ATLAS of PROTEIN SEQUENCE and STRUCTURE 1965
DEDICATION

To all the investigators who have developed the techniques necessary for the grand accomplishments represented by this tabulation, and to all those who have spent so much tedious effort in their application.

We would gratefully appreciate receiving suggestions, corrections, new data (even if fragmentary or provisional), and references to any data omitted from this volume.

M. O. D.
R. V. E.
M. A. C.
M. R. S.
This Atlas voluminously illustrates the triumph of experimental technique over the secretiveness of nature. Perhaps nowhere has the power of the scientific method been more brilliantly demonstrated than in the development of procedures for the study of the chemistry of life. As recently as twenty years ago, it was customary for biologists to have a hopeless attitude about biochemistry. Some details might be elicited, perhaps, but living things were thought to be so very complex and intricate that there surely was no hope of fully "understanding" them in all their chemical detail. Who, if he really comprehended the difficulty of the problem, would dare to think of man's ever knowing the detailed structure of a protein, for example, much less be able to synthesize it? Who would ever understand the mechanism of an enzyme as clearly as a chemist understands the details of an inorganic reaction? How could we ever hope to know the atomic details of a protein crystal?

Today some of these ambitions have already been attained, and the others no longer seem out of reach. We now rationally hope to be able to discover and understand the finest chemical details of living processes. These accomplishments and hopes have been made possible by the combined effect of several new approaches.

Techniques which permit the separation of chemically similar compounds have been developed for microgram samples. Among these are ion-exchange columns, paper chromatography, electrophoresis, and counter-current distribution. Radioactive tracer techniques and other micro-quantitative analytical procedures, often dependent on electronics and automation, have aided the analyses. X-ray crystallography, starting with the art of protein crystal production and ending with the processing of great numbers of experimental observations in the high-speed computer, has permitted a glimpse of three-dimensional structure.

Confidence in our understanding of experimental procedures and relationships among proteins has grown so great that sequences of amino acids are inferred from those found in homologous proteins. This technique requires only a small proportion of the analytical work needed to sequence a protein with no known relatives. The effectiveness of laboratory effort is thus magnified.

Some of the insights which have been developed cannot be attributed to any particular worker or school. Perhaps the greatest of these insights is that nature always uses "building blocks." A living cell is extremely complex and almost unimaginably intricate in detail. But it consists of a limited, understandable number of types of processes, reduplicated with variations. To understand the cell, we must have a few examples of each type of process, from which we can see the overall principles. For understanding, we need not work out the details of all the variations on these principles, although we may eventually choose to do so for medical or other practical reasons. Similarly, the analysis of such large, complex chemical molecules as proteins has been made possible by the recognition of their essential modularity, their building-block nature. Proteins are precise chemical structures built from regular subunits,
not statistical mixtures or hopelessly intricate molecular conglomerates as was once thought. It is by means of the discovery and utilization of such building block principles, combined with the large-scale application of new and improved techniques, that we now dare hope to make all of living nature accessible to our understanding.

Hidden in the amino acid sequence of a protein is the chemical information that produces its three-dimensional structure. In the case of an enzyme, this structure forms locks into which the proper keys—cell chemicals—fit. By these locks, the enzymes bring the proper reactants together quickly, efficiently, and selectively. Uncatalyzed reactions cannot compete with such specifically catalyzed reactions; therefore, the presence of enzymes determines which reactions can take place in living chemistry. In many cases, if not all, this three-dimensional structure is fully determined by the information in the one-dimensional sequence. The folding is the thermodynamically most stable result of all the possible intermolecular forces, such as hydrogen bonds and hydrophobic bonds, which can form between the various links of the chain. In principle, if we knew these forces in detail, and if we had appropriate computer routines, we should then be able to determine the three-dimensional structure of a protein, given only its amino acid sequence.

Also hidden in the sequences is information about the genes which directed their synthesis. For each amino acid there are a small number of possible corresponding nucleotide triplets in the gene. That is, each protein sequence corresponds to a limited number of possible nucleotide sequences. When nucleotide mutations occur, the substitution of alternative amino acids is not random. Analysis of amino acid sequence data, considered as a mathematical puzzle, can help elucidate both the mathematical details of the genetic code and the structural aspects of the genetic mechanism.

Hidden in each family of homologous sequences is the story of its evolution. Simple organisms, caught in their primitive ecological niches, still preserve even today enzymes performing primeval functions, held relatively fixed by natural selection. Even the older proteins of man are preserved as living "fossils" in his metabolism.

Enmeshed also in homologous sequences are the records of the many thousands of mutational steps by which we can quantify a phylogenetic tree. Each amino acid link is a trait by which we can trace species evolution. By comparison, the traditional taxonomic criteria are extremely vague and uncertain. In the case of distant relationships, they often break down completely. A truly quantitative and inclusive system of phylogenetic classification would be of great help to comparative physiologists and other students of evolution.

Conspicuous in comparative human protein sequences is information of great medical-diagnostic value. A long series of abnormalities has been found to be attributable to single amino acid replacements. One such tragically serious disease is sickle-cell anemia.

To facilitate the theoretical study of the protein sequences which have already been so ingeniously and laboriously determined, we have undertaken this compilation.
The information is kept in a compact, uniform format on punched cards. New information and corrections are easily inserted, while the text is kept accurate.

It is our intention to include the currently accepted amino acid sequence of every protein for which complete or substantial data is available. Usually, only the definitive report giving the complete sequence from each laboratory will be referenced. If a substantial amount of work has been done on the same protein in other laboratories, their reports will also be referenced. We have also included some smaller peptides that have come to our attention. Unusual polypeptides which are presumably not produced by the genetic code have been omitted.

The format in which the Atlas is kept on punched cards is suitable for direct use in our computer programs. We use a three-letter code, which is a slight modification of the conventional notation, and also a mnemonic one-letter code which is clearer and much more suitable for certain comparative studies. We use a system of punctuation to describe the degree of confidence in each bond. Brief remarks are also included about the nature and function of the protein, and additional structural information such as the attachment of prosthetic groups, the location of S-S bonds, amino acids involved in active sites, and three-dimensional structures. In later editions we intend to include a section in which the alignment of all sequences of each family is given. Possibly we will also have sections on alleles and on mathematical methods and computer programs to treat the information.

This first edition is incomplete and imperfect and is intended mainly for distribution to investigators who have published protein sequence analyses, to acquaint them with the existence of this Atlas. We would gratefully appreciate their cooperation in making corrections, additions and suggestions for future editions. Since sequences are being reported in great numbers, we plan to bring out supplementary material in six months and a second edition in a year.

We thank all those who have assisted with this compilation, particularly Mr. Javier Albarran for his help with the computer aspects and Miss Lorrie Goldstein for her design of the cover.

The tabulations and computations were made at the University of Maryland Computer Science Center, College Park.

This work was supported by Grants GM-08710 and GM-12168 from the National Institutes of Health to the National Biomedical Research Foundation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>PREFACE</th>
<th>CONTENTS</th>
<th>EXPLANATION OF NOTATION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.002</td>
<td>0.010</td>
<td></td>
</tr>
</tbody>
</table>

## I. CYTOCHROME C

<table>
<thead>
<tr>
<th>Baker's Yeast</th>
<th>Chicken</th>
<th>Horse</th>
<th>Human</th>
<th>Pig and Bovine</th>
<th>Pseudomonas</th>
<th>Tuna Fish</th>
<th>Bombyx Mori (Silkworm)</th>
<th>Rattlesnake</th>
<th>Rhodospirillum Rubrum</th>
<th>Salmon</th>
<th>CY BY</th>
<th>CY CH</th>
<th>CY HO</th>
<th>CY Hu</th>
<th>CY PG</th>
<th>CY PS</th>
<th>CY TF</th>
<th>CY SW</th>
<th>CY RS</th>
<th>CY RR</th>
<th>CY SM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.001</td>
<td>1.002</td>
<td>1.003</td>
<td>1.004</td>
<td>1.005</td>
<td>1.006</td>
<td>1.007</td>
<td>1.008</td>
<td>1.009</td>
<td>1.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## II. GLOBINS

<table>
<thead>
<tr>
<th>HEMOGLOBINS</th>
<th>GL HUHA</th>
<th>GL HUHB</th>
<th>GL HUHG</th>
<th>GL GOHB</th>
<th>GL HOHB</th>
<th>GL HOHA</th>
<th>GL LEBH</th>
<th>GL HUH</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Alpha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.000</td>
</tr>
<tr>
<td>Human Beta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.002</td>
</tr>
<tr>
<td>Human Gamma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.003</td>
</tr>
<tr>
<td>Gorilla Beta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.004</td>
</tr>
<tr>
<td>Horse Beta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.005</td>
</tr>
<tr>
<td>Horse Alpha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.006</td>
</tr>
<tr>
<td>Lemur Fulvus Beta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.007</td>
</tr>
<tr>
<td>Abnormal Human Hemoglobins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.020</td>
</tr>
<tr>
<td>Myoglobin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.100</td>
</tr>
<tr>
<td>Whale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.101</td>
</tr>
</tbody>
</table>

## III. OTHER RESPIRATORY PROTEINS

<table>
<thead>
<tr>
<th>DIHEME PEPTIDE - CHROMATIUM</th>
<th>DH CH</th>
<th>FE CP</th>
<th>AZ PS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferredoxin - Clostridium Pasteurianum</td>
<td></td>
<td></td>
<td></td>
<td>3.001</td>
</tr>
<tr>
<td>Azurin - Pseudomonas</td>
<td></td>
<td></td>
<td></td>
<td>3.002</td>
</tr>
</tbody>
</table>

## IV. RIBONUCLEASE

<table>
<thead>
<tr>
<th>Bovine</th>
<th>RN BO</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.001</td>
<td></td>
</tr>
</tbody>
</table>

## V. INHIBITORS

<table>
<thead>
<tr>
<th>Trypsin Inhibitor - Bovine</th>
<th>TI BOPA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.000</td>
<td></td>
</tr>
</tbody>
</table>
VI. TOBACCO MOSAIC VIRUS

TOBACCO MOSAIC VIRUS ...................................................... TM TM 6.001
TOBACCO MOSAIC VIRUS DAHLMENSE ........................................ TM TMD 6.002

VII. DIGESTIVE ENZYMES

CHYMOTRYPSINOGEN-A - BOVINE ............................................. TR BOCH 7.001
TRYPsinogen - BOVINE ........................................................ TR BOTR 7.002
PAPAIN ............................................................. PA PA 7.101
LYSOZYME - CHICKEN ..................................................... LS CH 7.201

VIII. HORMONES

GLUCAGON

BOVINE .................................................. GN BO 8.001

PRESSINS

VASOPRESSIN ARGinine .................................................. PR BOAR 8.101
VASOPRESSIN LYSINE .................................................... PR PGBS 8.102
OXYTOCIN .............................................................. PR BOOH 8.103
HYPERTENSIN ............................................................ PR BOHY 8.104

CORTICOIDs

ALPHA MELANOcyTE-STIMULATING HORMONE
BOVINE, PIG, AND HORSE .............................................. TN BPAM 8.201

BETA MELANOcyTE-STIMULATING HORMONE
BOVINE .............................................................. TN BOBM 8.202
PIG ................................................................. TN PGBM 8.203
HORSE ............................................................... TN HOBM 8.204
HUMAN ................................................................. TN HUBM 8.205

BETA CORTICOTROPIN - PIG ............................................. TN PGBM 8.206
ALPHA CORTICOTROPIN - SHEEP AND BOVINE ....................... TN SBAC 8.207

INSULINS

INSULIN A
BOVINE .............................................................. IS BOA 8.301
BONITO ............................................................... IS BNA 8.302
HORSE ............................................................... IS HOA 8.303
SHEEP ................................................................. IS SHA 8.304
SPERM WHALE, FIN-WHALE, PIG, AND HUMAN ......................... IS WPA 8.305
SEI-WHALE ............................................................ IH WHA 8.306

INSULIN B
BOVINE, SHEEP, HORSE, HUMAN, PIG, AND SPERM WHALE ............ IS BOB 8.321
BONITO ............................................................... IS BNB 8.322
IX. PLASMA PROTEINS

FIBRINOPEPTIDE A

<table>
<thead>
<tr>
<th>Species</th>
<th>FB BOA</th>
<th>FB SHA</th>
<th>FB GTA</th>
<th>FB RDA</th>
<th>FB PGA</th>
<th>FB HUA</th>
<th>FB RTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOVINE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.001</td>
</tr>
<tr>
<td>SHEEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.002</td>
</tr>
<tr>
<td>GOAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.003</td>
</tr>
<tr>
<td>REINDEER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.004</td>
</tr>
<tr>
<td>PIG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.005</td>
</tr>
<tr>
<td>HUMAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.006</td>
</tr>
<tr>
<td>RABBIT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.007</td>
</tr>
</tbody>
</table>

FIBRINOPEPTIDE B

<table>
<thead>
<tr>
<th>Species</th>
<th>FB BOB</th>
<th>FB SHB</th>
<th>FB GTB</th>
<th>FB RDB</th>
<th>FB PGB</th>
<th>FB HUB</th>
<th>FB RTB</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOVINE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.101</td>
</tr>
<tr>
<td>SHEEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.102</td>
</tr>
<tr>
<td>GOAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.103</td>
</tr>
<tr>
<td>REINDEER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.104</td>
</tr>
<tr>
<td>PIG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.105</td>
</tr>
<tr>
<td>HUMAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.106</td>
</tr>
<tr>
<td>RABBIT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.107</td>
</tr>
</tbody>
</table>

X. GLOBULINS

<table>
<thead>
<tr>
<th>Species</th>
<th>BJ HU</th>
</tr>
</thead>
<tbody>
<tr>
<td>BENCE-JONES</td>
<td></td>
</tr>
</tbody>
</table>

10.000

AUTHOR INDEX

100.000
THE MEANING OF THE PUNCTUATION IS AS FOLLOWS.

BLANK SEQUENCE OF AMINO ACIDS HAS BEEN DETERMINED.

( ) ENCLOSE PORTION OF SEQUENCE NOT SPECIFICALLY DETERMINED. TO PRESERVE PROPER SPACING, IS USED INSTEAD OF )

, SEPARATES AMINO ACIDS WITHIN PARENTHESES,

BUT * SEPARATES AMINO ACIDS, WHERE THE SEQUENCE IS PRESUMED BY HOMOLOGY WITH A KNOWN SEQUENCE.

/ OR /// FRAGMENT, CONNECTION UNDETERMINED

* OR ** CARBOXYL END OF PROTEIN

ASTERISK BEFORE REFERENCE INDICATES THAT THE SEQUENCE WAS COPIED FROM, AND PROOFREAD AGAINST, THE ORIGINAL ARTICLE.

= BEFORE REFERENCE INDICATES THAT WE HAVE NOT SEEN THE ORIGINAL ARTICLE.

NO MARK BEFORE REFERENCE INDICATES OTHER GROUPS WHICH HAVE ALSO REPORTED WORK ON THE SAME PROTEIN.
BOTH SINGLE- AND THREE-LETTER NOTATIONS ARE USED, AS FOLLOWS.

A = ALA = ALANINE
C = CYS = CYSTEINE
D = ASP = ASPARTIC ACID
E = GLU = GLUTAMIC ACID
F = PHE = PHENYLALANINE
G = GLY = GLYCINE
H = HIS = HISTIDINE
I = ILE = ISELEUCINE
K = LYS = LYSINE
L = LEU = LEUCINE
M = MET = METHIONINE
N = ASP = ASPARAGINE
O = GLN = GLUTAMINE
P = PRO = PROLINE
Q = GLN = GLUTAMINE
R = ARG = ARGinine
S = SER = SERINE
T = THR = THREONINE
V = VAL = VALINE

B = ASX = ASPARTIC ACID OR ASPARAGINE
Z = GLX = GLUTAMIC ACID OR GLUTAMINE
X = XXX = UNDETERMINED OR OTHERWISE UNUSUAL

MNEMONICS OF THE ONE-LETTER CODE

IF POSSIBLE, THE INITIAL LETTER OF THE AMINO ACID IS USED.
IF MORE THAN ONE AMINO ACID BEGINS WITH THE SAME LETTER,
THE MOST COMMONLY-OCCURRING ONE IS ASSIGNED THE INITIAL.

