General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

NATIONAL BIOMEDICAL RESEARCH FOUNDATION

GEORGETOWN UNIVERSITY MEDICAL CENTER

3900 RESERVOIR ROAD, N. W.

WASHINGTON, D. C. 20007

202-625-2121

FINAL TECHNICAL REPORT

NASA CONTRACT NASW3317

9/1/79 - 12/31/83

FINAL REPORT

for

NASA Contract NASW 3317

Investigation of Compounds Essential for the Origin of Life

covering the period

9/1/79 to 12/31/83

Principal Investigators

Margaret O. Dayhoff, Ph.D.	Sept. 1979 to Feb. 1983
Lois T. Hunt, Ph.D.	Feb. 1983 to Dec. 1983

TABLE OF CONTENTS

INTRODUCTION

The enclosed annual progress reports and publications describe and document the research performed by us with the support of NASA contract NASW 3317. This contract extended over a period characterized by intense activity and startling discoveries in the interrelated areas of molecular biology, genetics, and evolutionary studies of prokaryotes, eukaryotes, and their viruses.

PROPOSAL AND PROGRESS REPORT

INVESTIGATION OF COMPOUNDS ESSENTIAL FOR THE ORIGIN OF LIFE NASW 3317

National Biomedical Research Foundation
Georgetown University Medical Center
3900 Reservoir Road
Washington, D.C. 20007

August 28, 1980

Principal Investigator Margaret O. Dayhoff, Ph.D.

Co-principal Investigator Robert M. Schwartz, Ph.D.

TABLE OF CONTENTS

II. PROGRESS REPORT 1/1/78 - 8/25/80

Introduction

In early 1978, we published an article in Science that synthesized all of the available sequence data pertinent to bacterial and blue-green evolution and to the origin of eukaryote organelles. The articles in press from the 6th International Conference on the Origins of Life together with that in press from the International Colloquium on Endosymbiosis and Cell Research (see Section II-D) update that study. They suggest that the origin of the eukaryote organelles, the mitochondria and the chloroplasts, were not only endosymbiotic but also polyphyletic, i.e., organelles in different lines of descent arose from different bacteria or blue-greens.

The paper submitted to Nature with Dr. Barnabas initiated a new line of interest for us. In that work we correlated the metabolic capabilities of bacterial groups for which sequence data are available with their evolutionary position based on the sequence data. From this, we begin to infer the order in which a variety of metabolic pathways developed during the Precambrian. These metabolic capabilities include fermentation, anaerobic respiration, bacterial anoxygenic photosynthesis, sulfate reduction, aerobic respiration, and oxygenic photosynthesis.

Recent breakthroughs in DNA and RNA sequencing techniques have greatly speeded the elucidation of these sequence data. Much of the new data is a natural adjunct to our protein data collection, particularly the sequences of complete genomes, genes, and messenger RNAs. Other sequences, although less direct in their connection, are still extremely important, for example, control signals, ribosomal-binding sites, and origins of replication. During the last year, we have developed a computerized nucleic acid sequence data base and programs for data entry and retrieval as a demonstration project. Our demonstration project has as its goal showing what is necessary to make this new detailed genetic data intellectually accessible. In the first section of this report, we have included items describing the current state of our data base:

1. An editorial that appeared in Nature pointing out the need for such a

data base and our reply to that editorial.

2. A letter to be published in Science announcing the public availability of our data base.

3. Computer terminal display for our demonstraton system.

4. A table of contents of the data base as it currently stands as well as a sample of the data entries. Replaced by better printout with

IF ANY MEMBERS OF THE GROUP REVIEWING THIS PROPOSAL WOULD LIKE ACCESS TO OUR NUCLEIC ACID SEQUENCE REFERENCE DATA BASE, THEY MAY CALL EITHER DR. DAYHOFF OR DR. SCHWARTZ AT (202) 625-2121.

Clearly, making the data accessible is only the first step in the research process. Our NASA contract has supported that portion of this data collection bearing on origin of life studies. Additionally, we have requested supplemental funds to support one senior staff member during the four months the demonstration project will be on line in order to help update the retrieval system and modify our programs in response to user needs.

We have continued to maintain a reference data collection of protein sequences. Our NASA contract supports that part of the data collection and analysis that is of interest to the study of the origin and early evolution of life. In 1979, we published supplement 3 to volume 5 of the Atlas of Protein Sequence and Structure and a Protein Segment Dictionary (both submitted separately). We are currently working toward the publication of volume 6 of the Atlas at the end of 1982. This will be a comprehensive book including new data as well as combining and updating the information in volume 5 and its three supplements. A list of the new protein data arranged hierarchically by evolutionary relationship is shown in Section II-E.

List of Publications 1/1/78 - 8/25/80

Books Published:

Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, ed. M.O. Dayhoff, National Biomedical Research Foundation, Washington, D.C., 1978, 414 pp.

Protein Segment Dictionary 78, M.O. Dayhoff, L.T. Hunt, W.C. Barker, R.M. Schwartz and B.C. Orcutt, National Biomedical Research Foundation, Washington, D.C., 1978, 470 pp.

Other Output:

Protein sequence Data Tape, Atlas of Protein Sequence and Structure, M.O. Dayhoff, L.T. Hunt, W.C. Barker and R.M. Schwartz, National Biomedical Research Foundation, Washington, D.C., 1978. [119,006 residues from 1,081 sequences]

Papers Published:

An outline of biological evolution based on macromolecular sequences. R.M. Schwartz, M.O. Dayhoff. COMPARATIVE PLANETOLOGY, ed. by C. Ponnamperuma, pp. 225-242. Academic Press, N.Y., 1978.

Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. R.M. Schwartz and M.O. Dayhoff, Science 199: 395-403, January 27, 1978.

The point mutation process in proteins. R.M. Schwartz and M.O. Dayhoff, in: Origin of Life: Proceedings of the Second ISSOL Meeting, the Fifth ICOL Meeting, Haruhiko Noda, editor, Center for Academic Publications Japan/Japan Scientific Societies Press, 1978, pp. 457-469.

Evolution of early life inferred from protein and ribonucleic acid sequences. M.O. Dayhoff and R.M. Schwartz, in: Origin of Life: Proceedings of the Second ISSOL Meeting, the Fifth ICOL Meeting, Haruhiko Noda, editor, Center for Academic Publications Japan/Japan Scientific Societies Press, 1978, pp. 547-560.

Detection of distant relationships based on point mutation data. R.M. Schwartz and M.O. Dayhoff, Evolution of Protein Moleculaes, ed. H. Matsubara and T. Yamanaka, pp. 1-16. Center for Academic Publications Japan/Japan Scientific Societies Press, Tokyo, 1978.

Evolution of prokaryotes inferred from sequences. M.O. Dayhoff and R.M. Schwartz, Evolution of Protein Molecules, ed. H. Matsubara and T. Yamanaka, pp. 323-42. Center for Academic Publications Japan/Japan Scientific Societies Press, Tokyo, 1978.

