[14] Heller, C., and H. M. McConnell, *J. Chem. Phys.*, **32**, 1535 (1960).
[15] Pooley, D., and D. H. Whiffen, *Mol. Phys.*, **4**, 81 (1961).
[16] Jaseja, T. S., and R. S. Anderson, *J. Chem. Phys.*, **36**, 2727 (1962).
[17] Kwiram, A. L., and H. M. McConnell, these PROCEEDINGS, **48**, 499 (1962).
[18] Preliminary work on X-irradiated single crystals of maleic acid (cis $HO_2CCH{=}CHCO_2H$) indicates that radical III is also formed in this system.
[19] Meakins, R. J., *Trans. Faraday Soc.*, **51**, 953 (1955).
[20] Gilson, D. F. R., and C. A. McDowell, *Mol. Phys.*, **4**, 125 (1961).

THE FUSION OF TWO PEPTIDE CHAINS IN HEMOGLOBIN LEPORE AND ITS INTERPRETATION AS A GENETIC DELETION

CORRADO BAGLIONI*
DEPARTMENT OF BIOLOGY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated by S. E. Luria, August 22, 1962

Gerald and Diamond[1] have described an abnormal hemoglobin, called hemoglobin Lepore (Hb-Lepore$_{Boston}$). The present communication reports the results of a chemical investigation of Hb-Lepore$_{Boston}$ and an interpretation of the genetic events leading to the formation of a single peptide chain from the fusion of two different chains.

Abnormal hemoglobins electrophoretically identical to Hb-Lepore$_{Boston}$ have have been reported among Greeks,[2] Italians,[3] and Papuans[4] and designated Hb-Pylos, Hb-G, and Hb-Lepore$_{Hollandia}$ respectively. These hemoglobins are always found in concentrations of about 10–15 per cent of the total hemoglobin in the heterozygotes. One individual homozygous for the Hb-Pylos gene and two homozygous for the Hb-Lepore$_{Hollandia}$ gene have been reported.[2, 4] They were characterized by the absence of normal adult hemoglobin (Hb-A) and of the minor component hemoglobin A_2(Hb-A_2).[2, 4] Neeb *et al.*[4] have reported, however, the presence of traces of a hemoglobin component migrating like Hb-A_2 in Hb-Lepore$_{Hollandia}$ homozygotes; the identification of this trace component remains to be established.

The three normal human hemoglobins Hb-A, Hb-F (fetal hemoglobin), and Hb-A_2 consist of two α peptide chains which are under the control of a single structural gene and of two other peptide chains which differ in different hemoglobins: β chains in Hb-A, γ chains in Hb-F, and δ chains in Hb-A_2.[5] These are under the control of different structural genes.[5] The β gene is linked to the δ gene.[5] The β and δ chains have extremely similar amino acid compositions and sequences.[6–8] This similarity has suggested that the δ gene may have originated through a duplication of the β gene followed by independent evolution.[9]

The Hb-Pylos and Hb-Lepore$_{Hollandia}$ mutations seem to affect both the two linked genes β and δ, suppressing the synthesis of both Hb-A and Hb-A_2 in homozygotes.[2, 4] The α and γ peptide chains are synthesized normally.[2, 6]

Gerald *et al.*[10] have suggested that the Lepore abnormality is a "mutation possibly involving two cistrons." These authors fingerprinted Hb-Lepore$_{Boston}$ and Hb-Pylos and analyzed some of their tryptic peptides. The fingerprints were found to be indistinguishable from those of Hb-A_2. The composition of the tryptic peptides

analyzed was the same as that of homologous peptides of Hb-A_2.[10] Hb-$Lepore_{Boston}$ and Hb-Pylos appear to be identical.[2, 10] Barnabas and Muller[6] reported that Hb-$Lepore_{Hollandia}$ shows some differences from Hb-A_2. One peptide, called A_2a,[7] is absent in Hb-$Lepore_{Hollandia}$. A peptide which in Hb-$Lepore_{Boston}$ has the composition of a δ chain peptide ($\delta 5$)[11] has in Hb-$Lepore_{Hollandia}$ the composition of a β chain peptide ($\beta 5$).[6] This suggested that the non-α chain of Hb-$Lepore_{Hollandia}$ may have characteristics of both the β and δ peptide chains.

