Proc. Natl. Acad. Sci. USA
Vol. 78, No. 10, pp. 6076–6080, October 1981
Biochemistry

Structure of the ζ chain of human embryonic hemoglobin

(globin gene evolution/oxygen affinity/Bohr effect)

J. B. CLEGG* AND J. GAGNON†

*Medical Research Council Molecular Haematology Unit, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford, England; and †Medical Research Council Immunochemistry Unit, Department of Biochemistry, University of Oxford, England

Communicated by Sir Hans A. Krebs, June 29, 1981

ABSTRACT **The complete amino acid sequence of the ζ chain of human embryonic hemoglobin has been determined. It differs from human α globin at 57 of the 141 residues and several of the replacements are at positions of structural or functional importance, particularly in relationship to the Bohr effect and high intrinsic oxygen affinity which are characteristic of embryonic hemoglobins. The ζ-globin sequence is more closely related to other mammalian embryonic α-like globins than to human α, suggesting that there have been strong selective pressures to maintain these embryo-specific globins since their emergence several hundred million years ago.**

The nucleated erythrocytes derived from the mesothelium of the yolk sac of vertebrate embryos contain hemoglobins that differ both structurally and functionally from those found in the definitive erythrocytes produced later in fetal development (1). In man, three embryonic hemoglobins, Gower 1 ($\zeta_2\varepsilon_2$), Gower 2 ($\alpha_2\varepsilon_2$), and Portland ($\zeta_2\gamma_2$), have been characterized (2–6) and, whereas the α and γ chains are also produced in later fetal life, the ζ and ε chains (which are the embryonic counterparts of the adult α and β chains) appear to be synthesized mostly if not entirely in the yolk sac under physiological conditions.

Previous studies have established an amino acid sequence for the ε chain that shows a number of differences from the related β and γ chains of fetal ($\alpha_2\gamma_2$) and adult ($\alpha_2\beta_2$) hemoglobin (3, 7) although, in most cases, the changes are conservative and may have only relatively minor effects on the structure and function of the hemoglobin molecule. Here we present the complete amino acid sequence of the human ζ-globin chain. It differs at 40% of the residues from the human α-globin sequence, suggesting that the ζ and α genes on chromosome 16 diverged several hundred million years ago, following duplication of an ancestral α-like gene that substantially predated the duplications that later gave rise to the εγδβ gene complex on chromosome 11 (8).

Several of the amino acid replacements in the ζ chain occur at sites of particular structural or functional importance, which may in part account for the diminished Bohr effect and increased O_2 affinity characteristic of embryonic hemoglobins (9–11).

MATERIALS AND METHODS

Hemoglobin Gower 1 was purified from tissue obtained at vacuum abortion or from lysates of hemin-induced K562 erythroleukemia cells as described (3). Hemoglobin Portland was isolated from erythrocyte lysates prepared from cord bloods of stillborn fetuses having the Hb Bart's hydrops fetalis syndrome by isoelectric focusing in Ultrodex gel, pH 6–8, followed by DEAE-cellulose chromatography at 4°C in 0.2 M glycine/1.5 mM KCN (12) with a gradient of 0.03–0.09 M NaCl. Attempts to separate the constituent globin chains of these hemoglobins on a preparative scale were unsuccessful and subsequent operations were therefore carried out on unfractionated globin.

Cyanogen bromide cleavage of Gower 1 and Portland globins was carried out in 70% HCOOH for 24 hr at room temperature. After recovery of the material by lyophilization, the fragments were fractionated on a 95 × 2.5 cm column of Sephadex G-50 in 0.01 M HCl. Mild acid cleavage of the Asp-Pro bond at position 94–95 of the ζ chain was accomplished by incubating samples in 70% HCOOH at 37°C (13) for 48 hr. The cleavage products were separated on Sephadex G-50 in 0.01 M HCl.