Protein and nucleic acid sequence data and phylogeny. R.M. Schwartz and M.O. Dayhoff. Science 205 (4410): 1036-39, 7 Sept. 1979. [Exchange of Technical Comments with Vincent Demoulin]

Evolutionary relationships among photosynthetic prokaryotes inferred from protein and nucleic acid sequence data. R.M. Schwartz and M.O. Dayhoff. Third International Symposium on Photosynthetic Prokaryotes, Oxford, 1979. Abstracts. E7

Prokaryote evolution and the symbiotic origin of eukaryotes. M.O. Dayhoff and R.M. Schwartz. Proceedings of the International Colloquium in Endosymbiosis and Cell Research, April 11-15, 1980, Tubingen, Germany. Berlin: Walter deGruyter & Co., 1980. In press.

Phylogenetic sequence of metabolic pathways in precambrian cellular life. J. Barnabas, R.M. Schwartz, and M.O. Dayhoff. Proceedings of the 6th International Conference on the Origins of Life. Dordrecht, The Netherlands: Reidel, 1980. In press.

The evolution of blue-greens and the origins of chloroplasts. R.M. Schwartz and M.O. Dayhoff. Proceedings of the 6th International Conference on the Origins of Life. Dordrecht, The Netherlands: Reidel, 1980. In press.

Evolution of the rhodospirillaceae and mitochondria: a view based on sequence data. M.O. Dayhoff and R.M. Schwartz. Proceedings of the 6th International Conference on the Origins of Life. Dordrecht, The Netherlands: Reidel, 1980. In press.

Paper submitted for publication:

Evolution of major metabolic innovations in the precambrian. J. Barnabas, R.M. Schwartz, and M.O. Dayhoff. Submitted to Nature, June 1980.

New Entries and Their Protein Superfamilies

Up-date of May 1980

M.O. Dayhoff, H.R. Chen, B.C. Orcutt, W.C. Barker, L.T. Hunt, and R.M. Schwartz

NBR Report 08710-800515

National Biomedical Research Foundation

Georgetown University Medical Center

3900 Reservoir Road, N.W.

Washington, D.C. 20007

CONTENTS

2 Protein Superfamilies

M.O. Dayhoff, W.C. Barker, L.T. Hunt, and R.M. Schwartz

In the list that follows, we have organized all of the complete sequences reported in the *Atlas* volumes into groups of superfamilies, families, subfamilies, and entries. The number in each group, the criteria for clustering, and the method of identification of the hierarchical levels in the list are shown below.

Number of Groups	Group	Criteria for Clustering Sequences	Identification of Cluster
181	Superfamilies	Probability of similarity by chance $< 10^{-6}$	Number
314	Families	<50% different	Letter
537	Subfamilies	<20% different	Paragraph
793	*Atlas* entries	<5% different	Semicolon

Sequences in the same entry are separated by commas.

This list updates the one that appeared in Supplement 2,[1] in which there were 116 superfamilies, 197 families, 328 subfamilies, and 493 entries. There has been about a 60% increase in all categories in the intervening 2 years and 7 months.

Only complete or nearly complete sequences that are 20 or more residues in length are included. The constant and variable regions of immunoglobulins are counted as separate sequences. Sequences that can be considered complete in one sense but partial in another are generally included. Examples are active hormone and enzyme sequences that are derived from longer precursors, and sequences of entire homology regions from proteins with two or more such regions.

Proteins within a family usually differ at fewer than half of their amino acid positions and they are either homologs in various species or products of gene duplication; their similarity of function has usually been recognized before the sequences were known and they have identical or very similar names. Families are identified by letters in the list.

The sequences within a family have been divided into subfamilies, which are shown as paragraphs. Sequences within a subfamily usually differ from each other at fewer than 20% of their amino acid positions. Within a subfamily, sequences that differ by less than 5% and form a single *Atlas* entry are separated by commas, whereas sequences or groups that are more than 5% different are separated by semicolons.

In a clustering procedure such as this there will always be cases that are borderline, some pairs within a group being below the cutoff and some above. Where possible, we have grouped together proteins that fall on the same branch of an evolutionary tree.

The families are grouped into superfamilies,[2,3] identified by numbers, where similarity of sequences in different families can be recognized by statistical procedures. We have used two such methods to compare pairs of complete sequences: for sequences of comparable length we used a method based on the best alignment of the two sequences; for sequences of quite different length, we used a method based on the distribution of scores obtained on comparison of all segments of a given length from one sequence with those from the other. These methods are described in detail in chapter 1. Each method produces a probability that the scores from the comparison of two real sequences could have been derived from the distribution of scores produced by comparisons of pairs of randomly permuted sequences with the same amino acid compositions as the two real sequences.

A newly determined sequence is placed in an existing superfamily if, on comparison with the best conserved sequence from each family in that superfamily, at least one probability of $< 10^{-6}$ is obtained. For a collection of 314 families that might potentially be combined, $(314 \times 313)/2 = 49{,}141$ comparisons are possible. The probability of finding a score of 10^{-6} by chance in one or more of these is 5%. Thus, we have 95% confidence that all of the families that have been grouped together really share significant sequence similarity.

The ultimate superfamily list could be derived from sequence information alone, provided that at least one sequence was known from most subfamilies within each family. At present we do not have this much sequence information, but often we have information on chemical or physiological functions that reflect relationship. Where we know in advance that several proteins share a similar function, we have required that the probability for a single comparison within the group be $< 10^{-3}$ in order to cluster the sequences in the same superfamily.

It is also possible to establish relationships on the basis of search scores (see chapter 1). There are approximately 10^5 20-residue segments in the data collection. If a segment of 20 residues is compared with all of the 10^5 other segments in the collection, an approximately normal distribution of scores is obtained. From the mean and standard deviation of this distribution, the probability of finding a score equal to or greater than any given score can be calculated. In principle, 10^5 such searches, one for each 20-residue segment, could be performed, leading to the accumulation of $10^5 \times 10^5 = 10^{10}$ probabilities. If the probability associated with a given score is $\leq 0.5 \times 10^{-11}$, there is a probability of approximately $0.5 \times 10^{-11} \times 10^{10} = 0.05$ of finding one such score by chance in an exhaustive intercomparison of all segments. The probability of 0.5×10^{-11} calculated from the normal distribution corresponds to a confidence level of 95% that the sequence similarities discovered by search scores are unusual enough to indicate relationship. We feel that very low probabilities are a reflection of the common evolutionary origin of the proteins. Other similarities of structure, function, and control would therefore be predicted.

Superfamily Groups

Most of the family relationships in this list were pointed out in the papers describing the sequence work and are referenced in the data pages. Quantitative data on relationships are given in chapter 10 of the *Atlas*, Volume 5,[4] in the Survey of New Material of Supplement 1,[5] and in many tables in Supplement 2, as well as in this book. We have applied these quantitative criteria for defining relationship to the suggestions of others and to the hopeful leads that we have turned up by extensive searching of the data in organizing this list.