A = ALANINE
C = CYSTEINE
G = GLYCINE
H = HISTIDINE
I = ISELEUCINE
K = LYSINE
L = LEUCINE
M = METHIONINE
P = PROLINE
S = SERINE
T = THREONINE
V = VALINE

SOME OF THE OTHERS ARE PHONETICALLY SUGGESTIVE.

F = PHENYLALANINE
R = ARGinine
Q = TYROSINE

DOUBLE RING IN THE SIDE CHAIN.
W = Tryptophan

THE TWO ACIDS ARE ADJACENT, IN ALPHABETICAL ORDER.

D = ASPARTIC ACID
E = GLUTAMIC ACID

THE TWO AMINES HAVE LETTERS FROM THE MIDDLE OF THE ALPHABET.

N = ASPARAGINE (CONTAINS N)
Q = GLUTAMINE ('Q-TAMINE')

NON-INITIAL LETTER AS CLOSE AS POSSIBLE TO ITS INITIAL.

K = LYSINE
CYTOCHROME C – BAKER’S YEAST

HEME BONDED TO CYSTEINES AT POSITIONS 19 AND 22.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
31 H K V G P N L H G I F G R H S G Q A Q G O S O T D A N I K K
61 N V L W D E N N M S E O L T N P K K O I P G T K M A F G G L
91 K K E K D R N D L I T O L K K A C E

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 THR GLU PHE LYS ALA GLY SER ALA LYS LYS GLY ALA THR LEU PHE
LYS THR ARG CYS GLU LEU CYS HIS THR VAL GLU LYS GLY GLY PRO
31 HIS LYS VAL GLY PRO ASN LEU HIS GLY ILU PHE GLY ARG HIS SER
GLY GLN ALA GLN GLY TYR SER TYR THR ASP ALA ASN ILU LYS LYS
61 ASN VAL LEU TRP ASP GLU ASN ASN MET SER GLU TYR LEU THR ASN
PRO LYS LYS TYR ILU PRO GLY THR LYS MET ALA PHE GLY GLY LEU
91 LYS LYS GLU LYS ASP ARG ASN ASP LEU ILU THR TYR LEU LYS LYS
ALA CYS GLU ***

COMPOSITION

| 7 ALA A | 2 GLN Q | 8 LEU L | 4 SER S |
| 3 ARG R | 7 GLU E | 16 LYS K | 8 THR T |
| 7 ASN N | 12 GLY G | 2 MET M | 1 TRP W |
| 4 ASP D | 4 HIS H | 4 PHE F | 5 TYR O |
| 3 CYS C | 4 ILU I | 4 PRO P | 3 VAL V |

TOTAL NO. OF ACIDS = 108

© NARITA, K., TITANI, K., YAOI, Y., MURAKAMI, H., BIOCHIM. BIOPHYS. ACTA, VOL. 77, PP. 688-690, 1963
CYTOCHROME C - CHICKEN

ACETYL AT AMINO END.
HEME BONDED TO CYSTEINES AT POSITIONS 14 AND 17.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 G D I E K G K K I F V Q K C S Q C H T V E K G G K H K T G P
3 1 N L H G L F G R K T G Q A E G F S O T D A N K N K G I T W G
9 1 R V D L I A D L K K A T N S •

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLY ASP ILU GLU LYS GLY LYS LYS ILU PHE VAL GLN LYS CYS SER
GLN CYS HIS THR VAL GLU LYS GLY GLY LYS HIS LYS THR GLY PRO
3 1 ASN LEU HIS GLY LEU PHE GLY ARG LYS THR GLY GLN ALA GLU GLY
PHE SER TYR THR ASP ALA ASN LYS ASN LYS GLY ILU THR TRP GLY
6 1 GLU ASP THR LEU MET GLU TYR LEU GLU ASN PRO LYS LYS TYR ILU
PRO GLY THR LYS MET ILU PHE ALA GLY ILU LYS LYS LYS SER GLU
9 1 ARG VAL ASP LEU ILU ALA TYR LEU LYS LYS ALA THR ASN SER •••

COMPOSITION

5 ALA A
2 ARG R
5 ASN N
4 ASP D
2 CYS C
3 GLN Q
7 GLU E
13 GLY G
3 HIS H
7 ILU I
6 LEU L
18 LYS K
2 MET M
4 PHE F
3 PRO P
4 SER S
8 THR T
1 TRP W
3 VAL V

TOTAL NO. OF ACIDS = 104

• MARGOLIASH, E., NEEDLEMAN, S. B. AND STEWART, J. W., ACTA CHEM. SCAND.,
VOL. 17, SUPPL. 1, PP. 250-256, 1963
CYTOCHROME C - HORSE

ACETYL AT AMINO END.
HEME BONDED TO CYSTEINES AT POSITIONS 14 AND 17.
OXIDATION-REDUCTION POTENTIAL EQUALS .250 V.

123456789012345678901234567890
IGDVEKGGKIKFVQKCAQCHTVEKGGKHKTP
31NLHGLFGRKTQAPGFTDTDAKKNKGTW
61EETLMEOLENPKKOPGTKMIAGIKKKTE
91REDLIAOLKKATNE*

123456789101112131415
1GLYASPVALGLULYSGLYLYSLYSLIPHELVALGLNLYSLYSCSALA
GLN CYS HIS THR VAL GLU LYS GLY GLY LYS HIS LYS THR GLY PRO
31ASNLEUHISGLYLEUPHELGLYGLYARGLYSTHRGLYGLNLALAPROGLY
PHELTHRTHRALASNLYSASNGLYILUHTHTRGLYL
61GLULUTHTLUIMETGLUTYRELUGLUASNPROLYSLYSTYRILU
PRONGLYTHRLYMETILUPHELALAGLYILUFLYSLYSTHRLQ
91ARGGLUASPLEUALATHTLUFLYSLYSALATHRASNGLU***

COMPOSITION

<table>
<thead>
<tr>
<th>6 ALA</th>
<th>3 GLN</th>
<th>6 LEU</th>
<th>0 SER</th>
<th>5 ASN</th>
<th>12 GLY</th>
<th>2 MET</th>
<th>1 TRP</th>
<th>2 CYS</th>
<th>6 ILU</th>
<th>3 HIS</th>
<th>4 PHE</th>
<th>4 TYR</th>
<th>2 ARG</th>
<th>9 GLU</th>
<th>19 LYS</th>
<th>10 THR</th>
<th>3 ASP</th>
<th>3 HIS</th>
<th>4 PRO</th>
<th>3 VAL</th>
</tr>
</thead>
</table>

TOTAL NO. OF ACIDS = 104

CYTOCHROME C - HUMAN
ACETYL AT AMINO END.
HEME BONDED TO CYSTEINES AT POSITIONS 14 AND 17.
LEU (L) REPLACES MET (M) AT POSITION 65 IN 10 PERCENT
YIELD IN POOLED PROTEIN.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
31 N L H G L F G R K T G Q A P G O S O T A A N K N K G I I W G
91 R A D L I A O L K K A T N E

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLY ASP VAL GLU LYS GLY LYS LYS ILU PHE ILU MET LYS CYS SER
GLN CYS HIS THR VAL GLU LYS GLY GLY LYS HIS LYS THR GLY PRO
31 ASN LEU HIS GLY LEU PHE GLY ARG LYS THR GLY GLN ALA PRO GLY
TYR SER TYR THR ALA ALA ASN LYS ASN LYS GLY ILU ILU TRP GLY
61 GLU ASP THR LEU MET GLU TYR LEU GLU ASN PRO LYS LYS TYR ILU
PRO GLY THR LYS MET ILU PHE VAL GLY ILU LYS LYS GLU GLU
91 ARG ALA ASP LEU ILU ALA TYR LEU LYS LYS ALA THR ASN GLU ***

**COMPOSITION**

| 6 ALA A | 2 GLN Q | 6 LEU L | 2 SER S |
| 2 ARG R | 8 GLU E | 18 LYS K | 7 THR T |
| 5 ASN N | 13 GLY G | 3 MET M | 1 TRP W |
| 3 ASP D | 3 HIS H | 3 PHE F | 5 TYR O |
| 2 CYS C | 8 ILU I | 4 PRO P | 3 VAL V |

TOTAL NO. OF ACIDS = 104

* MATSUBARA, H., AND SMITH, E. L., J. BIOL. CHEM., VOL. 237, NO. 11, PC3575-PC3576, NOV., 1962*
CYTOCHROME - C PIG AND BOVINE

ACETYL AT AMINO END.
HEME BONDED TO CYSTEINES AT POSITIONS 14 AND 17.

```
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P
31 N L H G L F G R K T G Q A P G F S O T D A N K N K G I T W G
91 R E D L I A D L K K A T N E
```

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLY ASP VAL GLU LYS GLY LYS GLY LYS ILU PHE VAL GLN LYS CYS ALA
GLN CYS HIS THR VAL GLU LYS GLY GLY LYS HIS LYS THR GLY PRO
31 ASN LEU HIS GLY LEU PHE GLY ARG LYS THR GLY GLN ALA PRO GLY
PHE SER TYR THR ASP ALA ASN LYS ASN LYS GLY ILU THR TRP GLY
61 GLU GLU THR LEU MET GLU TYR LEU GLU ASN PRO LYS LYS TYR ILU
PRO GLY THR LYS MET ILU PHE ALA GLY ILU LYS LYS GLY GLU
91 ARG GLU ASP LEU ILU ALA TYR LEU LYS LYS ALA THR ASN GLU ***

**COMPOSITION**

| 6  | ALA | 3  | GLN | 6  | LEU | 1  | SER | S |
| 2  | ARG | 9  | GLU | 18 | LYS | 8  | THR | T |
| 5  | ASN | 14 | GLY | 2  | MET | 1  | TRP | W |
| 3  | ASP | 3  | HIS | 4  | PHE | 4  | TYR | D |
| 2  | CYS | 6  | ILU | 4  | PRO | 3  | VAL | V |

TOTAL NO. OF ACIDS = 104

* MARGOLIASH, E., NEEDLEMAN, S. B. AND STEWART, J. W., ACTA CHEM. SCAND., VOL. 17, SUPPL. 1, PP. 250-256, 1963 (PIG)
TUPPY H., AND PALEUS, S., ACTA CHEM. SCAND., VOL. 13, NO. 4, PP. 641-646, 1959, (HEME ATTACHMENT REGION ONLY - BOVINE)

YASUNOBU, K. T., NAKASHIMA, T., HIGA, H., MATSUBARA, H., AND BENSON, A., BIOCHIM. BIOPHYS. ACTA VOL. 78, PN1324, PP. 791-794, 1963 (BOVINE)
CYTOCHROME C - PSEUDOMONAS

HEME BONDED TO CYSTEINES AT POSITIONS 12 AND 15.
THE AMINO END IS NOT ACETYLATED.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
61 M P P N A V S D D E A Q T L A K W V L S Q K *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLU ASP PRO GLU VAL LEU PHE LYS ASN LYS GLY CYS VAL ALA CYS
HIS ALA ILU ASP THR LYS MET VAL GLY PRO ALA TYR LYS ASP VAL
31 ALA ALA LYS PHE ALA GLY GLN ALA GLY ALA GLU ALA GLU LEU ALA
GLN ARG ILU LYS ASN GLY SER GLN GLY VAL TRP GLY PRO ALU PRO
61 MET PRO PRO ASN ALA VAL SER ASP ASP GLU ALA GLN THR LEU ALA
LYS TRP VAL LEU SER GLN LYS ***

COMPOSITION

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>ALA A</td>
<td>5 GLN Q</td>
<td>4 LEU L</td>
<td>3 SER S</td>
</tr>
<tr>
<td>1</td>
<td>ARG R</td>
<td>5 GLU E</td>
<td>8 LYS K</td>
<td>2 THR T</td>
</tr>
<tr>
<td>3</td>
<td>ASN N</td>
<td>7 GLY G</td>
<td>2 MET M</td>
<td>2 TRP W</td>
</tr>
<tr>
<td>5</td>
<td>ASP D</td>
<td>1 HIS H</td>
<td>2 PHE F</td>
<td>1 TYR O</td>
</tr>
<tr>
<td>2</td>
<td>CYS C</td>
<td>3 ILU I</td>
<td>6 PRO P</td>
<td>7 VAL V</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 82

AMBLER, R. P., BIOCHEM. J., VOL.89, P.349-378. 1963
CYTOCHROME C - TUNA FISH

ACETYL AT AMINO END.
HEME BONDED TO CYSTEINES AT POSITIONS 14 AND 17.

1234567890123456789012345678901234567890
1GDVAKGKKTFTVQKCKAQ(C.H)TVENGKHK(V.G.P.
31NLWGFLGRK(T.G)AEOSDT(DA.N)KSKGIVVIN,
61N.DTLMEOLENPKDKO(I.P.G)TK(M.I)FGIKKKGE
91RQDL(V.A)O.LKSTASS

1234567890123456789012345678901234567890
1GLYASPVALALALYSGLYLYSLYSTHRPHEVALGLNLYS.CYSLA
GLN(CYS.HIS)THRVALGLUASNGLYLGLYLHSLYS.LYS(VAL.GLY.PRO.
31ASN)LEU TRP.GLLEU PHE.GLY ARG.LYS THRL(GLY.GLN)ALA GLU GLY
TYR.SER TYR THR(ASP.ALA.ASN)LYS.SER LYS.GLLY ILU VAL TRP(ASN,
61ASN,ASP)THR LEU MET GLU TYR.LEU GLU ASN PRO LYS LYS.TYR(ILI.
PRO.GLY)THR LYS.MET.ILU)PHE.ALA GLY ILU LYS LYS GLY GLU
91ARG.GLN ASP LEU(VAL.ALA)TYR.LEU LYS SER THR ALA SER ***

COMPOSITION

<table>
<thead>
<tr>
<th></th>
<th>7 ALA</th>
<th>4 GLN Q</th>
<th>6 LEU L</th>
<th>4 SER S</th>
<th>2 ARG R</th>
<th>5 GLU E</th>
<th>16 LYS K</th>
<th>7 THR T</th>
<th>6 ASN N</th>
<th>13 GLY G</th>
<th>2 MET M</th>
<th>2 TRP W</th>
<th>4 ASP D</th>
<th>2 HIS H</th>
<th>3 PHE F</th>
<th>5 TYR O</th>
<th>2 CYS C</th>
<th>4 ILU I</th>
<th>3 PRO P</th>
<th>6 VAL V</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL NO. OF ACIDS</td>
<td>= 103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*KREIL, G., Z. PHYSIOL. CHEM., BD. 334, PP. 154-166, 1963*
CYTOCHROME C = BOMBYX MORI (SILKWORM)

HEME BONDED TO CYSTEINES AT POSITIONS 4 AND 7 OF FRAGMENT.

```
1 2 3 4 5 6 7 8 9 10 11
/V Q R C A Q C H T(V,E)/
```

```
1 2 3 4 5 6 7 8 9 10 11
/// VAL GLN ARG CYS ALA GLN CYS HIS THR(VAL,GLU)///
```

**COMPOSITION OF FRAGMENT**

<table>
<thead>
<tr>
<th></th>
<th>ALA A</th>
<th>GLN Q</th>
<th>LEU L</th>
<th>SER S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARG R</td>
<td>GLU E</td>
<td>LYS K</td>
<td>THR T</td>
</tr>
<tr>
<td></td>
<td>ASN N</td>
<td>GLY G</td>
<td>MET M</td>
<td>TRP W</td>
</tr>
<tr>
<td></td>
<td>ASP D</td>
<td>HIS H</td>
<td>PHE F</td>
<td>TYR O</td>
</tr>
<tr>
<td></td>
<td>CYS C</td>
<td>ILU I</td>
<td>PRO P</td>
<td>VAL V</td>
</tr>
</tbody>
</table>

**TOTAL NO. OF ACIDS IN FRAGMENT = 11**

* TUPPY H., Z. NATURFORSCH., VOL.12, PP.784-788, 1957*
CYTOCHROME C - RATTLESNAKE

ACETYL AT AMINO END.