Glycinamidination of the COOH-terminal 46-residue fragment of the ζ chain was carried out as described by Brimhall *et al.* (14). Aminoethylation of ζ-chain fragments was carried out as described (15).

Digestions of ζ-chain fragments with trypsin (2 hr), chymotrypsin (2 hr), thermolysin (6 hr), *Staphylococcus aureus* V8 protease (16 hr), and mouse submaxillary arginine-specific protease (16 hr) were made in 1% NH_4HCO_3, pH 8.5, at 37°C for the times stated. Digests were fractionated by high voltage paper electrophoresis or chromatography as described (15). Carboxypeptidase Y digestions were performed at pH 6.5 on samples previously heated at pH 6.5 for 30 min at 95°C to ensure conversion of homoserine lactone to homoserine. At various times, aliquots of the reaction mixture were withdrawn and diluted with pH 2.2 loading buffer, and the released amino acids were determined on the amino acid analyzer without further treatment.

Manual sequence analysis of peptides was carried out by the method of Gray (16). After each degradation cycle, 5-nmol aliquots were taken for indentification of the NH_2-terminal residue by dansylation and for determination of the residual amino acid composition.

Peptides for amino acid analysis were hydrolyzed in 6 M HCl containing phenol at 2 mg/ml (17) in sealed capillaries for 16–24 hr at 105°C, and their amino acid compositions were determined on a Beckman 119 CL analyzer.

Automated sequence analysis of peptide fragments was carried out as described (18) in a Beckman 890 C spinning cup sequencer using a 0.3 M Quadrol program (19) in the presence of Polybrene (20).

RESULTS

An attempt to determine the sequence of intact Gower 1 globin produced a unique amino acid sequence characteristic of the

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

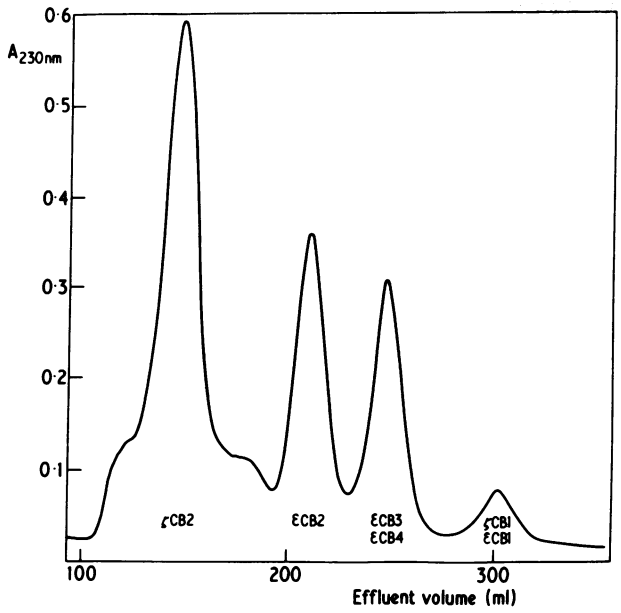

FIG. 1. Fractionation of CNBr digest of 6.5 mg of Gower 1 globin on a 2.5 × 90 cm Sephadex G-50 column in 0.01 M HCl.

ε chain only, suggesting that the NH_2 terminus of the ζ chain might be blocked.

Cyanogen bromide digestion of Hb Gower 1 globin followed by Sepadex G-50 filtration produced six major fragments (Fig. 1), the largest of which proved on subsequent analysis to contain the COOH-terminal fragment of the ζ chain (ζCB2; 128 residues). The NH_2-terminal tridecapeptide (ζCB1) coeluted from Sephadex G-50 with the NH_2-terminal ε-chain fragment εCB1 (Fig. 1), and attempts to fractionate these peptides on paper resulted in severe losses of ζCB1. However, this peptide was clearly resolved from other fragments on Sephadex G-50 fractionation of CNBr digest of Portland globin, so it was routinely prepared from this source. It had no detectable NH_2-terminal residue by dansylation. Furthermore, a peptide with the composition (Thr, Ser, Leu, Lys), which was neutral at pH 6.5, was isolated from tryptic digests of ζCB1, again suggesting a blocked NH_2-terminal residue. Subsequently, this peptide was shown to be derived from the NH_2-terminus of ζCB1 and thus of the intact ζ chain.

