THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 257, No. 2, Issue of January 25, pp. 651–658, 1982
Printed in U.S.A.

Minor Early Embryonic Chick Hemoglobin M

AMINO ACID SEQUENCES OF THE ε AND α^D CHAINS*

(Received for publication, June 3, 1981)

Barbara S. Chapman‡ and Leroy E. Hood§

From the Division of Biology, California Institute of Technology, Pasadena, California 91125

Allan J. Tobin¶

From the Department of Biology, University of California, Los Angeles, California 90024

Erythrocytes of the early chick embryo contain four hemoglobins, two major and two minor. In this paper, we present amino acid sequences for the β-like and α-like chains of HbM, the least abundant of the four early chick hemoglobins.

The complete amino acid sequence of the β-like chain of HbM is identical with that of the ε chain of HbE, the other minor early embryonic hemoglobin in the domestic chicken. Analysis of the α-like chain of HbM (92 of 141 residues) reveals a globin sequence closely related to the minor adult α^D chain. Comparison of our sequence data with the nucleotide sequence of the α^D globin gene suggests that a single gene encodes both the embryonic and adult α^D globin polypeptides. We discuss the structure, possible function, and evolution of the HbM globin chains.

Early embryos of chickens have four hemoglobins in their circulating erythrocytes. We have previously determined the amino acid sequences of the α-like and β-like globins of early hemoglobins P, P′, and E (Chapman *et al.*, 1980; Chapman *et al.*, 1981 and 1982). Hemoglobin M, which accounts for about 18% of the hemolysate from 5-day chick embryos, is the least abundant early embryonic hemoglobin. On the basis of electrophoretic mobility, immunological reactivity, and peptide map analysis, Brown and Ingram (1974) predicted that the β-like globin chain of HbM would be an ε chain similar to the β-like chain of the minor early embryonic hemoglobin E. We report here that the β-like chain of HbM is identical with the ε globin of HbE.

Genetic evidence supports the assignment of α^D as the α-like chain of HbM. Although hemoglobin variants are rare in the domestic chicken, there have been four reports of electrophoretic polymorphisms in the minor adult hemoglobin D (Washburn, 1968; Callegarini *et al.*, 1969; Kimura and Yokoyama, 1973; Keane *et al.*, 1974). It has been demonstrated that two variants behave as single co-dominant alleles that change the electrophoretic mobility of both HbD and HbM (Washburn, 1968; Keane *et al.*, 1974). The alteration in HbD can be assigned to the α^D chain, since HbA (the major adult hemoglobin) shares the adult β chain with HbD (Vandecasserie *et al.*, 1975), and HbA remains unaffected by this mutation. Since these two presumably independent α^D chain mutations affect both HbD and HbM, it follows that the same α^D globin is present in early embryonic HbM and adult HbD. However, amino acid sequence analysis of the α-like chain from HbM reported here differs at six amino acid positions from the adult α^D sequence (Takei *et al.*, 1975). Although this interesting result raises the possibility that the α^D globin may be switched during chick embryonic development, we favor alternative explanations for this discrepancy. We will discuss the sequences of the HbM globins in relation to their function and evolution.

EXPERIMENTAL PROCEDURES AND RESULTS[1]

DISCUSSION

Relationship of the Chicken ε Globin to the Mammalian ε Globins—The complete amino acid sequence of the ε chain from HbM (Fig. 1) is identical with the ε chain of HbE (Chapman *et al.*, 1982). This establishes that in the chicken there are two early embryonic β-like globins, ε and ρ, which are analogous to the ε and γ chains of mammals. The ε globin in domestic chickens is not directly related to the ε globins of mammals. The evolutionary tree shown in Fig. 2 illustrates the relationships among several ε globins and adult β globins. The rabbit β3 (Hardison *et al.*, 1979),[2] mouse εy (Jahn *et al.*, 1980; Gilman, 1976), and human ε (Baralle *et al.*, 1980) globin genes have probably arisen from an ancestral β-like globin gene. The chicken ε globin gene has an independent origin, having apparently been formed by a duplication within the avian β-like gene family.

Comparison of the chick ε globin with three other chicken β-like globins indicates that the ε globin gene was probably formed by duplication of the ρ globin gene. At the 79 positions in β, ρ, ε, and β^H for which there are comparable sequence data, ε chains differ from ρ chains at only four positions, but differ from the β and β^H chains at eight and nine positions, respectively (Matsuda *et al.*, 1973; Chapman *et al.*, 1981; Moss and Hamilton, 1974; Dolan *et al.*, 1981).

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of National Research Service Award HL05553. Present address, Department of Biochemistry, University of California, Berkeley, CA 94720. To whom inquiries may be addressed.

§ Recipient of Grant GM-06965 from the National Institutes of Health.

¶ Recipient of National Science Foundation Grants PCM 76-02859 and PCM 78-20767.

[1] Portions of this paper (including "Experimental Procedures," "Results," Figs. 1–9, and Table I) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 81M-1314, cite authors, and include a check or money order for $5.20 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

[2] R. Hardison, personal communication.

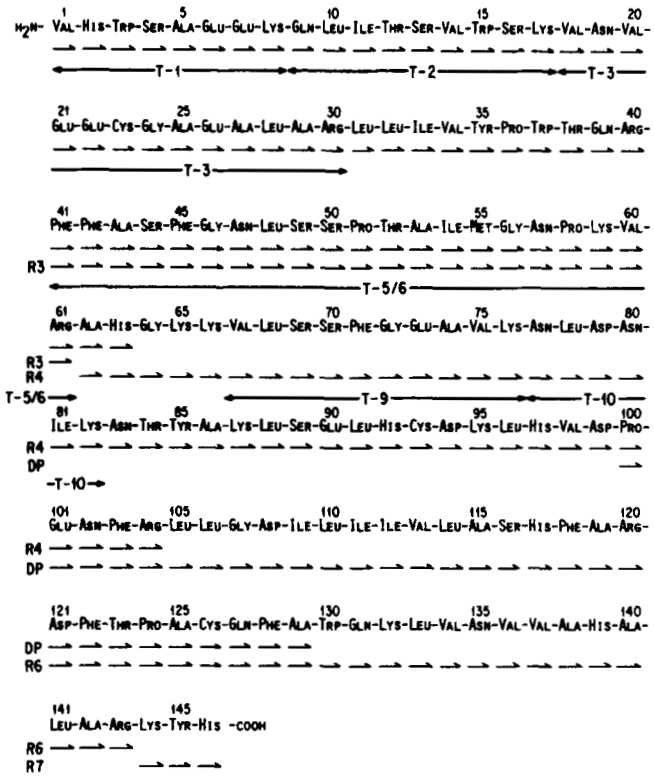


FIG. 1. **The complete amino acid sequence of the ϵ chain of minor early embryonic HbM.** →, automated Edman degradation; ←*T*→, tryptic peptides; *R*, arginine cleavage fragment; *DP*, aspartic acid-proline cleavage fragment. A peptide overlapping residues 143 and 144 was not isolated and sequenced. The COOH-terminal tripeptide was placed by homology.