123456789012345678901234567890
1GDEVKKIKFII.Y.KCNCSQCHTV.EKGGKHJKTG
31NLHGLFRKRTGQAVGOSOTAAANKNKGIIGW
61DDTLMEOLENPKKDIRPGTKMTGGLSKKKE
91RTNLIADOLKEKTAAG

1234567890112131415
1GLYASPAVGLULYSGLYLYSILUPHE(ILU.THR.LYS.CYS.SER.
GLN.CYS.HIS.THR.VAL.GLU.LYS.GLY.GLY.LYS.HIS)LYS
THR GLY PRO
31ASNLEUHISH.GLYLEUPHEGLYARGLYSTHRGLYGLNLALAVALGLY
TYR.SERTYR.THRALAALAASNLYSASNLYSLYULILUTRPGLY
61ASPASPTRLEUMETGLUTYRLEUGLUASNPROLYSTYR.ILU
PROGLYTHRLYSMET.VAL.PHE.THRGLYLAYSERLYSLYSLYGLU
91ARGTHRASNLEU.ILUALATYR.LEULYSLYLAYSLYSTHRALAALAA

COMPOSITION

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>GLN</td>
<td>LEU</td>
<td>SER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>GLU</td>
<td>LYS</td>
<td>THR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>GLY</td>
<td>MET</td>
<td>TRP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>HIS</td>
<td>PHE</td>
<td>TYR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYS</td>
<td>ILU</td>
<td>PRO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 104

*BAHL, O. P. AND SMITH, E. L., J. BIOL. CHEM., VOL. 240, NO. 9, PP. 3585-3593, SEPT., 1965*
CYTOCHROME - C  RHODOSPIRILLUM RUBRUM

HEME BONDED TO CYSTEINES AT POSITIONS 1 AND 4 OF FRAGMENT.

1 2 3 4 5 6 7 8 9 0 1 2 3
/ C L A C H T F B Z G A N K /

1 2 3 4 5 6 7 8 9 10 11 12 13
/// GYS LEU ALA GYS HIS THR PHE ASX GLX GLY ALA ASN LYS ///

<table>
<thead>
<tr>
<th>COMPOSITION OF FRAGMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ALA  A</td>
</tr>
<tr>
<td>0 ARG  R</td>
</tr>
<tr>
<td>1 ASN  N</td>
</tr>
<tr>
<td>0 ASP  D</td>
</tr>
<tr>
<td>2 CYS  C</td>
</tr>
<tr>
<td>1 ASX  B</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS IN FRAGMENT = 13

- TUPPY H., AND PALEUS, S., ACTA CHEM. SCAND., VOL. 13, NO.4, PP. 641-646, 1959
CYTOCHROME C - SALMON

HEME BONDED TO CYSTEINES AT POSITIONS 4 AND 7 OF FRAGMENT.

1 2 3 4 5 6 7 8 9 0 1
/ V Q K C A Q H C T(V,E)/

1 2 3 4 5 6 7 8 9 10 11
/// VAL GLN LYS CYS ALA GLN CYS HIS THR(VAL,GLU)///

COMPOSITION OF FRAGMENT

1 ALA A 2 GLN Q 0 LEU L 0 SER S
0 ARG R 1 GLU E 1 LYS K 1 THR T
0 ASN N 0 GLY G 0 MET M 0 TRP W
0 ASP D 1 HIS H 0 PHE F 0 TYR D
2 CYS C 0 ILE I 0 PRO P 2 VAL V

TOTAL NO. OF ACIDS IN FRAGMENT = 11

* TUPPY H., AND PALEUS, S., ACTA CHEM. SCAND., VOL. 9, P.353-364, 1955
HEMOGLOBIN ALPHA - HUMAN

123456789012345678901234567890
1VLSPADKTNVAWGVAGAHAGEOGAEALE
31RMFLSFPTTKTOFPFDLHSHSQAQVKGHGK
61KVAADALTNAHAVDDMPNALSALSDLHAHK
91LRVDPVNFKLLSCHLLVTLaAHLPÆFTPA
121VHASLDKFLASVSVTLTSKOR*

12345678901234567890123456789012345678901234567890
1VALLEUSERPRAALASPGLYSTHRASNVALGLYALAALAALATRPGLY
LYSVALGLYALAHISALAGLYGLTYRGLYALAGLUALAELUGLU
31ARGMETPHELLEUSERPHEPROTHRTHRLYSTHRTYRPHEPROHIS
PHEASPLEUSSHISGLYSERALAGLNAVGLYGLYHISGLYLYS
61LYSVALALASPALALEUTHRASNALAVALAHISVALASPASPMETPROASNALALEU
SERALALEUSERASPLEUHISALAHISHLYS
91LEUARGVALASPAGVVALASNPHELYSLEULEUSERHISCYSLEU
LEUVALELUAHAlAHISHLEUPROALAGLUPHETPROPROALA
121VALHISHALASERLEUASPLYPHELEUALASERVALSERTHRVAL
LEUTHRSERLYSTYRARG***

COMPOSITION

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>ALA</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ARG</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ASN</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ASP</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>CYS</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>LEU</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>SER</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>LYS</td>
<td>K</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>THR</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>MET</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>TRP</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PHE</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>TYR</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PRO</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>VAL</td>
<td>V</td>
<td></td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 141
THE SAME SEQUENCE, WITHOUT DISTINGUISHING AMINES, ALSO REPORTED IN THE ARTICLE.


FETAL ALPHA CHAIN IS VERY PROBABLY IDENTICAL WITH ADULT ALPHA CHAIN. TRYPtical AND CHYMOTRYPtical PEPTIDES, MOST OF WHICH WERE COMPLETELY SEQUENCED, WERE SHOWN TO FIT EXACTLY INTO THE ADULT ALPHA CHAIN SEQUENCE.
HEMOGLOBIN BETA - HUMAN

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 V H L T P E E K S A V T A L W G K V D V D E V G G E A L G R
31 L L V V O P W T E R F E S F G D L S T P D A V M G D P K V
91 L H C D K L H V D P E D F R L L G D V L V C V L A H H F G K
121 E F T P P V E A A O E K V V A G V A D A L A H K O H*

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 VAL HIS LEU THR PRO GLU GLU LYS SER ALA VAL THR ALA LEU TRP
GLY LYS VAL ASP VAL ASP GLU VAL GLY GLY GLU ALA LEU GLY ARG
31 LEU LEU VAL VAL TYR PRO TRP THR GLU ARG PHE PHE GLU SER PHE
GLY ASP LEU SER THR PRO ASP ALA VAL MET GLY ASP PRO LYS VAL
61 LYS ALA HIS GLY LYS LYS VAL LEU GLY ALA PHE SER ASP GLY LEU
ALA HIS LEU ASP ASP LEU LYS GLY THR PHE ALA THR LEU SER GLU
91 LEU HIS CYS ASP LYS LEU HIS VAL ASP PRO GLU ASP PHE ARG LEU
LEU GLY ASP VAL LEU VAL CYS VAL LEU ALA HIS HIS PHE GLY LYS
121 GLU PHE THR PRO PRO VAL GLU ALA ALA TYR GLU LYS VAL VAL ALA
GLY VAL ALA ASP ALA LEU ALA HIS LYS TYR HIS ***

COMPOSITION

15 ALA A 0 GLN Q 18 LEU L 5 SER S
3 ARG R 11 GLU E 11 LYS K 7 THR T
0 ASN N 13 GLY G 1 MET M 2 TRP W
13 ASP D 9 HIS H 8 PHE F 3 TYR O
2 CYS C 0 ILU I 7 PRO P 18 VAL V

TOTAL NO. OF ACIDS = 146

HEMOGLOBIN GAMMA - HUMAN

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
31 L L V V O P W T Q R F F D S F G N L S S A S A I M G N P K V
61 K A H G K O V L T S L G D A I K H L D D L K G T F A Q L S E
91 L H C D K L H V D P E N F K L L G N V L V T V L A I H F G K
121 E F T P E V Q A S W Q K M V T G V A S A L S S R D H *

GLY HIS PHE THR GLU GLU ASP LYS ALA THR ILU THR SER LEU TRP
GLY LYS VAL ASN VAL GLU ASP ALA GLY GLY GLU THR LEU GLY ARG
31 LEU LEU VAL VAL TYR PRO TRP THR GLN ARG PHE PHE ASP SER PHE
GLY ASN LEU SER SER ALA SER ALA ILU MET GLY ASN PRO LYS VAL
61 LYS ALA HIS GLY LYS LYS VAL LEU THR SER LEU GLY ASP ALA ILU
LYS HIS LEU ASP ASP LEU LYS GLY THR PHE ALA GLN LEU SER GLU
91 LEU HIS CYS ASP LYS LEU HIS VAL ASP PRO GLU ASN PHE LYS LEU
LEU GLY ASN VAL LEU VAL THR VAL LEU ALA ILU HIS PHE GLY LYS
121 GLU PHE THR PRO GLU VAL GLN ALA SER TRP GLN LYS MET VAL THR
GLY VAL ALA SER ALA LEU SER SER ARG TYR HIS ***

COMPOSITION

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>11</td>
</tr>
<tr>
<td>ARG</td>
<td>3</td>
</tr>
<tr>
<td>ASN</td>
<td>5</td>
</tr>
<tr>
<td>ASP</td>
<td>8</td>
</tr>
<tr>
<td>CYS</td>
<td>1</td>
</tr>
<tr>
<td>GLN</td>
<td>4</td>
</tr>
<tr>
<td>GLU</td>
<td>8</td>
</tr>
<tr>
<td>HIS</td>
<td>7</td>
</tr>
<tr>
<td>ILU</td>
<td>4</td>
</tr>
<tr>
<td>LYS</td>
<td>12</td>
</tr>
<tr>
<td>MET</td>
<td>2</td>
</tr>
<tr>
<td>PHE</td>
<td>8</td>
</tr>
<tr>
<td>PRO</td>
<td>4</td>
</tr>
<tr>
<td>SER</td>
<td>11</td>
</tr>
<tr>
<td>THR</td>
<td>10</td>
</tr>
<tr>
<td>TRP</td>
<td>3</td>
</tr>
<tr>
<td>TYR</td>
<td>2</td>
</tr>
<tr>
<td>VAL</td>
<td>13</td>
</tr>
<tr>
<td>V</td>
<td>4</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 146

HEMOGLOBIN BETA - GORILLA

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 VAL HIS LEU THR PRO GLU GLU LYS SER ALA VAL THR ALA LEU TRP
GLY LYS VAL ASP VAL ASP GLU VAL GLY LYS GLU ALA LEU GLY ARG
31 LEU LEU VAL VAL TYR PRO TRP THR GLU ARG PHE PHE GLU SER PHE
GLY ASP LEU SER THR PRO ASP ALA VAL MET GLY ASP PRO LYS VAL
61 LYS ALA HIS GLY LYS LYS VAL LEU GLY ALA PHE SER ASP GLY LEU
ALA HIS LEU ASP ASP LEU LYS GLY THR PHE ALA THR LEU SER GLU
91 LEU HIS CYS ASP LYS LEU HIS ASP ASP GLU ASP PHE LEU LEU
LEU GLY ASP VAL LEU VAL CYS VAL LEU ALA HIS HIS PHE GLY LYS
121 GLU PHE THR PRO PRO VAL GLU ALA ALA TYR GLU LYS VAL VAL ALA
GLY VAL ALA ASP ALA LEU ALA HIS LYS TYR HIS ***

COMPOSITION

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0</td>
<td>19</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>GLN</td>
<td>LEU</td>
<td>SER</td>
<td>S</td>
</tr>
<tr>
<td>2</td>
<td>ARG</td>
<td>GLU</td>
<td>LYS</td>
<td>THR</td>
</tr>
<tr>
<td>ASN</td>
<td>GLY</td>
<td>MET</td>
<td>TRP</td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>HIS</td>
<td>PHE</td>
<td>TYR</td>
<td></td>
</tr>
<tr>
<td>CYS</td>
<td>ILU</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 146

= ZUCKERKANDL, E., SCIENTIFIC AMERICAN, VOL. 212, NO. 5, PP. 110-118, MAY 1965
HEMOGLOBIN BETA - HORSE

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
121 D F T P E L E A S O E K V V A G V A D A L A H K O H

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 VAL GLU LEU SER GLY GLU GLU LYS ALA ALA LEU (V A L A LEU TRP ASP) LYS VAL ASP GLU GLU GLU VAL GLY (GLY GLU ALA) LEU GLY ARG
31 LEU LEU VAL VAL TYR PRO TRP THR GLU ARG PHE (PHE GLU SER PHE GLY ASP LEU SER GLY PRO ASP ALA VAL) MET (GLY ASP PRO) LYS VAL
61 LYS ALA HIS GLY LYS LYS VAL LEU HIS SER PHE GLY GLU GLY VAL
HIS HIS (LEU ASP ASP LEU) LYS GLY THR PHE ALA (ALA LEU SER GLU
91 LEU HIS CYS ASP LYS LEU HIS VAL ASP PRO GLU ASP PHE) ARG LEU
LEU GLY ASP VAL LEU ALA LEU VAL VAL ALA ARG HIS PHE GLY LYS
121 ASP PHE THR PRO GLU LEU GLU ALA SER TYR GLU LYS VAL VAL
GLY VAL ALA ASP ALA LEU ALA HIS LYS TYR HIS ***

COMPOSITION

15 ALA A 0 GLN Q 19 LEU L 6 SER S
4 ARG R 15 GLU E 11 LYS K 3 THR T
0 ASN N 14 GLY G 1 MET M 2 TRP W
13 ASP D 9 HIS H 8 PHE F 3 TYR O
1 CYS C 0 ILU I 5 PRO P 17 VAL V

TOTAL NO. OF ACIDS = 146

* SMITH, D. B., CAN. J. BIOCHEM., VOL. 42, NO. 5, PP. 755-762, 1964
HEMOGLOBIN ALPHA - HORSE

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 V L S A A D K T N V K A A W S K V G G H A G E O G A E A L E
31 R M F L G F P T T K T O F P H F D L S H G S A Q V K A H G K
61 K V A D G L T L A V G H L D D L P G A L S N L S D L H A H K
91 L R V D P V N F K L L S H C L L S T L A V H L P N D F T P A
121 V H A S L D K F L S S V S T V L T S K O R *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 VAL LEU SER ALA ALA ASP LYS THR ASN VAL LYS ALA ALA TRP SER
LYS VAL GLY GLY HIS ALA GLY GLU TYR GLY ALA GLU ALA LEU GLU
31 ARG MET PHE LEU GLY PHE PRO THR THR LYS THR TYR PHE PRO HIS
PHE ASP LEU SER HIS GLY SER ALA GLN VAL LYS ALA HIS GLY LYS
61 LYS VAL ALA ASP GLY LEU THR LEU ALA VAL GLY HIS LEU ASP ASP
LEU PRO GLY ALA LEU SER ASN LEU SER ASP LEU HIS ALA HIS LYS
91 LEU ARG VAL ASP PRO VAL ASN PHE LYS LEU LEU SER HIS CYS LEU
LEU SER THR LEU ALA VAL HIS LEU PRO ASN ASP PHE THR PRO ALA
121 VAL HIS ALA SER LEU ASP LYS PHE LEU SER SER VAL SER THR VAL

LEU THR SER LYS TYR ARG ***

COMPOSITION

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>GLN</td>
<td>Q</td>
<td>LEU</td>
<td>L</td>
<td>SER</td>
</tr>
<tr>
<td>ARG</td>
<td>GLU</td>
<td>E</td>
<td>LYS</td>
<td>K</td>
<td>THR</td>
</tr>
<tr>
<td>ASN</td>
<td>GLY</td>
<td>G</td>
<td>MET</td>
<td>M</td>
<td>TRP</td>
</tr>
<tr>
<td>ASP</td>
<td>HIS</td>
<td>H</td>
<td>PHE</td>
<td>F</td>
<td>TYR</td>
</tr>
<tr>
<td>CYS</td>
<td>ILU</td>
<td>I</td>
<td>PRO</td>
<td>P</td>
<td>VAL</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 141


THIS SEQUENCE WAS DETERMINED PARTIALLY BY HOMOLOGY WITH HUMAN ALPHA.
HEMOGLOBIN BETA - LEMUR FULVUS

123456789012345678901234567890
1TLLESEDEDAHTSLWGGKVNVKEKVGGEALGR
121XXGXVAGV(A,A,D,A,L,A,H,K,O,H)*

1THREULLETUERSALAGLUGLUSPALARATHISVALTHERSERLEUTRYP
GLYLYSVALASNVGLULYSVALGLYLGLULGLALAULEUGLYARG
31LEULEUVALEVITRY,PR0,TRP,THR,GLU,ARG,PHE,PHE,GLU,SER,PHE,
GLY,ASP=LEU,SER,SER,PRO,SER,ALA,VAL,MET,GLY,ASP,PRO,LYS,VAL,
61LYS,ALA,HIS,GLY,LYS,LYS,VAL,LEU,SER,ALA,PHE,SER,GLU,GLY=LEU,
HIS,HIS,LEU,ASP,ASP,LEU,LYS,GLY,THR,PHE,ALA,ALA,LEU,SER,GLU,
91LEU,HIS,CYS,VAL,ALA,LEU,HIS,VAL,ASP,PRO,GLU,ASP,PHE,LYS,LEU,
LEUGLY,ASP,SER,LEU,SER,ASP,VAL,LEU,ALA,ASP,HIS,PHE,GLY,LYS)
121XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX VAL VAL ALA
GLY VAL (ALA,ASP,ALA,LEU,ALA,HIS,LYS,TYR,HIS)***

COMPOSITION

14ALA A 0GLN Q 19LEU L 11SER S
2ARG R 9GLU E 10LYS K 4THR T
1ASN N 12GLY G 1MET M 2TRP W
11ASP D 9HIS H 7PHE F 2TYR D
1CYS C 0ILU I 4PRO P 15VAL V
12XXX X

TOTAL NO. OF ACIDS = 146

BUETTNER-JANUSCH, J. AND HILL, R. L., SCIENCE, VOL. 147, PP. 836-842, FEB. 19, 1965
ABNORMAL HUMAN HEMOGLOBIN

Normal adult human hemoglobin (hemoglobin A) contains two pairs of polypeptide chains, termed alpha and beta. Each pair is identical. Some modified beta chains have been given other Greek letters, for example, normal fetal hemoglobin is composed of two alpha chains and two "gamma" chains. Usually, however, altered hemoglobins are different in only a single amino acid. A number of hemoglobins bearing these altered amino acid sequences in their polypeptide chains have been described. For example, one of the early reports by Ingram (1957) shows the chemical difference between normal human hemoglobin and sickle cell hemoglobin. By comparison of amino acid sequences of tryptic peptide digests of the two hemoglobins, it was established that hemoglobin A (normal) contains a GLU residue in the locus where hemoglobin S (sickle cell) contains VAL. This replacement of two charged GLU residues for two uncharged VAL residues in the hemoglobin tetramer is sufficient to account for the "sickling" phenomenon in the abnormal hemoglobin. Listed below are a number of known amino acid replacements in abnormal human hemoglobins.

