J Mol Evol (1984) 21:33–41

Journal of
Molecular Evolution

Derivation of the Globins from Type b Cytochromes

B. Runnegar

Department of Geology and Geophysics, University of New England, Armidale, New South Wales 2351, Australia

Summary. Similarities in the amino acid sequences of vertebrate and invertebrate globins, b_5 and b_2 cytochromes and chicken sulfite oxidase point to a common ancestry for all of these proteins. The distal heme ligand (histidine or its equivalent) is common to both sets of proteins, but the proximal histidine ligand of the cytochromes is replaced by another histidine residue in the globins. This explains why the heme is reversed between globins and b_5 cytochromes. It seems likely that the genes for primitive globins contained three exons, the first two of which were derived from a cytochromelike DNA sequence. A model is presented to show how globins may have evolved from a pre-existing type b cytochrome; the complexity of the required changes is an indication that all globins are monophyletic.

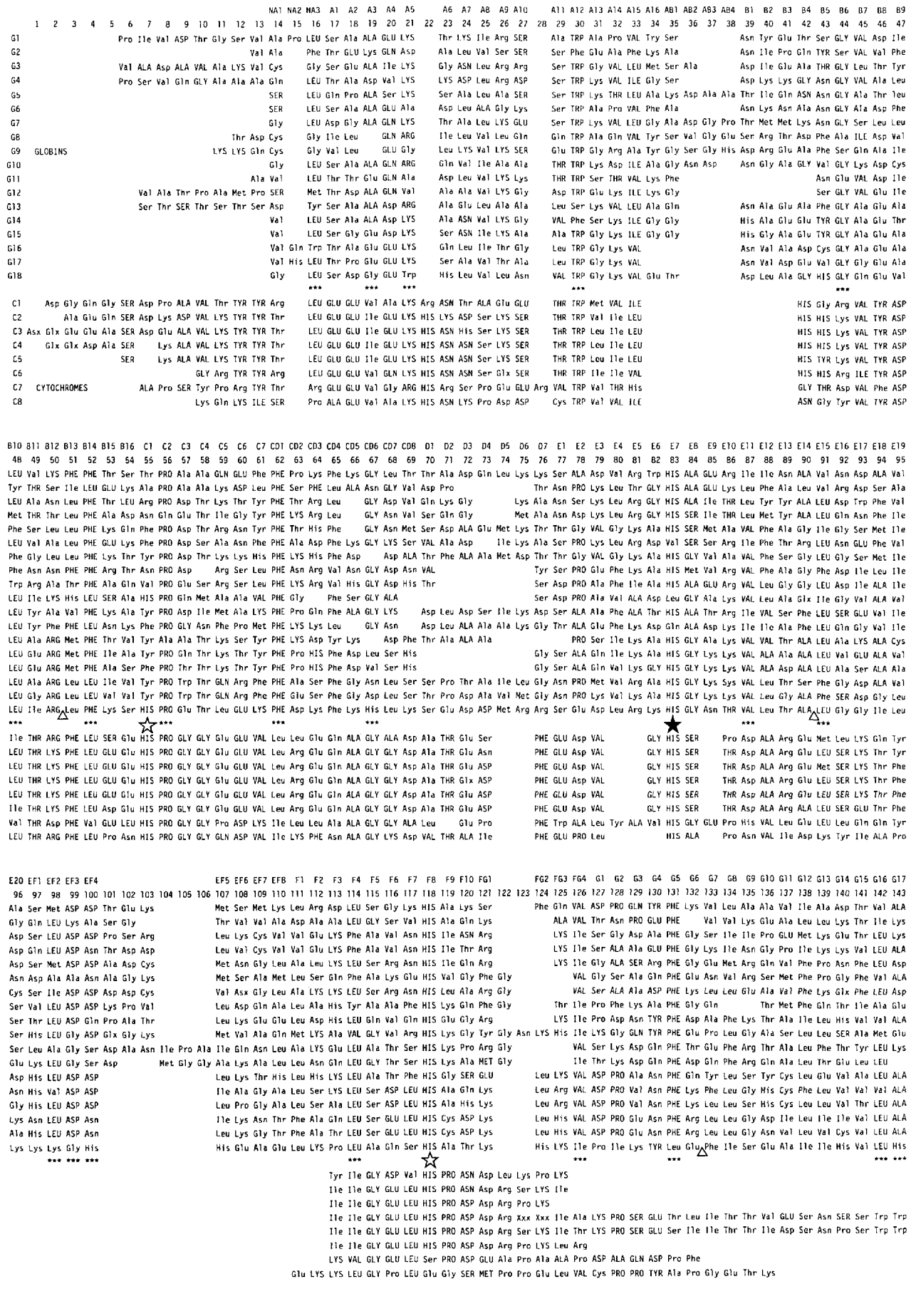
Key words: Protein evolution — Cytochrome b_5 — Myoglobin — Hemoglobin

Introduction

Anderson (1981) has suggested that multicelled animals that begin their development by means of a spiral cleavage of the fertilized egg may have attained the metazoan grade of organization independently from those animals (deuterostomes) that undergo an initial radial cleavage. However, many living members of both of these embryological groups manufacture monomeric or polymeric oxygen-binding proteins (globins) that are thought to be homologous because of significant similarities in their primary, secondary, and tertiary structures (Lesk and Chothia 1980; Thompson 1980). That the living ciliate *Paramecium* contains hemoglobin (Irie and Usuki 1980) suggests that the genes for these proteins may have been inherited from a remote common ciliate ancestor. However, it is also possible that the similarities seen in the structures of modern globins were inherited convergently from another heme-bearing molecule. In this article I attempt to show that whilst some of the features of the primary structure of all globins were inherited from an ancestral type b cytochrome, differences between the b cytochromes and the globins make it unlikely that the globins could have evolved more than once.

Type b cytochromes are electron-carrying proteins that, like the globins, contain protoheme IX as the prosthetic group (Hagihara et al. 1975; von Jagow and Sebald 1980). Cytochrome b is a hydrophobic molecule of about 380 amino acids that forms part of the ubiquinol : cytochrome c reductase complex (complex III) of the inner mitochondrial membrane (Nobrega and Tzagoloff 1980; Leonard et al. 1981). It is manufactured within the mitochondrion and its primary structure is known from DNA sequencing of the mitochondrial genes of humans, mouse, ox, yeast and *Aspergillus* (Widger et al. 1984). Saraste (1984) and Widger et al. (1984) have recently proposed similar models for the secondary structure of yeast cytochrome b.

Type b_5 cytochromes are microsomal and mitochondrial proteins of vertebrate livers that are encoded by nuclear genes. Each such cytochrome consists of a hydrophilic heme-bearing unit of about 95 amino acids and a hydrophobic C-terminal 'tail' of about 40 amino acids, which appears to anchor the protein to the membrane (Takagaki et al. 1983). The primary structure is known from several vertebrates (Hagihara et al. 1975; von Jagow and Sebald 1980; Lederer et al. 1983), and the secondary and tertiary structures have been determined at 2.0-Å resolution using calf liver b_5 (Mathews et al. 1972a). These data provide no evidence for a common ancestry of the b and b_5 cytochromes.