We have grouped together several proteins of similar function that get borderline probabilities of sequence similarity, including pancreatic hormone from chicken with glucagon and secretin, antibacterial substance A and neocarzinostatin from *Streptomyces*, ferredoxins with adrenodoxin and putidaredoxin, the fungal with the bacterial ribonucleases, and peanut protease inhibitor and bromelain inhibitor with the Bowman-Birk type protease inhibitors. In other instances we have chosen not to combine borderline cases. The four histones would be combined on the basis of comparisons using the identity matrix but would not be combined using the mutation data matrix. There are a number of short sequences that are repeated in at least two histone groups. However, there have been many insertions and deletions as well as point mutations, so we have left the four groups as separate superfamilies. Human epidermal growth factor (EGF) and a small part of bovine factor X are clearly related. We suspect that EGF may even be a degradation product of an as yet unsequenced serine protease. Because of its distinct function, we have left EGF as a distinct superfamily until the situation is clarified. Bird apovitellenins and the human lipid-binding proteins have been left in separate superfamilies. Additional groups of protease inhibitors may eventually be combined when more sequences are known.

Two groups with similar functions and three-dimensional structures do not display significant sequence similarity. The dehydrogenases (alcohol, lactate, glutamate, and glyceraldehyde 3-phosphate) are separate superfamilies, as are the constant and variable regions of the immunoglobulins. Presumably in both of these cases there have been too many insertions, deletions, and point mutations to deduce a common evolutionary origin from the sequences.

Relationships among some of the superfamilies may eventually be demonstrated as more extensive sequence information becomes available for each family, permitting the construction of ancestral sequences for which the mutability of each residue can be estimated. Additional information on relationships may be derived from the similarity of amino acid compositions.

A further organization of superfamilies reflecting common evolutionary origin may be possible based on additional nonsequence information; for example, the three-dimensional structures or the positions in a metabolic pathway. In this list, the superfamilies are grouped according to function or prosthetic group.

It has been estimated that in humans there are approximately 50,000 proteins of functional or medical importance. We conjecture that these will be grouped into about 500 superfamilies, each containing an average of 100 sequences that range from minor variants up to 85% or 90% different from one another. A similar number of superfamilies has been proposed by Zuckerkandl.[6] A landmark of molecular biology will occur when one member of each superfamily has been elucidated. At the present rate of 25 per year, this will take less than 15 years.

References

1. Dayhoff, M.O., Barker, W.C., and Hunt, L.T., *in* Atlas of Protein Sequence and Structure, Vol.5, Suppl.2, ed. Dayhoff, M.O., pp.9-19, Nat. Biomed. Res. Found., Washington, D.C., 1976
2. Dayhoff, M.O., McLaughlin, P.J., Barker, W.C., and Hunt, L.T., Naturwissenschaften 62, 154-161, 1975
3. Dayhoff, M.O., Fed. Proc. 35, 2132-2138, 1976
4. Barker, W.C., and Dayhoff, M.O., *in* Atlas of Protein Sequence and Structure, Vol.5, ed. Dayhoff, M.O., pp.101-110, Nat. Biomed. Res. Found., Washington, D.C., 1972
5. Dayhoff, M.O., *in* Atlas of Protein Sequence and Structure, Vol.5, Suppl.1, ed. Dayhoff, M.O., pp.S1-S8, Nat. Biomed. Res. Found., Washington, D.C., 1973
6. Zuckerkandl, E., J. Mol. Evol. 7, 1-57, 1975

Explanation of Computer Listings of New Entries

We have examined the relationships between all of the new sequences and the ones already in the collection. Each sequence in the Suppl. 3 superfamily list has been assigned five numbers, according to its superfamily number, its position among the families of the superfamily, its position among the subfamilies of its family (paragraphs), its position among the entries of its subfamily (strings separated by semicolons), and its position in the string of sequences in an entry. Each new item has been assigned five numbers that place it between two other entries in the list, where it belongs. We show the first four of these numbers on the updated superfamily list. Thus the first sequence on the list, Cytochrome c-Rice, belongs in the first superfamily and the first family of cytochrome c related proteins. It is in the 14th subfamily, in between the third and fourth entries, sesame and castor. Similarly, the C-phycocyanin alpha chains (No. 5.2) form a new superfamily coming between cytochrome b_{562} (No. 5) and ferredoxin (No. 6). Before publication the entire list can be renumbered using only integers, and a superfamily list similar to the one already published can be printed out by the computer.

Some of the new entries contain short sequences or fragments of longer sequences and have not been assigned superfamily numbers. These are listed separately in alphabetical order.

The two computer listings contain 396 items. Of these, 358 are totally new entries, whereas 38 are revisions to published Atlas entries, usually the completion of a sequence for which only fragmentary information was formerly known.

New, complete or almost complete sequences of 20 or more residues with their superfamily classifications

SUPFAM	FAM	SUBFAM	ENTRY	
1.0	1.0	14.0	3.5	Cytochrome c Rice
1.0	1.0	17.0	1.0	Cytochrome c2 Rhodopseudomonas globiformis
1.0	4.0	2.0	2.0	Cytochrome c2, iso 2 Rhodospirillum fulvum
1.0	10.5	1.0	1.0	Cytochrome C-554 Paracoccus sp.
1.0	12.0	4.0	1.0	Cytochrome c6 Synechococcus ATCC 27167
1.0	12.0	4.0	2.0	Cytochrome c6 Synechococcus lividus
1.0	15.5	1.0	1.0	Cytochrome c551 Ectothiorhodospira halophila
3.0	1.4	1.0	1.0	Cytochrome c556 Agrobacterium tumefaciens
3.0	1.5	1.0	1.0	Cytochrome c' Rhodopseudomonas gelatinosa
3.0	3.0	1.0	1.0	Cytochrome c' Rhodospirillum fulvum
3.0	3.0	1.0	2.0	Cytochrome c' Rhodospirillum molischianum
3.0	4.0	1.0	1.0	Cytochrome c' Rhodospirillum tenue
3.0	5.0	1.0	1.0	Cytochrome c' Chromatium vinosum
4.0	1.0	1.0	1.0	Cytochrome b5 Rabbit
5.	1.0	1.0	1.0	C-phycocyanin alpha chain Cyanidium caldarium
5.2	1.0	1.0	2.0	C-phycocyanin alpha chain Mastigocladus laminosus
5.2	2.0	1.0	1.0	C-phycocyanin beta chain Mastigocladus laminosus
5.2	2.0	2.0	1.0	C-phycocyanin beta chain Synechococcus sp. (=Anacystis nidulans)
6.0	2.0	1.5	1.0	Ferredoxin Chlorobium limicola f. sp. thiosulfatophilum
6.0	4.0	1.0	1.0	Ferredoxin Desulfovibrio gigas

