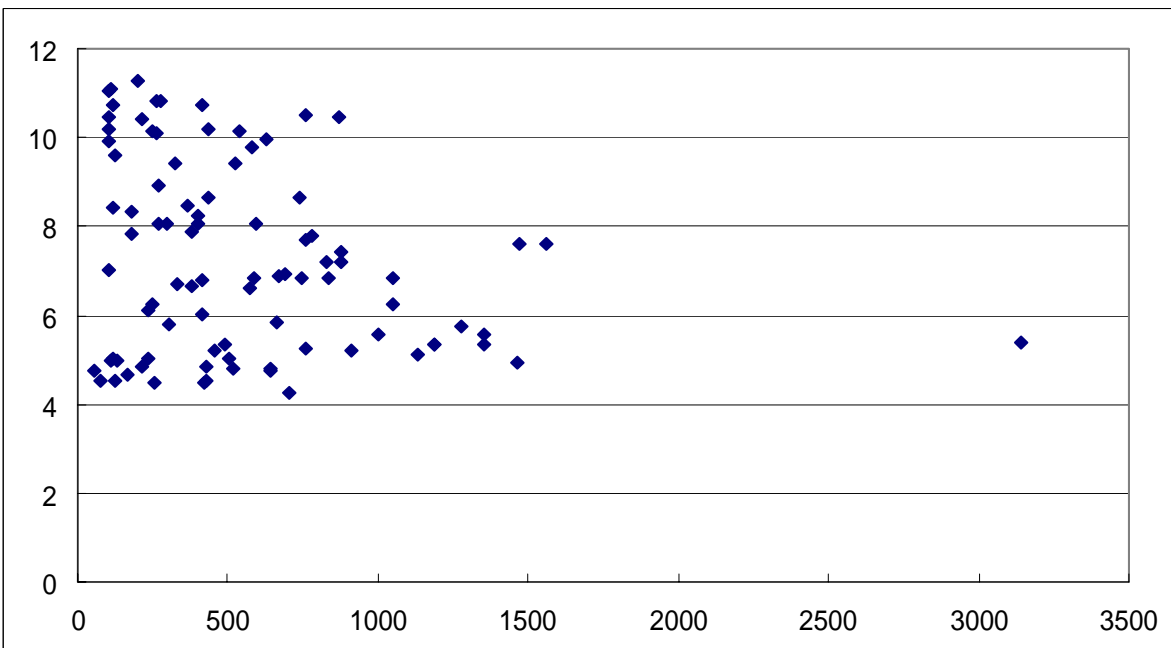
非必須  
7.26

# サイズと 等電点の 関係



必須

大きなタンパク質は等電点4-9の間にある。

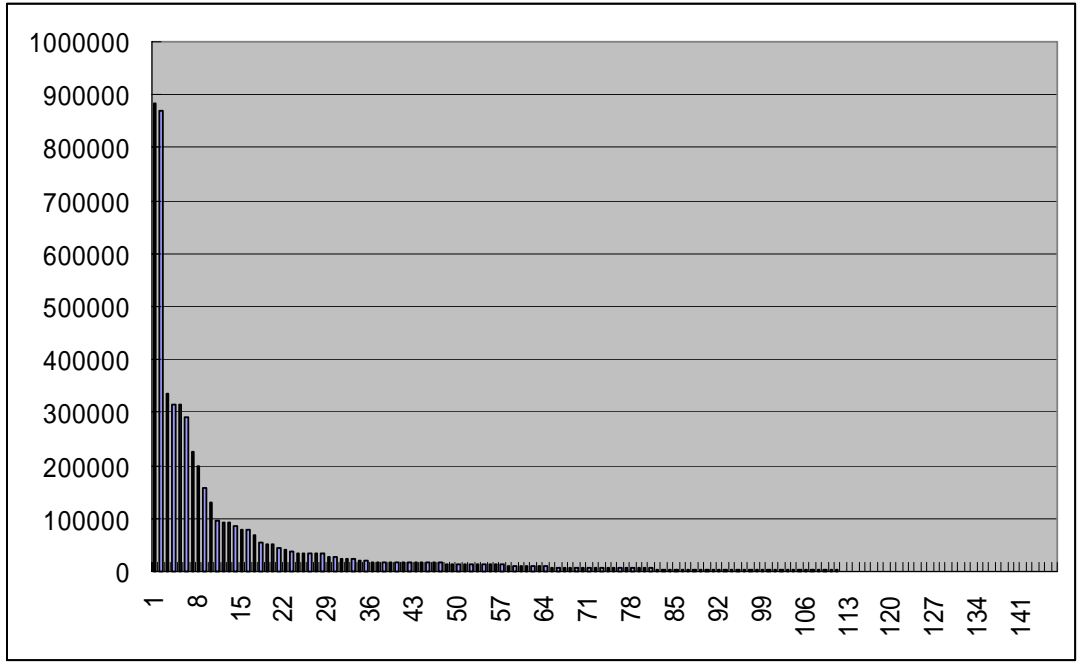