The sequence of ζCB1 was established from peptides isolated from trypsin, thermolysin, and staphylococcus digests and from carboxypeptidase Y digests of the intact fragment (Fig. 2). It has NH_2-terminal serine, which is thus the NH_2-terminal residue of the ζ chain. The nature of the blocking group has not been determined, although it is not cleaved under the mild acid conditions used to release the putative formyl group blocking the NH_2-terminal alanine residue in chicken embryonic α globin (21). Acetylserine has been reported as the NH_2-terminal residue in some fish and reptilian hemoglobins (22).

Treatment of ζCB2 by mild acid cleavage or with the arginine-specific protease gave smaller peptides, representing res-

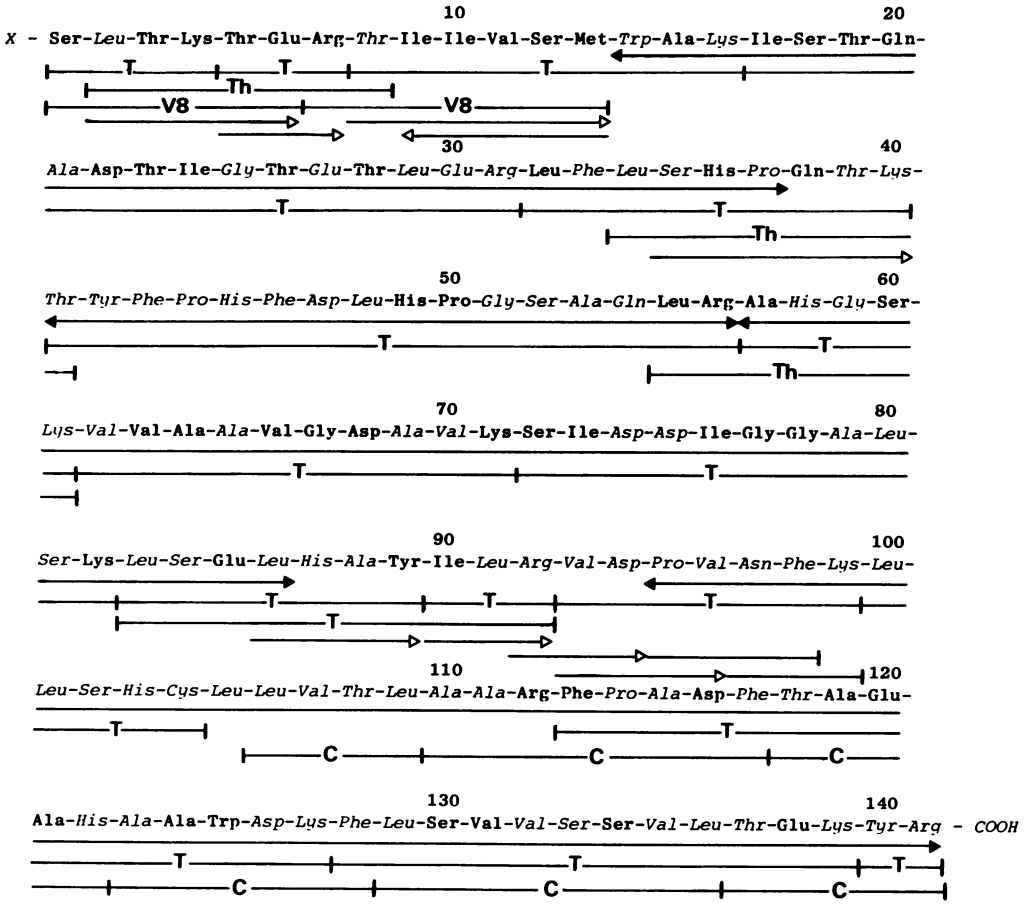

FIG. 2. Amino acid sequence of the human ζ-globin chain. Residues in italics are common to human α globin. ⟷, Residues determined automatically; ⟶, residues determined manually; ⟵, residues determined by carboxypeptidase Y digestion; ⊢⊣, peptides produced by trypsin (T), chymotrypsin (C), thermolysin (Th), and staphylococcal protease (V8) digestion.