SUPFAM	FAM	SUBFAM	ENTRY	
6.0	5.5	1.0	1.0	Ferredoxin Pseudomonas ovalis
6.0	5.5	2.0	1.0	Ferredoxin Mycobacterium smegmatis
6.0	7.0	2.0	2.0	Ferredoxin Elder
6.0	7.0	2.1	1.0	Ferredoxin Rape
6.0	7.0	2.2	1.0	Ferredoxin Wheat
6.0	7.0	2.3	1.0	Ferredoxin I Pokeweed
6.0	7.0	2.4	1.0	Ferredoxin II Pokeweed
6.0	7.0	2.5	1.0	Ferredoxin I Horsetail
6.0	7.0	2.6	1.0	Ferredoxin II Horsetail
6.0	7.0	3.0	2.0	Ferredoxin I Dunaliella salina
6.0	7.0	3.0	3.0	Ferredoxin II Dunaliella salina
6.0	7.0	3.1	1.0	Ferredoxin Cyanidium caldarium
6.0	7.0	3.2	1.0	Ferredoxin Porphyra umbilicalis
6.0	7.0	4.0	3.0	Ferredoxin Mastigocladus laminosus
6.0	7.0	5.5	1.0	Ferredoxin II Aphanothece sacrum
6.0	7.0	7.0	1.0	Ferredoxin Synechococcus lividus
6.0	8.0	1.0	2.0	Adrenodoxin Pig
7.0	2.5	1.0	1.0	High potential iron-sulfur protein Paracoccus sp.
7.0	4.0	1.0	1.0	High potential iron-sulfur protein Rhodospirillum tenue
8.0	3.0	1.0	1.0	Desulforedoxin Desulfovibrio gigas

SUPFAM	FAM	SUBFAM	ENTRY	
11.0	2.0	1.5	1.0	Plastocyanin Lombardy poplar
11.2	1.0	1.0	1.0	Stellacyanin Japanese lacquer tree
14.0	1.0	4.0	1.0	Lactate dehydrogenase C chain (EC 1.1.1.27) Mouse
16.0	1.0	3.0	2.0	Glyceraldehyde-3-phosphate dehydrogenase Baker's yeast
18.0	3.0	1.0	2.0	Dihydrofolate reductase (EC 1.5.1.3) Pig
18.0	4.0	1.0	1.0	Dihydrofolate reductase Lactobacillus casei
18.1	1.0	1.0	1.0	Dihydrofolate reductase (EC 1.5.1.3), type II Escherichia coli (plasmid R67)
18.4	1.0	1.0	1.0	Cytochrome oxidase (EC 1.9.3.1), polypeptide II Human mitochondrion
18.4	1.0	2.0	1.0	Cytochrome oxidase, polypeptide II Bovine
18.4	2.0	1.0	1.0	Cytochrome oxidase (EC 1.9.3.1), polypeptide II Baker's yeast mitochondrion
18.5	1.0	1.0	1.0	Cytochrome oxidase (EC 1.9.3.1), polypeptide IV Bovine
18.6	1.0	1.0	1.0	Cytochrome oxidase, heme a chain Bovine
18.7	1.0	1.0	1.0	Cytochrome oxidase (EC 1.9.3.1), polypeptide VII Bovine
18.8	1.0	1.0	1.0	Cytochrome oxidase, polypeptide VIIIa Bovine
18.9	1.0	1.0	1.0	Nitrogenase, iron protein component Clostridium pasteurianum
19.3	1.0	1.0	1.0	Protocatechuate 3,4-dioxygenase alpha chain (in EC 1.13.11.3) Pseudomonas aeruginosa
19.3	2.0	1.0	1.0	Protocatechuate 3,4-dioxygenase beta chain (in EC 1.13.11.3) Pseudomonas aeruginosa
19.6	1.0	1.0	1.0	Tyrosinase (EC 1.14.18.1) Neurospora crassa
20.0	1.0	2.0	1.0	Superoxide dismutase (Cu-Zn) (EC 1.15.1.1) Baker's yeast
20.1	1.0	1.0	1.0	Superoxide dismutase (Mn) Escherichia coli B

SUPFAM	FAM	SUBFAM	ENTRY	
20.5	1.0	1.0	1.0	Thymidylate synthase (EC 2.1.1.45) Lactobacillus casei
20.9	1.0	1.0	1.0	Chloramphenicol acetyltransferase (EC 2.3.1.28) Escherichia coli plasmids
21.2	1.0	1.0	1.0	Phosphorylase Rabbit
21.4	1.0	1.0	1.0	ATP-phosphoribosyltransferase Salmonella typhimurium
22.0	1.0	1.0	2.0	Aspartate aminotransferase (EC 2.6.1.1), cytoplasmic Chicken
22.0	1.0	2.0	1.0	Aspartate aminotransferase, mitochondrial Pig
22.5	1.0	1.0	1.0	Phosphoglycerate kinase Horse muscle.
28.0	1.0	1.0	2.0	Prophospholipase A2 (EC 3.1.1.4) Bovine
28.0	2.0	2.0	1.0	Phospholipase A2 alpha Eastern diamondback rattlesnake
28.0	3.0	2.5	1.0	Phospholipases A2 (EC 3.1.1.4) I and II Mozambique cobra
28.0	3.0	2.5	2.0	Phospholipase A2 (EC 3.1.1.4) III Mozambique cobra and blackneck spitting cobra
28.0	3.0	2.7	1.0	Phospholipase A2, beta-1 bungarotoxin A chain Formosan banded krait
28.0	3.0	3.5	1.0	Prophospholipase A2, taipoxin gamma chain Taipan
30.0	1.0	1.0	1.0	Nuclease (EC 3.1.31.1) precursor Staphylococcus aureus Foggi
31.0	1.0	1.0	2.0	Ribonuclease Bacillus intermedius 7P
32.0	1.0	6.0	1.0	Ribonuclease (EC 3.1.27.5) Golden hamster
32.0	1.0	12.0	1.0	Ribonuclease Red kangaroo
33.0	1.0	1.5	1.0	Lysozyme Rat
33.0	1.0	2.0	1.0	Lysozyme precursor Chicken
33.0	1.0	2.0	1.5	Lysozyme Ring-necked pheasant