非必須

進化的に強く保存されたタンパク質  
にはなんらかの傾向があるか？

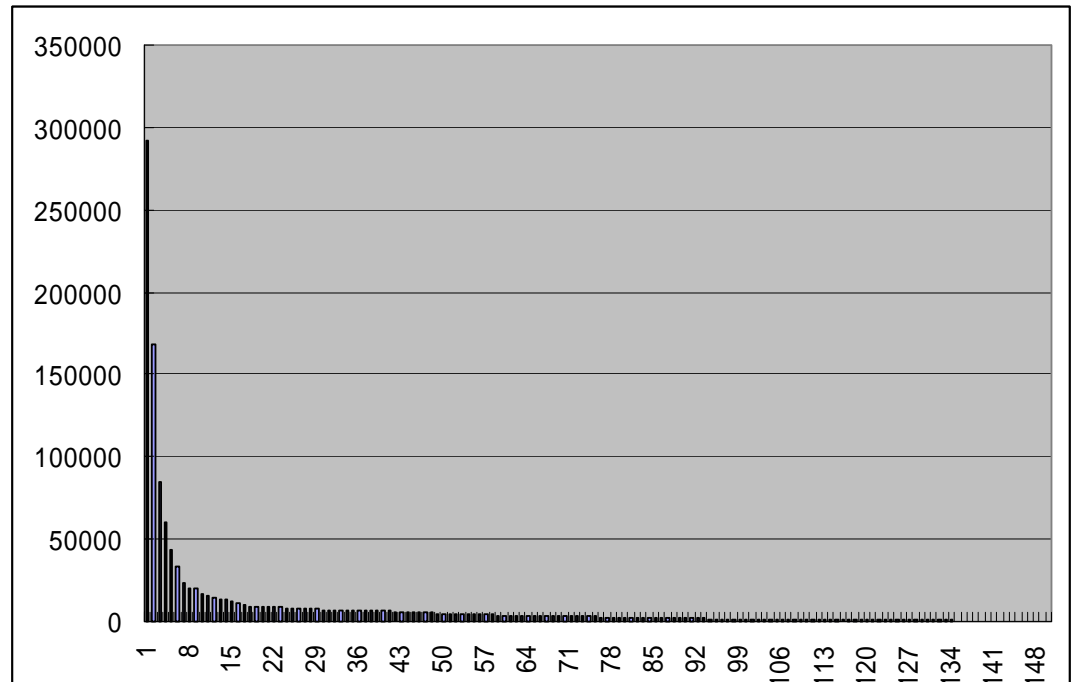
相同性とタンパク質のサイズ,  
等電点, および, タンパク量  
とを比較した