NA1 NA2 MA3 A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 A13 A14 A15 A16 AB1 AB2 AB3 AB4 B1 B2 B3 B4 B5 B6 B7 B8 B9
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
G1 Pro Ile Val ASP Thr Gly Ser Val Ala Pro LEU Ser Ala ALA GLU LYS Thr LYS Ile Arg SER Ala TRP Ala Pro VAL Try Ser Asn Tyr Glu Thr Ser GLY VAL Asp Ile
G2 Val Ala Phe Thr GLU Lys GLN Asp Ala Leu Val Ser SER Ser Phe Glu Ala Phe Lys Ala Asn Ile Pro Gln TYR Ser VAL Val Phe
G3 Val ALA Asp ALA VAL Ala LYS Val Cys Gly Ser Glu ALA Ile LYS Gly ASN Leu Arg Arg Ser TRP Gly VAL LEU Met Ser Ala Asp Ile Glu Ala THR GLY Leu Thr Tyr
G4 Pro Ser Val Gln GLY Ala Ala Ala Gln LEU Thr Ala Asp Val LYS LYS ASP Leu Arg ASP Ser TRP Lys VAL ILE Gly Ser Asp Lys Lys GLY Asn GLY VAL Ala Leu
G5 SER LEU Gln Pro ALA Ser LYS Ser Ala Leu Ala SER Ser TRP Lys THR LEU Ala Lys Asp Ala Ala Thr Ile Gln ASN Asn GLY Ala Thr Ile
G6 SER LEU Ser Ala ALA GLU Ala Asp Leu ALA Gly Lys Ser TRP Ala Pro VAL Phe Ala Asn Lys Asn Ala Asn GLY Ala Asp Phe
G7 Gly LEU Asp Gly ALA GLN LYS Thr Ala Leu LYS GLU Ser TRP Lys VAL LEU Gly Ala Asp Gly Pro Thr Met Met Lys Asn GLY Ser Leu Leu
G8 Thr Asp Cys Gly Ile Leu GLN ARG Ile Leu Val Leu Gln Gln TRP Ala Gln VAL Tyr Ser Val Gly Glu Ser Arg Thr Asp Phe Ala ILE Asp Val
G9 GLOBINS LYS LYS Gln Cys Gly Val Leu GLU Gly Leu LYS Val LYS SER Glu TRP Gly Arg Ala Tyr Gly Ser Gly His Asp Arg Glu Ala Phe Ser Gln Ala Ile
G10 Gly LEU Ser Ala ALA GLN ARG Gln Val Ile Ala Ala THR TRP Lys Asp ILE Ala Gly Asn Asp Asn Gly Ala GLY Val GLY Lys Asp Cys
G11 Ala Val LEU Thr Thr Glu GLN Ala Asp Leu Val LYS Lys THR TRP Ser THR VAL Lys Phe Asn Glu VAL Asp Ile
G12 Val Ala Thr Pro Ala Met Pro SER Met Thr Asp ALA GLN Val Ala Ala Val LYS Gly Asp TRP Glu Lys ILE Lys Gly Ser GLY VAL Glu Ile
G13 Ser Thr SER Thr Ser Thr Ser Asp Tyr Ser Ala ALA Asp ARG Ala Glu Leu Ala Ala Leu Ser Lys VAL LEU Ala Gln Asn Ala Glu Ala Phe GLY Ala Glu Ala
G14 Val LEU Ser Ala ALA Asp LYS Ala ASN Val LYS Gly VAL Phe Ser Lys ILE Gly Gly His Ala Glu Glu TYR GLY Ala Glu Thr
G15 Val LEU Ser Gly Glu Asp LYS Ser ASN Ile LYS Ala Ala TRP Gly Lys ILE Gly Gly His Gly Ala GLY TYR GLY Ala Glu Ala
G16 Val Gln Trp Thr Ala Glu GLU LYS Gln Leu Ile Thr Gly Leu TRP Gly Lys VAL Asn Val Ala Asp Cys GLY Ala Glu Ala
G17 Val His LEU Thr Pro Glu GLU LYS Ser Ala Val Thr Ala Leu TRP Gly Lys VAL Asn Val Asp Glu Val GLY Gly Glu Ala
G18 Gly LEU Ser Asp Gly GLU Trp His Leu Val Leu Asn VAL TRP Gly Lys VAL Glu Thr Asp Leu Ala GLY HIS GLY Gln Glu Val
*** *** *** *** ***
C1 Asp Gly Gln Gly SER Asp Pro ALA VAL Thr TYR TYR Arg LEU GLU GLU Val Ala LYS Arg ASN Thr ALA Glu GLU THR TRP Met VAL ILE HIS Gly Arg VAL TYR ASP
C2 Ala Glu Gln SER Asp Lys ASP VAL LYS TYR TYR Thr LEU GLU GLU Ile GLU LYS HIS LYS ASP Ser LYS SER THR TRP Val Ile LEU HIS HIS Lys VAL TYR ASP
C3 Asx Glx Glu Glu Ala SER Asp Glu ALA VAL LYS TYR TYR Thr LEU GLU GLU Ile GLU LYS HIS ASN His Ser LYS SER THR TRP Leu Ile LEU HIS HIS Lys VAL TYR ASP
C4 Glx Glx Asp Ala SER Lys ALA VAL LYS TYR TYR Thr LEU GLU GLU Ile GLU LYS HIS ASN ASN Ser LYS SER THR TRP Leu Ile LEU HIS HIS Lys VAL TYR ASP
C5 SER Lys ALA VAL LYS TYR TYR Thr LEU GLU GLU Ile GLU LYS HIS ASN ASN Ser LYS SER THR TRP Leu Ile LEU HIS TYR Lys VAL TYR ASP
C6 GLY Arg TYR TYR Arg LEU GLU GLU Val GLN LYS HIS ASN ASN Ser Glx SER THR TRP Ile Ile VAL HIS HIS Arg ILE TYR ASP
C7 CYTOCHROMES ALA Pro SER Tyr Pro Arg TYR Thr Arg GLU GLU Val Gly ARG HIS Arg Ser Pro Glu GLU Arg VAL TRP Val THR His GLY THR Asp VAL Phe ASP
C8 Lys Gln LYS ILE SER Pro ALA GLU Val Ala LYS HIS ASN LYS Pro Asp ASP Cys TRP Val VAL ILE ASN Gly Tyr VAL TYR ASP
B10 B11 B12 B13 B14 B15 B16 C1 C2 C3 C4 C5 C6 C7 CD1 CD2 CD3 CD4 CD5 CD6 CD7 CD8 D1 D2 D3 D4 D5 D6 D7 E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 E12 E13 E14 E15 E16 E17 E18 E19
48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95
LEU Val LYS PHE PHE Thr Ser Thr PRO Ala Ala GLN GLU Phe PHE Pro Lys Phe Lys GLY Leu Thr Thr Ala Asp Gln Leu Lys Lys Ser ALA Asp Val Arg Trp HIS ALA GLU Arg Ile Ile Asn ALA Val Asn Asp ALA Val
Tyr THR Ser Ile LEU GLU Lys Ala PRO Ala Ala Lys ASP Leu PHE Ser PHE Leu ALA Asn GLY Val Asp Pro Thr Asn PRO Lys Leu Thr GLY HIS ALA GLU Lys Leu Phe Ala Leu Val Arg Asp Ser Ala
LEU Ala Asn Leu PHE Thr LEU Arg PRO Asp Thr Lys Thr Tyr PHE Thr Arg Leu GLY Asp Val Gln Lys Gly Lys Ala Asn Ser Lys Leu Arg GLY HIS ALA Ile THR Leu Tyr Tyr ALA LEU Asp Trp Phe Val
Met THR Thr Leu PHE Ala Asp Asn Gln Glu Thr Ile Gly Tyr PHE LYS Arg Leu GLY Asn Val Ser Gln Gly Met Ala Asn Asp Lys Leu Arg GLY HIS SER Ile THR Leu Met Tyr ALA LEU Gln Asn Phe Ile
Phe Ser Leu Leu PHE Lys Gln Phe PRO Asp Thr Arg Asn Tyr PHE Thr His Phe GLY Asn Met Ser Asp ALA Glu Met Lys Thr Thr Gly VAL Gly Lys Ala HIS SER Met Ala VAL Phe Ala Gly Ile Gly Ser Met Ile
LEU Val Ala Leu PHE GLU Lys Phe PRO Asp Ser Ala Asn Phe PHE Ala Asp Phe Lys GLY LYS Ser VAL Ala Asp Ile Lys Ala Ser PRO Lys Leu Arg Asp Val SER Ser Arg Ile Phe Thr Arg LEU Asn GLU Phe Val
Phe Gly Leu Leu PHE Lys Thr Tyr PRO Asp Thr Lys Lys His PHE LYS His Phe Asp Asp ALA Thr Phe ALA Ala Met Asp Thr Thr Gly VAL Gly Lys Ala HIS GLY Val Ala VAL Phe Ser Gly LEU Gly Ser Met Ile
Phe Asn Asn PHE PHE Arg Thr Asn PRO Asp Arg Ser Leu PHE Asn Arg Val Asn GLY Asp Asn VAL Tyr Ser PRO Glu Phe Lys Ala HIS Met Val Arg VAL Phe Ala Gly Phe Asp Ile Leu Ile
Trp Arg Ala Thr PHE Ala Gln Val PRO Glu Ser Arg Ser Leu PHE LYS Arg Val His GLY Asp His Thr Ser Asp PRO Ala Phe Ile Ala HIS ALA GLU Arg VAL Leu Gly Gly LEU Asp Ile ALA Ile
LEU Ile LYS His LEU SER Ala HIS PRO Gln Met Ala Ala VAL PHE Gly Phe Ser GLY ALA Ser Asp PRO Ala Val ALA Asp Leu GLY Ala Lys VAL Leu Ala Glx Ile Gly Val ALA Val
LEU Tyr Ala Val PHE Lys Ala Tyr PRO Asp Ile Met Ala LYS PHE Pro Gln Phe ALA GLY LYS Asp Leu Asp Ser Ile Lys Asp Ser ALA Ala Phe ALA Thr HIS ALA Thr Arg Ile VAL Ser Phe LEU SER GLU Val Ile
LEU Tyr Phe PHE LEU Asn Lys Phe PRO GLY Asn Phe Pro Met PHE LYS Lys Leu GLY Asn Asp Leu ALA Ala Ala Lys Gly Thr ALA Glu Phe Lys Asp Gln ALA Asp Lys Ile Ile Ala Phe LEU Gln Gly Val Ile
LEU Ala ARG Met PHE Thr Val Tyr Ala Ala Thr Lys Ser Tyr PHE LYS Asp Tyr Lys Asp Phe Thr Ala ALA Ala PRO Ser Ile Lys Ala HIS GLY Ala Lys VAL VAL Thr ALA LEU Ala LYS ALA Cys