SUPFAM	FAM	SUBFAM	ENTRY	
33.0	2.0	4.0	1.0	Lactalbumin Rabbit
39.0	1.0	2.0	1.0	Prothrombin Human
39.0	1.5	1.0	1.0	Factor IX (Christmas factor) Bovine
39.0	2.5	1.0	1.0	Protein C Bovine
39.0	4.0	3.0	1.0	Trypsinogen Spiny dogfish
39.0	6.5	1.0	1.0	Group-specific protease Rat
41.0	2.0	1.0	1.0	Actinidin Chinese gooseberry
47.0	3.0	1.0	1.0	Penicillinase precursor Escherichia coli plasmids
47.2	1.0	1.0	1.0	Inorganic pyrophosphatase (EC 3.6.1.1) Baker's yeast
47.4	1.0	1.0	1.0	ATP pyrophosphatase, lipid-binding protein Bovine mitochondrion
47.4	1.0	2.0	1.0	ATP pyrophosphatase, lipid-binding protein Baker's yeast mitochondrion
47.4	1.0	3.0	1.0	ATP pyrophosphatase, lipid-binding protein Neurospora crassa mitochondrion
47.4	2.0	1.0	1.0	ATP pyrophosphatase, lipid-binding protein Spinach chloroplast
47.4	3.0	1.0	1.0	ATP pyrophosphatase, lipid-binding protein Escherichia coli
47.4	4.0	1.0	1.0	ATP pyrophosphatase, lipid-binding protein Thermophilic bacterium PS-3
47.6	1.0	1.0	1.0	Ribulose-1,5-bisphosphate carboxylase (EC 4.1.1.39) small chain Spinach
49.5	1.0	1.0	1.0	Enolase (EC 4.2.1.11) Baker's yeast
50.0	1.0	1.0	1.0	Tryptophan synthase (EC 4.2.1.20) alpha chain Escherichia coli
50.0	1.0	1.0	2.0	Tryptophan synthase (EC 4.2.1.20) alpha chain Salmonella typhimurium
52.2	1.0	1.0	1.0	Tryptophanyl tRNA synthetase Bacillus stearothermophilus

SUPFAM	FAM	SUBFAM	ENTRY	
52.8	1.0	1.0	1.0	D-erythrodihydroneopterin triphosphate synthetase Rat, guinea pig, and bovine
53.0	3.0	2.5	1.0	Venom basic protease inhibitor E Black mamba
53.0	3.5	1.0	1.0	Beta-1 bungarotoxin B chain, major component Formosan banded krait
53.0	3.5	1.0	2.0	Beta-1 bungarotoxin B chain, minor component Formosan banded krait
53.0	5.0	1.0	1.0	Basic protease inhibitor Red sea turtle
54.0	2.5	1.0	1.0	Acrosin inhibitor (PSTI type) Bovine
54.0	4.0	1.0	2.0	Ovomucoid (PSTI-type protease inhibitor) Turkey
54.0	4.0	1.0	3.0	Ovomucoid (PSTI-type protease inhibitor) precursor Chicken
54.0	5.0	2.0	1.0	Plasminostreptin (PSTI-type protease inhibitor) Streptomyces antifibrinolyticus
55.0	2.0	1.0	1.0	Ovalbumin Chicken
55.5	1.0	1.0	1.0	Protease B inhibitors 2 and 1 Baker's yeast
58.0	1.0	2.0	2.0	Protease inhibitors (Bowman-Birk) Adzuki bean
62.5	1.0	1.0	1.0	Proteinase inhibitor Eggplant
63.5	1.0	1.0	1.0	Somatomedin B Human
64.5	1.0	1.0	1.0	Pituitary glycopeptide Pig, bovine, and sheep
65.0	1.0	1.0	1.0	Corticotropin-lipotropin precursor Bovine
65.0	2.0	3.0	2.0	Beta-endorphin II Chum salmon
66.0	1.0	1.0	1.0	Thyrotropin alpha chain Human
66.0	1.0	1.0	1.0	Lutropin alpha chain Human
66.0	1.0	1.0	4.0	Follitropin alpha chain Horse

SUPFAM	FAM	SUBFAM	ENTRY	
66.0	1.0	1.0	4.0	Lutropin alpha chain Horse
67.0	1.0	1.0	1.0	Thyrotropin beta chain Human
67.0	2.0	1.0	2.0	Follitropin beta chain Horse
67.0	2.0	1.0	3.0	Follitropin beta chain Pig
68.0	1.0	0.5	1.0	Prolactin Human
68.0	1.0	3.0	1.0	Prolactin precursor Rat
68.0	2.0	1.0	2.5	Somatotropin precursor Rat
75.0	1.0	1.0	1.3	Insulin Chinchilla
75.0	1.0	1.0	1.6	Insulin Porcupine
75.0	1.0	3.0	2.0	Insulin Casiragua
77.5	1.0	1.0	1.0	Egg-laying hormone Sea hare
78.0	1.0	2.5	1.0	Long neurotoxin 1 Stokes' sea snake (Astrotia stokesii)
78.0	1.0	2.5	2.0	Long neurotoxin 2 Stokes' sea snake (Astrotia stokesii)
78.0	1.0	4.5	1.0	Long neurotoxin 1 Australian tiger snake
78.0	4.1	1.0	1.0	Venom protein S2C4 Jameson's mamba
78.0	4.1	1.0	2.0	Venom protein CM-11 Egyptian cobra
78.0	4.1	1.0	3.0	Short venom protein DE-1 King cobra
78.0	4.3	1.0	1.0	Short toxin 3 Black mamba
78.0	4.6	1.0	1.0	Short toxin CM-2 Egyptian cobra
78.0	5.0	3.0	7.5	Short neurotoxin 1 Mozambique cobra

SUPFAM	FAM	SUBFAM	ENTRY	
78.0	5.0	3.0	7.7	Short neurotoxin 3 Mozambique cobra
78.0	5.0	5.0	3.0	Short neurotoxin 1 Stokes' sea snake (Astrotia stokesii)
78.0	6.0	1.0	11.5	Cytotoxin 5 Egyptian cobra
79.0	1.0	1.0	2.0	Myotoxin a Prairie rattlesnake
79.0	1.0	1.0	3.0	Toxic peptide C Southern pacific rattlesnake
80.0	1.0	2.0	1.0	Neurotoxin III Scorpion (Androctonus australis)
80.0	2.0	2.0	1.0	Neurotoxin V Scorpion (Leiurus)
80.0	2.0	3.0	1.0	Mammalian neurotoxin M10 Scorpion (Buthus eupeus)
83.0	1.0	2.0	1.0	Toxin I Sea anemone (Anemonia sulcata)
83.2	1.0	1.0	1.0	Toxin III Sea Anemone (Anemonia sulcata)
84.0	2.0	1.0	3.0	Purothionin II Barley
88.0	1.0	2.0	9.5	Ig kappa chain V-I region Human Kue
88.0	1.0	3.5	1.0	Ig kappa chain V-II region Dog Gom
88.0	1.0	5.5	1.0	Ig kappa chain V region Mouse M167
88.0	1.0	6.5	1.0	Ig kappa chain precursor V region Mouse K2 gene translation
88.0	1.0	6.5	2.0	Ig kappa chain precursor V region Mouse K3 gene translation
88.0	1.0	9.0	2.0	Ig kappa chain precursor V region Mouse MOPC 321
88.0	1.0	9.0	3.0	Ig kappa chain V region Mouse CBPC.101
88.0	1.0	9.5	1.0	Ig kappa chain V regions Mouse X-44
88.0	1.0	16.0	2.0	Ig kappa chain V region Rabbit K29-213