相同性高い  
1.0E-261  
146個



901 aa  
6.63 pl  
38188

相同性低い  
1.0E-20~-02  
200個



430 aa  
7.22 pl  
8390

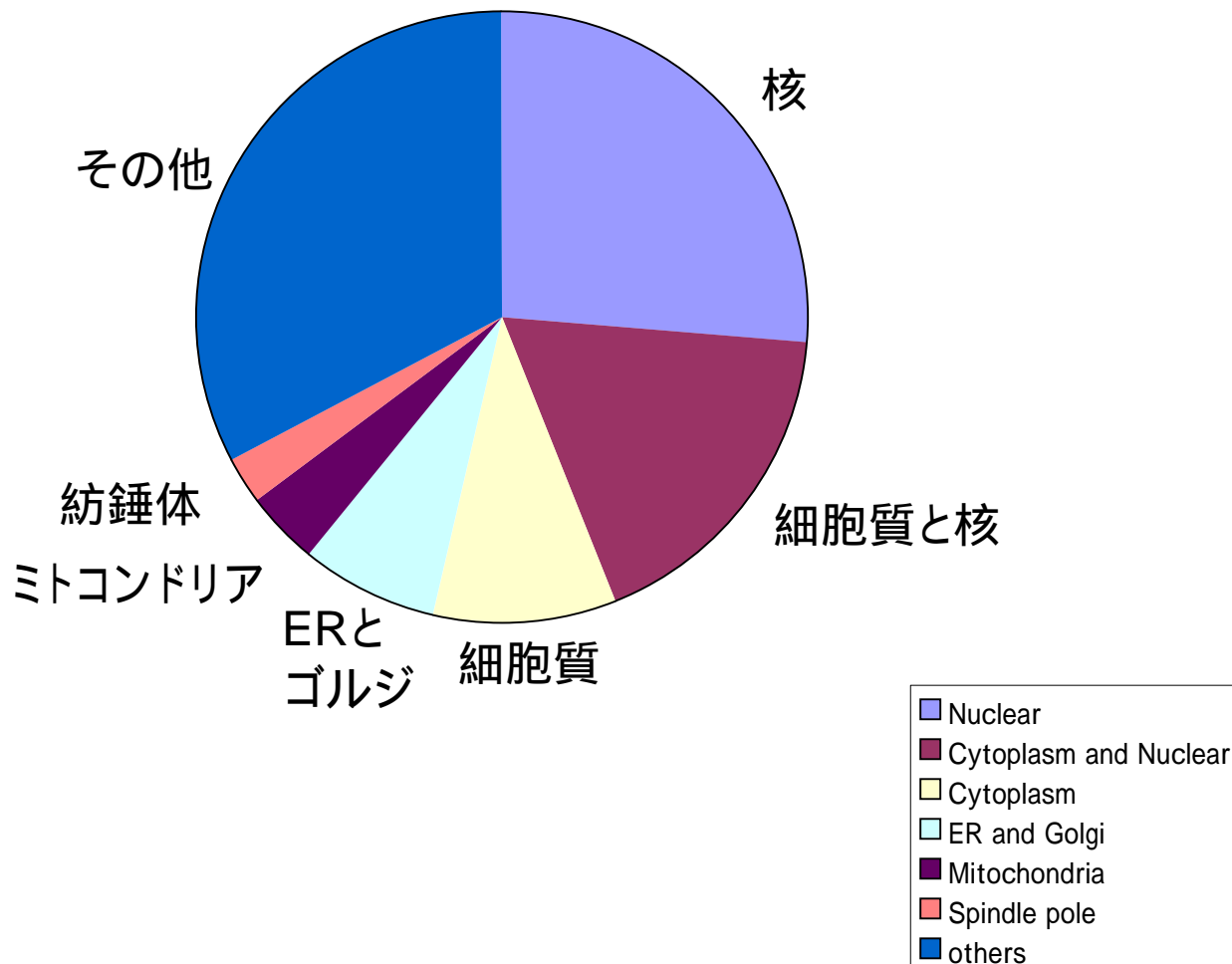
Nohomology  
411 aa  
7.39 pl  
4749

進化的に強く保存されたタンパク質  
にはなんらかの傾向があるか？

相同性の高いタンパク質は低いタン  
パク質よりもタンパクサイズが大きく、  
タンパク発現量も多い。

特に重要なタンパク質はサイズが大  
きく、細胞内での必要量も多いらしい

# 必須タンパク質の細胞内局在



## Localization of essential proteins

| Localization          | Number of genes | Length | pI          | No. of molecule |
|-----------------------|-----------------|--------|-------------|-----------------|