<table>
<thead>
<tr>
<th>HEMOGLOBIN NAME</th>
<th>CHANGES</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A NORMAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F NORMAL FETAL</td>
<td>BETA (CALLED GAMMA)</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>ALPHA 16 LYS-ASP</td>
<td>2</td>
</tr>
<tr>
<td>NORFOLK</td>
<td>ALPHA 57 GLY-ASP</td>
<td>3</td>
</tr>
<tr>
<td>M BOSTON</td>
<td>ALPHA 58 HIS-TYR</td>
<td>4</td>
</tr>
<tr>
<td>M SASKATOON</td>
<td>BETA 63 HIS-TYR</td>
<td>4</td>
</tr>
<tr>
<td>M MILWAUKEE</td>
<td>BETA 67 VAL-GLU</td>
<td>4</td>
</tr>
<tr>
<td>D PUNJAB</td>
<td>BETA 121 GLU-GLN</td>
<td>5</td>
</tr>
<tr>
<td>G SAN JOSE</td>
<td>BETA 7 GLU-GLY</td>
<td>6</td>
</tr>
<tr>
<td>ZURICH</td>
<td>BETA 63 HIS-ARG</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>BETA 6 GLU-LYS</td>
<td>8</td>
</tr>
<tr>
<td>D ARABIA</td>
<td>BETA 121 GLU-LYS</td>
<td>9</td>
</tr>
<tr>
<td>'O INDONESIA</td>
<td>ALPHA 116 GLU-LYS</td>
<td>9</td>
</tr>
<tr>
<td>X</td>
<td>ALPHA 68 ASN-LYS</td>
<td></td>
</tr>
<tr>
<td>and BETA 6 GLU-LYS</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>BETA 6 GLU-VAL</td>
<td>11</td>
</tr>
<tr>
<td>D IBADAN</td>
<td>BETA 87 THR-LYS</td>
<td>12</td>
</tr>
<tr>
<td>F TEXAS</td>
<td>GAMMA 5 or 6 GLU-LYS</td>
<td>13</td>
</tr>
<tr>
<td>KENWOOD</td>
<td>BETA 143 HIS-ASP</td>
<td>14</td>
</tr>
<tr>
<td>G</td>
<td>BETA 7 GLU-GLY</td>
<td>15</td>
</tr>
</tbody>
</table>
GL WHMY 2.101

MYOGLOBIN - WHALE

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0

1 V L S E 2 G E W Q L V L H V W A K V E A D V A G H G Q D I L I
31 R L F K S H P E T L E K F D R F K H L K T E A E M K A S E D
61 L K K H G V T V L T A L G A I L K K K G H H E A E L K P L A
91 Q S H A T K H K I P I K O L E F I S E A I I H V L H S R H P
121 G N F G A D A Q G A M N K A L E L F R K D I A A K O K E L G
151 O Q G *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 VAL LEU SER GLU GLY GLU TRP GLN LEU VAL LEU HIS VAL TRP ALA
LYS VAL GLU ALA ASP VAL ALA GLY HIS GLY GLN ASP ILU LEU ILU
31 ARG LEU PHE LYS SER HIS PRO GLU THR LEU GLU LYS PHE ASP ARG
PHE LYS HIS LEU LYS THR GLU ALA GLU MET LYS ALA SER GLU ASP
61 LEU LYS LYS HIS GLY VAL THR VAL LEU THR ALA LEU GLY ALA ILU
LEU LYS LYS GLY HIS HIS GLU ALA GLU LEU LYS PRO LEU ALA
91 GLN SER HIS ALA THR LYS HIS LYS ILU PRO ILU LYS TYR LEU GLU
PHE ILU SER GLU ALA ILU ILU HIS VAL LEU HIS SER ARG HIS PRO
121 GLY ASN PHE GLY ALA ASP ALA GLN GLY ALA MET ASN LYS ALA LEU
GLU LEU PHE ARG LYS ASP ILU ALA ALA LYS TYR LYS GLU LEU GLY
151 TYR GLN GLY ***

COMPOSITION

17 ALA A 5 GLN Q 18 LEU L 6 SER S
4 ARG R 14 GLU E 19 LYS K 5 THR T
2 ASN N 11 GLY G 2 MET M 2 TRP W
6 ASP D 12 HIS H 6 PHE F 3 TYR D
0 CYS C 9 ILU I 4 PRO P 8 VAL V

TOTAL NO. OF ACIDS = 153

* EDMUNDS0N, A. B., NATURE, VOL.205, NO.4974, PP.883-887,
FEBRUARY 27, 1965
DIHEME PEPTIDE - CHROMATIUM

THE PEPTIDE CONTAINS TWO HEME GROUPS. THE FIRST IS COVALENTLY BONDED TO CYSTEINES 5 AND 8. THERE IS ONLY ONE OTHER CYSTEINE AVAILABLE FOR THE OBSERVED COVALENT BONDING OF THE SECOND HEME.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7
/ F A G K C S Q C H T L V A D E G S A K C H T F D E G S /

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
/// PHE ALA GLY LYS CYS SER GLN CYS HIS THR LEU VAL ALA ASP GLU GLY SER ALA LYS CYS HIS THR PHE ASP GLU GLY SER ///

COMPOSITION

<table>
<thead>
<tr>
<th>Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>3</td>
</tr>
<tr>
<td>ARG</td>
<td>0</td>
</tr>
<tr>
<td>ASN</td>
<td>0</td>
</tr>
<tr>
<td>ASP</td>
<td>2</td>
</tr>
<tr>
<td>CYS</td>
<td>3</td>
</tr>
<tr>
<td>GLN</td>
<td>1</td>
</tr>
<tr>
<td>GLU</td>
<td>2</td>
</tr>
<tr>
<td>GLY</td>
<td>3</td>
</tr>
<tr>
<td>HIS</td>
<td>2</td>
</tr>
<tr>
<td>ILU</td>
<td>0</td>
</tr>
<tr>
<td>LEU</td>
<td>1</td>
</tr>
<tr>
<td>LYS</td>
<td>2</td>
</tr>
<tr>
<td>MET</td>
<td>0</td>
</tr>
<tr>
<td>PRO</td>
<td>0</td>
</tr>
<tr>
<td>PHE</td>
<td>2</td>
</tr>
<tr>
<td>SER</td>
<td>3</td>
</tr>
<tr>
<td>THR</td>
<td>2</td>
</tr>
<tr>
<td>TYR</td>
<td>0</td>
</tr>
<tr>
<td>VAL</td>
<td>1</td>
</tr>
<tr>
<td>TRP</td>
<td>0</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS IN FRAGMENT = 27

* DUS, K., BARTSCH, R.G., AND KAMEN, M.D., J. BIOL. CHEM., VOL. 237, NO. 10, PP. 3083-3093, OCT., 1962*
FERREDOXIN - CLOSTRIDIUM PASTEURIANUM

THE PROTEIN CONTAINS 7 SULPHIDE AND 7 IRON ATOMS PER MOLECULE. IT DOES NOT CONTAIN HEME.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 A O K I A D S C V S C/G A C/A S E C P V N A I S Q G D S I F/
31 V I D A D T C I D C G N C A N V C P V G A P V Q E *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ALA TYR LYS ILU ALA ASP SER CYS VAL SER CYS/GLY ALA CYS/ALA
SER GLU CYS PRO VAL ASN ALA ILU SER GLN GLY ASP SER ILU PHE/
31 VAL ILU ASP ALA ASP THR CYS ILU ASP CYS GLY ASN CYS ALA ASN
VAL CYS PRO VAL GLY ALA PRO VAL GLN GLU ***

COMPOSITION

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>ALA</td>
<td>2</td>
<td>GLN</td>
<td>Q</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>ARG</td>
<td>2</td>
<td>GLU</td>
<td>E</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>ASN</td>
<td>4</td>
<td>GLY</td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>ASP</td>
<td>0</td>
<td>HIS</td>
<td>H</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>CYS</td>
<td>5</td>
<td>ILU</td>
<td>I</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>SER</td>
<td>5</td>
<td>LEU</td>
<td>L</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>THR</td>
<td>1</td>
<td>LYS</td>
<td>K</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>TRP</td>
<td>0</td>
<td>MET</td>
<td>M</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>TYR</td>
<td>1</td>
<td>PHE</td>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>VAL</td>
<td>6</td>
<td>VAL</td>
<td>V</td>
<td>6</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 55

AZURIN - PSEUDOMONAS FLUORESCENS

THE BLUE PROTEIN CONTAINS ONE COPPER ATOM PER MOLECULE.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 A E C S V D I Q G N D Q M Q F N T N A I T V D K S C K Q F T
3 1 V N L S H P G N L P K N V M G H N W V L S T A A D M Q G V V
6 1 T D G M A S G L D K D O L K P D D S R V I A H T K L I G S G
9 1 E K D S V T F D V S K L K E G E Q D M F F C T F P G H S A L
121 M K G T L T L K *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ALA GLU CYS SER VAL ASP ILU GLN GLY ASN ASP GLN MET GLN PHE
ASN THR ASN ALA ILU THR VAL ASP LYS SER CYS LYS GLN PHE THR
31 VAL ASN LEU SER HIS PRO GLY ASN LEU PRO LYS ASN VAL MET GLY
HIS ASN TRP VAL LEU SER THR ALA ALA ASP MET GLN GLY VAL VAL
61 THR ASP GLY MET ALA SER GLY LEU ASP LYS ASP TYR LEU LYS PRO
ASP ASP SER ARG VAL ILU ALA HIS THR LYS LEU ILU GLY SER GLY
91 GLU LYS ASP SER VAL THR PHE ASP VAL SER LYS LEU LYS GLU GLY
GLU GLN TYR MET PHE PHE CYS THR PHE PRO GLY HIS SER ALA LEU
121 MET LYS GLY THR LEU THR LEU LYS ***

COMPOSITION

<table>
<thead>
<tr>
<th></th>
<th>7 ALA</th>
<th>6 GLN</th>
<th>10 LEU</th>
<th>10 SER</th>
<th>1 ARG</th>
<th>4 GLU</th>
<th>11 LYS</th>
<th>10 THR</th>
<th>7 ASN</th>
<th>11 GLY</th>
<th>6 MET</th>
<th>1 TRP</th>
<th>11 ASP</th>
<th>4 HIS</th>
<th>6 PHE</th>
<th>2 TYR</th>
<th>3 CYS</th>
<th>4 ILU</th>
<th>4 PRO</th>
<th>10 VAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>G</td>
<td>L</td>
<td>S</td>
<td>R</td>
<td>E</td>
<td>K</td>
<td>T</td>
<td>N</td>
<td>G</td>
<td>M</td>
<td>W</td>
<td>D</td>
<td>H</td>
<td>F</td>
<td>O</td>
<td>C</td>
<td>I</td>
<td>P</td>
<td>V</td>
</tr>
</tbody>
</table>
| TOTAL NO. OF ACIDS = 128

RIBONUCLEASE - BOVINE

Disulfide bonds are formed between cysteines at positions 26 and 84, 40 and 95, 58 and 110, and 65 and 72.

1 KETAAAKFERQHMDSSSTSSAASSSNDOCNQMM
2 KSRNLTKDRCVKPVTFTPHELQADVVQAVCSQ
3 KNVACKNGQTNOQSOSTMSITDCRETTGSS
4 KOPNCADKTQANKHIIVACEGNPOVPVF

121 DASV*

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 LYS GLU THR ALA ALA ALA LYS PHE GLU ARG GLN HIS MET ASP SER SER THR SER ALA ALA SER SER SER ASN TYR CYS ASN GLN MET MET
31 LYS SER ARG ASN LEU THR LYS ASP ARG CYS LYS PRO VAL ASN THR PHE VAL HIS GLU SER LEU ALA ASP VAL GLN ALA VAL CYS SER GLN
61 LYS ASN VAL ALA CYS LYS ASN GLY GLN THR ASN CYS TYR GLN SER TYR SER THR MET SER ILU THR ASP CYS ARG GLU THR GLY SER SER
91 LYS TYR PRO ASN CYS ALA TYR LYS THR THR GLN ALA ASN LYS HIS ILU ILU VAL ALA CYS GLU GLY ASN PRO TYR VAL PRO VAL HIS PHE
121 ASP ALA SER VAL ***

Composition

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>12</td>
</tr>
<tr>
<td>GLU</td>
<td>7</td>
</tr>
<tr>
<td>ARG</td>
<td>4</td>
</tr>
<tr>
<td>ASN</td>
<td>10</td>
</tr>
<tr>
<td>ASP</td>
<td>5</td>
</tr>
<tr>
<td>CYS</td>
<td>8</td>
</tr>
<tr>
<td>LEU</td>
<td>2</td>
</tr>
<tr>
<td>LYS</td>
<td>10</td>
</tr>
<tr>
<td>MET</td>
<td>4</td>
</tr>
<tr>
<td>HIS</td>
<td>3</td>
</tr>
<tr>
<td>PRO</td>
<td>4</td>
</tr>
<tr>
<td>15 SER</td>
<td></td>
</tr>
<tr>
<td>10 THR</td>
<td></td>
</tr>
<tr>
<td>0 TRP</td>
<td></td>
</tr>
<tr>
<td>6 TYR</td>
<td></td>
</tr>
<tr>
<td>9 VAL</td>
<td></td>
</tr>
</tbody>
</table>

Total no. of acids = 124

TRYPSIN INHIBITOR - BOVINE

DISULPHIDE BONDS ARE FORMED BETWEEN CYSTEINES AT POSITIONS 5-55, 14-38, AND 30-51.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
Q T F V O G G C R A K R N N F K S A E D C M R T C G G A *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ARG PRO ASP PHE CYS LEU GLU PRO PRO TYR THR GLY PRO CYS LYS
ALA ARG ILU ILU ARG TYR PHE TYR ASN ALA LYS ALA GLY LEU CYS
31 GLN THR PHE VAL TYR GLY GLY CYC ARG ALA LYS ARG ASN PHE
LYS SER ALA GLU ASP CYS MET ARG THR CYC GLY GLY ALA ***

COMPOSITION

<table>
<thead>
<tr>
<th>6 ALA</th>
<th>1 GLN</th>
<th>2 LEU</th>
<th>1 SER</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 ARG</td>
<td>2 GLU</td>
<td>4 LYS</td>
<td>3 THR</td>
</tr>
<tr>
<td>3 ASN</td>
<td>6 GLY</td>
<td>1 MET</td>
<td>0 TRP</td>
</tr>
<tr>
<td>2 ASP</td>
<td>0 HIS</td>
<td>4 PHE</td>
<td>4 TYR</td>
</tr>
<tr>
<td>6 CYC</td>
<td>2 ILU</td>
<td>4 PRO</td>
<td>1 VAL</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 58

- KASSELL, B., RADICEVIC, M., ANSFIELD, M. J., AND LASKOWSKI, M., BIOCHEM. BIOPHYS. RES. COMMUN., VOL. 18, NO. 2, PP. 255-258, 1965
- DLOUHA, V., POSPISILOVA, D., MELOUN, B. AND SORM, F., COLLECTION CZECH. CHEM. COMMUN., VOL. 30, PP. 1311-1325, 1965

THE SEQUENCE REPORTED HERE DIFFERS FROM THE ABOVE IN HAVING THE ILU (I) DELETED AT POSITION 19.
THE SEQUENCE REPORTED HERE DIFFERS FROM THE ABOVE IN THE FOLLOWING RESPECTS.
THE ARG (R) FROM POSITION 42 HAS BEEN REMOVED AND INSERTED BETWEEN POSITIONS 20 AND 21. THE GLN (Q) AT POSITION 31 HAS BEEN DELETED AND A GLU (E) ADDED BETWEEN POSITIONS 32 AND 33.