Materials and Methods.—Hemoglobins: Hb-A and Hb-A_2 were purified by column chromatography according to Schnek and Schroeder.[12] The hemoglobin solutions were concentrated as previously described[13] and were heat-denatured and digested with trypsin according to Ingram.[14] The tryptic digest of 20 mg of purified Hb-$Lepore_{Boston}$ was kindly donated by Park S. Gerald of the Children's Hospital (Boston, Mass.).

Fingerprinting: The original procedure of Ingram was followed for the ionophoresis,[14] while different chromatographic solvents were used: *#1*, pyridine/*iso*amyl/alcohol water (35/35/30);[15] *#2*, butanol/acetic acid/pyridine/water (75/15/50/60);[16] *#3*, butanol/acetic acid/pyridine/water (75/15/50/50). When elution of the peptides from the fingerprints was desired, the ionophoresis was prolonged for 4 hr, allowing the most basic peptides to run off the paper. The fingerprints were then developed in solvent #3.

Analysis of tryptic peptides: The fingerprints were stained with 0.01% ninhydrin and then washed with acetone. The peptides were eluted with 1 *N* acetic acid. The eluates were hydrolyzed and the amino acid composition was determined as previously indicated.[13]

Results.—Fingerprints of Hb-$Lepore_{Boston}$ were compared to fingerprints of Hb-A and of Hb-A_2. The fingerprints of Hb-$Lepore_{Boston}$ and Hb-A developed in solvent #3 are shown in Figure 1. The peptides of Hb-$Lepore_{Boston}$ indicated

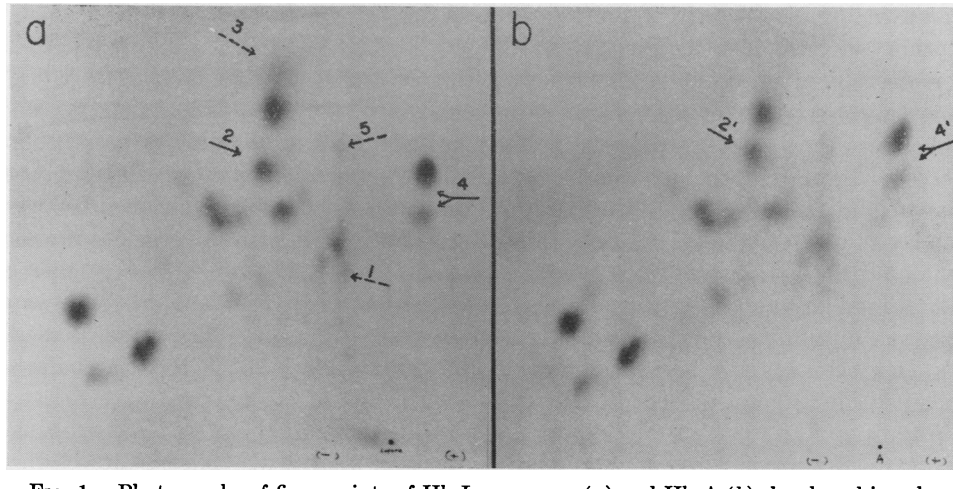

FIG. 1.—Photographs of fingerprints of Hb-$Lepore_{Boston}$ (*a*) and Hb-A (*b*) developed in solvent *#3* (see *Methods*). The solid arrows in *a* and *b* indicate corresponding peptides that show different R_f. The arrows in *a* indicate the peptides of Hb-$Lepore_{Boston}$ which have been eluted from fingerprints and analyzed. The numbers in *a* refer to the peptides reported in Table 1: 1 = x_B1; 2 = x_B2; 3 = x_B4; 4 = x_B5 + x_B5 oxidized; 5 = x_B9. 2′ in *b* indicate peptide $\beta 2$ and 4′ peptides $\beta 5$ + $\beta 5$ oxidized.[15]

by solid arrows in Figure 1*a* were found to have R_f's different from those of the corresponding peptides of Hb-A, indicated by solid arrows in Figure 1*b*. The fingerprints of Hb-Lepore$_{Boston}$ developed in solvent #1 or #3 were found to be indistinguishable from fingerprints of Hb-A_2, confirming the observation of Gerald *et al.*[10]