| Cytoplasm             | 110             | 610.5  | 6.71        | <b>51000</b>    |
| Nuclear               | 301             | 606.3  | 6.98        | 7248            |
| Cytoplasm and Nuclear | 201             | 527.9  | 6.78        | 13161           |
| ER and Golgi          | 83              | 513.1  | 7.04        | 12746           |
| Mitochondria          | 44              | 378.2  | <b>8.68</b> | 15781           |
| Spindle pole          | 30              | 461.0  | 6.51        | 1796            |

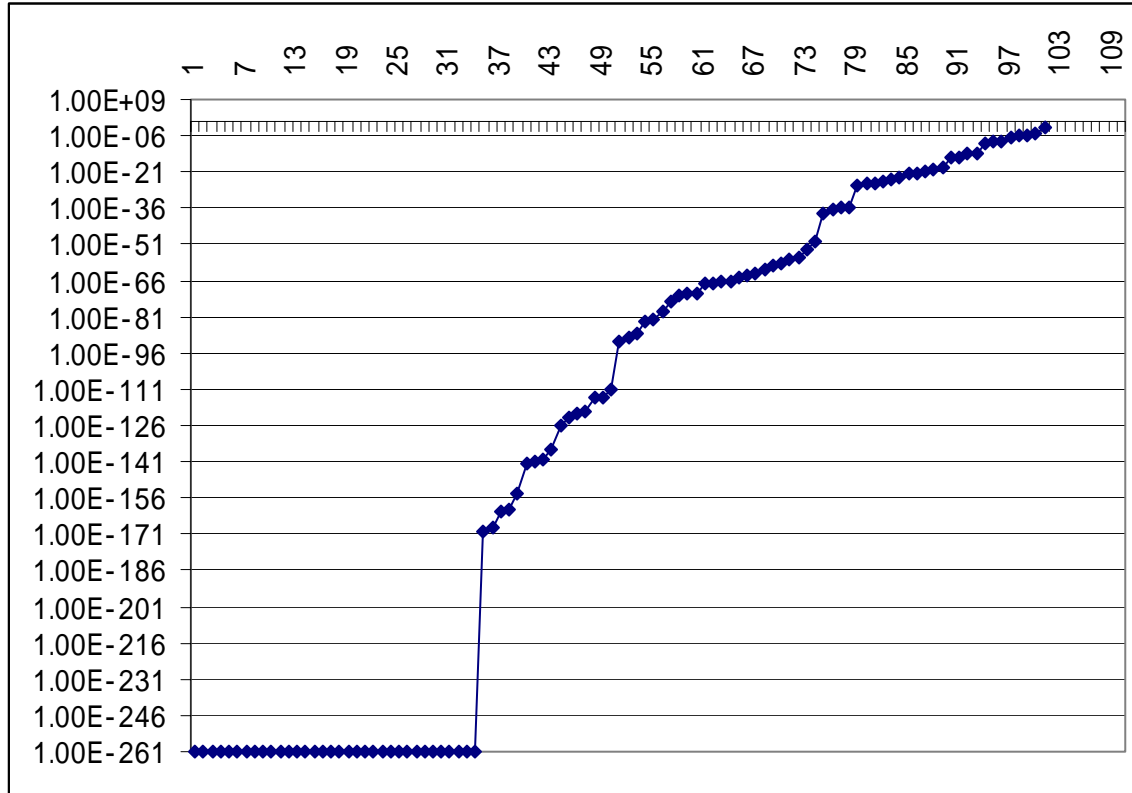
細胞質タンパク質は発現量が高いものが多い  
ミトコンドリアタンパク質の等電点は高い