KASSELL, B., AND LASKOWSKI, M., BIOCHEM. BIOPHYS. RES. COMMUN. VOL 20, NO.4, PP.463-468, 1965
TOBACCO MOSAIC VIRUS

ACETYL - AT AMINO END

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 S O S I T T P S Q F V F L S S A W A D P I E L I N L C T N A
3 L G N Q F Q T Q Q A R T V Q V R Q F S Q V W K P S P Q V T V
9 N R I I Q V Q D Q A N P T T A Q T L D A T R V D D A T V A
12 I R S A D I N L I V E L I R G T G S Q N R S S F E S S S G L
15 V W T S G P A T •

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 SER TYR SER ILU THR THR PRO SER GLN PHE VAL PHE LEU SER SER
ALA TRP ALA ASP PRO ILU GLU LEU ILU ASN LEU CYS THR ASN ALA
31 LEU GLY ASN GLN PHE GLN THR GLN GLN ALA ARG THR VAL GLN VAL
ARG GLN PHE SER GLN VAL TRP LYS PRO SER PRO GLN VAL THR VAL
61 ARG PHE PRO ASP SER ASP PHE LYS VAL TYR ARG TYR ASN ALA VAL
LEU ASP PRO LEU VAL THR ALA LEU LEU GLY ALA PHE ASP THR ARG
91 ASN ARG ILU ILU GLN VAL GLN ASP GLN ALA ASN PRO THR THR ALA
GLN THR LEU ASP ALA THR ARG ARG VAL ASP ASP ALA THR VAL ALA
121 ILU ARG SER ALA ASP ILU ASN LEU ILU VAL GLU LEU ILU ARG GLY
THR GLY SER TYR ASN ARG SER SER PHE GLU SER SER SER SER GLY LEU
151 VAL TRP THR SER GLY PRO ALA THR •••

COMPOSITION

14 ALA 13 GLN 12 LEU 16 SER 14 VAL
11 ARG 3 GLU 2 LYS 16 THR
8 ASN 6 GLY 0 MET 3 TRP
10 ASP 0 HIS 8 PHE 4 TYR
1 CYS 9 ILU 8 PRO 14 VAL

TOTAL NO. OF ACIDS = 158

STRUCTURE REVISIONS AND CONFIRMATIONS.


FUNATSU, G., TSUGITA, A., AND FRAENKEL-CONRAT, H., ARCH. BIOCHEM. BIOPHYS., VOL. 105, NO. 1, PP. 25-41, APR. 1964
TOBACCO MOSAIC VIRUS STRAIN DAHMENSE

ACETYL-AT AMINO END

I 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 S O S I T S P S Q F V F L S S V W A D P I E L L N V C T S S
3 I L G N Q F Q T Q Q A R T T Q V Q Q F S E V W K P F P Q S T V
9 I N R I I E V E N Q Q S P T T A E T L D A T R R V D D A T V A
12 I R S A N I N L V N E L V R G T G L O N Q N T F E S M S G L
15 I V W T S A P A S *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 SER TYR SER ILU THR SER PRO SER GLN PHE VAL PHE LEU SER SER VAL TRP ALA ASP PRO ILU GLU LEU LEU ASN VAL CYS THR SER SER
31 LEU GLY ASN GLN PHE GLN THR GLN GLN ALA ARG THR THR GLN VAL GLN GLN PHE SER GLU VAL TRP LYS PRO PHE PRO GLN SER THR VAL
61 ARG PHE PRO GLY ASP VAL TYR LYS VAL TYR ARG TYR ASN ALA VAL LEU ASP PRO LEU ILU THR ALA LEU LEU GLY THR PHE ASP THR ARG
91 ASN ARG ILU ILU GLU VAL GLU ASN GLN GLN SER PRO THR THR ALA GLU THR LEU ASP ALA THR ARG ARG VAL ASP ASP ALA THR VAL ALA
121 ILU ARG SER ALA ASN ILU ASN LEU VAL ASN GLU LEU VAL ARG GLY THR GLY LEU TYR ASN GLN ASN THR PHE GLU SER MET SER GLY LEU
151 VAL TRP THR SER ALA PRO ALA ALA SER ***

COMPOSITION

11 ALA A 12 GLN Q 13 LEU L 16 SER S
9 ARG R 7 GLU E 2 LYS K 17 THR T
10 ASN N 6 GLY G 1 MET M 3 TRP W
7 ASP D 0 HIS H 8 PHE F 5 TYR O
1 CYS C 7 ILU I 8 PRO P 15 VAL V

TOTAL NO. OF ACIDS = 158
WITTMANN-LIEBOLD, B. AND WITTMA NN, H. G., Z. VERERBUNGS.,
VOL. 94, PP. 427-435, 1963
CHYMOTRYPSINOGEN-A - BOVINE

1 CGVPAIQPVLSSGLSRLTVGDDEAEAPGSWPWQ
23456789012345678901234567890
31VSLQDKTFHGCGSSLINENWVTAAHCGV
61TTSDVVVAEGFDQSSEKIQKLIKIAKVKF
91NSKONSTTINNNITLKLSTAAASFSQTVSA
121VCLPSASDDFAATCTCVTTGGLTLRONTAN
151TPDRLQQASLPLLLNTNCKKOWGTKIKDAM
181ICAGASGVSSTCMGDSSGLPLVCKKNGAWTLY
211GIVSWGSSSTCSTSTPGVDAVTAIVNWWVQ
241TLLAN

1 CYS GLY VAL PRO ALA ILU GLN PRO VAL LEU SER GLY LEU SER ARG
2 ILU VAL GLY ASP GLU GLU ALA VAL PRO GLY SER TRP PRO TRP GLN
3 VAL SER LEU GLN ASP LYS THR GLY PHE HIS PHE CYS GLY GLY SER
4 LEU ILU ASN GLU ASN TRP VAL VAL THR ALA ALA HIS CYS GLY VAL
6 THR THR SER ASP VAL VAL ALA GLY GLU PHE ASP GLN GLY SER
7 SER SER GLU LYS ILU GLN LYS LEU LYS ILU ALA GLN LYS VHE LYS
9 ASN SER LYS TYR ASN SER LEU THR ILU ASN ASN ASN ILU THR LEU
10 LEU LYS LEU SER THR ALA ALA SER PHE SER GLN THR VAL SER ALA
12 VAL CYS LEU PRO SER ALA SER ASP ASP PHE ALA ALA GLY THR THR
13 CYS VAL THR THR GLY TRP GLY LEU THR ARG TYR THR ASN ALA ASN
15 THR PRO ASP ARG LEU GLN GLN ALA SER LEU PRO LEU LEU SER ASN
17 THR ASN CYS LYS TYR TRP GLY THR LYS ILU LYS ASP ALA MET
18 ILU CYS ALA GLY ALA SER GLY VAL SER SER CYS MET GLY ASP SER
19 GLY GLY PRO LEU VAL CYS LYS LYS ASN GLY ALA TRP THR LEU VAL
21 GLY ILU VAL SER TRP GLY SER SER THR CYS SER THR SER THR PRO
22 GLY VAL TYR ALA ARG VAL THR ALA LEU VAL ASN TRP VAL GLN GLN
24 THR LEU ALA ALA ASN ***
**COMPOSITION**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>22 ALA A</td>
<td>10 GLN Q</td>
<td>19 LEU L</td>
<td>28 SER S</td>
<td></td>
</tr>
<tr>
<td>4 ARG R</td>
<td>5 GLU E</td>
<td>14 LYS K</td>
<td>23 THR T</td>
<td></td>
</tr>
<tr>
<td>14 ASN N</td>
<td>23 GLY G</td>
<td>2 MET M</td>
<td>8 TRP W</td>
<td></td>
</tr>
<tr>
<td>9 ASP D</td>
<td>2 HIS H</td>
<td>6 PHE F</td>
<td>4 TYR O</td>
<td></td>
</tr>
<tr>
<td>10 CYS C</td>
<td>10 ILE I</td>
<td>9 PRO P</td>
<td>23 VAL V</td>
<td></td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 245

---


THIS SEQUENCE HAS BEEN CORRECTED BY DELETING SER (S) WHICH WAS AT POSITION 215.


THE ACTIVE SITE SERINE IS AT POSITION 195.

KEIL, B., PRUSIK, Z., AND SORM, F., BIOCHIM. BIOPHYS. ACTA, VOL. 78, P. 559-561, 1963


HARTLEY, B.S., NATURE, VOL. 201, NO. 4962, PP. 1284-1287, MARCH 28, 1964
TRYPsinogen

1 2 3 4 5 6 7 8 9 0
1 V O D D D K I V G G O T C G A N T V P O Q V S L N S G O H F
31 C G G S L I N S Q W V V S A A H C O K S G I Q V R L G E D N
91 I M L I K L K S A A S L N S R V A S I S L P T S C A S A G T
121 Q C L I S G W G N T K S S G T S O P D V L K C L K A P I L S
151 D S S C K S A O P G Q I T S N M F C A G O L E G G K N S C Q
181 G D S G G P V V C S G K L Q G I V S W G S G C A Q K N K P G
211 V O T K V C N O V S W I K Q T I A S N •

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 VAL ASP ASP ASP ASP LYS ILU VAL GLY GLY TYR THR CYS GLY ALA
ASN THR VAL PRO TYR GLN VAL SER LEU ASN SER GLY TYR HIS PHE
31 CYS GLY GLY SER LEU ILU ASN SER GLN TRP VAL VAL SERALA ALA
HIS CYS TYR LYS SER GLY ILU GLN VAL ARG LEU GLY GLU ASP ASN
61 ILU ASN VAL VAL GLU GLY ASP GLU GLN PHE ILU SER ALA SER LYS
SER ILU VAL HIS PRO SER TYR ASN(PRO,LEU,THR,ASN)ASN ASN ASP
91 ILU MET LEU ILU LYS LEU LYS SER ALA ALA SER LEU ASN SER ARG
VAL ALA SER ILU SER LEU PRO THR SER CYS ALA SER ALA GLY THR
121 GLN CYS LEU ILU SER GLY TRP GLY ASN THR LYS SER SER GLY THR
SER TYR PRO ASP VAL LEU LYS CYS LEU LYS ALA PRO ILU LEU SER
151 ASP SER SER CYS LYS SER ALA TYR PRO GLY GLN ILU THR SER ASN
MET PHE CYS ALA GLY TYR LEU GLU GLY GLY LYS ASN SER CYS GLN
181 GLY ASP SER GLY GLY PRO VAL VAL CYS SER GLY LYS LEU GLN GLY
ILU VAL SER TRP GLY SER GLY CYS ALA GLN LYS ASN LYS PRO GLY
211 VAL TYR THR LYS VAL CYS ASN TYR VAL SER TRP ILU LYS GLN THR
ILU ALA SER ASN •••
### COMPOSITION

<table>
<thead>
<tr>
<th>Residue</th>
<th>Count</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA A</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>GLN Q</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>LEU L</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>SER S</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>ARG R</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>GLU E</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>LYS K</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>THR T</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>ASN N</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>GLY G</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>MET M</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>TRP W</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>ASP D</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>HIS H</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>PHE F</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>TYR D</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>CYS C</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>ILU I</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>PRO P</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>VAL V</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

**TOTAL NO. OF ACIDS = 229**

---

*WALSH, K., AND NEURATH, H.,* PROG. NATL. ACAD. SCI. U.S., VOL. 52, NO. 4, PP. 884-889, 1964

*KAUFFMAN, D. L., J. MOL. BIOL.,* VOL. 12, PP. 929-932, 1965

Disulphide bridges were found between links 13-143, 31-47, 115-216, 122-189, 154-168, and 179-203. The active serine is at link 183.
PA PA 7.101

PAPAIN

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1/ I P E O V D W R Q K G A V T P V K N Q G S C G S C W/A F/I I/
31 R N T P O D E G V Q R O C R S R E K G P O A A K T O G V R Q
61 V Q P O N Q G A L L O S I A N Q P S V V L Q A A G K D F Q L
121 K N S W G T G W G E N G O I R I K T G N L N Q O S E Q E L L
151 D C D R R S O G C O P G D G W/S A L/V A Q O G I H O R G T G
181 N S O G V C G L O T S S F O P V K N •

1 2 3 4 5 6 7 8 9 0 1 1 1 1 1 1 1 1 1
V A L L Y S A S N ***
### COMPOSITION

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>13</td>
</tr>
<tr>
<td>Glu (E)</td>
<td>12</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>10</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>12</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>11</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>9</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>7</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>12</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>27</td>
</tr>
<tr>
<td>Met (M)</td>
<td>0</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>5</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>7</td>
</tr>
<tr>
<td>His (H)</td>
<td>2</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>4</td>
</tr>
<tr>
<td>Tyr (O)</td>
<td>19</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>7</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>10</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>10</td>
</tr>
<tr>
<td>Val (V)</td>
<td>15</td>
</tr>
</tbody>
</table>

**TOTAL NO. OF ACIDS = 198**

---


DISULPHIDE BRIDGES ARE FORMED BETWEEN CYSTEINES AT POSITIONS 43 AND 152, 100 AND 186, AND 22 AND 159.

THE ACTIVE SULPHYDRYL GROUP IS AT POSITION 25.
LYSOZYME - CHICKEN

LYSOZYME HAS A BETA (1-4) GLUCOSAMINIDASE ACTIVITY WITH THE ABILITY TO HYDROLYSE A MUCOPOLYSACCHARIDE COMPONENT OF SOME BACTERIAL CELL WALLS RELEASING N-ACETYLATED AMINO SUGARS DERIVED FROM GLUCOSAMINE AND MURAMIC ACID.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 K V F G R C E L A A A M K R H G L D N O R G O S L G N W V C
31 A A K F E S N F N T Q A T N R N T D G S T D O G I L Q I N S
61 R W W C N D G R T P G S R N L C N I P C S A L L S S D I T A
121 Q A W I R G C R L *

I 1 LYS VAL PHE GLY ARG CYS GLU LEU ALA ALA ALA MET LYS ARG HIS
GLY LEU ASP ASN TYR ARG GLY TYR SER LEU GLY ASN TRP VAL CYS
31 ALA ALA LYS PHE GLU SER ASN PHE ASN THR GLN ALA THR ASN ARG
ASN THR ASP GLY SER THR ASP TYR GLY ILU LEU GLN ILU ASN SER
61 ARG TRP TRP CYS ASN ASP GLY ARG THR PRO GLY SER ARG ASN LEU
CYS ASN ILU PRO CYS SER ALA LEU LEU SER SER ASP ILU THR ALA
91 SER VAL ASN CYS ALA LYS LYS ILU VAL SER ASP GLY ASP GLY MET
ASN ALA TRP VAL ALA TRP ARG ASN ARG CYS LYS GLY THR ASP VAL
121 GLN ALA TRP ILU ARG GLY CYS ARG LEU ***

**COMPOSITION**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>Q</th>
<th>L</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>ALA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>ARG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>ASN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ASP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CYS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 129
ABOVE SEQUENCE CONFIRMED IN THIS WORK. DISULPHIDE BONDS ARE FOUND BETWEEN 6 AND 127, 30 AND 115, 64 AND 80, AND 76 AND 94.