idues 95–141 and 57–92 respectively. Sequences for residues 95–141 and 57–86 of these fragments, the tryptic peptide representing residues 41–56, and the NH_2-terminal 24 residues (14–37) of intact ζCB2 were determined directly by automated Edman degradation. Overlaps between these sequences and, in the cases of residues 38–40 and 87–94, extensions of the sequences were established from compositions and partial sequences of tryptic, chymotryptic, or thermolytic peptides straddling the junctions.

For each of these large fragments sequenced, a contiguous set of tryptic, chymotryptic, or thermolytic peptides was isolated and characterized by amino acid composition and, in some cases, NH_2-terminal analysis to corroborate the direct sequence. No peptides corresponding to residues 100–139 were found in tryptic digests of ζCB2. However, some were isolated from aminoethylated ζCB2 (for residues 100–104) or the glycinamidated COOH-terminal fragment produced by mild acid cleavage of ζCB2 (for residues 113–139). The amide status of residues 6, 38, and 94 was deduced from peptide electrophoretic mobilities. The overlap between ζCB1 and ζCB2 was established by a tryptic peptide (residues 8–16) isolated from a digest of unfractionated Gower 1 globin.

The complete ζ-chain sequence, together with details of the sequencing strategy, is given in Fig. 2. Comparison with peptide compositions and alignments published previously (23) shows a number of discrepancies. Although material from three different sources was used to establish the sequence given here, the compositions of equivalent ζ-chain peptides were identical irrespective of the nature of the starting material. All of the large fragments (representing 83% of the sequence) used in the automatic sequence determination were prepared from hemoglobin Gower 1 isolated from normal human tissue, as was ζCB1. Only a few of the overlap peptides were isolated from K562 cell hemoglobin Gower 1, and their compositions were the same as those isolated from other sources. It thus seems unlikely that this could be responsible for the differences. However, the embryonic hemoglobins used in previous investigations (usually hemoglobin Portland) were not always completely purified, and contamination with non-heme proteins or other globins may be a cause of some of the discrepancies.

DISCUSSION

The sequence of the human ζ-globin chain is clearly that of an α-like chain, even though it differs substantially from that of human adult α globin. It has the characteristic Lys-Tyr-Arg COOH terminus, is 141 residues long, and lacks the residues of the D helix. However, it differs from the α chain at 57 out of the 141 positions (Fig. 2) and shows a greater resemblance to the two chicken embryonic α-like globins (21) than to human adult α globin. Furthermore, of the 57 variant sites, no fewer than 16 are found in carp α globin, and a few others are found in carp β globin or in the hemoglobins of other fish or amphib-