SUPFAM	FAM	SUBFAM	ENTRY	
88.0	2.0	5.0	2.0	Ig lambda chain V-II region Human Bur
88.0	2.0	7.5	1.0	Ig lambda chain V-II region Human Win
88.0	2.0	9.5	1.0	Ig lambda chain V-IV region Human Hil
88.0	2.0	11.5	1.0	Ig lambda chain V-VI region Human Nig-48
88.0	3.0	1.0	2.0	Ig lambda-2 chain precursor V region Mouse
88.0	7.0	1.5	1.0	Ig heavy chain V-III region Human Hil
88.0	7.0	8.0	2.0	Ig heavy chain V region Mouse X24
88.0	7.0	8.0	3.0	Ig heavy chain V region Mouse J539
88.0	7.0	8.0	3.0	Ig heavy chain V region Mouse X44
88.0	7.0	8.0	4.0	Ig heavy chain V region Mouse T601
88.0	7.0	5.0	1.0	Ig heavy chains V-III regions Human Til
88.0	7.0	10.5	1.0	Ig heavy chain V-III region Human Dob
88.0	7.0	12.5	1.0	Ig heavy chain V-III region Dog Gom
89.0	1.0	3.0	1.0	Ig kappa chain V and C regions Mouse MOPC 21
89.0	2.0	3.0	1.0	Ig lambda-2 chain precursor V and C regions Mouse MOPC 315
89.0	4.0	1.0	2.0	Ig gamma-3 heavy chain disease protein Human Wis
89.0	4.0	3.5	1.0	Ig gamma-2b chain C region Mouse
89.0	4.0	4.0	2.0	Ig gamma-1 chain C region Mouse
89.0	5.0	2.0	1.0	Ig mu chain V and C regions Mouse MOPC 104E
89.0	5.0	3.0	1.0	Ig mu chain V-III region and fragment of C region Dog Moo (fragments)

SUPFAM	FAM	SUBFAM	ENTRY	
89.0	6.0	1.0	2.0	Ig alpha-2 chain V-III and C regions, A2m(2) allotype Human But
89.0	6.0	1.0	2.0	Ig alpha-2 chain C region, A2m(1) allotype Human Lan
89.0	6.0	2.0	1.0	Ig alpha-1 chain V and C regions Mouse MOPC 47A
89.0	7.0	2.0	1.0	Beta-2 microglobulin Rabbit
89.0	8.0	1.0	1.0	Histocompatability antigen HLA-B7 heavy chain Human (fragment)
90.0	1.0	1.0	2.5	Hemoglobin alpha chain Brown lemur
90.0	1.0	1.0	4.5	Hemoglobin alpha chain Treeshrew
90.0	1.0	1.0	6.5	Hemoglobin alpha chain Guinea pig
90.0	1.0	1.0	7.5	Hemoglobin alpha chain European hedgehog
90.0	1.0	1.0	8.5	Hemoglobin alpha chain Badger
90.0	1.0	1.0	10.0	Hemoglobin alpha chain Llama and Arabian camel
90.0	1.0	1.0	11.0	Hemoglobin alpha chain Pig
90.0	1.0	5.0	1.5	Hemoglobin alpha chain (stress-induced) Chicken (plasmid pHb1003)
90.0	1.0	5.0	2.0	Hemoglobin alpha chain Greylag goose
90.0	1.0	5.5	1.0	Hemoglobin pi chains Chicken embryo
90.0	3.0	1.0	2.0	Hemoglobin delta chain Human, chimpanzee, gorilla, and gibbon
90.0	3.0	1.0	4.5	Hemoglobin beta chain Brown lemur
90.0	3.0	1.0	5.5	Hemoglobin beta chain Treeshrew
90.0	3.0	1.0	9.5	Hemoglobin beta chain Guinea pig
90.0	3.0	1.0	9.5	Hemoglobin beta chain European hedgehog

SUPFAM	FAM	SUBFAM	ENTRY	
90.0	3.0	1.0	14.0	Hemoglobin beta chain Pig
90.0	3.0	2.0	1.0	Hemoglobin gamma chain Pig-tailed macaque
90.0	3.0	3.5	1.0	Hemoglobin beta chain Opossum
90.0	3.0	5.0	2.0	Hemoglobin beta chain Greylag goose
90.0	3.1	1.0	1.0	Hemoglobin beta chain Carp
90.0	3.2	1.0	1.0	Hemoglobin beta chain Port Jackson shark
90.0	4.0	1.0	7.5	Myoglobin Fruit bat
90.0	4.0	1.0	40.0	Myoglobin Echidna
90.0	4.5	1.0	1.0	Myoglobin Port Jackson shark
90.0	5.0	1.0	1.0	Globin River lamprey
90.0	5.5	1.0	1.0	Globin III Hagfish
90.0	8.5	1.0	1.0	Globin CTT-I Midge larva
90.0	9.0	1.0	1.0	Globin CTT-II beta Midge larva
90.0	9.0	2.0	1.0	Globin CTT-IX Midge larva
90.0	9.0	3.0	1.0	Globin CTT-VI Midge larva
90.0	9.0	4.0	1.0	Globin CTT-VIIB Midge larva
90.0	9.0	5.0	1.0	Globin CTT-X Midge larva
90.0	10.0	1.0	1.0	Globin CTT-III Midge larva
90.0	11.0	3.0	2.0	Leghemoglobin c Soybean
90.0	11.0	4.0	2.0	Leghemoglobin II Yellow lupin

SUPFAM	FAM	SUBFAM	ENTRY	
91.5	1.0	1.0	1.0	Histone H1.3 Rabbit
91.5	1.0	2.0	1.0	Histone H1 Trout
91.5	1.0	3.0	1.0	Histone H1, gonadal Sea urchin (Parechinus angulosus)
91.5	2.0	1.0	1.0	Histone H5 Goose
92.0	1.0	1.0	1.0	Histone H2A Chicken
92.0	1.0	1.0	3.0	Histone H2A, gonadal Sea urchin (Psammechinus miliaris)
92.0	1.0	1.0	4.0	Histone H2A, embryonic Sea urchins
93.0	1.0	1.0	1.0	Histone H2B Bovine and human
93.0	1.0	1.0	2.0	Histone H2B Brown trout
93.0	1.0	2.0	1.0	Histone H2B, gonadal Limpet (Patella granatina)
93.0	1.0	2.0	2.0	Histone H2B Fruit fly
93.0	1.0	3.0	1.0	Histone H2B, embryonic Sea urchin (Strongylocentrotus purpuratus)
93.0	1.0	4.0	1.0	Histone H2B(1), sperm Sea urchin (Parechinus angulosus)
93.0	1.0	5.0	1.0	Histone H2B(2), sperm Sea urchin (Parechinus angulosus)
93.0	1.0	5.0	1.0	Histone H2B(3), sperm Sea urchin (Parechinus angulosus)
94.0	1.0	1.0	1.0	Histone H3, embryonic Sea urchins
95.8	1.0	1.0	1.0	Nonhistone chromosomal protein HMG-17 Bovine
95.8	1.0	1.0	2.0	Nonhistone chromosomal protein HMG-17 Chicken
95.8	1.0	2.0	1.0	Nonhistone chromosomal protein H6 Rainbow trout
99.0	1.0	5.5	1.0	Protamine (stellin.A) Sturgeon (Acipenser stellatus)