LEU Glu ARG Met PHE Ile Ala Tyr PRO Gln Thr Lys Thr Tyr PHE Pro HIS Phe Asp Leu Ser His Gly Ser ALA Gln Ile Leu Ala HIS GLY Lys Lys VAL Ala Ala ALA LEU Val GLU ALA Val
LEU Glu ARG Met PHE Ala Ser Phe PRO Thr Thr Lys Thr Tyr PHE Pro HIS Phe Asp Val Ser His Gly Ser ALA Gln Val Lys GLY HIS GLY Lys Lys VAL ALA Asp ALA LEU Ala Ser ALA Ala
LEU Ala ARG Leu LEU Ile Val Tyr PRO Trp Thr GLN Arg Phe PHE Ala Ser Phe Gly Asn Leu Ser Ser Pro Thr Ala Ile Leu Gly Asn PRO Met Val Arg Ala HIS GLY Lys Sys VAL Leu Thr Ser Phe Gly Asp ALA Val
LEU Gly ARG Leu LEU Val Val Tyr PRO Trp Thr GLN Arg Phe PHE Glu Ser Phe Gly Asp Leu Ser Thr Pro Asp Ala Val Met Gly Asn PRO Lys Val Lys Ala HIS GLY Lys Lys VAL Leu Gly ALA Phe SER Asp Gly Leu
LEU Ile ARG Leu PHE Lys Ser HIS PRO Glu Thr Leu GLU LYS PHE Asp Lys Phe Lys His Leu Lys Ser Glu Asp ASP Met Arg Arg Ser Glu Asp Leu Arg Lys HIS GLY Asn THR VAL Leu Thr ALA LEU Gly Gly Ile Leu
*** *** *** *** *** *** *** *** *** *** ***
Ile THR ARG PHE LEU SER Glu HIS PRO GLY GLY Glu GLU VAL Leu Leu Glu Gln ALA GLY ALA Asp Ala THR Glu Ser PHE GLU Asp VAL GLY HIS SER Pro Asp ALA Arg Glu Met Leu LYS Gln Tyr
LEU THR LYS PHE LEU GLU Glu HIS PRO GLY GLY Glu GLU VAL Leu Arg Glu Gln ALA GLY GLY Asp Ala THR Glu Asn PHE GLU Asp VAL GLY HIS SER THR Asp ALA Arg Glu LEU SER LYS Thr Tyr
LEU THR LYS PHE LEU GLU Glu HIS PRO GLY GLY Glu GLU VAL Leu Arg Glu Gln ALA GLY GLY Asp Ala THR Glu ASP PHE GLU Asp VAL GLY HIS SER THR Asp ALA Arg Glu Met SER LYS Thr Phe
LEU THR LYS PHE LEU GLU Glu HIS PRO GLY GLY Glu GLU VAL Leu Arg Glu Gln ALA GLY GLY Asp Ala THR Glx ASP PHE GLU Asp VAL GLY HIS SER THR Asp ALA Arg Glu LEU SER LYS Thr Phe
LEU THR LYS PHE LEU GLU Glu HIS PRO GLY GLY Glu GLU VAL Leu Arg Glu Gln ALA GLY GLY Asp Ala THR Glu ASP PHE GLU Asp VAL GLY HIS SER THR Asp ALA Arg Glu LEU SER LYS Thr Phe
Ile THR LYS PHE LEU Asp Glu HIS PRO GLY GLY Glu GLU VAL Leu Arg Glu Gln ALA GLY GLY Asp Ala THR Glu ASP PHE GLU Asp VAL GLY HIS SER THR Asp ALA Arg ALA LEU SER GLU Thr Phe
Val THR Asp PHE Val GLU LEU HIS PRO GLY GLY Pro ASP LYS Ile Leu Leu Ala ALA GLY GLY ALA Leu Glu Pro PHE Trp ALA Leu Tyr ALA Val HIS GLY GLU Pro His VAL Leu Glu LEU Leu Gln Gln Tyr
LEU THR ARG PHE LEU Pro Asn HIS PRO GLY GLY GLN ASP VAL Ile LYS PHE Asn ALA GLY LYS Asp VAL THR ALA Ile PHE GLU PRO Leu HIS ALA Pro Asn VAL Ile Asp Lys Tyr Ile ALA Pro
E20 EF1 EF2 EF3 EF4 EF5 EF6 EF7 EF8 F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 FG1 FG2 FG3 FG4 G1 G2 G3 G4 G5 G6 G7 G8 G9 G10 G11 G12 G13 G14 G15 G16 G17
96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143
Ala Ser Met ASP ASP Thr Glu Lys Met Ser Met Lys Leu Arg Asp LEU Ser Gly Lys HIS Ala Lys Ser Phe Gln VAL ASP PRO GLN TYR PHE Lys Val Leu Ala Ala Val Ile Ala Asp Thr Val ALA
Gly Gln LEU Lys Ala Ser Gly Thr Val Val Ala Asp Ala Ala LEU GLY Ser Val HIS Ala Gln Lys ALA VAL Thr Asn PRO GLU PHE Val Val Lys Glu Ala Leu Leu Lys Thr Ile Lys
Asp Ser LEU ASP ASP Pro Ser Arg Leu Lys Cys Val Val Glu LYS Phe Ala Val Asn HIS Ile ASN Arg LYS Ile Ser Gly Asp Ala PHE GLY Ser Ile Ile Pro GLU Met Lys Glu Thr LEU Lys
Asp Gln LEU ASP Asn Thr Asp Asp Leu Val Cys Val Val Glu LYS Phe Ala Val Asn HIS Ile Thr Arg LYS Ile Ser ALA Ala GLU PHE Gly Lys Ile Asn Gly Pro Ile Lys Lys Val LEU ALA
Asp Ser Met ASP ASP Ala Asp Cys Met Asn Gly Leu Ala Leu LYS LEU Ser Arg Asn HIS Ile Gln Arg LYS Ile Gly ALA SER Arg PHE Gly Glu Met Arg Gln Val Phe Pro Asn Phe LEU Asp
Asn Asp Ala Ala Asn Ala Gly Lys Met Ser Ala Met Leu Ser Gln Phe Ala Lys Glu HIS Val Gly Phe Gly VAL Gly Ser Ala Gln PHE Glu Asn Val Arg Ser Met Phe Pro Gly Phe Val ALA
Cys Ser Ile ASP ASP Asp Asp Cys Val Asx Gly Leu Ala LYS LYS LEU Ser Arg Asn HIS Leu Ala Arg Gly VAL Ser ALA Ala ASP PHE Lys Leu Leu Glu Ala Val Phe Lys Glx Phe LEU Asp
Ser Val LEU ASP ASP Lys Pro Val Leu Asp Gln Ala Leu Ala His Tyr Ala Ala Phe HIS Lys Gln Phe Gly Thr Ile Pro Phe Lys Ala PHE Gly Gln Thr Met Phe Gln Thr Ile Ala Glu
Ser Thr LEU ASP Gln Pro Ala Thr Leu Lys Glu Glu Leu Asp His LEU Gln Val Gln HIS Glu Gly Arg LYS Ile Pro Asp Asn TYR PHE Asp Ala Phe Lys Thr Ala Ile Leu His Val Val ALA
Ser His LEU Gly ASP Glx Gly Lys Met Val Ala Gln Met LYS Ala VAL GLY Val Arg HIS Lys Gly Tyr Gly Asn LYS His Ile LYS Gly GLN TYR PHE Glu Pro Leu Gly Ala Ser Leu Leu SER Ala Met Glu
Ser Leu Ala Gly Ser Asp Ala Asn Ile Pro Ala Ile Gln Asn Leu Ala LYS Glu LEU Ala Thr Ser HIS Lys Pro Arg Gly VAL Ser Lys Asp Gln PHE Thr Glu Phe Arg Thr Ala Leu Phe Thr Tyr LEU Lys
Glu Lys LEU Gly Ser Asp Met Gly Gly Ala Lys Ala Leu Leu Asn Gln LEU GLY Thr Ser HIS Lys Ala MET Gly Ile Thr Lys Asp Gln PHE Asp Gln Phe Arg Gln Ala Leu Thr Glu Leu LEU
Asp His LEU ASP ASP Leu Lys Thr His Leu His LYS LEU Ala Thr Phe HIS Gly SER GLU Leu LYS VAL ASP PRO Ala Asn PHE Gln Tyr Leu Ser Tyr Cys Leu Glu Val Ala LEU ALA
Asn His Val ASP ASP Ile Ala Gly Ala Leu Ser LYS LEU Ser ASP LEU HIS Ala Gln Lys Leu Arg VAL ASP PRO Val Asn PHE Lys Phe Leu Gly His Cys Phe Leu Val Val Ala
Gly His LEU ASP ASP Leu Pro Gly Ala Leu Ser Ala LEU Ser ASP LEU HIS Ala His Lys Leu Arg VAL ASP PRO Val Asn PHE Lys Leu Leu Ser His Cys Leu Leu Val Thr LEU ALA
Lys Asn LEU ASP Asn Ile Lys Asn Thr Phe Ala Gln LEU Ser GLU LEU HIS Cys ASP Lys Leu His VAL ASP PRO Glu Asn PHE Arg Leu Leu Gly Asp Ile Leu Ile Ile Val LEU ALA
Ala His LEU ASP Asn Leu Lys Gly Thr Phe Ala Thr LEU Ser GLU LEU HIS Cys ASP Lys Leu His VAL ASP PRO Glu Asn PHE Arg Leu Leu Gly Asn Val Leu Val Cys Val LEU ALA
Lys Lys Lys Gly His His Glu Ala Glu Leu LYS Pro LEU Ala Gln Ser HIS Ala Thr Lys His LYS Ile Pro Ile Lys TYR Leu Glu Phe Ile Ser Glu Ala Ile Ile His Val LEU His
*** *** *** *** *** *** *** *** *** *** ***
Tyr Ile GLY ASP Val HIS PRO ASN Asp Leu Lys Pro LYS
Ile Ile GLY GLU LEU HIS PRO ASN Asp Arg Ser LYS Ile
Ile Ile GLY GLU LEU HIS PRO ASP Asp Arg Pro LYS
Ile Ile GLY GLU LEU HIS PRO ASP Asp Arg Xxx Xxx Ile Ala LYS PRO SER GLU Thr Leu Ile Thr Thr Val GLU Ser Asn SER Ser Trp Trp
Ile Ile GLY GLU LEU HIS PRO ASP Asp Arg Ser LYS Ile Thr LYS PRO SER GLU Ser Ile Ile Thr Thr Ile Asp Ser Asn Pro Ser Trp Trp
Ile Ile GLY GLU LEU HIS PRO ASP Asp Arg Pro LYS Leu Arg
LYS VAL GLY GLU LEU Ser PRO ASP GLU Ala Pro Ala ALA Pro ASP ALA GLN ASP Pro Phe
Glu LYS LYS LEU GLY Pro LEU Glu Gly SER MET Pro Pro Glu Leu VAL Cys PRO PRO TYR Ala Pro Gly Glu Thr Lys