JOLLES, J., JAUREGUI-ADELL, J., BERNIER, I., AND JOLLES, P., BIOCHIM. BIOPHYS. ACTA, VOL. 78, PP. 668-689, 1963


GLUCAGON - BOVINE

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
1 HIS Q G T F T S D O S K O L D S R R A Q D F V Q W L M N T

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 HIS SER GLN GLY THR PHE THR SER ASP TYR SER LYS TYR LEU ASP
SER ARG ARG ALA GLN ASP PHE VAL GLN TRP LEU MET ASN THR ***

COMPOSITION

<table>
<thead>
<tr>
<th>1</th>
<th>ALA A</th>
<th>3</th>
<th>GLN Q</th>
<th>2</th>
<th>LEU L</th>
<th>4</th>
<th>SER S</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>ARG R</td>
<td>0</td>
<td>GLU E</td>
<td>1</td>
<td>LYS K</td>
<td>3</td>
<td>THR T</td>
</tr>
<tr>
<td>1</td>
<td>ASN N</td>
<td>1</td>
<td>GLY G</td>
<td>1</td>
<td>MET M</td>
<td>1</td>
<td>TRP W</td>
</tr>
<tr>
<td>3</td>
<td>ASP D</td>
<td>1</td>
<td>HIS H</td>
<td>2</td>
<td>PHE F</td>
<td>2</td>
<td>TYR O</td>
</tr>
<tr>
<td>0</td>
<td>CYS C</td>
<td>0</td>
<td>ILU I</td>
<td>0</td>
<td>PRO P</td>
<td>1</td>
<td>VAL V</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 29

ARGinine Vasopressin - BOVINE

The C terminal glycine is present as the amide. The two cysteines are linked by a disulphide bond.

1 2 3 4 5 6 7 8 9
1 C O F Q N C P R G *

1 2 3 4 5 6 7 8 9
1 CYS TYR PHE GLN ASN CYS PRO ARG GLY ***

Composition

<table>
<thead>
<tr>
<th></th>
<th>0 Ala A</th>
<th>1 Gln Q</th>
<th>0 Leu L</th>
<th>0 Ser S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Arg R</td>
<td>0 Glu E</td>
<td>0 Lys K</td>
<td>0 Thr T</td>
<td></td>
</tr>
<tr>
<td>1 Asn N</td>
<td>1 Gly G</td>
<td>0 Met M</td>
<td>0 Trp W</td>
<td></td>
</tr>
<tr>
<td>0 Asp D</td>
<td>0 His H</td>
<td>1 Phe F</td>
<td>1 Tyr O</td>
<td></td>
</tr>
<tr>
<td>2 Cys C</td>
<td>0 Ilu I</td>
<td>1 Pro P</td>
<td>0 Val V</td>
<td></td>
</tr>
</tbody>
</table>

Total no. of acids = 9


This work confirmed the sequence above, however Glu (E) and Asp (D) were not distinguished from Gln (Q) and Asn (N).
LYSINE VASOPRESSIN - PIG

THE C TERMINAL GLYCINE IS PRESENT AS THE AMIDE.
THE TWO CYSTEINES ARE LINKED BY A DISULPHIDE BOND.

1 2 3 4 5 6 7 8 9
1 C.O.F.Q.N.C.P.K.G.*

1 2 3 4 5 6 7 8 9
1 CYS.TYR.PHE/GLN.ASN.CYS.PRO.LYS.GLY ***

COMPOSITION

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>ALA</td>
<td>GLN</td>
<td>LEU</td>
<td>SER</td>
</tr>
<tr>
<td>ARG</td>
<td>ARG</td>
<td>GLU</td>
<td>LYS</td>
<td>THR</td>
</tr>
<tr>
<td>ASN</td>
<td>ASN</td>
<td>GLY</td>
<td>MET</td>
<td>TRP</td>
</tr>
<tr>
<td>ASP</td>
<td>ASP</td>
<td>HIS</td>
<td>PHE</td>
<td>TYR</td>
</tr>
<tr>
<td>CYS</td>
<td>CYS</td>
<td>ILU</td>
<td>PRO</td>
<td>VAL</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 9

* POPENOE, E. A., LAWLER, H. C., AND DU VIGNEAUD, V.,
J. AM. CHEM. SOC., VOL. 74, P. 3713, JULY 20, 1952
OXYTOCIN - BOVINE

THE C TERMINAL GLYCINE IS PRESENT AS THE AMIDE. THE TWO CYSTEINES ARE LINKED BY A DISULPHIDE BOND. OXYTOCIN IS THE PRINCIPAL UTERINE CONTRACTING AND MILK EJECTING HORMONE OF THE POSTERIOR PITUITARY.

1 2 3 4 5 6 7 8 9
1 C O I Q N C P L G•

1 2 3 4 5 6 7 8 9
1 CYS TYR ILU GLN ASN CYS PRO LEU GLY •••

COMPOSITION

| 0 ALA A | 1 GLN Q | 1 LEU L | 0 SER S  |
| 0 ARG R | 0 GLU E | 0 LYS K | 0 THR T  |
| 1 ASN N | 1 GLY G | 0 MET M | 0 TRP W  |
| 0 ASP D | 0 HIS H | 0 PHE F | 1 TYR O  |
| 2 CYS C | 1 ILU I | 1 PRO P | 0 VAL V  |

TOTAL NO. OF ACIDS = 9

• DU VIGNEAUD, V., RESSLER, C., TRIPPETT, S., J. BIOL. CHEM., VOL.205, PP.949-957, 1953

TUPPY, H. AND MICHL, H., MONATSH. CHEM., VOL.84, PP.1011-1020, 1953
HYPERTENSIN - BOVINE

1 2 3 4 5 6 7 8 9 0
1 D R V O V H P F H L *

1 2 3 4 5 6 7 8 9 10
1 ASP ARG VAL TYR VAL HIS PRO PHE HIS LEU ***

COMPOSITION

<table>
<thead>
<tr>
<th></th>
<th>ALA</th>
<th>GLN</th>
<th>LEU</th>
<th>SER</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>ARG</td>
<td>GLU</td>
<td>LYS</td>
<td>THR</td>
</tr>
<tr>
<td>0</td>
<td>ASN</td>
<td>GLY</td>
<td>MET</td>
<td>TRP</td>
</tr>
<tr>
<td>1</td>
<td>ASP</td>
<td>HIS</td>
<td>PHE</td>
<td>TYR</td>
</tr>
<tr>
<td>0</td>
<td>CYS</td>
<td>ILE</td>
<td>PRO</td>
<td>VAL</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 10

* ELLIOT, D. F., AND PEART, W. S., BIOCHEM. J., VOL. 65, PP. 246-254, 1957
**Alpha Melanocyte-Stimulating Hormone** — Bovine, Pig, and Horse

Acetyl at amino end. C-terminal valine is aminated.

```
1 2 3 4 5 6 7 8 9 0 1 2 3
S O S M E H F R W G K P V *
```

```
1 2 3 4 5 6 7 8 9 10 11 12 13
SER TYR SER MET GLU HIS PHE ARG TRP GLY LYS PRO VAL ***
```

**Composition**

<table>
<thead>
<tr>
<th>0 Ala</th>
<th>0 Glu</th>
<th>0 Leu</th>
<th>2 Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Arg</td>
<td>1 Glu</td>
<td>1 Lys</td>
<td>0 Thr</td>
</tr>
<tr>
<td>0 Asn</td>
<td>1 Gly</td>
<td>1 Met</td>
<td>1 Trp</td>
</tr>
<tr>
<td>0 Asp</td>
<td>1 His</td>
<td>1 Phe</td>
<td>1 Tyr</td>
</tr>
<tr>
<td>0 Cys</td>
<td>0 Illu</td>
<td>1 Pro</td>
<td>1 Val</td>
</tr>
</tbody>
</table>

Total no. of acids = 13

---


*Li, C. H., Laboratory Investigation, Vol. 8, No. 2, PP. 574-587, 1959 (Bovine)*

BETA MELANOCYTE-STIMULATING HORMONE—BOVINE

123456789012345678
1D5GPDOKMEHFRGWG5PPKD*

123456789101112131415
1ASP SER GLY PRO TYR LYS MET GLU HIS PHE ARG TRP GLY SER PRO
PRO LYS ASP ***

COMPOSITION

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ALA A</td>
<td>0 GLN Q</td>
<td>0 LEU L</td>
<td>2 SER S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 ARG R</td>
<td>1 GLU E</td>
<td>2 LYS K</td>
<td>0 THR T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 ASN N</td>
<td>2 GLY G</td>
<td>1 MET M</td>
<td>1 TRP W</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 ASP D</td>
<td>1 HIS H</td>
<td>1 PHE F</td>
<td>1 TYR O</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 CYS C</td>
<td>0 ILU I</td>
<td>3 PRO P</td>
<td>0 VAL V</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 18

BETA MELANOCYTE-STIMULATING HORMONE - PIG

DEGPOKMEHFRWGSPPKD

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8

1 ASP GLU GLY PRO TYR LYS MET GLU HIS PHE ARG TRP GLY SER PRO

PRO LYS ASP ***

COMPOSITION

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Ala A</td>
<td>0 Glu Q</td>
<td>0 Leu L</td>
<td>1 Ser S</td>
</tr>
<tr>
<td>1 Arg R</td>
<td>2 Glu E</td>
<td>2 Lys K</td>
<td>0 Thr T</td>
</tr>
<tr>
<td>0 Asn N</td>
<td>2 Gly G</td>
<td>1 Met M</td>
<td>1 Trp W</td>
</tr>
<tr>
<td>2 Asp D</td>
<td>1 His H</td>
<td>1 Phe F</td>
<td>1 Tyr D</td>
</tr>
<tr>
<td>0 Cys C</td>
<td>0 Ile I</td>
<td>3 Pro P</td>
<td>0 Val V</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 18

\* HARRIS, J. I. AND ROOS, P., NATURE, VOL. 178, NO. 4524, P. 90, JULY 14, 1956

BETA MELANOCYTE-STIMULATING HORMONE - HORSE

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8
1 D E G P O K M E H F R W G S P R K D

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ASP GLU GLY PRO TYR LYS MET GLU HIS PHE ARG TRP GLY SER PRO
ARG LYS ASP ***

<table>
<thead>
<tr>
<th>COMPOSITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ALA A 0 GLN Q 0 LEU L 1 SER S</td>
</tr>
<tr>
<td>2 ARG R 2 GLU E 2 LYS K 0 THR T</td>
</tr>
<tr>
<td>0 ASN N 2 GLY G 1 MET M 1 TRP W</td>
</tr>
<tr>
<td>2 ASP D 1 HIS H 1 PHE F 1 TYR O</td>
</tr>
<tr>
<td>0 CYS C 0 ILU I 2 PRO P 0 VAL V</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 18

* DIXON, J. S. AND LI, C. H., GEN. COMP. ENDOCRINOL., VOL.1, PP.161-169, 1961
BETA MELANOCYTE-STIMULATING HORMONE - HUMAN

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2
1 A E K K D E G P O R M E H F R W G S P P K D *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ALA GLU LYS LYS ASP GLU GLY PRO TYR ARG MET GLU HIS PHE ARG TRP GLY SER PRO PRO LYS ASP ***

COMPOSITION

<table>
<thead>
<tr>
<th>1 ALA</th>
<th>A</th>
<th>0 GLN</th>
<th>Q</th>
<th>0 LEU</th>
<th>L</th>
<th>1 SER</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ARG</td>
<td>R</td>
<td>3 GLU</td>
<td>E</td>
<td>3 LYS</td>
<td>K</td>
<td>0 THR</td>
<td>T</td>
</tr>
<tr>
<td>0 ASN</td>
<td>N</td>
<td>2 GLY</td>
<td>G</td>
<td>1 MET</td>
<td>M</td>
<td>1 TRP</td>
<td>W</td>
</tr>
<tr>
<td>2 ASP</td>
<td>D</td>
<td>1 HIS</td>
<td>H</td>
<td>1 PHE</td>
<td>F</td>
<td>1 TYR</td>
<td>O</td>
</tr>
<tr>
<td>0 CYS</td>
<td>C</td>
<td>0 I LU</td>
<td>I</td>
<td>3 PRO</td>
<td>P</td>
<td>0 VAL</td>
<td>V</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 22

* HARRIS, J. I., NATURE, VOL. 184, NO. 4681, PP. 167-169, JULY 18, 1959
BETA CORTICOTROPIN - PIG

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 S O S M E H F R W G K P V G K K R R P V K V O P G A E D D Q
31 L A E A F P L E F

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 SER TYR SER MET GLU HIS PHE ARG TRP GLY LYS PRO VAL GLY LYS
LYS ARG ARG PRO VAL LYS VAL TYR PRO GLY ALA GLU ASP ASP GLN
31 LEU ALA GLU ALA PHE PRO LEU GLU PHE ***

COMPOSITION

<table>
<thead>
<tr>
<th>3 ALA</th>
<th>1 GLN</th>
<th>2 LEU</th>
<th>2 SER</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 ARG</td>
<td>4 GLU</td>
<td>4 LYS</td>
<td>0 THR</td>
<td>T</td>
</tr>
<tr>
<td>0 ASN</td>
<td>3 GLY</td>
<td>1 MET</td>
<td>1 TRP</td>
<td>W</td>
</tr>
<tr>
<td>2 ASP</td>
<td>1 HIS</td>
<td>3 PHE</td>
<td>2 TYR</td>
<td>O</td>
</tr>
<tr>
<td>0 CYS</td>
<td>0 ILE</td>
<td>4 PRO</td>
<td>3 VAL</td>
<td>V</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 39


BELL, P. H., J. AM. CHEM. SOC., VOL.76, PP.5565-5567, NOV. 1954

THIS SEQUENCE DIFFERS FROM THAT SHOWN ABOVE BY REMOVING THE ASP (D) FROM POSITION 29 AND INSERTING IT BETWEEN POSITIONS 24 AND 25.
ALPHA CORTICOTROPIN - SHEEP AND BOVINE

123456789012345678901234567890
1S0SMEHFRWGRKVPGKRRPVKVOPDGEAED
31SAQAFPLEF

1234567890112345678901234567890
1SER TYR SER MET GLU HIS PHE ARG TRP GLY LYS PRO VAL GLY LYS
LYS ARG ARG PRO VAL LYS VAL TYR PRO ASP GLY GLU ALA GLU ASP
31 SER ALA GLN ALA PHE PRO LEU GLU PHE ***

COMPOSITION

3 ALA A
3 ARG R
0 ASN N
2 ASP D
0 CYS C
1 GLN Q
4 GLU E
3 GLY G
1 HIS H
0 ILU I
1 LEU L
4 LYS K
1 MET M
3 PHE F
4 PRO P
3 SER S
0 THR T
1 TRP W
2 TYR O
3 VAL V

TOTAL NO. OF ACIDS = 39

* LI, C.H,, GESCHWIND, I. I,, COLE, D., RAACK, I. D,, HARRIS, J.I,,
AND DIXON, J. S,, NATURE, VOL.176, NO.4484, PP.687-689,
OCT. 8, 1955 (SHEEP)

LI, C. H,, DIXON, J. S,, AND CHUNG, D,, J. AM. CHEM. SOC.,
VOL. 80, P.2587, 1958 (BOVINE)
INSULIN A - BOVINE

123456789012345678901
1GIVEQCCASVCSSLOQLENOCN*

123456789101112131415
1GLYILUVALGLUNCYSCYSALASERVALCYSSEULEUTYRGLN
LEUGLUASNTYRCYCYSASN***

COMPOSITION

\[
\begin{align*}
1 & \text{ALA} & 2 & \text{GLN} & 1 & \text{GLY} & 2 & \text{LEU} & 2 & \text{SER} \\
0 & \text{ARG} & 2 & \text{GLU} & 0 & \text{HIS} & 0 & \text{LYS} & 0 & \text{THR} \\
2 & \text{ASN} & 1 & \text{GLY} & 0 & \text{MET} & 0 & \text{LYS} & 0 & \text{TRP} \\
0 & \text{ASP} & 0 & \text{HIS} & 0 & \text{PHE} & 2 & \text{TYR} & 0 & \text{LYS} \\
4 & \text{CYS} & 1 & \text{ILU} & 0 & \text{PRO} & 2 & \text{VAL} & & \\
\end{align*}
\]

TOTAL NO. OF ACIDS = 21


THE AMIDE GROUPS WERE SUBSEQUENTLY DETERMINED.