Table 1. Invariant residues in hemoglobin α chain

Residue in α_1	Comments
2(NA2)Leu	Internal contact with helix H
7(A5)Lys	Salt bridge or H bond with α_1(EF3) Asp or Asn (Arg in human ζ)
27(B8)Glu	Interaction with α_1(B12) Arg
29(B10)Leu	Internal contacts with (E7)His and (G8)Leu
31(B12)Arg	H bonds with C=O of β_1(GH5)Phe, α_1(H5)His, and α_1(B8)Glu
39(C4)Thr	H bond with C=O of α_1(C1)Phe
40(C5)Lys	Salt bridge to COO^- of β_2(HC3)His in deoxy Hb
42(C7)Tyr	H bond with β_2(G1)Asp in deoxy Hb
43(CD1)Phe	Heme contact
59(E8)Gly	Short contact with α_1(β6)Gly
65(E14)Ala	Heme contact
69(E18)Ala	Internal contact with helix F
75(EF4)Asp	Salt bridge to α_1(EF1)Lys or His
83(F4)Leu	Heme contact
87(F8)His	Bond to Fe
91(FG3)Leu	$\alpha_1\beta_2$ contact in HbO_2
94(G1)Asp	H bond with β_2(G4)Asn in HbO_2
95(G2)Pro	$\alpha_1\beta_2$ contact
97(G4)Asn	H bond with α_1(C7)Tyr
98(G5)Phe	Heme contact
101(G8)Leu	Heme contact
126(H9)Asp	Salt bridge to α_2(HC3)Arg in deoxy Hb; H bond to OH of β_1(C1)Tyr
127(H10)Lys	Salt bridge to α_2(HC3)Arg in deoxy Hb
128(H11)Phe	Internal contacts with helices A and E
136(H19)Leu	Heme contact
140(HC2)Tyr	H bond with C=O of α_1(FG5)Val in deoxy Hb; $\alpha_1\beta_2$ contact
141(HC3)Arg	Salt bridges to α_2(H9)Asp and α_2(H10)Lys in deoxy Hb

Detailed discussion and illustration of the α-chain structure is given in ref. 24.

Table 2. Comparison of ζ and α residues at functional sites

Residue	α→ζ	Functions in α	Comments
1(NA1)	Val→X-Ser	Bohr cooperativity; CO_2 binding; α–α contact; salt bridge to α-COOH of 141(HC3)Arg in other α chain	Not possible in ζ
20(B1)	His→Gln	Salt bridge to 23(B4)Glu of the same α chain	Not possible in ζ
23(B4)	Glu→Thr	Salt bridge to 20(B1)His of the same α chain	Not possible in ζ
32(B13)	Met→Leu	Heme contact	Probably conservative
36(C1)	Phe→His	$\alpha_1\beta_1$ Contact; van der Waals' contact to β131(H9)Gln	ε131(H9) also Gln
38(C3)	Thr→Gln	$\alpha_1\beta_2$ Contact; van der Waals' contacts to β145(HC2)Tyr, β100(G2)Pro, and β99(G1)Asp in deoxy Hb; van der Waals' contacts to β98(FG5)Val, and β97(FG4)His in HbO_2	See text
50(CD8)	His→Pro	Salt bridge to 30(B11)Glu of the same α chain	Not possible in ζ
119(H2)	Pro→Ala	$\alpha_1\beta_1$ contact; van der Waals' contacts to β55(D6)Met, β51(D2)Pro, and β33(B15)Val	ε55(D6) is Leu
82(F3)	Ala→Lys		Salt bridge to heme propionate in ζ

Other changes are either conservative or occur at external, surface crevice, or central cavity sites, with no anticipated major effects on structure of function. Detailed discussion and illustration of subunit structures is given in ref. 24.

ians, reinforcing the general impression of a "primitive" sequence. Nevertheless, with the exception of a conservative Lys→Arg change at position 7(A5), all of the α-chain sites considered to be invariant (Table 1) are maintained in the ζ chain. From the physiological point of view, 4 of the 57 changes relative to the α chain involve positions specifically associated with functions such as the Bohr effect and subunit cooperativity and interaction (Table 2). Perhaps most noteworthy is the lack of a free α-NH_2-terminal group (as also occurs in the chicken π chain), which is presumably responsible, at least in part, for the diminished Bohr effect in human embryonic hemoglobin (9, 11, 25), and the Thr→Gln substitution at position 38(C3) [also found in mouse, rabbit (26), and chicken (21) embryonic α-like globins] which, because this residue is in the $\alpha_1\beta_2$ contact (21, 24), may be responsible for the higher intrinsic O_2 affinity (9–11, 27–29) of the embryonic hemoglobins of these species. The human ζ chain also lacks two of the external salt bridges that stabilize the human α chain, 20(B1)His to 23(B4)Glu and 30(B11)Glu to 50(CD8)His; on the other hand, 82(F3)Lys provides an additional salt bridge to one of the heme propionates.