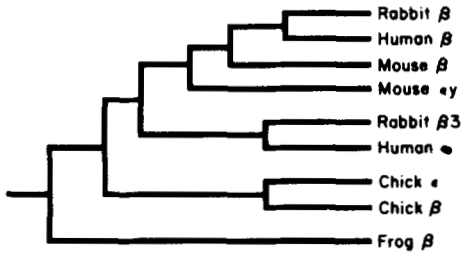


FIG. 2. **Evolutionary tree of ϵ globin chains.** Sequences were obtained from Dayhoff (1976), Hardison *et al.* (1979),[2] Gilman (1976), Jahn *et al.* (1980), Konkel *et al.* (1979), Watt *et al.* (1980), and Baralle *et al.* (1980). Computer analysis used the method of Farris (1970) and a matrix of differences.

Unusual Distribution of Amino Acid Substitutions—When the ϵ chain is aligned with the adult β chain, there are 18 amino acid replacements, randomly distributed. Compared to the ρ chain and its allele ρ′, there are 12 and 13 substitutions. However, three-quarters of these are clustered in the COOH-terminal third of the ϵ chain (Table I). At least three explanations for this uneven arrangement are worth discussing. 1) The COOH termini of the ϵ and ρ globins have undergone accelerated evolution to adapt them for different roles in the early embryo. 2) Avian β globins normally evolve at a much slower rate than mammalian β globins, but the COOH termini of the early embryonic forms have been released from adult constraints and are evolving at the rate of mammalian β globins. 3) The NH_2 termini of the ϵ and ρ globins are similar because the first two coding blocks of their genes have recently been corrected against each other.

Accelerated Evolution and the Roles of ϵ and ρ Globins—The α-like and β-like chains of hemoglobin can be divided into three functional regions. In the COOH-terminal third of the β-globin molecule, defined as residues 105–146, are many intersubunit contact residues ($\alpha_1\beta_1$ contacts). Organophosphate-binding and Bohr-effect residues are also found in this segment (Eaton, 1980). Having a complete set of embryonic and adult globin sequences from a single species provides an opportunity to test the hypothesis that "amino acid replacements signal the evolution of function at those positions" (Eaton, 1980).

The ϵ chains are very similar in structure to ρ chains, having fixed only 12 amino acid changes. Eight of these differences occur after amino acid 105, involving a minimum of 10 nucleotide substitutions. Oddly, the COOH-terminal third of the ϵ globin is more divergent from the adult β globin than is the ρ globin in this region (Table I). Consideration of functional constraints leads us to expect the opposite. We expect selective pressures to have modified $\alpha_1\beta_1$ contact positions clustered in the COOH-terminal region of the ρ chain because it associates with the highly divergent π and π′ chains to form the major early embryonic hemoglobins (Ladner *et al.*, 1977; Eaton, 1980; Brown and Ingram, 1974; Chapman *et al.*, 1980 and 1981). The π and π′ globins differ from one or the other of the adult α-like globins at 11 of 18 $\alpha_1\beta_1$ contact positions (Table II). The ϵ globin residues 105 to 146, on the other hand, should be like those of the adult β globin, since these chains combine with adult-type α chains. If adaptive selection is driving changes in the COOH terminus of the ϵ globin, it is not clear how or why.

TABLE I

Amino acid substitutions in the carboxyl-terminal regions of the chicken ϵ, ρ′, ρ, and β globin chains

Amino acid sequences were obtained from Roninson and Ingram (1981), Chapman *et al.* (1981), and Matsuda *et al.* (1973). Locations of the residues in the three-dimensional structure of hemoglobin were determined by Arnone and Perutz (1974) and Ladner *et al.* (1977).

Position	Helix	ϵ	ρ′	ρ	β	Function	Interacts with α
108	G10	Asp	Asn	Asn	Asp	α1β1 contact	103
116	G18	Ser	Ala	Ala	Ala	α1β1 contact	117, 114, 110
119	GH2	Ala	Thr	Thr	Ser	α1β1 contact	111
120	GH3	Arg	Lys	Lys	Lys		
125	H3	Ala	Thr	Glu	Glu		
128	H6	Phe	Ala	Ala	Ala	α1β1 contact	34, 35
129	H7	Ala	Val	Ala	Ala		
135	H13	Asn	Ser	Ser	Arg	IPP contact	
139	H17	His	His	Lys	His	IPP contact	
143	H21	Arg	Tyr	Arg	Arg	DPG contact	

TABLE II

Substitutions in αβ contacts in chicken α-like globins

Sources of amino acid sequence data: the π chain (Chapman *et al.*, 1980), the α^E chain (Chapman *et al.*, 1982), and the α^A chain (Matsuda *et al.*, 1971). Three dimensional positions were determined by Ladner *et al.* (1977).

Function	Position	Helix	α^D	π	α^E	α^A
α1β2 contact	38	C3	Gln	Gln	Pro	Thr
α1β1 contact	34	B15	Thr	Ala	Thr	Ile
	35	B16	Thr	Ser	Thr	Gly
	36	C1	Tyr	Tyr	Tyr	Phe
	103	G10	Gln	His	Gln	Gln
	107	G14	Cys	Cys	Val	Val
	110	G17	Ala	Ala	Ala	Val
	111	G18	Val	Ala	Ile	Leu
	114	GH2	Gly	Pro	Pro	Pro
	115	GH3	Lys	Ser	Ala	Ala
	117	GH5	Tyr	Phe	Leu	Leu
	118	H1	Thr	Thr	Thr	Ala

Recently, nucleotide sequencing of an early embryonic cDNA recombinant clone has revealed an allele of the ρ globin that differs from the ρ globin by four amino acid substitutions. These are clustered in the carboxyl-terminal region (Table I; Roninson and Ingram, 1981). We have found quantities of the ρ' polypeptide in our ρ globin preparations (Chapman *et al.*, 1981), but only this ρ' allele has been identified in the chicken population sampled by Roninson and Ingram (1981). Of the four additional substitutions in the ρ' COOH terminus, one is a Tyr for Arg change at DPG-binding position 143, which could change the affinity of HbP and HbP' for ATP. Nevertheless, the ρ' globin seems not to be detrimental to developing chick embryos, and we conclude that the ρ and ρ' globins are interchangeable. These results are not consistent with the fixation of amino acid replacements by strong adaptive selection.