**INSULIN A - BONITO**

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
1 G I(H,E,E,C(C,K,P,H))G,D,L)F E L E D O C N

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLY ILU(HIS,GLU,GLU,CYS,CYS,LYS,PRO,HIS)CYS,ASP,LEU)PHE GLU
LEU GLU ASP TYR CYS ASN ***

**COMPOSITION**

<table>
<thead>
<tr>
<th>0 ALA</th>
<th>0 GLN</th>
<th>2 LEU</th>
<th>0 SER</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ARG</td>
<td>4 GLU</td>
<td>1 LYS</td>
<td>0 THR</td>
</tr>
<tr>
<td>1 ASN</td>
<td>1 GLY</td>
<td>0 MET</td>
<td>0 TRP</td>
</tr>
<tr>
<td>2 ASP</td>
<td>2 HIS</td>
<td>1 PHE</td>
<td>1 TYR</td>
</tr>
<tr>
<td>4 CYS</td>
<td>1 ILU</td>
<td>1 PRO</td>
<td>0 VAL</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 21

* KOTAKI, A., J. BIOCHEM. (TOKYO), VOL.53, NO.1, PP.61-70, 1963
INSULIN A - HORSE

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
I G I V E Q C C T G I C S L O Q L E N O C N *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLY ILU VAL GLU GLN CYS CYS THR GLY ILU CYS SER LEU TYR GLN
LEU GLU ASN TYR CYS ASN ***

COMPOSITION

0 ALA A 2 GLN Q 2 LEU L 1 SER S
0 ARG R 2 GLU E 0 LYS K 1 THR T
2 ASN N 2 GLY G 0 MET M 0 TRP W
0 ASP D 0 HIS H 0 PHE F 2 TYR D
4 CYS C 2 ILU I 0 PRO P 1 VAL V

TOTAL NO. OF ACIDS = 21

* HARRIS, J. I., SANGER, F., AND NAUGHTON, M. A., ARCH. BIOCHEM. BIOPHYS., VOL. 65, PP. 427-438, 1956

SOME EVIDENCE FOR THE SEQUENCE WAS DERIVED FROM HOMOLOGY WITH BOVINE INSULIN.
INSULIN A - SHEEP

IS SHA 8.304

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
1 G I V E Q C C A G V C S L O Q L E N D C N

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLY ILU VAL GLU GLN CYS CYS ALA GLY VAL CYS SER LEU TYR GLN
LEU GLU ASN TYR CYS ASN ***

COMPOSITION

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ALA A</td>
<td>2 GLN Q</td>
<td>2 LEU L</td>
<td>1 SER S</td>
<td></td>
</tr>
<tr>
<td>0 ARG R</td>
<td>2 GLU E</td>
<td>0 LYS K</td>
<td>0 THR T</td>
<td></td>
</tr>
<tr>
<td>2 ASN N</td>
<td>2 GLY G</td>
<td>0 MET M</td>
<td>0 TRP W</td>
<td></td>
</tr>
<tr>
<td>0 ASP D</td>
<td>0 HIS H</td>
<td>0 PHE F</td>
<td>2 TYR D</td>
<td></td>
</tr>
<tr>
<td>4 CYS C</td>
<td>1 ILU I</td>
<td>0 PRO P</td>
<td>2 VAL V</td>
<td></td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 21


SOME EVIDENCE FOR THE SEQUENCE WAS DERIVED FROM HOMOLOGY WITH BOVINE INSULIN.
INSULIN A - SPERM WHALE, FIN-WHALE, PIG, AND HUMAN

123456789012345678901
1GIVEQCCTSCISLOQLENOCN*

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1GLY ILU VAL GLU GLN CYS CYS THR SER ILU CYS SER LEU TYR GLN
LEU GLU ASN TYR CYS ASN ***

COMPOSITION

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 21

* BROWN, H., SANGER, F., AND KITAI, R., BIOCHEM. J., VOL. 60, PP. 556-565, 1955 (PIG)

SOME EVIDENCE FOR THE SEQUENCE WAS DERIVED FROM HOMOLOGY WITH BOVINE INSULIN.

HARRIS, J. I., SANGER, F., AND NAUGHTON, M. A., ARCH. BIOCHEM. BIOPHYS., VOL. 65, PP. 427-438, 1956 (SPERM WHALE)

HAMA, H., TITANI, K., SAKAKI, S., AND NARITA, K., J. BIOCHEM. (TOKYO), VOL. 56, NO. 3, PP. 285-293, 1964 (FIN-WHALE)

THIS WORK CONFIRMED THE SEQUENCE ABOVE, EXCEPT GLU (E) AND GLN (Q) WERE INTERCHANGED AT POSITIONS 15 AND 17.

INSULIN A – SEI-WHALE

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
1 G I V E Q C C A S T C S L O Q L E N O C N *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLY ILU VAL GLU GLN CYS CYS ALA SER THR CYS SER LEU TYR GLN
LEU GLU ASN TYR CYS ASN ***

COMPOSITION

| 1 ALA | 2 GLN | 2 LEU | 2 SER | S   |
| 0 ARG | 2 GLU | 0 LYS | K    |
| 2 ASN | 1 GLY | 0 MET | M    |
| 0 ASP | 0 HIS | 0 PHE | F    |
| 4 CYS | 1 ILU | 0 PRO | P    | 1 VAL | V |

TOTAL NO. OF ACIDS = 21

ISHIHARA, Y., SAITO, T., ITO, Y., AND FUJINO, M., NATURE, VOL. 181, NO. 4621, PP. 1461-1469, MAY 24, 1958 (SEI-WHALE)
INSULIN B - BOVINE, SHEEP, HORSE, HUMAN, PIG, AND SPERM WHALE

TWO DISULPHIDE BONDS CONNECT THE A AND B CHAINS. A7 IS BONDED TO B7 AND A20 IS BONDED TO B19. IN ADDITION THERE IS A BOND FROM A6 TO A11.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 F V N Q H L C G S H L V E A L O L V C G E R G F F O T P K A

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 PHE VAL ASN GLN HIS LEU CYS GLY SER HIS LEU VAL GLU ALA LEU
TYR LEU VAL CYS GLY GLU ARG GLY PHE PHE TYR THR PRO LYS ALA

***

COMPOSITION

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>ALA</td>
<td>A</td>
<td>1</td>
<td>GLN</td>
<td>Q</td>
<td>4</td>
<td>LEU</td>
<td>L</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>ARG</td>
<td>R</td>
<td>2</td>
<td>GLU</td>
<td>E</td>
<td>1</td>
<td>LYS</td>
<td>K</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>ASN</td>
<td>N</td>
<td>3</td>
<td>GLY</td>
<td>G</td>
<td>0</td>
<td>MET</td>
<td>M</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>ASP</td>
<td>D</td>
<td>2</td>
<td>HIS</td>
<td>H</td>
<td>3</td>
<td>PHE</td>
<td>F</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>CYS</td>
<td>C</td>
<td>0</td>
<td>ILU</td>
<td>I</td>
<td>1</td>
<td>PRO</td>
<td>P</td>
<td>3</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 30


THE AMIDE GROUPS WERE SUBSEQUENTLY DETERMINED.


HUMAN INSULIN B CHAIN IS IDENTICAL WITH ABOVE EXCEPT THAT POSITION 30 IS THR (T).
**INSULIN B - BONITO**

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ALA ALA ASN(PRO,HIS,LEU)CYS(GLY,SER,HIS,LEU,VAL,GLU,ALA,LEU)
TYR LEU(VAL,CYS,GLY,GLU)ARG GLY PHE PHE TYR GLN PRO LYS ***

**COMPOSITION**

<table>
<thead>
<tr>
<th></th>
<th>3 ALA A</th>
<th>1 GLN Q</th>
<th>4 LEU L</th>
<th>1 SER S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ARG R</td>
<td>2 GLU E</td>
<td>1 LYS K</td>
<td>0 THR T</td>
<td></td>
</tr>
<tr>
<td>1 ASN N</td>
<td>3 GLY G</td>
<td>0 MET M</td>
<td>0 TRP W</td>
<td></td>
</tr>
<tr>
<td>0 ASP D</td>
<td>2 HIS H</td>
<td>2 PHE F</td>
<td>2 TYR O</td>
<td></td>
</tr>
<tr>
<td>2 CYS C</td>
<td>0 ILU I</td>
<td>2 PRO P</td>
<td>2 VAL V</td>
<td></td>
</tr>
</tbody>
</table>

**TOTAL NO. OF ACIDS = 29**

FIBRINOPEPTIDE A - BOVINE

FIBRINOPEPTIDES ARE THOSE PORTIONS OF VERTEBRATE FIBRINOGEN MOLECULES WHICH ARE PROTEOLYTICALLY REMOVED BY THE ENZYME THROMBIN. THEIR REMOVAL PERMITS SPONTANEOUS POLYMERIZATION OF THE PARENT MOLECULES TO FORM AN INSOLUBLE FIBRINOGL. SINCE THE FUNCTION OF THE FIBRINOPEPTIDES IS RATHER NON-SPECIFIC, LARGE SEQUENCE CHANGES ARE OBSERVED AMONG CLOSELY RELATED SPECIES.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
1 E D G S D P P S G D F L T E G G G V R /

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
1 GLU ASP GLY SER ASP PRO PRO SER GLY ASP PHE LEU THR GLU GLY
GLY GLY VAL ARG ///

COMPOSITION

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>0</td>
</tr>
<tr>
<td>ARG</td>
<td>1</td>
</tr>
<tr>
<td>ASN</td>
<td>0</td>
</tr>
<tr>
<td>ASP</td>
<td>3</td>
</tr>
<tr>
<td>CYS</td>
<td>0</td>
</tr>
<tr>
<td>GLN</td>
<td>0</td>
</tr>
<tr>
<td>GLU</td>
<td>2</td>
</tr>
<tr>
<td>GLY</td>
<td>5</td>
</tr>
<tr>
<td>HIS</td>
<td>0</td>
</tr>
<tr>
<td>ILU</td>
<td>0</td>
</tr>
<tr>
<td>LYS</td>
<td>0</td>
</tr>
<tr>
<td>MET</td>
<td>0</td>
</tr>
<tr>
<td>PHE</td>
<td>1</td>
</tr>
<tr>
<td>PRO</td>
<td>2</td>
</tr>
<tr>
<td>THR</td>
<td>1</td>
</tr>
<tr>
<td>TRP</td>
<td>0</td>
</tr>
<tr>
<td>TYR</td>
<td>0</td>
</tr>
<tr>
<td>VAL</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL ACIDS</td>
<td>19</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 19

- DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964
FIBRINOPEPTIDE A - SHEEP

1234567890123456789
1 A D D S D P V G G E F L A E G G G V R /

123456789101112131415
1 ALA ASP ASP SER ASP PRO VAL GLY GLY GLU PHE LEU ALA GLU GLY
GLY GLY VAL ARG ///

COMPOSITION

<table>
<thead>
<tr>
<th></th>
<th>ALA</th>
<th>ASP</th>
<th>ASP</th>
<th>SER</th>
<th>ASP</th>
<th>PRO</th>
<th>VAL</th>
<th>GLY</th>
<th>GLY</th>
<th>GLU</th>
<th>PHE</th>
<th>LEU</th>
<th>ALA</th>
<th>GLU</th>
<th>GLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 19

- DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964
FIBRINOPePTIDE A - GOAT

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
1 A D D S D P V G G E F L A E G G G V R /

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ALA ASP ASP SER ASP PRO VAL GLY GLY GLU PHE LEU ALA GLU GLY GLY GLY VAL ARG ///

COMPOSITION

<table>
<thead>
<tr>
<th>2 ALA</th>
<th>0 GLN</th>
<th>1 LEU</th>
<th>1 SER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ARG</td>
<td>2 GLU</td>
<td>0 LYS</td>
<td>0 THR</td>
</tr>
<tr>
<td>0 ASN</td>
<td>5 GLY</td>
<td>0 MET</td>
<td>0 TRP</td>
</tr>
<tr>
<td>3 ASP</td>
<td>0 HIS</td>
<td>1 PHE</td>
<td>0 TYR</td>
</tr>
<tr>
<td>0 CYS</td>
<td>0 ILU</td>
<td>1 PRO</td>
<td>2 VAL</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 19

• DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964
FIBRINOPEPTIDE A - REINDEER

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 1 ADGSDPPAGGEF(L,A,E,G,G,G,V)R /

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 1 ALA ASP GLY SER ASP PRO ALA GLY GLY GLU PHE(LEU,ALA,GLU,GLY,
GLY,GLY,VAL)ARG ///

COMPOSITION

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>ALA</td>
<td>0</td>
<td>GLN</td>
<td>Q</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>ARG</td>
<td>2</td>
<td>GLU</td>
<td>E</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>ASN</td>
<td>6</td>
<td>GLY</td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>ASP</td>
<td>0</td>
<td>HIS</td>
<td>H</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>CYS</td>
<td>0</td>
<td>ILU</td>
<td>I</td>
<td>1</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 19

* DOULITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964*
PIBRINOPePTIDE A - PIG

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7
1 A E V Q D K G E F L A E G G G V R /

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ALA GLU VAL GLN ASP LYS GLY GLU PHE LEU ALA GLU GLY GLY GLY

VAL ARG ///

COMPOSITION

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>GLN</td>
<td>LEU</td>
<td>SER</td>
<td>GLU</td>
</tr>
<tr>
<td>ARG</td>
<td>GLU</td>
<td>LYS</td>
<td>THR</td>
<td>MET</td>
</tr>
<tr>
<td>ASN</td>
<td>G1Y</td>
<td>HIS</td>
<td>PSP</td>
<td>MET</td>
</tr>
<tr>
<td>ASP</td>
<td>HIS</td>
<td>PHE</td>
<td>THR</td>
<td>PRO</td>
</tr>
<tr>
<td>CYS</td>
<td>ILU</td>
<td>PRO</td>
<td>VAL</td>
<td>VAL</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 17

* DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964
FIBRINOPEPTIDE A - HUMAN

1234567890123456
1ADSGEGDFLAE GG V R /

123456789101112131415
1 ALA ASP SER GLY GLU GLY ASP PHE LEU ALA GLU GLY GLY GLY VAL ARG ///

COMPOSITION

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ALA</td>
<td>0 GLN</td>
<td>1 LEU</td>
<td>1 SER</td>
<td></td>
</tr>
<tr>
<td>1 ARG</td>
<td>2 GLU</td>
<td>0 LYS</td>
<td>0 THR</td>
<td></td>
</tr>
<tr>
<td>0 ASN</td>
<td>5 GLY</td>
<td>0 MET</td>
<td>0 TRP</td>
<td></td>
</tr>
<tr>
<td>2 ASP</td>
<td>0 HIS</td>
<td>1 PHE</td>
<td>0 TYR</td>
<td></td>
</tr>
<tr>
<td>0 CYS</td>
<td>0 ILE</td>
<td>0 PRO</td>
<td>1 VAL</td>
<td></td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 16


PHOSPHO-SERINE OCCURS AT POSITION 3 IN ABOUT HALF THE MOLECULES. A MINOR COMPONENT FRAGMENT, WITH THE N TERMINAL ALANINE MISSING, HAS BEEN DETECTED IN ALL INDIVIDUALS.
FIBRINOPEPTIDE A - RABBIT

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6
1 V D P G E T S F L(T,E,G,G)D A R /

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 VAL ASP PRO GLY GLU THR SER PHE LEU(THR,GLU,GLY,GLY)ASP ALA
ARG ///