When, however, the fingerprints of Hb-A_2 and Hb-Lepore$_{Boston}$ were developed in solvent #2 (see Fig. 2) and stained for sulfur-containing amino acids and for

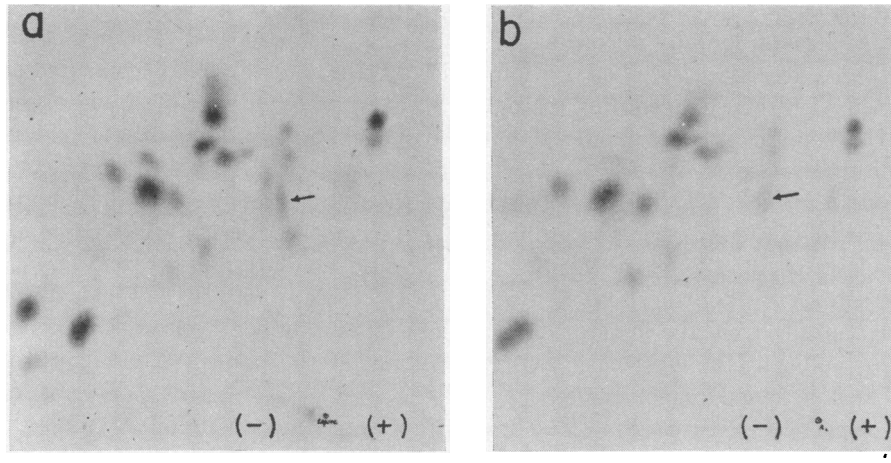

FIG. 2.—Photographs of fingerprints of Hb-Lepore$_{Boston}$ (*a*) and Hb-A_2 (*b*) developed in solvent *#2* (see *Methods*). The arrow in *b* indicates peptide δ14; the arrow in *a* indicates the corresponding peptide of Hb-Lepore$_{Boston}$ (x_B13) that gives the same staining reaction as β13.

tyrosine, a difference between them became apparent. A Hb-A_2 peptide indicated by the arrow in Figure 2*b* gave a positive reaction for sulfur and tyrosine. A peptide which occupies an identical position in the fingerprints of Hb-Lepore$_{Boston}$ (arrow in Fig. 2*a*) showed a positive reaction for tyrosine but not for sulfur. This peptide has been identified as β13 in Hb-A.[15] In Hb-A_2, there is a peptide, δ14 (see later),[11] which occupies the same position in fingerprints as β13 in Hb-A but which gives a positive reaction for sulfur and tyrosine.[7] Analysis of peptide δ14 has shown the presence of a methionine residue.[8] No methionine is present in β13.[15] This difference between Hb-A and Hb-A_2 does not show in fingerprints developed in solvent #1 or #3 since the Hb-A_2 peptide δ14 overlaps peptide α9, which also contains methionine.[7, 15] The difference is noticed only in fingerprints run at different pH.[7]

A second difference between Hb-A_2 and Hb-Lepore$_{Boston}$ tryptic digests was observed by running a one-dimensional ionophoresis at pH 6.4.[18] A peptide, called A_2a, has been observed in ionograms of the tryptic digest of isolated δ chain.[18] This peptide was not present in Hb-Lepore$_{Boston}$[19] and in Hb-Lepore$_{Hollandia}$.[6]