SUPFAM	FAM	SUBFAM	ENTRY	
100.2	1.0	1.0	1.0	DNA-binding protein NS1 Escherichia coli
100.2	1.0	2.0	1.0	DNA-binding protein NS2 Escherichia coli
100.2	2.0	1.0	1.0	DNA-binding protein Thermoplasma acidophilum
100.3	1.0	1.0	1.0	Initiation factor IF-1 Escherichia coli
100.4	1.0	1.0	1.0	Initiation factor IF-3 Escherichia coli
100.6	1.0	1.0	1.0	Ricin D, A chain Castor bean
100.7	1.0	1.0	1.0	Ricin D, B chain Castor bean
100.9	1.0	1.0	1.0	30S ribosomal protein S3 Escherichia coli
101.5	1.0	1.0	1.0	30S ribosomal protein S5 Escherichia coli
102.5	1.0	1.0	1.0	30S ribosomal protein S7 Escherichia coli
108.2	1.0	1.0	1.0	30S ribosomal protein S17 Escherichia coli
109.5	1.0	1.0	1.0	30S ribosomal protein S19 Escherichia coli
111.7	1.0	1.0	1.0	50S ribosomal protein L1 Escherichia coli
111.9	1.0	1.0	1.0	50S ribosomal protein L3 Escherichia coli
112.5	1.0	1.0	1.0	50S ribosomal protein L6 Escherichia coli
113.0	1.0	2.0	1.0	50S ribosomal protein L9 Bacillus subtilis
113.5	1.0	1.0	1.0	60S ribosomal protein 44 Baker's yeast
114.1	1.0	1.0	1.0	50S ribosomal protein L11 Escherichia coli
114.4	1.0	1.0	1.0	50S ribosomal protein L13 Escherichia coli
114.5	1.0	1.0	1.0	50S ribosomal protein L14 Escherichia coli

SUPFAM	FAM	SUBFAM	ENTRY	
114.8	1.0	1.0	1.0	50S ribosomal protein L15 Escherichia coli
116.2	1.0	1.0	1.0	50S ribosomal protein L19 Escherichia coli
116.3	1.0	1.0	1.0	50S Ribosomal protein L20 Escherichia coli
116.4	1.0	1.0	1.0	50S ribosomal protein L21 Escherichia coli
116.7	1.0	1.0	1.0	50S ribosomal protein L23 Escherichia coli
116.8	1.0	1.0	1.0	50S ribosomal protein L24 Escherichia coli
118.5	1.0	1.0	1.0	50S ribosomal protein L28 Escherichia coli
120.5	1.0	1.0	1.0	50S ribosomal protein L31 Escherichia coli
123.1	1.0	1.0	1.0	80S Ribosomal protein eL12 Brine shrimp
123.2	1.0	1.0	1.0	Structural protein VP1 Simian virus 40
123.3	1.0	1.0	1.0	Structural protein VP2 Simian virus 40
123.3	1.0	1.0	1.0	Structural protein VP3 Simian virus 40
123.4	1.0	1.0	1.0	Viral T (tumor) antigen Simian virus 40
123.4	2.0	1.0	1.0	Viral t (tumor) antigen Simian virus 40
123.8	1.0	1.0	1.0	Hemagglutinin precursor Fowl influenza virus
127.0	3.0	1.0	1.0	Coat protein Bacteriophage PRR1
128.0	3.0	1.0	1.0	Coat protein Bacteriophage Xf (Xanthomonas oryzae)
128.2	1.0	1.0	1.0	Regulatory protein O Bacteriophage lambda
128.3	1.0	1.0	1.0	Regulatory protein cII Bacteriophage lambda
128.4	1.0	1.0	1.0	Regulatory protein Cro Bacteriophage lambda

SUPFAM	FAM	SUBFAM	ENTRY	
128.5	1.0	1.0	1.0	Repressor protein Bacteriophage lambda
129.4	1.0	1.0	1.0	Inclusion body protein Nuclear polyhedrosis virus
131.2	1.0	1.0	1.0	Small outer capsid protein Bacteriophage T4
131.3	1.0	1.0	1.0	Internal peptide VII Bacteriophage T4, T2, And T6
131.8	1.0	2.0	1.0	Gene A and A* proteins Bacteriophage phi-X174
131.8	1.0	2.0	1.0	Gene A and A* proteins Bacteriophage G4
132.0	1.0	2.0	1.0	Gene B protein Bacteriophage G4
132.1	1.0	1.0	1.0	Gene C protein Bacteriophage phi-X174
132.1	1.0	2.0	1.0	Gene C protein Bacteriophage G4
133.0	1.0	1.0	2.0	Gene D protein Bacteriophage G4
134.0	1.0	2.0	1.0	Gene E protein Bacteriophage G4
134.5	1.0	1.0	1.0	Gene F protein Bacteriophage phi-X174
134.5	1.0	2.0	1.0	Gene F protein Bacteriophage G4
135.0	2.0	1.0	1.0	Gene G protein Bacteriophage G4
135.5	1.0	1.0	1.0	Gene H protein Bacteriophage phi-X174
135.5	1.0	2.0	1.0	Gene H protein Bacteriophage G4
136.0	2.0	1.0	1.0	Gene J protein Bacteriophage G4
136.1	1.0	1.0	1.0	Gene K protein Bacteriophage phi-X174
136.1	1.0	2.0	1.0	Gene K protein Bacteriophage G4
138.0	1.0	1.0	2.5	Alpha crystallin A chain, minor component Rat