Adult Avian β Globins Evolve Slowly. Do the Embryonic Globins?—The avian adult β globins have evolved at a much slower rate than the globins of mammals or the α-like globins of birds (Oberthur *et al.*, 1980). Table III shows the number of amino acid substitutions accumulated in the α^A and β chains of chickens relative to other birds and a reptile. Rates of amino acid substitution can be calculated from the estimated divergence times of these orders (Prager and Wilson, 1980). The data show a 4-fold difference in evolutionary rate between the adult α and β globins of birds. The substitution rate of the avian α globins is comparable to that of the mammalian globins.

Oberthur *et al.* (1980) have suggested that the evolution of the adult bird β chains has been slowed by the constraints that binding of IPP[3] places on the tertiary structure of hemoglobin. Adult birds use IPP to regulate their hemoglobin oxygen affinity. IPP is a larger molecule than DPG, it binds with very high affinity to bird hemoglobin, and there are two additional pairs of basic residues in the hemoglobin central cavity (Arnone and Perutz, 1974; Rollema and Bauer, 1979). Both of these IPP-binding residues are in the COOH-terminal segment of chicken β globin. There is, however, no detectable IPP in early embryonic chick red cells. Instead, ATP is the primary organophosphate regulator in developing chickens and ducks (Bartlett and Borgese, 1976). Without the constraints of IPP binding, the COOH termini of the early embryonic ϵ and ρ globins may be free to evolve at the same rate as the globins of mammals. This might explain the large number of differences accumulated in the COOH termini of the ϵ and ρ chains.

Using the amino acid replacements in the COOH-terminal segments and the evolutionary rate for mammalian β globins, we can compute the divergence time of the ϵ and ρ globins to be roughly 75 million years (a 20% divergence at 3.8 myrs/1% difference). Since biochemical evidence suggests that the various lineages of modern birds originated about 80 million years ago, both ϵ and ρ globins could be present in all bird species (Prager and Wilson, 1980).

The Effect of Gene Correction on Interpretations of ϵ Globin Structure—One explanation for the uneven distribution of amino acid substitutions in the ϵ and ρ chains entails a gene conversion event. Nonreciprocal DNA sequence conversion has been well documented as the mechanism for maintaining similarity between the human γ globin genes (Slightom *et al.*, 1980). Although the ϵ and ρ genes are known to lie at opposite ends of the chicken β globin gene cluster, the β-like pseudogene of the Brown lemur is a precedent for hybrid gene formation involving distal sequences (Barrie *et al.*, 1981). Another possible example of gene conversion within the chicken β globin gene family is suggested by the remarkable ρ and ρ' alleles. The distribution and nature of amino acid replacements suggest that these alleles could have been formed by a recent gene conversion event involving the third coding block of the β^H gene. Frequent deletions of the third exon of the β^H gene in recombinant DNA clones may be evidence for a recombination "hot spot" in this region (Dolan *et al.*, 1981; Villeponteau and Martinson, 1981).

[3] The abbreviations used are: IPP, inositol pentaphosphate; CAM, carboxyamidomethylated; PTH, phenylthiohydantoin; HPLC, high pressure liquid chromatography.

TABLE III

Different evolutionary rates in adult α and β globin chains of birds

The number of amino acid substitutions between the chicken globins and the globins of the other species was calculated from published sequence data: chicken (*Gallus gallus*) β globin (Matsuda *et al.*, 1973), duck (*Anas platyrhynchos*) β globin (Hampe *et al.*, 1981), goose (*Anser anser*) and ostrich (*Struthio camelus*) β chains (Oberthur *et al.*, 1980), goose α chain (Debouverie, 1975), ostrich α chain (Oberthur *et al.*, 1980), and viper (*Vipera aspis*) α chain (Dayhoff, 1976). Divergence times for these species were obtained from Prager and Wilson (1980) and are expressed in Myrs.

	Duck	Goose	Ostrich	Viper	Myr/substitution
Chicken β chain	3	6	7	...	14
Chicken α^A chain	...	17	28	57	3.4
Divergence time	60	60	80	225	

The α-like Globin of HbM—We have determined 92 amino acid residues of the α-like globin from HbM. In Fig. 3, this sequence is shown in alignment with the adult α^D globin sequence (Takei *et al.*, 1975) and the polypeptide encoded by the α^D globin gene (Dodgson *et al.*, 1981). There are six positions at which the early embryonic α^D chain appears to differ from the adult form of α^D (22, 38, 53, 54, 107, and 111).

The observed differences can be reconciled as technical errors and a possible allelic polymorphism. Four of these differences involve assignment of the acid or amide form of glutamic or aspartic acid (22 Gln to Glu, 38 Glu to Gln, 53 Asn to Asp, and 54 Glu to Gln). Although many glutamine and asparagine residues become partially deamidated during protein preparation and sequenator degradation, HPLC data for these residues in our analyses are unequivocal. The apparent change of Cys for Gln at position 107 might also be explained as a technical artifact. Since cysteine residues are difficult to detect and glutamine is the previous residue, glutamine might have been assigned to position 107 in the absence of a cysteine signal. Similarly, the Ala for Val substitution at position 111 could result if residual Ala from the previous cycle were 8–10-fold greater than the Val signal (cycle 17, Fig. 6, Miniprint). The nucleotide sequence of Dodgson *et al.* (1981) confirms our residue assignments at five of the six positions and shows no differences from the adult α^D sequence through residues 64–93 and 121–141, which we did not analyze. However, the DNA sequence differs from both polypeptides at residues 109 and 110, and identifies Val at position 111. The partial α^D sequence determined by Paul *et al.* (1974) shows glutamic acid at position 22, as does our sequence, but their data do not extend to the remaining residues in question. The amino acid sequence difference at position 111 could represent allelic polymorphism. The likelihood of a previously undetected genetic polymorphism in the α^D globin is enhanced by the