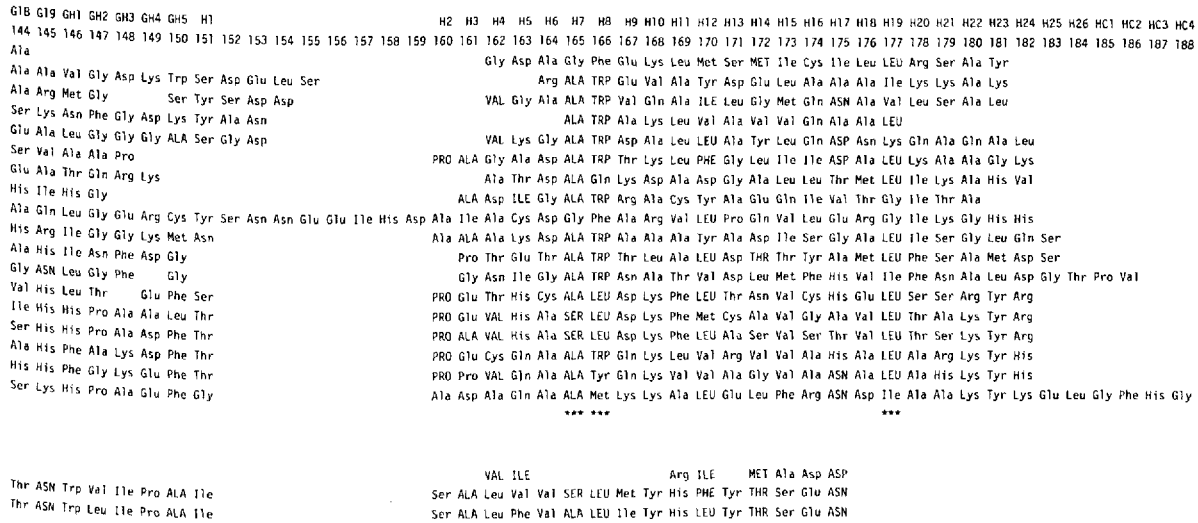


Fig. 1 *(facing page and above)*. Alignment of the amino acid sequences of 18 globins with the sequences of b_2 and b_5 cytochromes and CSO. Residues common to both sets of sequences are shown in capital letters; residues that are greatly conserved in globins are indicated by asterisks; open triangles mark the positions of introns in globin genes; and heme ligands are indicated by stars—the solid star is the common distal histidine and the open stars the alternate proximal histidines. The sequences illustrated are as follows: G1, *Petromyzon* hemoglobin (Hombrados et al. 1983); G2, soybean leghemoglobin Lbc (Sievers et al. 1978); G3, *Anadara trapezia* hemoglobin (Como and Thompson 1980); G4, *Anadara broughtonii* hemoglobin (Furuta and Kajita 1983), G5, *Cerithidea rhizophorarum* myoglobin (Takagi et al. 1983); G6, *Aplysia limacina* myoglobin (Suzuki et al. 1981); G7, *Busycon canaliculatum* myoglobin (Bonner and Laursen 1977); G8, *Tylorrhynchus heterochaetus* hemoglobin (Suzuki et al. 1982); G9, *Lumbricus terrestris* hemoglobin (Garlick and Riggs 1982); G10, *Glycera dibranchiata* hemoglobin (Imamura et al. 1972); G11, *Chironomus thummi* hemoglobin VI (Aschauer et al. 1981); G12, *Chironomus thummi* hemoglobin IIIa (Steer and Braunitzer 1981); G13, shark α-hemoglobin (Thompson 1980); G14, duck α-hemoglobin (Paddock and Gaubatz 1981); G15, *Rhea americana* β-hemoglobin (Oberthür et al. 1983); G16, human β-hemoglobin (Lawn et al. 1980); G18, seal myoglobin (Blanchetot et al. 1983); C1, rat outer membrane cytochrome b_5 (Lederer et al. 1983); C2, rat mitochondrial cytochrome b_5 (Lederer et al. 1983); C3–C6, human, pig, ox and chicken cytochromes b_5, respectively (Ozols 1974; Hagihara et al. 1975; van Jagow and Sebald 1980); C7, CSO (Guiard and Lederer 1979); C8, yeast cytochrome b_2 (Guiard and Lederer 1979)

Yeast cytochrome b_2 (yeast lactate dehydrogenase) and chicken sulfite oxidase (CSO) are composite proteins of the mitochondrial intermembranous space. They consist of N-terminal heme-bearing units attached to domains that contain other prosthetic groups (flavins and molybdenum; Guiard and Lederer 1979). These proteins are thought to have evolved by the fusion of originally separate genes, and the heme-bearing unit of each has a primary structure that is clearly homologous with that of the heme-bearing hydrophilic domain of vertebrate b_5 cytochromes (Guiard and Lederer 1979; Ghrir et al. 1984). A partial amino acid sequence of *Neurospora* assimilatory nitrate reductase (NNR) shows that it also belongs to this family of proteins (Lê and Lederer 1983).

Soon after the primary structure of the water-soluble segment of calf liver cytochrome b_5 was determined, Ozols and Strittmatter (1967) noted similarities in the amino acid sequences of cytochrome b_5 and vertebrate globins. Subsequent studies of Rossmann and Argos (1975) and Argos and Rossmann (1979) showed some similarities also between the secondary structures of these two kinds of proteins (see also Murthy 1984), but these similarities could not be used to exclude the possibility that the apparent relationship was due to convergence.

When Ozols and Strittmatter published their study, the secondary and tertiary structures of cytochrome b_5 were unknown. They therefore incorrectly identified a histidine residue at position 80 of the amino acid sequence (Mathews et al. 1972a) as the proximal heme ligand. This unsuccessful prediction led Mathews et al. (1972b) to conclude that the similarity in the amino acid sequences must be coincidental; consequently, Ozols and Strittmatter's proposal that b_5 and the vertebrate globins share a common ancestry has received little further attention.

In calf liver b_5 the heme group is supported by histidine side chains at positions 39 and 63 of the water-soluble segment (Mathews et al. 1972a). A similar situation seems to exist in yeast b_2 and CSO, as both proteins have histidine residues in positions homologous to those in b_5 (Guiard and Lederer 1979). By analogy with the globins, His-39 of b_5 may be termed the proximal histidine, since it faces the 'front' of the heme (Mathews 1980), and His-63 the distal histidine. This orientation of the heme was unknown to Argos and Rossmann, and it there-

fore diminishes the significance of their comparisons between the secondary structures of globins and cytochrome b_5 (Mathews 1980).