細胞内局在性と相同性に関係があるか？

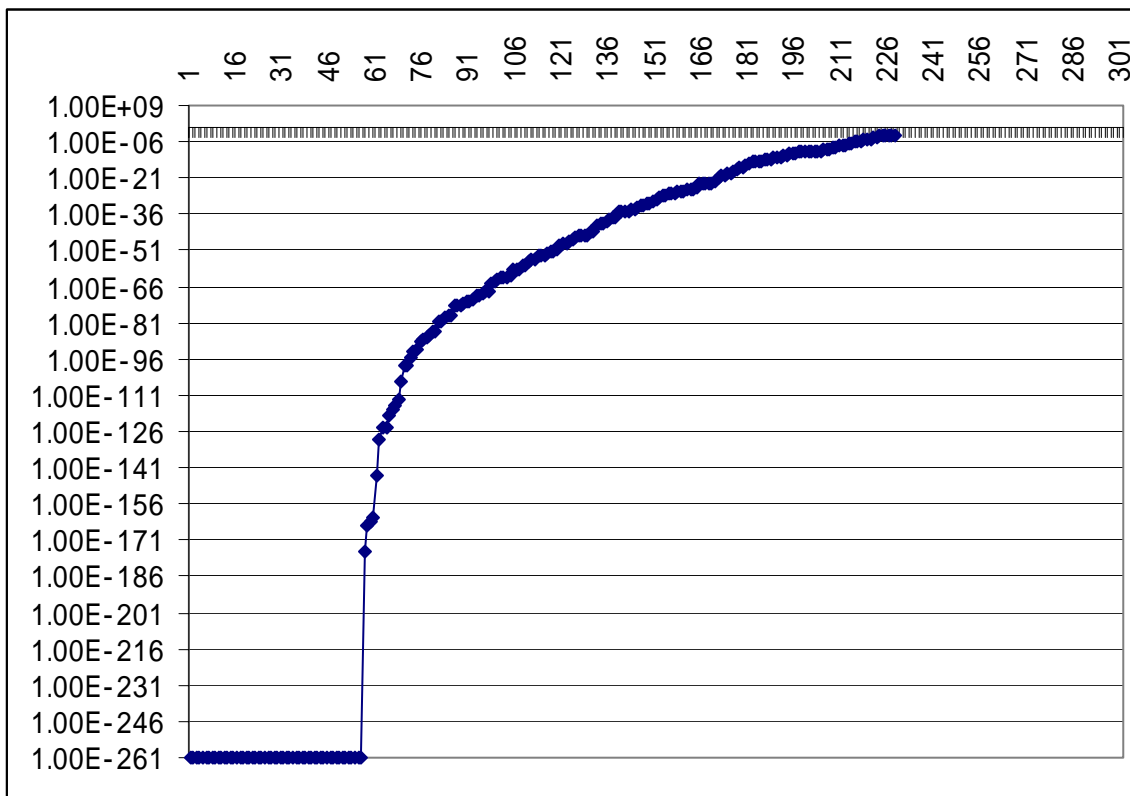
Cytoplasm

110個



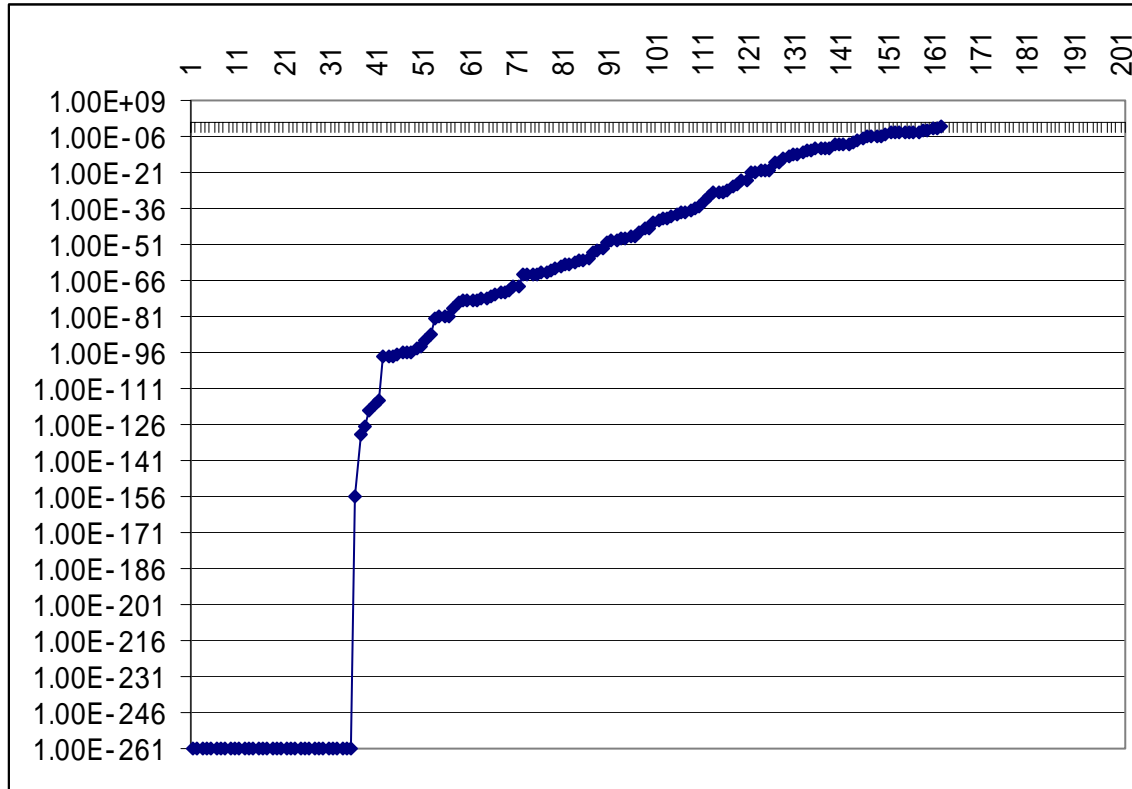