	1	5	10	15	20
EMBRYO α^D	MET-LEU-THR-ALA-GLU-ASP-LYS-LYS-LEU-ILE-GLN-GLN-ALA-TRP-GLU-LYS-ALA-ALA-SER-HIS-				
ADULT α^D	----------				
α^D GENE	----------				
	21	25	30	35	40
EMBRYO α^D	GLN-GLU-GLU-PHE-GLY-ALA-GLU-ALA-LEU-THR-ARG-MET-PHE-THR-THR-TYR-PRO-GLN-THR-LYS-				
ADULT α^D	----GLN----------				----GLU---------
α^D GENE	----------				
	41	45	50	55	60
EMBRYO α^D	THR-TYR-PHE-PRO-HIS-PHE-ASP-LEU-SER-PRO-GLY-SER-ASP-GLN-VAL-ARG-GLY-HIS-GLY-LYS-				
ADULT α^D	----------			ASN-GLU----------	
α^D GENE	----------				
	61	65	70	75	80
EMBRYO α^D	LYS-VAL-LEU-				
ADULT α^D	----------GLY-ALA-LEU-GLY-ASN-ALA-VAL-LYS-ASN-VAL-ASP-ASN-LEU-SER-GLN-ALA-MET-				
α^D GENE	---------- ----------				
	81	85	90	95	100
EMBRYO α^D				(ASP)PRO-VAL-ASN-PHE-LYS-LEU-	
ADULT α^D	ALA-GLU-LEU-SER-ASN-LEU-HIS-ALA-TYR-ASN-LEU-ARG-VAL-ASP-PRO-VAL-ASN-PHE-LYS-LEU-				
α^D GENE	----------				
	101	105	110	115	120
EMBRYO α^D	LEU-SER-GLN-CYS-ILE-GLN-CYS-VAL-LEU-ALA-ALA-HIS-MET-GLY-LYS-ASP-TYR-THR-PRO-GLU-				
ADULT α^D	----------GLN----------VAL----------				
α^D GENE	----------ARG-LEU-VAL----------				
	121	125	130	135	140
EMBRYO α^D	VAL-HIS-				
ADULT α^D	--------ALA-ALA-PHE-ASP-LYS-PHE-LEU-SER-ALA-VAL-SER-ALA-VAL-LEU-ALA-GLU-LYS-TYR-ARG				
α^D GENE	----------				

FIG. 3. **Partial amino acid sequence of the α-like chain of HbM.** This sequence was prepared from automated Edman degradation of the α globin together with the ε globin. A small cyanogen bromide cleavage fragment was analyzed by sequenator to determine residues 114–122 (data not shown). For comparison, the amino acid sequence of adult α^D globin (Takei *et al.*, 1975) and the sequence encoded by the α^D globin gene (Dodgson *et al.*, 1981) are shown.

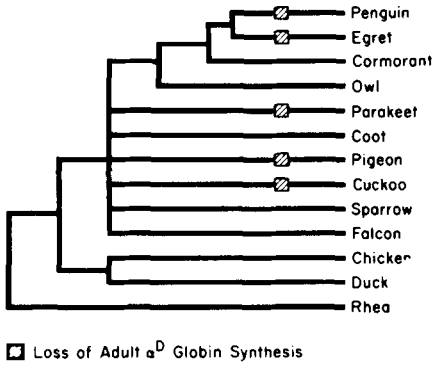


FIG. 4. **Evolutionary tree of some birds whose hemoglobin type has been determined by gel electrophoresis.** Hemoglobin analyses were obtained from Saha and Ghosh (1965) and Paul *et al.* (1978). The phylogenetic tree is adapted from Prager and Wilson (1980). Loss of adult α^D globin gene expression is inferred from the presence of a single adult hemoglobin having an α^A-type globin subunit (Vandecasserie *et al.*, 1973).

existence of several electrophoretic α^D globin variants in domestic chickens (Washburn, 1968; Callegarini *et al.*, 1969; Kimura and Yokoyama, 1973; Keane *et al.*, 1974). Classical genetic analyses indicate that adult and embryonic α^D chains are the products of a single gene (Washburn, 1968; Keane *et al.*, 1974). Only one α^D globin gene has been identified in recombinant chicken chromosomal DNA (Engel and Dodgson, 1980). In sum, the available evidence indicates that a single α^D globin gene encodes the α-like chains of HbM and HbD, and there may be allelic forms in domestic chickens.

What Is the Function of the α^D Globin?—There is conflicting evidence regarding the physiological necessity for α^D globin during avian development. Keane *et al.* (1974) observed fewer homozygotes carrying the $\alpha^{D_{Davis}}$ mutation than expected and concluded that reduced embryonic fitness might be associated with this mutant form of α^D. On the other hand, many bird species do not have HbD in adult circulation (Saha and Ghosh, 1965; Paul *et al.*, 1978). Penguins, egrets, pigeons, cuckoos, and parakeets have all lost the ability to produce α^D chains in adult life (Fig. 4). Since different orders of birds are represented, loss of adult α^D globin gene expression has occurred at least twice (Fig. 4; Prager and Wilson, 1980). Moreover, mammals express no globin equivalent to α^D, suggesting that the gene may have been lost after divergence of birds and mammals about 350 million years ago. There seems to be no compelling evidence for or against the supposition that production of α^D globin is essential.

Some of the structural and functional characteristics of α^D globin are consistent with adaptation to embryonic physiology. Functional studies have revealed that isolated embryonic hemoglobins exhibit higher oxygen affinities than adult hemoglobins (Bauer *et al.*, 1975; Cirotto and Geraci, 1975; Tuchinda *et al.*, 1975; Isaacks *et al.*, 1976; Borgese and Nagel, 1977; Jelkmann and Bauer, 1977). The presence of α^D subunits raises the oxygen affinity of the minor adult HbD relative to HbA (Vandecasserie *et al.*, 1971; Isaacks *et al.*, 1976). Although there are 65 amino acid differences between α^D and α^A globins (Takei *et al.*, 1975), position 38 is the only predicted $\alpha_1\beta_2$ contact position at which α^A and α^D globins differ (based on the three-dimensional structure of Ladner *et al.*, 1977) (Table II). In the early embryonic π and π' globins, a glutamine substitution for the usual threonine residue at this position may contribute substantially to their elevated oxygen affinity. At position 63 of the α^D globin, where Leu replaces Ala, exchange of a large aliphatic residue for the small, polar Ala residue may result in elevated oxygen affinity by perturbing the tertiary structure of the α globin (Oberthur *et al.*, 1980). We have found unusual amino acid substitutions at positions 38 and 63 in all four early embryonic α-like globins of the chicken (Table II).

Amino acid sequence analysis of early embryonic globins has provided information useful in understanding the structure, function, and evolution of developmentally regulated hemoglobin chains (Chapman, 1981). Protein sequence data have been used in the identification of the genes encoding

these polypeptides and can predict some aspects of gene arrangement and evolution (Hood *et al.*, 1975; Dolan *et al.*, 1981; Dodgson *et al.*, 1981). We believe that this work will serve as a foundation for analysis of the α and β globin gene families in birds.

Acknowledgments—We thank Dr. Allan Wilson for sharing his stock of bird lore, and for giving us comments that greatly improved this manuscript. We also thank Sandra Martin for useful discussions. Margaret Kowalszyk's careful illustrations are greatly appreciated. We are grateful to Drs. J. D. Engel and J. B. Dodgson for many lively discussions and for sharing with us their unpublished manuscripts.