A few peptides, which were separated in a sufficiently pure form by fingerprinting (indicated by arrows in Fig. 1*a*), were eluted and analyzed. The results of the amino acid analysis are shown in Table 1. The peptides analyzed have amino acid composition identical to homologous peptides of Hb-A_2.[7, 18]

Discussion.—In the following discussion, the position of the tryptic peptides along the amino acid sequence of the δ chain of Hb-A_2 and the non-α chain of Hb-

TABLE 1

AMINO ACID ANALYSIS OF Hb-LEPORE$_{Boston}$ PEPTIDES

Amino Acid†	Peptide* x_B^1	x_B^2	x_B^4	x_B^5	x_B^9
Lysine	1.02	1.01		1.13	1.11
Arginine			1.10		
Histidine	0.98				1.22
Glycine		0.87		1.97	2.07
Alanine		1.81		1.24	2.07
Valine	0.95	1.13	1.94	1.30	1.10
Serine				§3.57	§4.38
Leucine	1.25	1.25	2.33		
Threonine	0.82	0.74	0.91		
Methionine				0.76	
Glutamic acid	1.98		1.03	1.28	
Phenylalanine				2.02	1.12
Aspartic acid		1.19		2.64	2.93
Tyrosine			0.69		
Proline‡	++		++	++	

* The identification of the peptides is based on their positions in fingerprints, identical with those of β or δ chain peptides.[7, 15] The peptides are designated as indicated.[11] The x_B peptide chain is the non-α chain of Hb-Lepore$_{Boston}$.

† The molar ratio of the amino acids is given.

‡ Proline cannot be estimated quantitatively by this method of analysis; the presence of this amino acid, judged by the visual examination of the ionograms, is indicated by ++.

§ Serine and leucine, when present in high relative amounts, are incompletely resolved. The values reported represent the sum of the leucine plus serine molar ratios.

Lepore$_{Boston}$ (here called x_B chain) is arbitrarily assigned according to the principle of homology.[20] This principle postulates that peptides having identical or very similar amino acid sequence or composition occupy corresponding positions along homologous peptide chains, which have differentiated from a common ancestor chain either in the course of evolution in different species or following gene duplications and independent evolution within the same species.[9, 20]

Following this principle of homology it has been possible to predict the amino acid sequence of animal hemoglobins from the composition of their tryptic peptides[21, 22] and the sequence of the human hemoglobin chains.[23, 24]

Thus, it may be justified to assume, until the complete amino acid sequence of the δ and x_B chains is established, that the order of the tryptic peptides in these chains corresponds to the order in the β chain. The similarities in the amino acid composition of the tryptic peptides are striking.[6, 7, 10, 18] Moreover, the amino acid sequences so far determined within δ chain peptides are identical to those of β chain peptides, except for a few amino acid substitutions.[8, 18]

The β chain differs from the δ chain in four positions: 9, β serine $\rightarrow$ δ threonine; 12, β threonine $\rightarrow$ δ asparagine; 22, β glutamic acid $\rightarrow$ δ alanine; and 50, β threonine $\rightarrow$ δ serine.[7, 18] Preliminary evidence has been obtained by Stretton for four more differences.[8] One of these has been localized in residue 126, β valine $\rightarrow$ δ methionine.[8] Additional differences are revealed by the presence of peptide A_{2a} in the tryptic digest of the δ chain.[18] This peptide has the sequence[25] *Asp*(NH_2). *Phe.Gly.Lys*.[8, 18] The only sequence homologous to this one in the β chain is the C-terminal ... *His.His.Phe.Gly.Lys* sequence of peptide β12 (from residue 116 to 120).[23] This has suggested[8] two further amino acid differences between the β and the δ chains: 117, β histidine $\rightarrow$ δ asparagine and 116, β histidine $\rightarrow$ δ arginine or lysine, to account for the splitting of peptide A_{2a} (provisionally designated δ13) by trypsin. Stretton[8] has obtained an indication that residue 116 of the δ chain is arginine. The δ chain has, therefore, one more tryptic peptide than the β chain.