In yeast b_2, CSO, NNR and all b_5 cytochromes so far sequenced, there is a tryptophan residue 17 positions to the N-terminal side of the proximal histidine. A highly conserved tryptophan residue occurs in the A helices of almost all globins. When the globin and cytochrome b_5 sequences are aligned in the way proposed by Ozols and Strittmatter (1967), these two tryptophan residues are close to one another, and it was therefore thought that they might provide a suitable 'tie-line' for a comparison of the primary structures of the two sets of proteins.

Methods

The amino acid sequences of 18 distantly related globins were aligned as shown in Fig. 1 by using the well-conserved residues Leu-16, Lys-21, Trp-30, Gly-44, Pro-56, Phe-62, Gly-67, Ala/Pro-78, His-83, Leu-91, Leu-98, Asp-99, Asp-100, Leu-114, His-118, Phe-131, Leu-142, Ala-143, Ala-165, Trp-166, Leu-177 and Ala/Lys-180 (numbering as in Fig. 1). With the exception of small differences in the NA region and some inter-helical segments, the alignments shown in Fig. 1 correspond to those adopted by Lesk and Chothia (1980) in their comparative study of 11 globins in which the crystalline structure has been determined to 1.5–2.5-Å resolution. Because the inter-helical segments are of variable length (Fig. 4) and poorly conserved (Fig. 1), it is largely a matter of opinion as to how these parts of the molecules should be aligned; I have used an arrangement that minimizes both the total length of the aligned sequences and the number of gaps within them. In the N-terminal regions, I have assumed that the extra residue found in vertebrate β-globins but not in the vertebrate globins or myoglobins was inserted after the α- and β-genes diverged from each other. It is probable that invertebrate hemoglobins were derived independently from myoglobins that lacked this extra residue.

Insofar as it was possible, I chose the globins for this study at random from the sequences available. However, I considered it desirable to include, where possible, those globins for which DNA sequence data are available, and to provide some numerical balance among sequences from different major higher taxa. Because a very limited number of invertebrate globins have been sequenced, the within-taxon sampling is far from ideal (e.g., two species of *Anadara* from the Bivalvia). Nevertheless, the 18 globin sequences of Fig. 1 embrace the polychaete (*Glycera*) and oligochaete Annelida (*Tylorrhynchus* and *Lumbricus*) the uniramian arthropods (*Chironomus*), the bivalved Mollusca (*Anadara*), the gastropod Mollusca (*Aplysia, Cerithidea* and *Busycon*), the jawless fish (*Petromyzon*), cartilaginous fish, birds and mammals, plus a globin from the root nodules of the soybean. They also sample distantly related members of the three main vertebrate globin families, the α- and β-hemoglobins and the myoglobins.

The globin sequences were compared with an analogous but smaller set of cytochrome b_5 sequences and the sequences of yeast cytochrome b_2 and CSO (Fig. 1). The alignments adopted for these sequences are those given by Ozols (1974), Hagihara et al. (1975), Guiard and Lederer (1979), von Jagow and Sebald (1980) and Lederer et al. (1983), and thus are independent of this study. Again, the taxonomic distribution of available sequences is far from ideal, as it may be assumed that all mammalian cytochrome b_5 genes have descended from a common ancestral gene within the last few hundred million years. Thus although several mammalian sequences are shown in Fig. 1 for illustrative purposes, they are given relatively little weight in the statistical analysis of the similarity of cytochrome and globin sequences.

The tryptophan residue (Trp-30, Fig. 1) that is found in most globins and in each of the b_5, b_2 and CSO sequences was used to align the N-terminal ends of both sets of sequences. Further tie-lines were provided by three other apparently conserved residues, Pro-56, His-83 and His-118 (Fig. 2). To correlate these four sets of residues and others within the chains it was necessary to separate the cytochrome and CSO sequences into four parts (segments 1–4, Fig. 4), each of which is tied to the globin sequences by one of the four residues mentioned above. To achieve a better fit of segment 1 with the N-terminal part of the globin set, it was also assumed that His/Arg-22 of the cytochrome sequences is missing in the globins (Figs. 1 and 4). This is the only deletion postulated to have occurred in the production of a globin from a b cytochrome.

To test the statistical significance of the alignments shown in Fig. 1, I used the following methods. The probability (ρ) of any particular amino acid being found at any site was assumed to be equal to 0.01 times the average percentage abundance of that amino acid in the 18 globin sequences (2655 residues) in the case of the globins, or to 0.01 times the average percentage abundance of that amino acid in 118 different proteins (18,115 residues; Cornish-Bowden 1983) in the case of the b cytochromes and CSO. Using the expression

$$P = \frac{n!}{x!(n-x)!} \cdot \rho^x \cdot (1-\rho)^{n-x},$$

where x is the observed frequency of any particular amino acid in n sequences and ρ is the expected probability, the probability (P) that the presence of a common or shared amino acid at any position is due to chance was calculated for each set of sequences separately. When two or more common residues were found at the same position, the probabilities calculated for each kind of residue were multiplied together. Distributions that are significant at the 0.05, 0.01 and <0.001 levels are shown by one, two or three asterisks, respectively, in Fig. 2. Some of the calculated probabilities are exceedingly small. For example, the probability of leucine occurring at position 16 with the observed frequencies in both sets of proteins is about 4×10^{-10}. Some other values are 1×10^{-12} (Lys/Arg-21), 3×10^{-10} (Leu-48), 1×10^{-29} (Pro-56) and 4×10^{-14} (Gly-67).

All 18 globin sequences were given equal weight in the analysis because the data set consists of sequences or pairs of sequences that are distantly related. However, it was felt that the vertebrate b_5 sequences would bias the analysis of the cytochrome data set if each was given equal weight, and therefore the very similar rat, human, pig and ox sequences were treated as a single unit. Thus the values of n used were 18 for the globins and 5 for the cytochromes.

Results

It can be seen from Fig. 2 that statistically significant similarities are widespread in all four segments (positions 7–35, 39–73, 76–95 and 111–132). In segment 1, 37% of the positions have statistically significant occurrences of the same residues in both sets of sequences; in segments 2–4 the values are 29%, 25% and 32%, respectively (mean, 31%). These values are much higher than those expected on the basis of chance, and they are improved (mean, 35%)