REFERENCES

Arnone, A., and Perutz, M. F. (1974) *Nature* **249,** 34–36
Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980) *Cell* **21,** 621–626
Barrie, P. A., Jeffreys, A. J., and Scott, A. F. (1981) *J. Mol. Biol.* **149,** 319–336
Bartlett, G. R., and Borgese, T. A. (1976) *Comp. Biochem. Physiol. A Comp. Physiol.* **55,** 207–210
Bauer, C., Tamm, R., Petschow, D., Bartels, R., and Bartels, H. (1975) *Nature* **257,** 333–334
Borgese, T., and Nagel, R. (1977) *Comp. Biochem. Physiol. A Comp. Physiol.* **56,** 539–543
Brown, J., and Ingram, V. (1974) *J. Biol. Chem.* **249,** 3960–3972
Callegarini, C., Baragellesi, A., and Conconi, F. (1969) *Experientia* **25,** 537–538
Chapman, B. S. (1981) in *Hemoglobins in Development and Differentiation* (Stamatoyannopoulos, G., and Nienhuis, A., eds) pp. 193–201, Alan R. Liss, Inc., New York
Chapman, B. S., and Tobin, A. J. (1979) *Dev. Biol.* **69,** 375–387
Chapman, B. S., Tobin, A. J., and Hood, L. E. (1980) *J. Biol. Chem.* **255,** 9051–9059
Chapman, B. S., Tobin, A. J., and Hood, L. E. (1981) *J. Biol. Chem.* **256,** 5524–5531
Chapman, B. S., Hood, L. E., and Tobin, A. J. (1982) *J. Biol. Chem.* **257,** 643–650
Cirotto, C., and Geraci, G. (1975) *Comp. Biochem. Physiol. A Comp. Physiol.* **51,** 159–163
Dayhoff, M. O. (1976) *Atlas of Protein Sequence and Structure,* National Biomedical Research Foundation, Silver Spring, MD
Debouverie, D. (1975) *Biochimie* **57,** 569–578
Dodgson, J. B., McCune, K. C., Rusling, D. J., Krust, A., and Engel, J. D. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 5998–6002
Dolan, M., Sugarman, B. J., Dodgson, J. B., and Engel, J. D. (1981) *Cell* **24,** 669–677
Eaton, W. (1980) *Nature* **284,** 183–185
Engel, J. D., and Dodgson, J. B. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 2596–2600
Farris, J. S. (1970) *Syst. Zool.* **19,** 83–92
Gilman, J. G. (1976) *Biochem. J.* **155,** 231–241
Hampe, A., Therwath, A., Soriano, P., and Galibert, F. (1981) *Gene* **14,** 11–21
Hardison, R., Butler, E., Lacy, E., and Maniatis, T. (1979) *Cell* **18,** 1285–1297
Hood, L., Campbell, J., and Elgin, S. (1975) *Annu. Rev. Genet.* **9,** 305–353
Hunkapiller, M., and Hood, L. (1978) *Biochemistry* **17,** 2124–2133
Isaacks, R., Harkness, D., Adler, J., and Goldman, P. (1976) *Arch. Biochem. Biophys.* **173,** 114–120
Jahn, C. L., Hutchinson, C. A., Phillips, S. J., Weaver, S., Haigwood, N. L., Voliva, C. F., and Edgell, M. H. (1980) *Cell* **21,** 159–168
Jelkmann, W., and Bauer, C. (1977) *Pfleugers Arch. Eur. J. Physiol.* **372,** 149–156
Johnson, N., Hunkapiller, M., and Hood, L. (1979) *Anal. Biochem.* **100,** 335–339
Keane, R., Abbott, U., Brown, J., and Ingram, V. (1974) *Dev. Biol.* **38,** 229–236
Kimura, M., and Yokoyama, Y. (1973) *Jpn. Poultry Sci.* **10,** 169–171
Konkel, D. A., Maizel, J. V., and Leder, P. (1979) *Cell* **18,** 865–873
Ladner, C., Heidner, E., and Perutz, M. (1977) *J. Mol. Biol.* **114,** 385–414
Matsuda, G., Maita, T., Mizumo, K., and Ota, H. (1973) *Nat. New Biol.* **244,** 244
McKean, D., Potter, M., and Hood, L. (1973) *Biochemistry* **12,** 749–759
Moss, B., and Hamilton, E. (1974) *Biochim. Biophys. Acta* **371,** 379–391
Oberthur, W., Voelter, W., and Braunitzer, G. (1980) *Hoppe-Seyler's Z. Physiol. Chem.* **361,** 967–975
Paul, C., Vandecasserie, C., Schnek, A., and Leonis, J. (1974) *Biochim. Biophys. Acta* **371,** 155–158
Paul, C., Vandecasserie, C., Fraboni, A., Depreter, J., Leonis, J., and Schnek, A. (1978) in *Molecular Interactions of Hemoglobin* (Labie, D., Poyart, F., and Rosa, J., eds) *INSERM Symposium,* **70,** 183–200
Piszkiewicz, D., Landon, M., and Smith, E. (1970) *Biochem. Biophys. Res. Commun.* **40,** 1173–1178
Prager, E., and Wilson, A. (1980) in *Proceedings of the XVII International Ornithological Congress* (Immelman, K., ed) pp. 1209–1214, Verlag Der Deutschen Ornithologen-Gesellschaft, Berlin
Rollema, H. S., and Bauer, C. (1979) *J. Biol. Chem.* **254,** 12038–12043
Roninson, I., and V. M. Ingram (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4782–4785
Rossi-Fanelli, A., Antonini, E., Caputo, A. (1958) *Biochim. Biophys. Acta* **30,** 608–615
Saha, A., and Ghosh, J. (1965) *Comp. Biochem. Physiol.* **15,** 217–235
Slightom, J., Blechl, A., and Smithies, O. (1980) *Cell* **21,** 627–638
Takei, H., Ota, Y., Wu, K., Kiyohara, T., and Matsuda, G. (1975) *J. Biochem. (Tokyo)* **77,** 1345–1347
Tuchinda, S., Nagai, K., and Lehmann, H. (1975) *FEBS Lett.* **49,** 390–391
Vandecasserie, C., Schnek, A., and Leonis, J. (1971) *Eur. J. Biochem.* **24,** 284–287
Vandecasserie, C., Paul, C., Schnek, A., and Leonis, J. (1973) *Comp. Biochem. Physiol. A Comp. Physiol.* **44,** 711–718
Vandecasserie, C., Paul, C., Schnek, A., and Leonis, J. (1975) *Biochimie* **57,** 843–844
Villeponteau, B., and Martinson, H. (1981) *Nucleic Acids Res.* **9,** 3731–3746
Washburn, K. (1968) *Poultry Sci.* **47,** 561–564
Watt, K. W. K., Maruyama, T., and Riggs, A. (1980) *J. Biol. Chem.* **255,** 3294–3301

Supplementary material to "Minor Early Embryonic Chick Hemoglobin M: Amino Acid Sequences of the ε and α^D Chains," B.S. Chapman, L.E. Hood, and A.J. Tobin.