No difference in composition between other β and δ chain peptides has been observed.[8, 18] It so happens that $\beta 1$ is identical in composition to $\delta 1$, $\beta 4$ to $\delta 4$, $\beta 6$ to $\delta 6$, $\beta 7$ to $\delta 7$, $\beta 8$ to $\delta 8$, $\beta 9$ to $\delta 9$, $\beta 14$ to $\delta 15$, and $\beta 15$ to $\delta 16$ (because of the presence of the extra peptide $\delta 13$ in the δ chain).

The fingerprints of Hb-Lepore$_{\text{Boston}}$ and the amino acid analyses of the peptides isolated indicate that the x_B chain of this hemoglobin is made up of peptides characteristic of both the β and the δ chains. The N-terminal portion of the x_B chain is made up of δ-like peptides. $x_B 2$, $x_B 3$, and $x_B 5$ are identical in amino acid composition to the corresponding δ chain peptides[10] and $x_B 1$ and $x_B 4$ could not be distinguished from $\delta 1$ and $\delta 4$ (= $\beta 1$ and $\beta 4$ respectively). The C-terminal portion of the x_B chain is made up of β-like peptides (see Fig. 3). $x_B 13$ is identical in the fingerprinting analysis and in the specific staining reaction to $\beta 13$; $x_B 14$ and $x_B 15$ could not be distinguished from $\beta 14$ and $\beta 15$ (= $\delta 15$ and $\delta 16$ respectively). Moreover, the absence in Hb-Lepore$_{\text{Boston}}$ of peptide $\delta 13$ (A_2a), suggests that the x_B chain peptide corresponding to $\beta 12$ is similar if not identical to this β chain peptide.

Similarly, the non-α chain of Hb-Lepore$_{\text{Hollandia}}$ (here indicated as x_H) seems to have a δ-like N-terminal portion and a β-like C-terminal portion. The N-terminal δ-like portion of the x_H chain, however, seems to be shorter than the δ-like portion of the x_B chain. $x_H 5$ has indeed the composition of $\beta 5$.[6] Thus, the x_H chain may be δ-like from residue 1 to 40 approximately and β-like in the remaining portion of the chain (see Fig. 3).

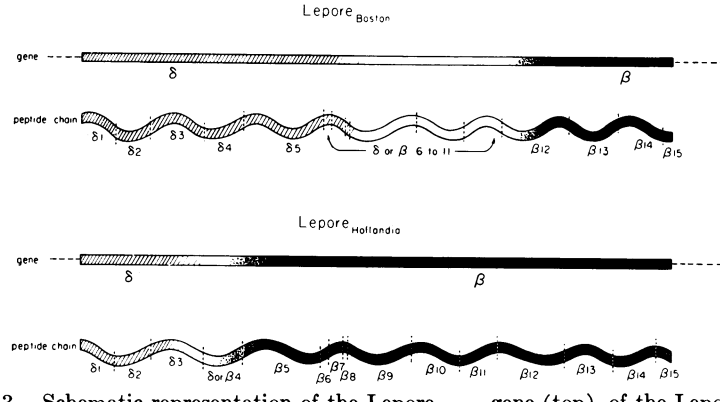

FIG. 3.—Schematic representation of the Lepore$_{\text{Boston}}$ gene (top), of the Lepore $_{\text{Hollandia}}$ gene (bottom), and of the corresponding peptide chains. The shaded areas indicate that part of the Lepore genes which appears to be derived from the δ gene and that part of the corresponding peptide chains which is δ-like. The area in solid color indicates that part of the Lepore genes which appears to be derived from the β gene and that part of the corresponding peptide chains which is β-like. The empty areas between shaded and solid color areas indicate those portions of the genes and of the peptide chains that may be derived either from β or from δ. The postulated joining of part of the β gene to part of the δ gene has occurred presumably in this area. The dotted lines along the peptide chains indicate schematically the peptide bonds which may be split by trypsin. The resulting peptides are indicated below and are numbered consecutively starting from the N-terminus. The symbols β or δ which precede the peptide numbers indicate whether the corresponding peptides are β-like or δ-like.