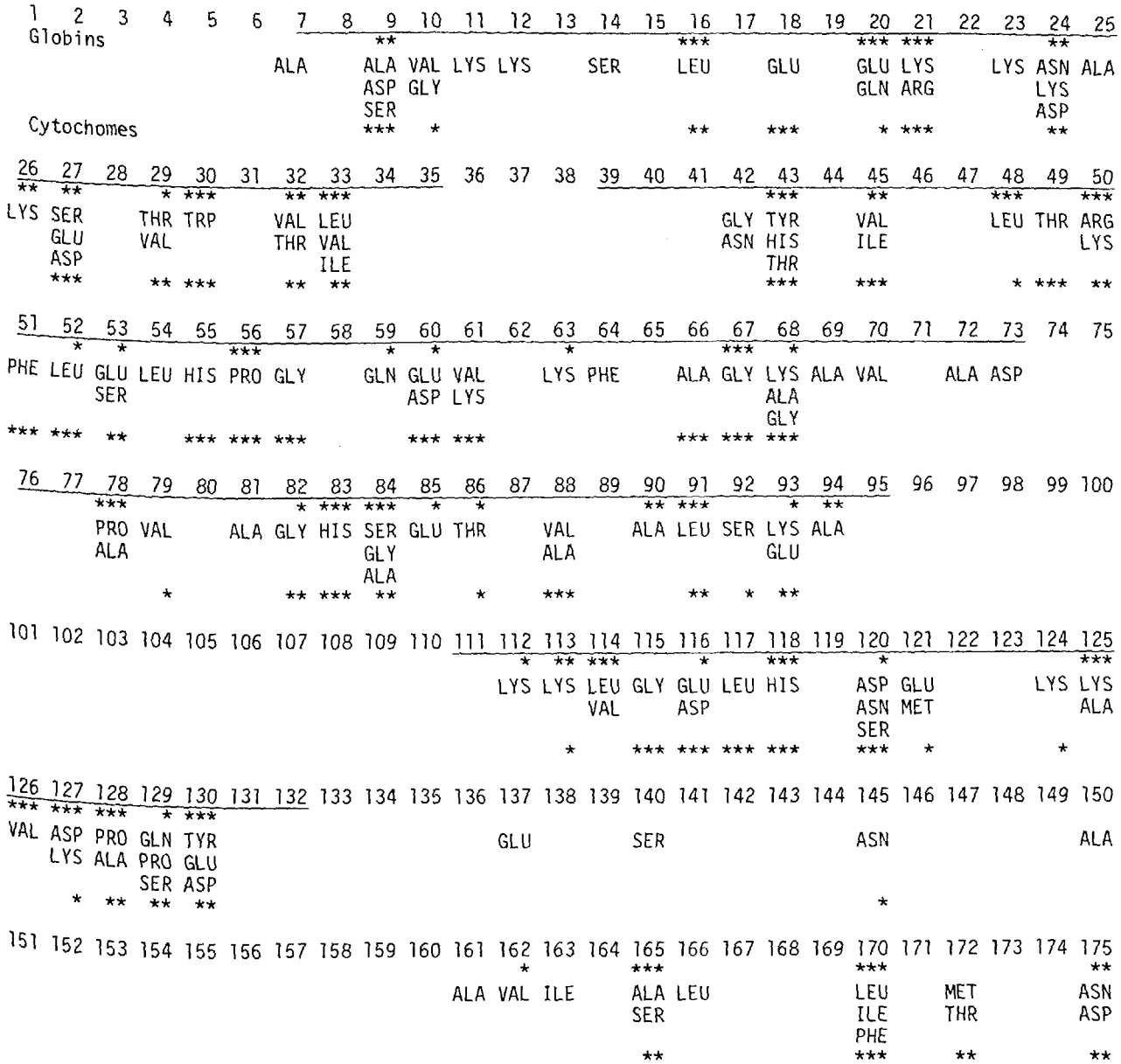


Fig. 2. Statistical comparison of the similarities between the 18 globin and 8 cytochrome sequences shown in Fig. 1. The residue positions are numbered as in Fig. 1; underlined segments are those numbered 1–4 in Fig. 4. Residues present in the same positions in each set of sequences are listed, and those that have statistically significant distributions in each set of sequences are marked by asterisks. See text for additional explanation

if the residues used to tie each segment to the globin sequences are included. For example, the expected frequency for segment 1 is approximately 29/400 (7%), where 29 is the number of positions compared and 400 (20 × 20) is an estimate of the average probability that any one amino acid will occur at the same position in each set of sequences.

As might be expected, some globin sequences resemble the cytochrome sequences more closely than do others. The partial sequences shown in Fig. 3 illustrate that close similarities can be demonstrated between parts of individual polypeptide chains.

Discussion

The first conclusion to be drawn from these comparisons is that the distal heme ligand (His-83 or its equivalent in some globins) is common to both sets of proteins, but the proximal ligand of the cytochromes, CSO and NNR is replaced by the invariant histidine residue at position 118 (F8) in the globins. This explains why the heme is reversed between globins and b_5 cytochromes (Mathews 1980). As might be expected, all cytochromes and all globins have a histidine in the proximal position (His-

	NA1	NA3	A1	A2	A3	A4	A5		A6	A7	A8	A9	A10	A11	A12	A13	A14	A15
Globin (1)	SER	LEU	Ser	ASP	Lys	ASP	LYS		Ala	ASP	Val	LYS						
Cytochrome	SER	LEU	Glu	GLU	Val	GLU	LYS	His	Asn	ASP	Ser	LYS	GLU	Thr	TRP	Val	VAL	LEU
Globin (2)												LYS	GLU	Ser	TRP	Lys	VAL	LEU

	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	B16	C1	C2	C3	C4	C5
Globin (3)	GLY	Ser	Gly	VAL	Glu	Ile	LEU	Tyr	Phe	PHE	LEU							
Cytochrome	GLY	TYR	Lys	VAL	Tyr	Asp	LEU	THR	ARG	PHE	LEU	GLU	GLU	HIS	PRO	Gly	Gly	GLN
Globin (4)									ARG	Leu	LEU	Gln	GLU	HIS	PRO	Glu	Thr	GLN
Globin (5)	Gln	TYR	Ser	VAL	Val	Phe	Tyr	THR	Ser	Ile	LEU	GLU	Lys	Ala	PRO	Ala	Ala	GLN

Fig. 3. Segments of globin sequences that closely resemble the greatly conserved vertebrate cytochrome b_5 sequences (center). 1, *Catostomus clarkii* α-hemoglobin (Croft 1980); 2, *Busycon* myoglobin; 3, *Chironomus* hemoglobin IIIa; 4, alligator myoglobin (Dene et al. 1980); 5, soybean leghemoglobin Lbc

55 and His-118, respectively), but only some cytochromes have a histidine at position 118 and only some globins a histidine at position 55. The presence of His-118 in some cytochromes is considered to represent the ancestral condition that existed before the cytochromes and the globins diverged, and the presence of His-55 to be a relic that similarly has survived in an unmodified condition in some globins.

The only other invariant residue of normal globins is the phenylalanine at position 62 (CD1), which lies parallel to the heme on its distal side and wedges it into its crevice. In calf liver b_5 a leucine residue at this position is in van der Waals contact with the proximal histidine (Hagihara et al. 1975). This residue is conserved as either leucine or isoleucine in all of the b_5 cytochromes and in b_2 and CSO.

There are 23 positions marked by asterisks in Fig. 1. These represent sites (excluding the heme ligands) at each of which a single residue is conserved in at least 9 of the 18 globin sequences. These positions should, more than any others, reflect the condition in the ancestral globin. It is therefore of interest that the same residue is found in a corresponding position in at least one of the cytochrome sequences at the following sites: Leu-16 (invariant in vertebrate b_5); Lys-21 (Arg in CSO, Lys in the other sequences); Trp-30 (invariant in the cytochromes); Leu-48; Phe-52 (frequently Leu in globins and Leu in all cytochrome sequences except CSO); Pro-56 (invariant in the cytochromes); Phe-62 (Leu or Ile in the cytochromes); Leu-91 (Leu or Met in all cytochrome sequences except b_2); Leu-114 (Leu, Val or Ile in the cytochromes); Val-126 (Val in b_2); and Ser-Leu-165-166. Thus most of the residues that are conserved in most globins could have been inherited from an ancestral b_5-like cytochrome.

The tryptophan residue at position 30 (A12) must have a significant and perhaps unrecognized function, because it is conserved in most globins, b_5 and b_2 cytochromes, CSO and NNR. In the latter group of proteins (the cytochrome group) it may support the N-terminal helix through a π–π interaction with His-22 (Guiard and Lederer 1979), but its role in globins is not well understood, even though its structural relationships are known precisely (Lesk and Chothia 1980).

The similarity between the globin and cytochrome sequences stops rather abruptly near position 131 (Figs. 1 and 2). To a certain extent this is because few cytochrome sequences are available [the b_2 sequence for this region (Ghrir et al. 1984) appeared after Figs. 1 and 2 were prepared]. However, there is little similarity between the dominantly hydrophobic 'tail' of calf and other b_5 cytochromes (von Jagow and Sebald 1980) and the G and H helical regions of the globins. This is not entirely unexpected, because only the N-terminal parts of b_2 and b_5 cytochromes and CSO are homologous; the hydrophobic membrane anchor of b_5 proteins and the analogous regions of CSO and b_2 (and presumably NNR) are thought to have been independently joined to the homologous N-terminal regions by gene fusion (Guiard and Lederer 1979).

The presumed positions of the junctions between the homologous and analogous sequences of b_5, b_2 and CSO sequences correspond almost exactly with the position of the second intron of vertebrate globin genes (Fig. 4). It is therefore possible that globins are also composite proteins constructed from a homolog of the water-soluble b_5 segment plus a C-terminal module coded by an exotic exon. Conversely, it is possible to predict that the cytochrome b_5 gene sequences will prove to have an intron at the end of the region coding for the hydrophilic domain (Fig. 4).

Seal myoglobin lacks the third central intron of the leghemoglobin genes (Jensen et al. 1981; Blanchetot et al. 1983), thus diminishing the likelihood that the presence of three introns in globin genes represents the primitive state (Blake 1981, 1983; Lewin 1981, 1983). Although Gō's successful prediction of the existence and location of the third intron in the leghemoglobin genes (Blake 1981; Gō 1981) shows that the positions of introns may be related to the symmetry and topology of the finished

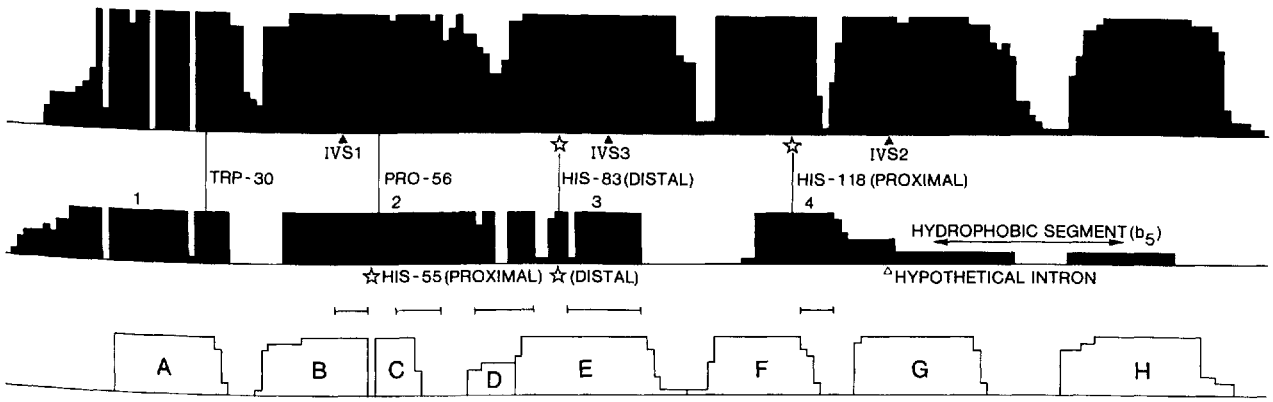


Fig. 4. Histograms of the residues/position of the globin (top) and cytochrome sequences (middle) of Fig. 1. The histograms at the bottom show the distribution of globin residues that are incorporated in α-helices (Lesk and Chothia 1980); the horizontal bars show the α-helical regions of calf liver cytochrome b_5 (Mathews et al. 1972). Note that most insertions/deletions in globins (low points in upper histogram) lie between the conserved helical regions, and that there is an approximate correspondence between the globin helical regions and the four segments (1–4) of the cytochrome sequences. Tie-lines between the globin and cytochrome sequences are indicated by the named residues; IVS1, IVS2 and IVS3 are the positions of the globin introns. The location of a hypothetical intron in the cytochrome sequences is shown at the junction of the hydrophilic and hydrophobic domains of the b_5 sequences. Stars indicate heme ligands