EXPERIMENTAL PROCEDURES

Preparation of HbM globins

Hemolysates of peripheral blood from five day embryos were prepared and fractionated as previously described (Chapman and Tobin, 1979; Brown and Ingram, 1974). Approximately 18% of the hemoglobin recovered from ion exchange chromatography of the hemolysates on CM-Sephadex (Pharmacia) was found to elute between the HbP,P' peak and HbE (data not shown). HbM-containing fractions were pooled, concentrated against solid sucrose, dialyzed against distilled water, and stored in liquid nitrogen.

Globin was prepared by acid acetone precipitation (Rossi-Fanelli et al., 1958), then reduced and carboxyamidomethylated (McKean et al., 1973). The ε and α^D chains were separated by ion exchange chromatography on CM-52 cellulose (Whatman) in 8M urea (Fig. 1 and Moss and Hamilton, 1974). Pooled fractions were dialyzed exhaustively against distilled water and lyophilized.

Preparation and analysis of HbM globin fragments

Tryptic peptides of ε globin were separated by electrophoresis and chromatography on thin layer sheets (Brown and Ingram, 1974; Chapman et al., 1980). Amino acid compositions (Table I) were obtained using a Durrum D-500 amino acid analyzer modified for quantitative analysis of subnanomole amounts of material.

Succinylated ε globin was digested with trypsin to produce cleavage at arginine residues. The resulting fragments were separated by gel filtration on Biogel P-10 (Chapman et al., 1980). The separation is shown in Fig. 2. Mild acid hydrolysis was used to cleave ε and α^D globins at their aspartic acid-proline bonds (Piszkiewicz et al., 1970; Chapman et al., 1980). Large NH_2-terminal and small C-terminal mild acid cleavage fragments were separated by gel filtration on Biogel P-10 in 0.5 % formic acid (data not shown). Sequenator analyses were performed as previously described (Hunkapiller and Hood, 1978; Johnson et al., 1979; Chapman et al., 1980).

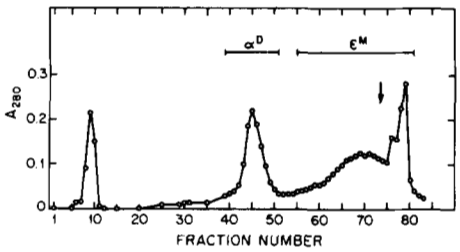


Figure 1. Separation of the α-like and β-like chains of HbM by ion exchange chromatography. A 1.6 x 10 cm column of Whatman CM-52 was equilibrated with 8 M urea containing 50 mM NaCl and adjustad to pH 4.3 with formic acid. Approximately 10 mg of CAM-HbM globin was dissolved in this starting buffer, applied to the colum and washed in with 50 ml of the same buffer. The globin chains were eluted with a 300 ml linear gradient prepared from 8 M urea containing 50 mM and 150 mM NaCl. 350 mM NaCl in 8 M urea was added as indicated by the arrow to finish elution of the ε chains. Five ml fractions were collected at a flow rate of 46 ml/hr. Pooled fractions are indicated by the bars.

Table I. Amino acid compositions of tryptic peptides from the ε globin of HbM [a]

	T-1 1-8	T-2 9-17	T-3 18-30	T-5/6 41-61	T-9 67-76	T-10 77-82
Asp			1.1 (1)	2.3 (2)		3.3 (3)
Thr		0.9 (1)		1.0 (1)		
Ser	1.0 (1)	1.8 (2)		2.8 (3)	2.2 (2)	
Glu	1.7 (2)	1.1 (1)	2.7 (3)		1.1 (1)	
Pro				2.0 (2)		
Gly			1.2 (1)	2.3 (2)	1.1 (1)	
Ala	1.1 (1)		3.2 (3)	2.3 (2)	1.2 (1)	
Cys			b (1)			
Val	1.0 (1)	1.2 (1)	1.5 (2)	0.9 (1)	1.6 (2)	
Met				0.9 (1)		
Ile		0.7 (1)		0.9 (1)		0.7 (1)
Leu		0.7 (1)	0.7 (1)	1.2 (1)	0.7 (1)	1.2 (1)
Tyr						
Phe				2.4 (3)	0.6 (1)	
His	0.6 (1)					
Lys	0.7 (1)	0.9 (1)		0.9 (1)	0.8 (1)	0.8 (1)
Arg			0.6 (1)	0.8 (1)		
Trp	ND[c] (1)	ND[c] (1)				
nmol yield	0.1	0.1	0.1	0.1	0.2	0.1
charge	0	+1	-2	+2	0	0
NH_2-term. residues	Val 8	Gln 9	Val 13	Phe 21	Val 10	Asn 6

[a] Tryptic peptides from 8 nmol of pure ε^M globin were separated by electrophoresis at pH 6.4 and chromatography on polyamide sheets (Brown and Ingram, 1974). Peptides identified by ninhydrin staining were hydrolyzed in 6 N HCl. Values shown are molar ratios (when greater than 0.2) of amino acids recovered from each peptide. Numbers in parentheses are integer values determined by sequenator analysis. Peptides are numbered from the NH_2 terminus, taking into account those expected but not isolated on the thin layer map.

[b] Cysteine residues were detected in the amino acid analysis but not quantified.

[c] Tryptophan was destroyed during hydrolysis

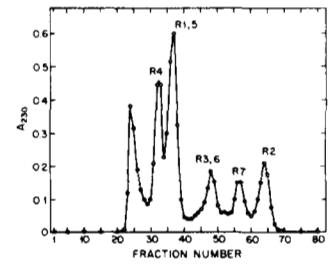


Figure 2. Separation of ε globin arginine cleavage fragments by gel filtration. Two mg of ε globin prepared as shown in Fig. 1, was succinylated and digested with trypsin. This digest was applied to a 1.6 x 72 cm column of Biogel P-10 equilibrated with 0.1 M ammonium bicarbonate. Two ml fractions were collected at a flow rate of 8 ml/hr. Three fractions from each peak were pooled for further analysis. Peptides present in each peak were identified by amino acid composition and sequenator analysis.

RESULTS

Sequencing strategy

The complete primary structure of the ε chain of HbM was determined by automated sequenator analysis using the scheme outlined in Fig. 3. A two dimensional thin layer peptide map of ε globin purified from HbM presented a pattern essentially identical to that of ε globin prepared from HbE (data not shown). Tryptic peptides eluted from this map provided compositional data confirming the accuracy of the sequenator analysis (Table I).

Sixty-three amino acids from the NH_2-termini of ε and α^D chains were characterized by sequenator degradation of whole HbM globin. Likewise, 30 residues from the C-terminal regions of ε and α^D globins were analyzed following cleavage of unseparated HbM globins at their aspartic acid-proline bonds and removal of NH_2-terminal peptides. Additional fragments were prepared by cleavage of purified ε globin at arginine residues. A short sequence for α^D residues 114-122 was obtained by cyanogen bromide cleavage at methionine 113 (data not shown).