Conclusions.—There is now a large body of evidence which suggests that the structural gene for a protein determines its primary amino acid sequence.[26] The structural genes seem to be separated by regions, which establish boundaries between them and allow their function to be expressed individually.[27]

In the well-studied region r_{II} of phage *T4* there are two adjacent cistrons, called A and B.[28] In the mutant *r1589* of *T4*, there is a deletion of the region separating the A from the B cistron and the two cistrons appear to be joined, with loss of the A function but with preservation of the B function.[28] The insertion of certain deletions in the A cistron of *r1589* prevents the B cistron from functioning.[27] This suggests that in *r1589* a protein may be produced "which consists of part of the protein from the A cistron and part of the B cistron, joined together in the same polypeptide chain." [27]

The x_B peptide chain of Hb-Lepore$_{Boston}$ seems to consist of part of the β chain and part of the δ chain joined together. That part of the δ gene corresponding to the N-terminal portion of the δ chain is presumably joined to that part of the β gene which determines the C-terminal portion of the β chain. The fact that the β and δ genes are linked[5] suggests that the joining may be due to a deletion of parts of the β and δ genes and of the region separating them.

If this postulated joining of part of the β gene with part of the δ gene were due to a random deletion, it would be difficult to understand why the peptide chain made by the hybrid β-δ gene would have the same length as the β or δ chain. The evidence so far suggests that there is no amino acid sequence in the x_B chain that is not also present either in the β or in the δ chain. One would not expect to find such a situation if the corresponding genes were randomly joined by a deletion. A specific mechanism that accounts for the joining of the β to the δ gene in such an exactly complementary way is a nonhomologous crossing-over between corresponding points of the β and the δ genes, resulting in the formation of unequal products (see Fig. 4).

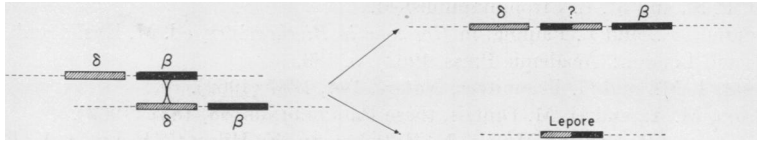

FIG. 4.—Schematic representation of the postulated nonhomologous crossing-over leading to the formation of the Lepore genes. The shaded area indicates the nucleotide sequence corresponding to the δ gene. The nonhomologous crossing-over is supposed to have occurred in the areas shown as empty in Figure 3.

Unequal crossing-over is known to involve frequently chromosomal duplications, such as the *bar* duplication of *Drosophila melanogaster*,[30] leading to recombinants with deletion of the duplication (*wild type*) or insertion of one extra *bar* duplication (*double-bar*). It has been suggested that the δ gene originated from a duplication of the β gene, because the primary sequence of the δ chain is extremely similar to the primary sequence of the β chain.[9] The similarity of the β and δ genes may increase the probability of nonhomologous crossing-overs. It is suggestive that Hb-Lepore$_{Boston}$ and Hb-Lepore$_{Hollandia}$ have been observed in different populations.[1-2] The corresponding genes may have been formed by different nonhomologous crossing-overs, which have involved different points of the β and δ genes (see Figs. 3 and 4).

Note added in proof: Unequal crossing-over has recently been invoked as the interpretation of variations in the length of the α chain of human haptoglobin.[31]

I wish to thank Vernon M. Ingram for help, advice, and criticism during this work and in the preparation of this manuscript, Park S. Gerald and Mary L. Efron for providing a sample of Hb-

Lepore$_{Boston}$ and for discussions and suggestions, and Antony O. W. Stretton for making available unpublished results. This work has been supported by research grants from the Medical Foundation, Inc., Boston, and from the National Science Foundation.

* Present address: International Laboratory for Genetics and Biophysics, Via Claudio 1, Naples, Italy.

[1] Gerald, P. S., and L. K. Diamond, *Blood*, **13**, 835 (1958).