The human α-like globin genes are localized in a 40-kilobase cluster on chromosome 16, and it has been established by heteroduplex mapping and blot hybridization experiments that there are two ζ-like genes lying ≈11 kilobases apart at the 5′ extremity of this complex (30, 31). There is firm evidence that the 5′ ζ2 gene is a functional gene (32), but its DNA sequence has not yet been determined. On the other hand, the DNA sequence of part of the 3′ ζ1 gene is known (31), but there are preliminary findings suggesting that ζ1 is in fact a pseudogene (N. J. Proudfoot, personal communication) and thus is not expressed. The ζ-protein sequence must therefore correspond to the nucleotide sequence of the ζ2 gene. Interestingly, comparison of the protein sequence predicted from the known nucleotide sequence of ζ1 with the authentic ζ-globin sequence given in Fig. 2 shows complete identity over the 24 codons known, suggesting, as does restriction endonuclease mapping (D. R. Higgs and S. Goodbourn, personal communication), that the ζ1 and ζ2 genes may have very similar structures notwithstanding the inactivity of the former. This in turn implies that the ζ1 and ζ2 genes probably diverged only recently, with sequence identity between the two loci having been previously maintained by some sort of gene conversion mechanism, as postulated for the duplicated human α and γ genes (33, 34). Alternatively, the two ζ-like genes may only recently have duplicated, followed by the almost immediate mutation and inactivation of ζ1.

The recent determination of the nucleotide sequences of the human ε, γ, δ, and β genes (7, 34–36) and their comparison with the sequences of other mammalian globin genes has, when taken in conjunction with available phylogenetic evidence, enabled fairly precise estimates to be made for the times of divergence of these genes (8). Although the DNA sequence of the human ζ chain gene is not at present available, the ≈40% protein sequence difference from α globin would, in the light of these comparisons, place the estimated divergence of the ζ and α genes at ≈400 million years ago, about the time that the line leading to bony fish became established. Similar estimates have been derived from the sequences of the chicken embryonic and adult α-like globins, which differ at ≈45% of their amino acid residues (21). The ζ–α divergence thus clearly preceded those of the non-α gene cluster, which successively gave rise to the ε, γ, and δ genes 200, 100, and 40 million years ago (8).

It is clear that selection has operated on the individual α-like genes, resulting in widely divergent sequences and producing structural changes that presumably suit the adaptation of the ζ and α globins to their respective embryonic and adult environments. Supporting this are the findings that there is more overall similarity between the ζ chains of chicken, mouse, rabbit (21, 26), and man, than between ζ and α chains in these same widely divergent species. Although hemoglobin function can be altered by a wide variety of cellular effectors, it appears that this mechanism has insufficient versatility to cope with the changing physiological demands during development. The evolution of the hierarchy of diverse globin polypeptides is, presumably, a result of this inadequacy.

We thank Lynne Pressley for her encouragement and support, Max Perutz and E. R. Huehns for their advice and for much of the information in Tables 1 and 2, Rosemary Gale and Julie Bennett for gifts of embryonic hemoglobin, Malcolm Shepherd for his technical assistance, and Elizabeth Gunson for her excellent secretarial work. This project was supported by the Cancer Research Campaign.