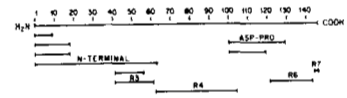


Figure 3. Scheme for the automated sequence analysis of the ε globin chain. The extent of sequenator analysis for each fragment is shown by a bar. Analyses indicated on the same line were done simultaneously, e.g. R3 and R6. Overlaps of at least two residues were obtained for the NH_2 and C termini of each fragment except for R7, which was placed by homology. Details of fragment preparation and residue assignment are discussed in the text.

NH_2-terminal analysis

In Figure 4, sequenator analysis of 63 residues from the NH_2 termini of the α-like and β-like chains of HbM is presented. Two amino acid residues were identified at each cycle, one representing the α^D-globin (Takei et al., 1975) and the other representing the ε globin. The α^D chain sequence determined in this analysis differs from the published sequence at four positions (22, 38, 53, and 54).

Residues assigned to the ε globin were compared with the compositions, NH_2 termini, and charges of tryptic peptides T-1, T-2, T-3 and T-5/6 isolated from purified ε globin (Table I). Tryptic peptide T-4, containing residues 31-40 was mixed with another peptide, and was not useful as a check on the sequenator data. An arginine cleavage fragment from purified ε globin, containing residues 41-61, was analyzed twice, confirming the sequence assignments for this region of the ε chain. The ε globin residues 62 and 63 also were found in the overlapping R4 fragment. By the 50th NH_2-terminal cycle, three proline residues were encountered in the α^D chain and two in the ε chain. Slow cleavage of these proline residues caused the sequence to "lag". After going through a proline residue, a proportion of each principal PTH-amino acid was cleaved in cycles following its initial appearance. This is particularly noticable in the glutamine and isoleucine residues of cycle 54 (Fig. 4), where larger yields of these residues are found in cycle 55. The residues are properly assigned to the first cycle in which they appear.

Analysis of internal fragments

An arginine cleavage fragment extending from residue 62 to 104 was isolated and subjected to sequenator analysis (Fig. 5). This fragment, uncontaminated with other arginine cleavage products, yielded 43 unambiguous residues. Compositions of tryptic peptides T-9 and T-10 (Table I) correspond to amino acid sequenator data. This fragment overlaps two residues of the NH_2-terminal sequence (Fig. 4), and five residues of the aspartic acid proline cleavage fragment. Predicted tryptic peptides T-7 and T-8, and peptides from the C-terminal third of the ε globin chain, were not isolated in sufficient quantity to provide accurate compositional data. Residues 83-99 in the R4 fragment (cycles 22-38) were clearly identifiable above a very low background.

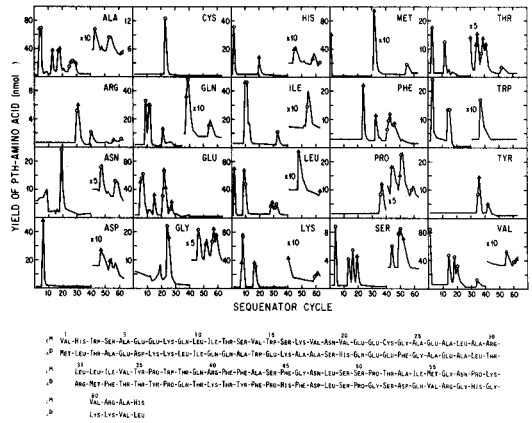


Figure 4. NH_2-terminal sequence of the α-like and β-like chains of HbM. One hundred nmol of CAM-HbM were degraded through 63 cycles at a repetitive yield of 95%. PTH-amino acids were analyzed by HPLC. At each cycle, peak heights of each PTH amino acid were measured, nmol yields calculated from standard chromatograms, and the values plotted. The following symbols are used: (Δ) α^D chain and (o) β chain. In order to show the analysis on a linear scale, nmol yields for later cycles are multiplied by the factor indicated. Assignment of residues to the α-like or β-like chain is discussed in the text.

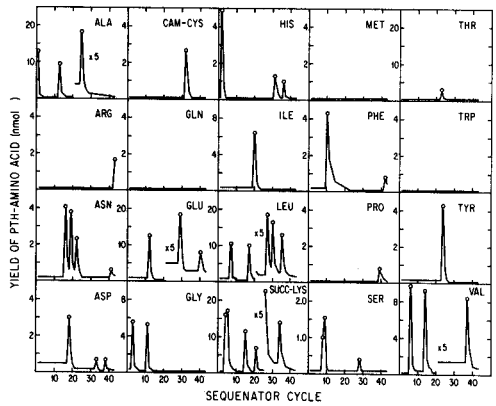
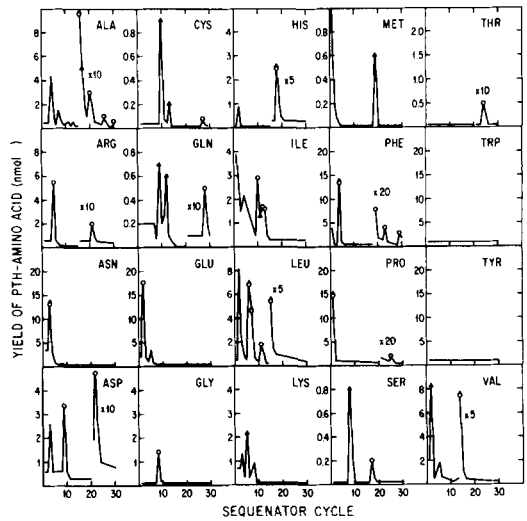


Figure 5. Automated Edman degradation of arginine cleavage fragment R4. Thirty nmol of fragment prepared as shown in Figure 2 were degraded through 44 cycles at a repetitive yield of 93%. The background is extremely small in this analysis.

Primary Structure of the C-terminal region

Approximately 10 moles of ε globin were used in the preparation of an ε globin peptide map. Although this map showed a ninhydrin staining pattern essentially identical to that of the ε globin from HbE, we were unable to recover C-terminal peptides in sufficient quantity for an accurate compositional analysis. We found that analysis by sequenator and HPLC, both modified for microsequencing, was much more sensitive and reliable than compositional data from peptide mapping and amino acid analysis. The sequence of the C-terminal region of ε globin was therefore completed using automated sequenator analysis alone.