COMPOSITION

1 ALA A 0 GLN Q 1 LEU L 1 SER S
1 ARG R 2 GLU E 0 LYS K 2 THR T
0 ASN N 3 GLY G 0 MET M 0 TRP W
2 ASP D 0 HIS H 1 PHE F 0 TYR O
0 CYS C 0 ILU I 1 PRO P 1 VAL V

TOTAL NO. OF ACIDS = 16

• DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964
FIBRINOPEPTIDE B - BOVINE

PYRROCIDONE CARBOXYLIC ACID - AT AMINO END
S04 ATTACHED TO TYROSINE AT POSITION 5

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 FPTDODEGQDDRPKVGLGAR/

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 PHE PRO THR ASP TYR ASP GLU GLY GLN ASP ASP ARG PRO LYS VAL
GLY LEU GLY ALA ARG ///

COMPOSITION

| 1 ALA A | 1 GLN Q | 1 LEU L | 0 SER S |
| 2 ARG R | 1 GLU E | 1 LYS K | 1 THR T |
| 0 ASN N | 3 GLY G | 0 MET M | 0 TRP W |
| 4 ASP D | 0 HIS H | 1 PHE F | 1 TYR D |
| 0 CYS C | 0 ILU I | 2 PRO P | 1 VAL V |

TOTAL NO. OF ACIDS = 20

* DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202,
NO. 4928, PP. 147-152, APRIL 11, 1964
FIBRINOPEPTIDE B - SHEEP

SU4 ATTACHED TO TYROSINE AT POSITION 5

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 G O L D O D E V D D N R A K L P L D A R /

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLY TYR LEU ASP TYR ASP GLU VAL ASP ASN ARG ALA LYS LEU
PRO LEU ASP ALA ARG ///

COMPOSITION

2 ALA A 0 GLN Q 3 LEU L 0 SER S
2 ARG R 1 GLU E 1 LYS K 0 THR T
1 ASN N 1 GLY G 0 MET M 0 TRP W
5 ASP D 0 HIS H 0 PHE F 2 TYR O
0 CYS C 0 I LU I 1 PRO P 1 VAL V

TOTAL NO. OF ACIDS = 20

* DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202,
NO. 4928, PP. 147-152, APRIL 11, 1964
FIBRINOPEPTIDE B - GOAT

SO4 ATTACHED TO TYROSINE AT POSITION 5

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
1 G O L D O D E V D D N R A K L P L D A R /

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 GLY TYR LEU ASP TYR ASP GLU VAL ASP ASP ASN ARG ALA LYS LEU
PRO LEU ASP ALA ARG ///

COMPOSITION

2 ALA A 0 GLN Q 3 LEU L 0 SER S
2 ARG R 1 GLU E 1 LYS K 0 THR T
1 ASN N 1 GLY G 0 MET M 0 TRP W
5 ASP D 0 HIS H 0 PHE F 2 TYR O
0 CYS C 0 ILU I 1 PRO P 1 VAL V

TOTAL NO. OF ACIDS = 20

FIBRINOPEPTIDE B – REINDEER

PYRROLIDONE CARBOXYLIC ACID - AT AMINO END
SO4 ATTACHED TO TYROSINE AT POSITION 4
A MUTANT HAS BEEN FOUND WHERE GLYCINE REPLACES HISTIDINE
IN POSITION 9.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
1 L A D O D E V(E,H,D)R A K L H L D A R /

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 LEU ALA ASP TYR ASP GLU VAL (GLU,HIS,ASP) ARG ALA LYS LEU HIS
LEU ASP ALA ARG ///

COMPOSITION

<table>
<thead>
<tr>
<th>3 ALA</th>
<th>0 GLN</th>
<th>3 LEU</th>
<th>0 SER</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ARG</td>
<td>0 GLY</td>
<td>1 LYS</td>
<td>0 THR</td>
<td>T</td>
</tr>
<tr>
<td>0 ASN</td>
<td>2 GLU</td>
<td>0 MET</td>
<td>0 TRP</td>
<td>W</td>
</tr>
<tr>
<td>4 ASP</td>
<td>0 HIS</td>
<td>0 PHE</td>
<td>1 TYR</td>
<td>D</td>
</tr>
<tr>
<td>0 CYS</td>
<td>0 ILU</td>
<td>0 PRO</td>
<td>1 VAL</td>
<td>V</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 19

* DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964
FIBRINOPEPTIDE B - PIG

SO4 ATTACHED TO TYROSINE AT POSITION 4

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
1 A I D O D E D E D G R P K V H V D A R /

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ALA ILU ASP TYR ASP GLU ASP GLU ASP GLY ARG PRO LYS VAL HIS
VAL ASP ALA ARG ///

COMPOSITION

2 ALA A 0 GLN Q 0 LEU L 0 SER S
2 ARG R 2 GLU E 1 LYS K 0 THR T
0 ASN N 1 GLY G 0 MET M 0 TRP W
5 ASP D 1 HIS H 0 PHE F 1 TYR Y
0 CYS C 1 ILE I 1 PRO P 2 VAL V

TOTAL NO. OF ACIDS = 19

• DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964
FIBRINOPEPTIDE B - HUMAN

PYRROLIDONE CARBOXYLIC ACID - AT AMINO END
PHOSPHO-SERINE OCCURS IN POSITION 11.

1 2 3 4 5 6 7 8 9 0 1 2 3
1 G V N D N E E G F F S A R /

1 2 3 4 5 6 7 8 9 10 11 12 13
1 GLY VAL ASN ASP ASN GLU GLY PHE PHE SER ALA ARG ///

COMPOSITION

<table>
<thead>
<tr>
<th>1</th>
<th>ALA</th>
<th>0</th>
<th>GLN</th>
<th>Q</th>
<th>0</th>
<th>LEU</th>
<th>L</th>
<th>1</th>
<th>SER</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ARG</td>
<td>R</td>
<td>2</td>
<td>GLU</td>
<td>E</td>
<td>0</td>
<td>LYS</td>
<td>K</td>
<td>0</td>
<td>THR</td>
</tr>
<tr>
<td>2</td>
<td>ASN</td>
<td>N</td>
<td>2</td>
<td>GLY</td>
<td>G</td>
<td>0</td>
<td>MET</td>
<td>M</td>
<td>0</td>
<td>TRP</td>
</tr>
<tr>
<td>1</td>
<td>ASP</td>
<td>D</td>
<td>0</td>
<td>HIS</td>
<td>H</td>
<td>2</td>
<td>PHE</td>
<td>F</td>
<td>0</td>
<td>TYR</td>
</tr>
<tr>
<td>0</td>
<td>CYS</td>
<td>C</td>
<td>0</td>
<td>ILU</td>
<td>I</td>
<td>0</td>
<td>PRO</td>
<td>P</td>
<td>1</td>
<td>VAL</td>
</tr>
</tbody>
</table>

TOTAL NO. OF ACIDS = 13

* DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964*
FIBRINOPEPTIDE B - RABBIT

SG4 ATTACHED TO TYROSINE AT POSITION 4

1 2 3 4 5 6 7 8 9 0 1 2 3
1 A D D O I D E, P, L, D, V I D A R /

1 2 3 4 5 6 7 8 9 10 11 12 13
1 ALA ASP ASP TYR (ASP, GLU, PRO, LEU, ASP, VAL) ASP ALA ARG ///

COMPOSITION

2 ALA A 0 GLN Q 1 LEU L 0 SER S
1 ARG R 1 GLU E 0 LYS K 0 THR T
0 ASN N 0 GLY G 0 MET M 0 TRP W
5 ASP D 0 HIS H 0 PHE F 1 TYR O
0 CYS C 0 ILU I 1 PRO P 1 VAL V

TOTAL NO. OF ACIDS = 13

• DOOLITTLE, R. F. AND BLOMBACK, B., NATURE, VOL. 202, NO. 4928, PP. 147-152, APRIL 11, 1964
Immunoglobulins are serum proteins distinguishable by electrophoretic mobilities, sedimentation coefficients and differential solubilities in variable ethanol-salt solutions. Of these, the gamma globulins are associated with normal antibody function. A proposal for the structure of gamma globulin has been made by Porter (1959) and Fleishman et al., (1963).

\[ \text{L} \]
\[ \text{H} \]
\[ \text{H} \]
\[ \text{L} \]

Gamma globulin is thought to be a tetramer consisting of two pairs of identical polypeptide chains held in a particular configuration by disulfide bonds. There are two L (m.w. 20-25,000 each) and two H chains (m.w. 50,000-55,000 each). Because of the chemical problems associated with elucidation of gamma globulin structure, attention has turned to the abundantly produced, structurally similar globulins found in multiple myeloma.

Bence-Jones proteins are found exclusively in the urine of all multiple myeloma patients, and probably represent abnormal protein synthesized by the multiple myeloma tumor cell. They are thought to be made exclusively of L chains, related to gamma globulins, (Edelman and Gally, 1962, S. Cohen, 1963, Putnam 1962). It is thought that determination of the amino acid sequence of a particular individual's Bence-Jones protein would reflect a homologous sequence in that individual's antibody structure, thereby partially elucidating the structure of gamma globulin.

BENCE-JONES PROTEIN

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
121 P,F)L K S(T,S,G,A)V(V,C) L L D(D,P,F)D R E A K V E W K V(D,D,
211 E C *

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 ASP(THR, SER, SER, SER, GLU, GLU, PRO, MET, ILE) LEU SER(SER, GLY, ALA,
31 VAL)ASP ARG(ASP, THR, THR, SER, SER, GLU, GLU, ALA, VAL, ILE, ILE, ILE,
PHE, CYS) LEU(TYR, THR, ASP, TRP, GLU, GLU, PRO, GLY) LYS LYS ALA PRO LYS
LEU LEU ILE TYR ASP ALA SER LYS LEU GLU(SER, PRO, GLY, ALA, VAL)
61 ARG PHE SER(ASP, THR, THR, SER, GLY, GLY, GLY) PHE THR(ASP, SER, SER,
91 GLU, GLU, PRO, ILE, LEU) ILE ALA THR TYR(ASP, ASP, THR, GLU, GLU, PRO,
121 LEU, LEU, CYS, TYR, PHE, PHE) GLY(THR, GLY, GLY) LYS VAL ASP PHE LYS
ARG THR(SER, PRO, ALA, ALA, VAL) VAL PHE ILE(ASP, SER, GLU, GLU, PRO,
151 PRO, PHE) LEU LYS SER(THR, SER, GLY, ALA) VAL(VAL, CYS) LEU LEU ASP
(ASP, PRO, PHE) TYR ARG GLU ALA LYS VAL GLU TRP LYS VAL(ASP, ASP,
181 ASP, SER, SER, GLU, GLU, GLY, ALA, LEU) GLU SER(ASP, THR, SER, GLU, GLU,
VAL) LYS ASP(THR, SER) TYR SER SER SER THR LEU LEU THR LEU SER
211 LYS ALA ASP TYR GLU LYS HIS LYS LEU TYR ALA CYS GLU VAL(THR,
GLU, GLY, HIS) LEU SER(THR, SER, PRO, VAL) LYS SER PHE ASP ARG GLY
211 GLU CYS **
### COMPOSITION

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (AL)</td>
<td>13</td>
</tr>
<tr>
<td>Glu (GL)</td>
<td>24</td>
</tr>
<tr>
<td>Leu (LE)</td>
<td>17</td>
</tr>
<tr>
<td>Ser (SE)</td>
<td>29</td>
</tr>
<tr>
<td>Arg (AR)</td>
<td>5</td>
</tr>
<tr>
<td>Lys (LY)</td>
<td>14</td>
</tr>
<tr>
<td>Thr (TH)</td>
<td>17</td>
</tr>
<tr>
<td>Trp (TR)</td>
<td>2</td>
</tr>
<tr>
<td>Asn (AS)</td>
<td>0</td>
</tr>
<tr>
<td>Gly (GL)</td>
<td>13</td>
</tr>
<tr>
<td>Met (ME)</td>
<td>1</td>
</tr>
<tr>
<td>His (HS)</td>
<td>2</td>
</tr>
<tr>
<td>Phe (PH)</td>
<td>10</td>
</tr>
<tr>
<td>Tyr (TY)</td>
<td>8</td>
</tr>
<tr>
<td>Cys (CY)</td>
<td>5</td>
</tr>
<tr>
<td>Ile (IL)</td>
<td>8</td>
</tr>
<tr>
<td>Pro (PR)</td>
<td>11</td>
</tr>
<tr>
<td>Val (VL)</td>
<td>13</td>
</tr>
</tbody>
</table>

*HILSCHMANN, N. AND CRAIG, L.C., PROC. NATL. ACAD. SCI. U.S., VOL. 53, NO. 6, PP. 1403-1409, 1965*
<table>
<thead>
<tr>
<th>Author Index</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACHER, R.</td>
<td></td>
</tr>
<tr>
<td>AMBLER, R.</td>
<td></td>
</tr>
<tr>
<td>ANDERER, F.</td>
<td></td>
</tr>
<tr>
<td>ANSFIELD, M.</td>
<td></td>
</tr>
<tr>
<td>BAHL, O.</td>
<td></td>
</tr>
<tr>
<td>BAGLIONI, C.</td>
<td></td>
</tr>
<tr>
<td>BARNAFI, L.</td>
<td></td>
</tr>
<tr>
<td>BARTSCH, R.</td>
<td></td>
</tr>
<tr>
<td>BEALE, D.</td>
<td></td>
</tr>
<tr>
<td>BEHRENS, O.</td>
<td></td>
</tr>
<tr>
<td>BELL, P.</td>
<td></td>
</tr>
<tr>
<td>BENSON, A.</td>
<td></td>
</tr>
<tr>
<td>BERNIER, I.</td>
<td></td>
</tr>
<tr>
<td>BLAKE, C.</td>
<td></td>
</tr>
<tr>
<td>BLOMBACK, B.</td>
<td></td>
</tr>
<tr>
<td>BRAUNITZER, G.</td>
<td></td>
</tr>
<tr>
<td>BROMER, W.</td>
<td></td>
</tr>
<tr>
<td>BROWN, H.</td>
<td></td>
</tr>
<tr>
<td>BROWN, J.</td>
<td></td>
</tr>
<tr>
<td>BUETTNER-JANUSCH, J.</td>
<td></td>
</tr>
<tr>
<td>CANFIELD, R.</td>
<td></td>
</tr>
<tr>
<td>CHAUVET, J.</td>
<td></td>
</tr>
<tr>
<td>CHUNG, D.</td>
<td></td>
</tr>
<tr>
<td>COHEN, S.</td>
<td></td>
</tr>
<tr>
<td>COLE, D.</td>
<td></td>
</tr>
<tr>
<td>CORMICK, J.</td>
<td></td>
</tr>
<tr>
<td>CRAIG, L.</td>
<td></td>
</tr>
<tr>
<td>DAVIES, D.</td>
<td></td>
</tr>
<tr>
<td>DIXON, J.</td>
<td></td>
</tr>
<tr>
<td>DLOUHA, V.</td>
<td></td>
</tr>
</tbody>
</table>
DOOLITTLE, R. F.

DUS, K.
DU VIGNAUD, V.

EDELMAN, G. M.
EDMUNDSON, A. B.
EFRON, M. L.
EIGNER, E. A.
ELLiot, D. F.
FLEISHMAN, J. B.
FRAENKEL-CONRAT, H.
FRATER, R.
FUJINO, M.

FUNATSU, G.
GALLY, J. A.
GEHRING-MULLER, R.

GERALD, P. S.
GESCHWIND, I. I.

HAMA, H.
HARRIS, J. I.

HARTLEY, B. S.
HIGA, H.
HILL, R. J.
HILL, R. L.

HILSCHMANN, N.

HILSE, K.
SHELTON, Jo B.
SHELTON, Jo R.
SHEPHERD, Re G.
SINN, Le G.
SMILLIE, Le B.
SMITH, De B.
SMITH, E. L.
SMITH, L. F.
SMYTH, De G.
SORM, F.
STEIN, We H.
STEWART, Jo W.
SWENSON, Re T.
TANAKA, M.
THOMPSON, E. O. Pa.
TITANI, K.
TRIPPETT, S.
TSUGITA, A.
TUPPY, H.

UHLIG, H.
WALSH, K.
WATSON-WILLIAMS, E. J.
WEBER, E.
WHITE, We F.
WITTMAN, M. G.
WITTMAN-LIEBOLD, B.

YAOI, Y.
YASUNOBU, K. T.
ZUCKERKANDL, E.