# Nuclear Homology

301個



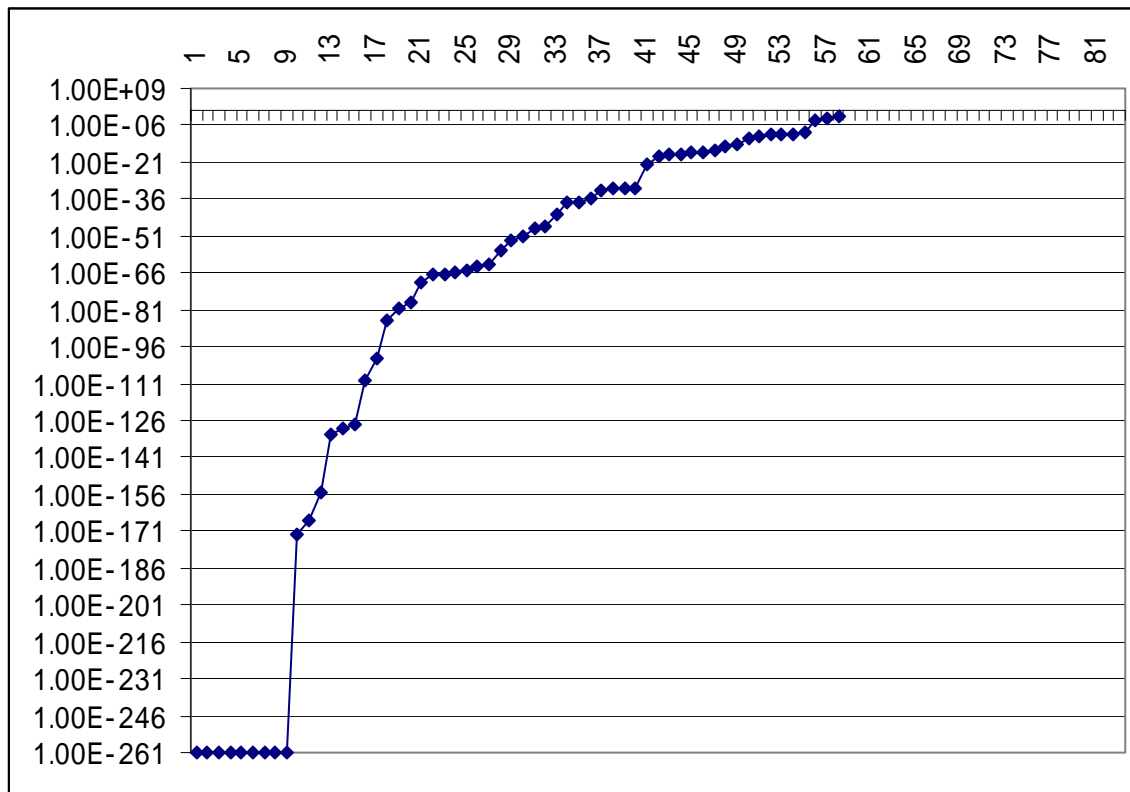
Nuclear-cytoplasm

201個