Fig. 6 illustrates a sequenator analysis of aspartic acid-proline cleavage fragments from unfractionated HbM globin. Both α-like and β-like globins have highly conserved aspartic acid-proline dipeptides at homologous positions 48 residues from their C termini. Acid hydrolysis under mild conditions specifically cleaves this bond (Piszkiewicz et al., 1970), yielding fragments of approximately five and ten thousand daltons for each globin chain. Although we separated the large and small fragments by gel filtration before sequencing (data not shown), the C-terminal fragments were about 25% contaminated with NH_2-terminal peptides. The contaminating material appeared to be in short pieces that washed out during sequence analysis. The background improved during these sequenator runs at the expense of a poor repetitive yield (79%). An attempt to block contaminating NH_2 termini with fluorescamine was accompanied by a very poor repetitive yield (data not shown), but the data confirmed residue assignments to positions 100-119 for the ε chain and 95-113 for the α^D chain. The aspartic acid-proline cleavage fragment of the ε chain overlapped five residues of the R4 fragment and nine residues of the R6 fragment. A short peptide of the α^D chain, isolated from a cyanogen bromide preparation, yielded data for α^D residues 114-122 (data not shown). Analysis of the α^D aspartic acid-proline fragment revealed two differences from the previously reported adult α^D sequence at positions 107 (Cys for Gln) and 111 (Ala for Val). These residues were assigned to the α^D chain rather than the ε chain for several reasons: 1) the Gln 107 and Val 111 predicted by Takei et al. (1975) were not found in cycles 13 and 17, respectively (Fig. 6). 2) Ile 112 and Ser 116 are found in the ε chain of HbE (Chapman et al., ms submitted) and are likely residues to be found in these positions in β-like chains (Dayhoff, 1976). 3) Cys 107 and Ala 111 occur in other chicken α-like globins (Chapman et al., 1980; Chapman et al., ms submitted). In either case the alternative residue assignments would still lead to a difference in the α^D chains analyzed by us and by Takei et al. (1975).



εM 1 5 10 15 20 25
PRO-GLU-ASN-PHE-ARG-LEU-LEU-GLY-ASP-ILE-LEU-ILE-ILE-VAL-LEU-ALA-SER-HIS-PHE-ALA-ARG-ASP-PHE-THR-PRO-

αD
PRO-VAL-ASN-PHE-LYS-LEU-LEU-SER-GLN-CYS-ILE-GLN-CYS-VAL-LEU-ALA-ALA-HIS-MET-

εM 26 30
ALA-CYS-GLN-PHE-ALA

αD

Figure 6. Sequenator analysis of the two C-terminal cleavage fragments of the α-like and β-like globins of HbM. The CAM-globin was hydrolyzed in 10% acetic acid, 7 M guanidine HCl, pH 2.5 at 46° C for four days. The small C-terminal fragments were separated from the large NH_2-terminal fragments on a column of Biogel P-10 in 0.5% formic acid. Fifty nmol of small fragments were degraded through 31 cycles at a repetitive yield of 79%. The following symbols identify residues of the two chains: (Δ) α^D and (o) ε. Assignment of residues to each chain is discussed in the text. Peaks without symbols are from contaminating NH_2-terminal fragments.

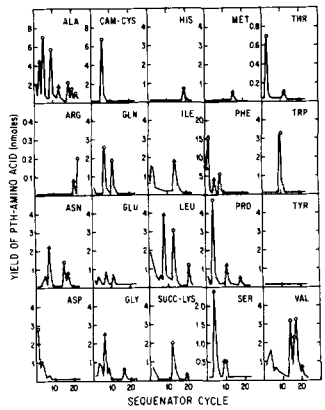


Figure 7. Sequenator analysis of arginine cleavage fragments R3 and R6. Twenty-five nmol of a mixture of R3 and R6 fragments were degraded through 23 cycles at repetitive yields of 88% and 89%, respectively. Assignment of residues is discussed in the text.

Two arginine fragments were used to complete the ε chain sequence. Sequenator analysis of R6 (121-143) mixed with R3 (41-61) is shown in Fig. 7. The R3 sequence, identified in three separate preparations, was subtracted to give an unambiguous amino acid sequence for residues 121-143. R6 residue assignments were in agreement with expectations based on the difference in repetitive yield of R3 and R6 residues (Fig. 8). The R6 fragment overlaps the aspartic acid-proline cleavage fragment by nine residues. The R7 fragment, sequenced together with the R3 fragment (Fig. 9), was the only arginine cleavage product lacking a C-terminal arginine, and was therefore assumed to be the C-terminal tripeptide of the ε globin chain. Although we have no data showing overlap of this tripeptide to the rest of the chain, Lys-Tyr-His tripeptides are found at the C termini of most β-like globin polypeptides (Dayhoff, 1976).

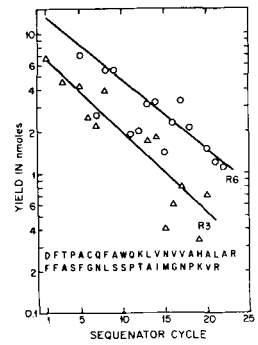
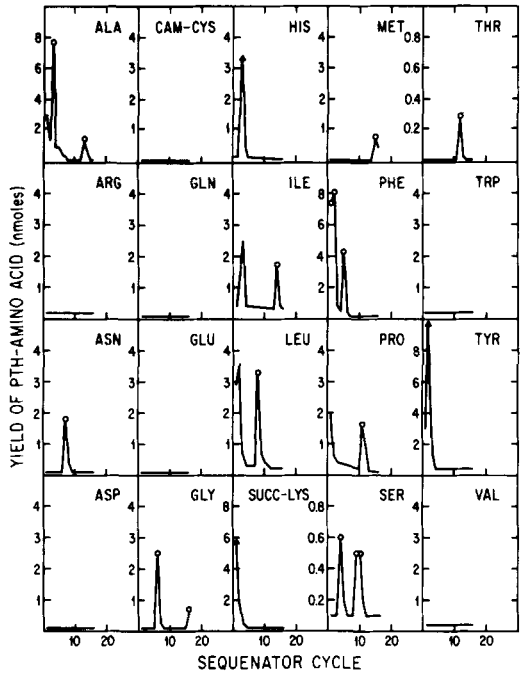


Figure 8. Semilog plot of yields of PTH-amino acids determined to be residues of R3 and R6 (based on data shown in Fig. 7). Yields of PTH-Thr, -Ser, -His, -Arg, and -Pro were not used in this calculation because of their poor recovery relative to other PTH-amino acids. Repetitive yields were calculated from the slopes of lines fitted by linear regression analysis. Sequences for R3 and R6 are shown below.



R7 1
LYS-TYR-HIS

R3 1 5 10 15
PHE-PHE-ALA-SER-PHE-GLY-ASN-LEU-SER-SER-PRO-THR-ALA-ILE-MET-GLY

Figure 9. Sequenator analysis of arginine cleavage fragments R3 and R7. Approximately 25 nmol of peptide from the peak labeled R7 (Fig. 2) were degraded through 16 cycles at a repetitive yield of 84%. The following symbols are used: (Δ) R7 and (o) R3.