product, it is difficult to find a functional explanation for the presence of an intron in the middle of the region that encodes the heme-bearing domain (Craik et al. 1980). If an equivalent intron exists in invertebrate globins, it may be an historical relic reflecting the fact that in a b_5 cytochrome such an intron would *not* lie within the sequence coding for the heme-bearing domain. Its absence from vertebrate globins could then be explained as being due to selection for an uninterrupted coding sequence in this critical region.

In previous attempts to relate b_5 cytochromes to the globins it was assumed that the primary heme ligands would remain unchanged and that there should be a close similarity between the secondary structures of the two kinds of molecules (Ozols and Strittmatter 1967; Rossman and Argos 1975; Argos and Rossmann 1979). However, it is also possible that the evolution of new kinds of proteins results from small but critical changes in the primary structure that cause dramatic changes in both the tertiary structure (e.g. Dijkstra et al. 1983) and the function of the molecules.

In calf liver b_5 the heme lies in a cavity that is walled by short α-helical segments and floored by a β structure (Mathews et al. 1972b). The N-terminus lies beneath the floor of the heme crevice and the C-terminal chain of the hydrophilic domain lies to one side (Fig. 5). Such a molecule could have been converted into a 'protoglobin' in the following way: If the C-terminal segment had been lengthened by about ten residues (positions 96–112) at its N-terminal end, the C-terminal helix could have come to lie across the mouth of the heme crevice (Fig. 5). Given the development of this typology, it is not difficult to imagine some of the other modifications that would have been required to convert a type b_5 cytochrome into an ancestral globin. The heme would have needed to be detached from its proximal ligand (His-55) and attached to the histidine (His-118) in the middle of the C-terminal helix. Given the fact that apomyoglobin reconstituted with deuteroheme derivatives contains a disordered fraction in which the heme is rotated through 180° about the α–γ axis (Ahmad and Kincaid 1983), this development could have required little more than the formation of an appropriate tertiary structure. Mutations or insertions could then have lengthened the helical part of this chain to develop the globin F helix (Fig. 5), and sequential changes elsewhere could have converted the β floor of the cytochrome heme pocket into parts of the A, B and D helices of the globins. As can be seen in Fig. 4, there is an approximate correspondence between the four cytochrome segments identified from the amino acid sequence comparisons and the major helical regions of the globins. If the gene fusion model proposed above is correct, the globin G and H helices were inherited from another, unrelated gene.

If the globins were derived by the fusion of two gene segments as outlined above, it is unlikely that the transition could have occurred more than once. Both regions are clearly homologous in all globins so far studied, and the probability that two unrelated gene segments fused twice to form a single functional unit must be vanishingly small. When this observation is coupled with the complex structural changes that appear to have been required to develop the globin tertiary structure, it is even more improbable that the globins evolved by convergent evolution. On the other hand, it may no longer be possible to use the presence of greatly conserved residues or the

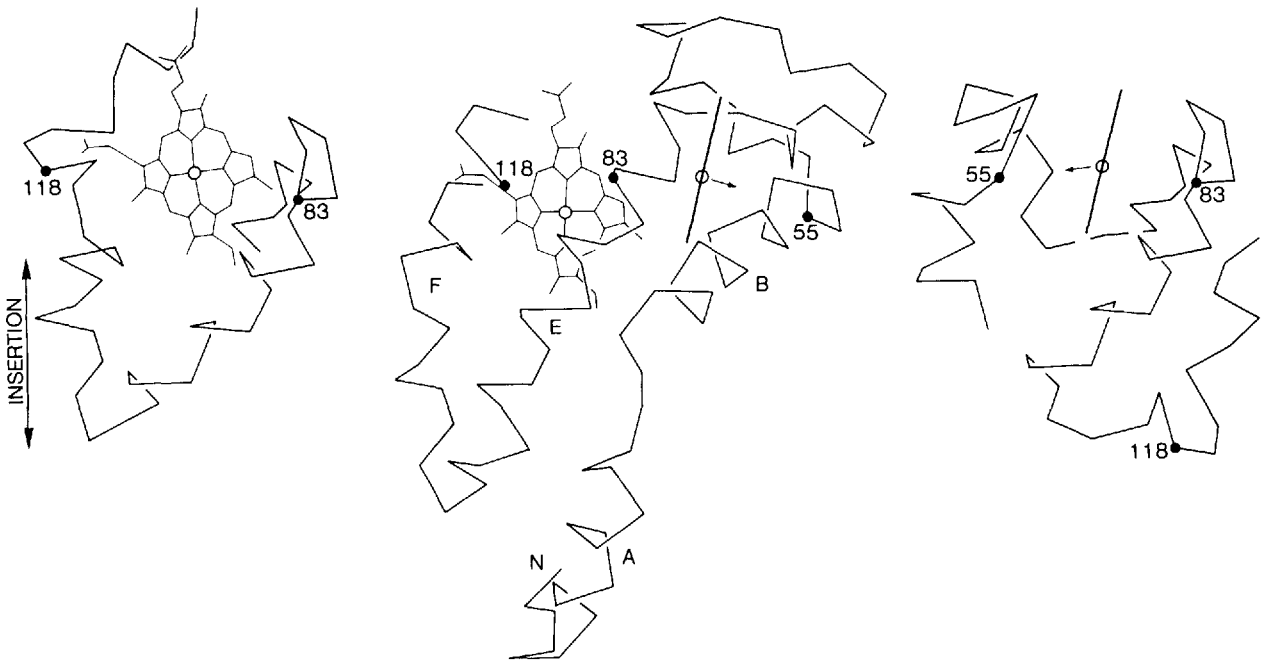


Fig. 5. Schematic views of the structural changes that would be required to convert the heme-bearing domain of a cytochrome b_5 molecule into an analogous region in a globin. In calf liver b_5 [right, redrawn from Mathews et al. (1972b)], the heme is held between His-55 and His-83 (numbering as in Fig. 1). The central helix of this structure is thought to be equivalent to the E helix of a globin [center, redrawn from Richardson (1981)], and the C-terminal chain of b_5 to the upper part of the globin F helix. An insertion of about 10 residues (positions 96–112 of Fig. 1) would be required to place His-118 opposite the conserved distal histidine, His-83 (left). The hemes are reversed in the two kinds of structure, but the reversal seems to result from a reorientation of the helical segments rather than a movement of the heme across the E helix (center). The arrows point to the histidine that is proximal in cytochrome b_5. [See Poulos and Mauk (1983) for computer-drawn stereoscopic views of both structures at an equivalent scale]

positions of introns as evidence that the globins evolved from a single ancestral gene (Lee et al. 1983), as all of these features could have been inherited from a b_5-like cytochrome. This matter may be clarified when the nucleic acid sequences of b_5 cytochromes and invertebrate globins become available.

It is, however, more difficult to use the evidence presented above to argue that all invertebrate phyla that synthesize globins are monophyletic. The presence of leghemoglobins in leguminous (Lee et al. 1983) and non-leguminous (Appleby et al. 1983) plants may be explained in two ways: Either the genes for these proteins were inherited from a common ancestor of animals and plants, or the genes were transferred to angiosperms from an animal during the Mesozoic (Lewin 1981; Lee et al. 1983). The former is possible in that the taxonomic distribution of globins may be inadequately known; the latter is likely in view of the unusual and specific association of *Rhizobium* in nitrogen-fixing root nodules of the plants. Thus either process could explain the distribution of globins in the animal kingdom.

Acknowledgments. I thank M.T.C. Runnegar for providing most of the literature, F.A. Shaw for the computing and V.J. Bofinger for advice on the statistics.