[2] Fessas, P., G. Stamatoyannopoulos, and A. Karaklis, *Blood*, **9**, 1 (1962).

[3] Silvestroni, E., and I. Bianco, *Policlinico*, **65**, 203 (1958).

[4] Neeb, H., J. L. Beiboer, J. H. P. Jonxis, J. A. Kaars-Sijpesteijn, and C. J. Muller, *Trop. Geogr. Med.*, **13**, 207 (1961).

[5] Baglioni, C., in *Progresses in Molecular Genetics*, ed. H. Taylor (New York and London: Academic Press), vol. 1, in press.

[6] Barabas, J., and C. J. Muller, *Nature*, **194**, 931 (1962).

[7] Ingram, V. M., and A. O. W. Stretton, *Nature*, **190**, 1079 (1961).

[8] Stretton, A. O. W., unpublished.

[9] Ingram, V. M., *Nature*, **189**, 704 (1961).

[10] Gerald, P. S., M. L. Efron, and L. K. Diamond, *Am. J. Dis. Child.*, **102**, 514 (1961).

[11] The peptides obtainable by tryptic digestion are indicated as follows: the symbol of the peptide chain precedes arabic numerals indicating the order in which the peptides occur in the chain starting from the N-terminus.[15] This system of nomenclature is similar to the one recently recommended.[17]

[12] Schnek, A. G., and W. A. Schroeder, *J. Am. Chem. Soc.*, **83**, 1472 (1961).

[13] Baglioni, C., *J. Biol. Chem.*, **237**, 69 (1962).

[14] Ingram, V. M., *Biochim. Biophys. Acta*, **28**, 539 (1958).

[15] Baglioni, C., *Biochim. Biophys. Acta*, **48**, 392 (1961).

[16] Hill, R. L., R. T. Swenson, and H. C. Schwartz, *Blood*, **19**, 573 (1962).

[17] Gerald, P. S., and V. M. Ingram, *J. Biol. Chem.*, **236**, 2155 (1961).

[18] Ingram, V. M., and A. O. W. Stretton, *Biochim. Biophys. Acta*, **62**, 456 (1962) and in press.

[19] Gerald, P. S., and M. L. Efron, unpublished.

[20] Zuckerkandl, E., and L. Pauling, in *Horizons in Biochemistry*, ed. M. Kasha and B. Pullman (New York and London: Academic Press, 1962), p. 189.

[21] Diamond, J. M., and G. Braunitzer, *Nature*, **194**, 1287 (1962).

[22] Naughton, M. A., and H. M. Dintzis, these PROCEEDINGS, **48**, 1822 (1962).

[23] Braunitzer, G., R. Gehring-Muller, N. Hillschmann, K. Hilse, G. Hobman, V. Rudloff, and B. Wittman-Liebold, *Hoppe Seyler's Z. physiol. Chem.*, **325**, 283 (1961).

[24] Konigsberg, W., G. Guidotti, and R. G. Hill, *J. Biol. Chem.*, **236**, PC55 (1961).

[25] The amino acids are indicated by the first three letters of their names. Asparagine is indicated as Asp(NH_2).

[26] Yanofsky, C., D. R. Helinski, and B. D. Maling, in *Cellular Regulatory Mechanisms*, Cold Spring Harbor Symposia on Quantitative Biology, vol. 26, (1961), p. 11.

[27] Crick, F. H. C., L. Barnett, S. Brenner, and R. J. Watts-Tobin, *Nature*, **192**, 1227 (1961).

[28] Champe, S. P., and S. Benzer, *J. Mol. Biol.*, **4**, 288 (1962).

[29] Kunkel, H. G., R. Ceppellini, V. Muller-Eberhardt, and J. Wolf, *J. Clin. Invest.*, **36**, 1615 (1957).

[30] Bridges, C. B., *Science*, **83**, 210 (1936).

[31] Smithies, O., G. E. Connell, and G. D. Dixon, *Nature*, **196**, 232 (1962).