1. Kitchen, H. & Brett, I. (1974) *Ann. N.Y. Acad. Sci.* **241**, 653–671.
2. Huehns, E. R., Flynn, F. V., Butler, E. A. & Beaven, G. H. (1961) *Nature (London)* **189**, 496–497.
3. Gale, R. E., Clegg, J. B. & Huehns, E. R. (1979) *Nature (London)* **280**, 162–164.
4. Kaltsoya, A., Fessas, P. & Stavropoulos, A. (1966) *Science* **153**, 1417–1418.
5. Capp, G., Rigas, D. I. & Jones, R. T. (1967) *Science* **157**, 65–66.
6. Capp, G., Rigas, D. I. & Jones, R. T. (1970) *Nature (London)* **288**, 278–280.
7. Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21**, 621–626.
8. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21**, 653–668.
9. Tuchinda, S., Nagai, K. & Lehmann, H. (1975) *FEBS. Lett.* **49**, 390–391.
10. Huehns, E. R. & Farooqui, A. (1975) *Nature (London)* **254**, 335–337.
11. Gale, R. E. (1978) Dissertation (Univ. London, London, England).
12. Abraham, E. C., Reese, A., Stallings, M. & Huisman, T. H. J. (1977) *Hemoglobin* **1**, 27–44.
13. Terhorst, C., Robb, R., Jones, C. & Strominger, J. L. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4002–4006.
14. Brimhall, B., Vedvick, T. S., Jones, R. T., Ahern, E., Palomino, E. & Ahern, V. (1974) *Br. J. Haematol.* **27**, 313–318.
15. Clegg, J. B., Naughton, M. A. & Weatherall, D. J. (1966) *J. Mol. Biol.* **19**, 91–108.
16. Gray, W. R. (1967) *Methods Enzymol.* **11**, 469–475.
17. Sanger, F. & Thompson, E. O. P. (1963) *Biochim. Biophys. Acta.* **71**, 468–471.
18. Campbell, D. G., Gagnon, J., Reid, K. B. M. & Williams, A. F. (1981) *Biochem. J.* **195**, 15–30.
19. Hunkapiller, M. W. & Hood, L. E. (1978) *Biochemistry* **17**, 2124–2133.
20. Klapper, D. G., Wilde, C. E. & Capra, J. D. (1978) *Anal. Biochem.* **85**, 126–131.
21. Chapman, B. S., Tobin, A. J. & Hood, L. E. (1980) *J. Biol. Chem.* **225**, 9051–9059.
22. Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure* (Nat. Biomed. Res. Found., Washington, D.C.), Vol. 5, Suppl. 3.
23. Kamuzora, H. & Lehmann, H. (1975) *Nature (London)* **256**, 511–513.
24. Ladner, R. C., Heidner, E. J. & Perutz, M. F. (1977) *J. Mol. Biol.* **114**, 385–413.
25. Kilmartin, J. (1976) *Br. Med. Bull.* **32**, 209–212.
26. Melderis, H., Steinheider, G. & Ostertag, W. (1974) *Nature (London)* **250**, 774–776.
27. Huehns, E. R. & Farooqui, A. (1976) *Ciba Found. Symp.* **37**, 69–83.
28. Jelkmann, W. & Bauer, C. (1978) *Pfleugers Arch. Eur. J. Physiol.* **377**, 75–80.

29. Bauer, C., Tamm, R., Petschow, D., Bartels, R. & Bartels, H. (1975) *Nature (London)* **257**, 333–334.
30. Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W. F., Ruddle, R., Lawrence, J., Creagan, R. & Kucherlapati, R. (1977) *Cell* **12**, 205–218.
31. Lauer, J., Shen, C. J. & Maniatis, T. (1980) *Cell* **20**, 119–130.
32. Pressley, L., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3586–3589.
33. Zimmer, E. A., Martin, S. L., Beverly, S. M., Kan, Y. W. & Wilson, A. C. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2158–2162.
34. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21**, 627–638.
35. Spritz, R. A., DeRiel, J. K., Forget, B. G. & Weissman, S. M. (1980) *Cell* **21**, 639–646.
36. Lawn, R. M., Efstratiadis, A., O'Connell, C. & Maniatis, T. (1980) *Cell* **21**, 647–651.