SUPFAM	FAM	SUBFAM	ENTRY	
146.0	1.0	1.0	1.0	Tropomyosin alpha chain, skeletal muscle Rabbit
146.0	1.0	1.0	2.0	Tropomyosin beta chain, skeletal muscle Rabbit
147.0	1.0	1.0	1.0	Actin Rabbit
147.0	1.0	1.0	2.0	Actin Physarum polycephalum
147.2	1.0	1.0	1.0	Profilin Bovine
148.0	4.0	1.0	1.0	Myosin DTNB light chain, skeletal muscle Rabbit
148.0	4.0	1.0	2.0	Myosin DTNB light chain, skeletal muscle Chicken
148.0	4.5	1.0	1.0	Myosin EDTA light chain Scallop
148.0	5.0	6.5	1.0	Parvalbumin beta Coelacanth
148.0	5.0	8.0	1.0	Parvalbumin alpha Coelacanth
148.0	6.0	1.0	1.0	Calcium-binding protein, intestinal Pig
148.0	7.0	1.0	1.0	S-100 protein Bovine
150.0	1.0	1.0	2.0	Troponin I, skeletal muscle Chicken
150.0	1.0	1.5	1.0	Troponin I, slow skeletal muscle Rabbit
151.0	3.5	1.0	1.0	Lipid-binding protein C-II Human
152.0	1.0	2.0	2.0	Apovitellenin I Turkey
160.0	1.0	1.0	2.0	Neurophysin 1 Pig
167.0	1.0	3.0	1.0	Complement C3a anaphylatoxin Rat
167.0	2.0	1.0	1.0	Complement C5a anaphylatoxin Human
167.2	1.0	1.0	1.0	J chain Human

SUPFAM	FAM	SUBFAM	ENTRY	
169.2	1.0	1.0	1.0	Beta-thromboglobulin Human
169.2	2.0	1.0	1.0	Platelet factor 4 Human
169.4	1.0	1.0	1.0	Uteroglobin precursor Rabbit
170.0	2.0	1.0	1.0	Gamma-carboxyglutamic acid-containing protein Swordfish
171.0	1.0	1.0	0.5	Metallothionein-1A Horse
171.0	1.0	1.0	2.0	Metallothionein-2 Human
171.0	1.0	2.0	1.0	Metallothionein-I Mouse
171.0	2.0	1.0	1.0	Metallothionein Neurospora crassa
172.1	1.0	1.0	1.0	Proline-rich phosphoprotein A Human
172.2	1.0	1.0	1.0	Retinol-binding protein Human
173.5	1.0	1.0	1.0	Chorion class B protein pc401 precursor Silkmoth (Antheraea polyphemus)
173.5	1.0	2.0	1.0	Chorion class B protein pc10 Silkmoth (Antheraea polyphemus)
173.6	1.0	1.0	1.0	Proline-rich peptide P-B Human
174.2	1.0	1.0	1.0	Secapin Honey bee
174.5	1.0	1.0	1.0	Sillucin Mucor pusillus
175.0	2.0	1.0	1.0	Favin alpha chain Broad bean
175.0	2.0	1.0	2.0	Lectin alpha chain Pea
175.0	3.0	1.0	1.0	Favin beta chain Broad bean
179.5	1.0	1.0	1.0	Thaumatin I Thaumatococcus daniellii
180.1	1.0	1.0	1.0	Citrate lyase acyl carrier protein Klebsiella aerogenes

SUPFAM	FAM	SUBFAM	ENTRY	
180.2	1.0	1.0	1.0	Biotin carboxyl carrier protein (in EC 2.1.3.1) Propionibacterium shermanii
180.4	1.0	1.0	1.0	Phosphocarrier protein HPr Staphylococcus aureus
180.6	1.0	1.0	1.0	LIV-binding protein Escherichia coli
180.7	1.0	1.0	1.0	Outer membrane protein Ia Escherichia coli B/r
180.8	1.0	1.0	1.0	L-Arabinose-binding protein Escherichia coli
181.2	1.0	1.0	1.0	Bacteriorhodopsin Halobacterium halobium
181.4	1.0	1.0	1.0	Spore protein A Bacillus megaterium
182.1	1.0	1.0	1.0	Lactose permease Escherichia coli

Total Number of Sequences: 348

Alphabetical listing of other new entries

2-Keto-3-deoxy-6-phosphogluconic aldolase
Pseudomonas putida (fragment)

3-Oxoadipate enol-lactonase II
Acinetobacter calcoaceticus (fragment)

Alpha-1-antitrypsin
Human (fragment)

Alpha-amylase (EC 3.2.1.1)
Bacillus amyloliquefaciens (fragment)

Aminopeptidase I alpha chain
Bacillus stearothermophilus (fragment)

Anthranilate phosphoribosyltransferase
Serratia marcescens (fragment)

Anthranilate phosphoribosyltransferase
Erwinia caratovora (fragment)

Anthranilate synthase, component I
Escherichia coli (fragment)

Anthranilate synthase, component II
Escherichia coli (fragment)

Aspartate carbamoyltransferase C chain
Escherichia coli (fragment)

Beta-endorphin I
Chum salmon

Biotin carboxyl carrier protein (in EC 6.4.1.2)
Escherichia coli (fragment)

Carboxyl protease
Mucor miehei (fragments)

Ceruloplasmin
Human (fragment)

Coagulogen
Japanese horseshoe crab (fragment)

Corticotropin-lipotropin precursor
Pig (fragment)

Creatine kinase
Rabbit (fragment)

D-Serine dehydratase
Escherichia coli (fragments)

Factor XIII (EC 2.3.2.13) a chain
Human (fragment)

Fructose bisphosphatase
Rabbit (fragment)

Glucose-1-phosphate adenylyltransferase
Escherichia coli B (fragment)

Gonadotropin alpha chain
Carp (fragments)

Gonadotropin beta chain
Carp (fragments)

Hemoglobin epsilon chain
Human (fragments)

Histidine decarboxylase small chain
Micrococcus sp. (fragment)

Histocompatability antigen H-2Kb heavy chain
Mouse (fragment)

Histocompatability antigen HLA-A2 heavy chain
Human (fragment)

Histone H2A1
Wheat germ (fragment)

Histone H2A2
  Wheat germ (fragment)

Histone H2A3
  Wheat germ (fragment)

Histone H2B
  African crocodile (fragments)

Histone H2B
  Chicken (fragments)

Histone H2B
  South African toad (fragments)

Histone H4
  Tetrahymena thermophila (fragments)

Histone H5
  Pigeon (fragment)

Indoleglycerol phosphate synthase
  Escherichia coli (fragment)

Inter-alpha-trypsin inhibitor (BPI-type)
  Human (fragment)

Isomaltase
  Rabbit (fragment)

Lectin
  Lentil (fragments)

Limulin
  Horseshoe crab (fragment)

Lipotropin beta (and beta-endorphin)
  Mouse (fragment)

Macromomycin
  Streptomyces macromomyceticus (fragment)

Myelin P2 protein
  Rabbit (fragment)

Ribonuclease (colicin E3, A chain)
  Escherichia coli plasmid (fragment)

Ribulose-1,5-bisphosphate carboxylase (EC 4.1.1.39)
  Chlamydomonas reinhardtii (fragment)

Serum albumin precursor
  Rat (fragments)

Transaminase B (EC 2.6.1.32)
  Escherichia coli (fragment)

Triacylglycerol lipase (EC 3.1.1.3)
  Pig (fragment)

Total Number of Sequences: 48