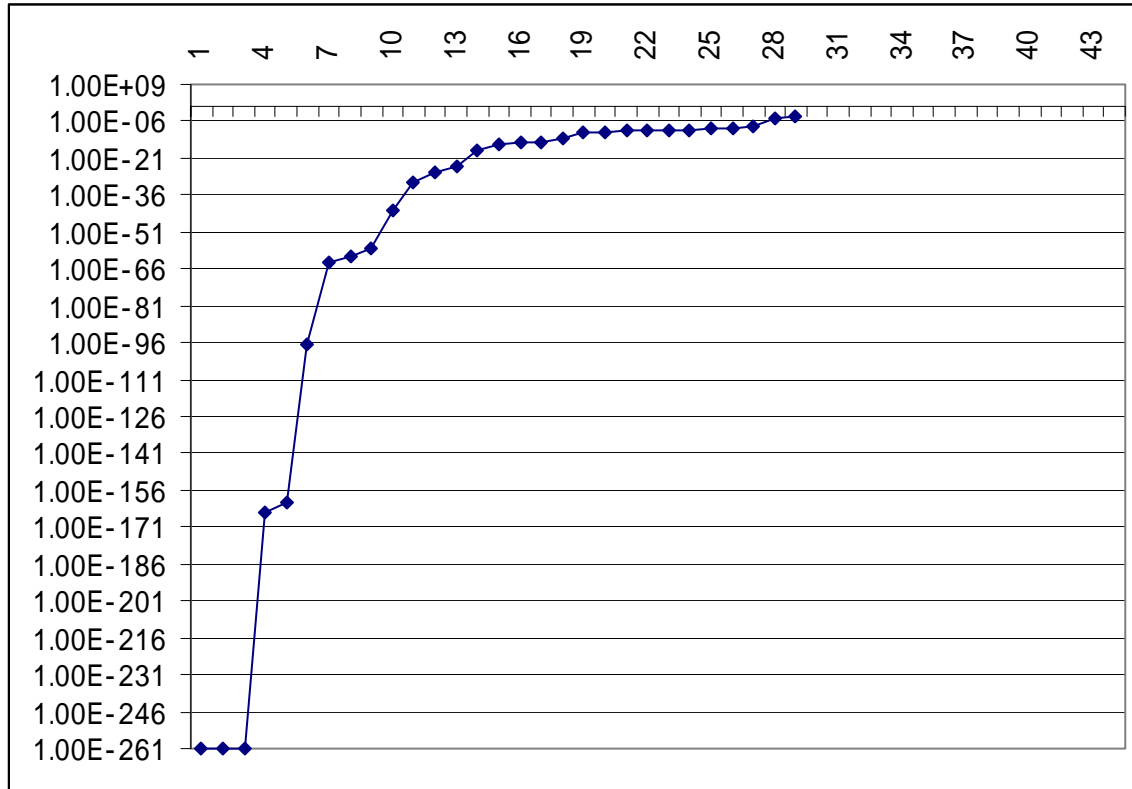
# ER and Golgi

83個



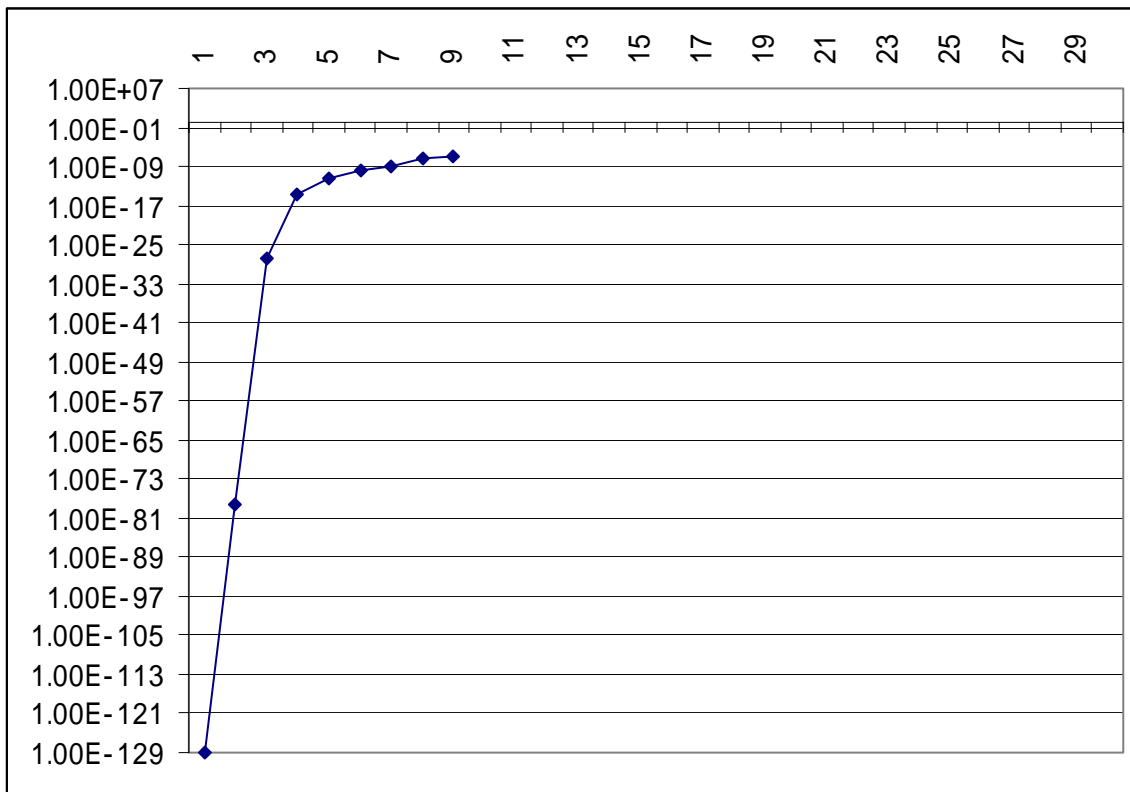
Mitochondria

44個



Spindle pole

30個



細胞質タンパク質は他に局在するタンパク質に比べて相同性が高いものが多い

細胞質タンパク質が保存性が高い

タンパク質間の相互作用が多いから？  
細胞内環境のせい？

ゲノムバイオロジー  
システムバイオロジーへ