References

Ahmad MB, Kincaid JR (1983) Haem disorder in modified myoglobins. Biochem J 215:117–122

Anderson DT (1981) Origins and relationships among the animal phyla. Proc Linn Soc NSW 106:151–166

Appleby CA, Tjepkema JD, Trinick MJ (1983) Hemoglobin in a nonleguminous plant, *Parasponia*: possible genetic origin and function in nitrogen fixation. Science 220:951–953

Argos P, Rossmann MG (1979) Structural comparisons of heme binding proteins. Biochemistry 18:4951–4960

Aschauer H, Zaidi ZH, Braunitzer G (1981) Amino acid sequence of a dimeric hemoglobin (erythrocruorin), component VI from *Chironomus thummi thummi* (CCT VI). Hoppe Seylers Z Physiol Chem 362:261–273

Blake CCF (1981) Exons and the structure, function and evolution of haemoglobin. Nature 291:616

Blake CCF (1983) Exons and the evolution of proteins. Tr Biochem 8:11–13

Blanchetot A, Wilson V, Wood D, Jeffreys AJ (1983) The seal myoglobin gene: an unusually long globin gene. Nature 301: 732–734

Bonner AG, Laursen RA (1977) The amino acid sequence of a dimeric myoglobin from the gastropod mollusc, *Busycon canaliculatum* L. FEBS Lett 73:201–203

Como PF, Thompson EOP (1980) Amino acid sequence of the α-chain of the tetrameric haemoglobin of the bivalve mollusc *Anadara trapezia*. Aust J Biol Sci 33:653–664

Cornish-Bowden A (1983) The amino acid compositions of

proteins are correlated with their molecular sizes. Biochem J 213:271–274

Craik CS, Buchman SR, Beychok S (1980) Characterization of globin domains: heme binding to the central exon product. Proc Natl Acad Sci USA 77:1384–1388

Croft LR (1980) Handbook of protein sequence analysis, 2nd edn. John Wiley & Sons, New York

Dene H, Sazy J, Goodman M, Romero-Herrera AE (1980) The amino acid sequence of alligator (*Alligator mississippiensis*) myoglobin. Biochim Biophys Acta 624:397–408

Dijkstra BW, Weijer WJ, Wierenga RK (1983) Polypeptide chains with similar amino acid sequences but a distinctly different conformation. FEBS Lett 164:25–27

Furuta H, Kajita A (1983) Dimeric hemoglobin of the bivalve mollusc *Anadara broughtonii*: complete amino acid sequence of the globin chain. Biochemistry 22:917–922

Garlick RL, Riggs AF (1982) The amino acid sequence of a major polypeptide chain of earthworm hemoglobin. J Biol Chem 257:9005–9015

Ghrir R, Becam A-M, Lederer F (1984) Primary structure of flavocytochrome b_2 from baker's yeast. Eur J Biochem 139: 59–74

Gō M (1981) Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90–92

Guiard B, Lederer F (1979) The "cytochrome b_5 fold": structure of a novel protein superfamily. J Mol Biol 135:639–650

Hagihara B, Sato N, Yamanaka T (1975) Type b cytochromes. In: Boyer PD (ed) The enzymes, vol 11A. Academic Press, New York, San Francisco, London, pp 549–593

Hombrados I, Rodewald K, Neuzil E, Braunitzer G (1983) Haemoglobins, LX. Primary structure of the major haemoglobin of the sea lamprey *Petromyzon marinus* (var. Garonne, Loire). Biochimie 65:247–257

Imamura T, Baldwin TO, Riggs A (1972) The amino acid sequence of the monomeric hemoglobin component from the bloodworm, *Glycera dibranchiata*. J Biol Chem 247:2785–2797

Irie T, Usuki I (1980) Disparity of native oxyhemoglobin components isolated from *Paramecium caudatum* and *Paramecium primaurella*. Comp Biochem Physiol [B] 67:549–554

Jensen EØ, Paludin K, Hyldig-Nielsen JJ, Jørgensen P, Marcker KA (1981) The structure of a chromosomal leghaemoglobin gene from soybean. Nature 291:677–679

Lawn RM, Efstratiadis A, O'Connell C, Maniatis T (1980) The nucleotide sequence of the human β-globin gene. Cell 21:647–651

Lê KHD, Lederer F (1983) On the presence of a heme-binding domain homologous to cytochrome b_5 in *Neurospora crassa* assimilatory nitrate reductase. EMBO J 2:1909–1914

Lederer F, Ghrir R, Guiard B, Cortial S, Ito A (1983) Two homologous cytochromes b_5 in a single cell. Eur J Biochem 132:95–102

Lee JS, Brown GG, Verma DPS (1983) Chromosomal arrangement of leghemoglobin genes in soybean. Nucleic Acids Res 11:5541–5553

Leonard K, Wingfield P, Arad T, Weiss H (1981) Three-dimensional structure of ubiquinol : cytochrome c reductase from *Neurospora* mitochondria determined by electron microscopy of membrane crystals. J Mol Biol 149:259–274

Lesk AM, Chothia C (1980) How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. J Mol Biol 136: 225–270

Lewin R (1981) Evolutionary history written in globin genes. Science 214:426–429

Lewis R (1983) Myoglobin gene is a big surprise. Science 219: 1312

Mathews FS (1980) The orientation of the heme group in crystalline cytochrome b_5. Biochim Biophys Acta 622:375–379

Mathews FS, Argos P, Levine M (1972a) The structure of cytochrome b_5 at 2.0 Å resolution. Cold Spring Harbor Symp Quant Biol 36:387–395

Mathews FS, Levine M, Argos P (1972b) Three-dimensional Fourier synthesis of calf liver cytochrome b_5 at 2.8 Å resolution. J Mol Biol 64:449–464

Murthy MRN (1984) A fast method of comparing protein structures. FEBS Lett 168:97–102

Nobrega FG, Tzagoloff A (1980) Assembly of the mitochondrial membrane system. J Biol Chem 255:9828–9837

Oberthür W, Braunitzer G, Baumann R, Wright PG (1983) The primary structures of α- and β-chains from the major hemoglobin component of ostrich (*Struthio camelus*) and American rhea (*Rhea americana*). Hoppe Seylers Z Physiol Chem 364: 119–134

Ozols J (1974) Cytochrome b_5 from microsomal membranes of equine, bovine, and porcine livers. Isolation and properties of preparation containing the membranous segment. Biochemistry 13:426–434

Ozols J, Strittmatter P (1967) The homology between cytochrome b_5, hemoglobin, and myoglobin. Proc Natl Acad Sci USA 58:264–267

Paddock GV, Gaubatz J (1981) Nucleotide sequence for a novel duck alpha-globin gene. Eur J Biochem 117:269–273

Poulos TL, Mauk AG (1983) Models for the complexes formed between cytochrome b_5 and the subunits of methemoglobin. J Biol Chem 258:7369–7373

Richardson JS (1981) The anatomy and taxonomy of protein structure. Adv Protein Chem 34:167–339

Rossmann MG, Argos P (1975) A comparison of the heme binding pocket in globins and cytochrome b_5. J Biol Chem 250:7525–7532

Saraste M (1984) Location of haem-binding sites in the mitochondrial cytochrome b. FEBS Lett 166:367–372

Sievers G, Huhtala M-L, Ellfolk N (1978) The primary structure of soybean (*Glycine max*) leghemoglobin c. Acta Chem Scand [B] 32:380–386

Steer W, Braunitzer G (1981) Die Primarstruktur eines monomeren Insektenhamoglobins (Erythrocruorin), Komponente CTT IIIa von *Chironomus thummi thummi*. Hoppe Seylers Z Physiol Chem 362:73–80

Suzuki T, Takagi T, Shikama K (1981) Amino acid sequence of myoglobin from *Aplysia kurodai*. Biochim Biophys Acta 669:79–83

Suzuki T, Takagi T, Gotoh T (1982) Amino acid sequence of the smallest polypeptide chain containing heme of extracellular hemoglobin from the polychaete *Tylorrhynchus heterochaetus*. Biochim Biophys Acta 708:253–258

Takagaki Y, Radhakrishnan R, Gupta CM, Khorana HG (1983) The membrane-embedded segment of cytochrome b_5 as studied by cross-linking with photoactivatable phospholipids. J Biol Chem 258:9128–9135

Takagi T, Tobita M, Shikama K (1983) Amino acid sequence of dimeric myoglobin from *Cerithidea rhizophorarum*. Biochim Biophys Acta 745:32–36

Thompson EOP (1980) Amino acid sequences of globin chains and their use in phylogenetic divergence point estimations. In: Sigman DS, Brazier MAB (eds) The evolution of protein structure and function. Academic Press, New York, London, Toronto, Sydney, San Francisco, pp 267–298

von Jagow G, Sebald W (1980) Type b cytochromes. Annu Rev Biochem 49:281–314

Widger WR, Cramer WA, Herrmann RG, Trebst A (1984) Sequence homology and structural similarity between cytochrome b of mitochondrial complex III and the chloroplast b_6–f complex: position of the cytochrome b hemes in the membrane. Proc Natl Acad Sci USA 81:674–678

Received September 26, 1983/Revised July 31, 1984