Evolution of Globins

- family of proteins
- best known – haemoglobin, myoglobin
  - involved in oxygen transport
  - also bind other ligands (CO)
- first crystal structures of proteins
- many studies of evolution
  - sequence, structure, function
- allosteric change of haemoglobin
- links with clinical practice
  - sickle-cell anaemia = first 'molecular disease'

Evolution of globin structure

- mammals, insects, worms, plants
- *Chironomus* erythrocruorin
- lupin leghaemoglobin

- what is conserved?
- how do structural changes reflect sequence changes?

Investigations 'finished' years ago

- divergence of sequences
- divergence of structures
- effects of mutations

- But – new set of homologues found

Evolution of globins – highlights

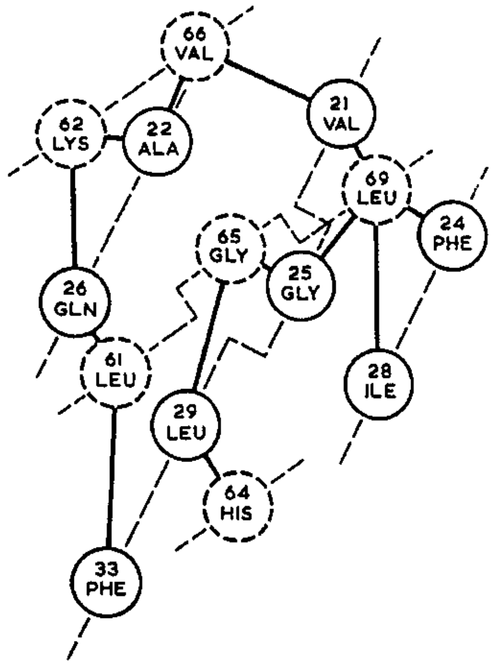
- 1909 – Reichert and Brown
- 1949 – sickle-cell anaemia is a molecular disease
- 1956 – HbS mutation
- 1959 – myoglobin structure
- 1961 – haemoglobin sequences
- 1970 – nonmammalian structures
- 1985 – phycocyanin structure
- 1989 – truncated structure

Haemoglobin crystals

- first discovered in 1840
- Steno's law (1669):
  constancy of interfacial angles
- Reichert and Brown compared angles of Hb crystals from different fishes
- Divergence of interfacial angles consistent with evolutionary tree

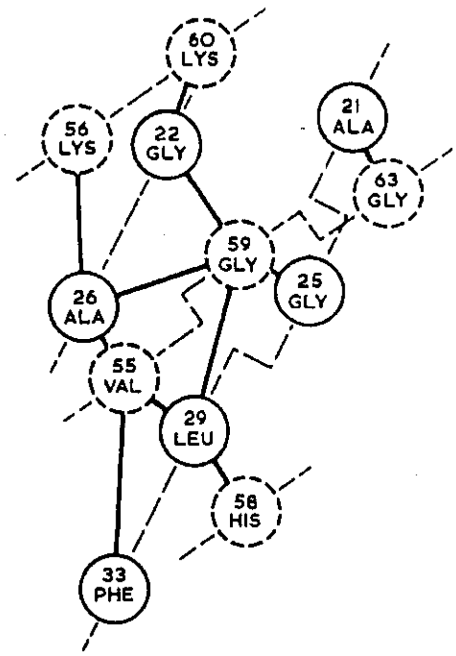
Sperm whale myoglobin

- 'classical' globin structure
- about 153 residues
- 8 or 9 α-helices
- standard folding pattern



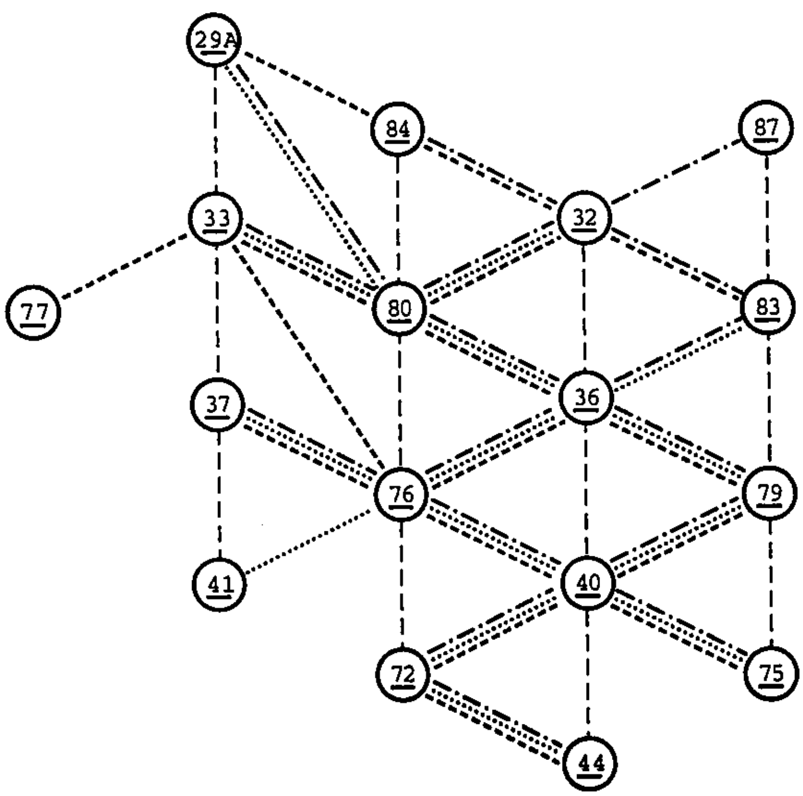
66 VAL
62 LYS
22 ALA
21 VAL
69 LEU
24 PHE
65 GLY
25 GLY
26 GLN
61 LEU
28 ILE
29 LEU
64 HIS
33 PHE

SWMb

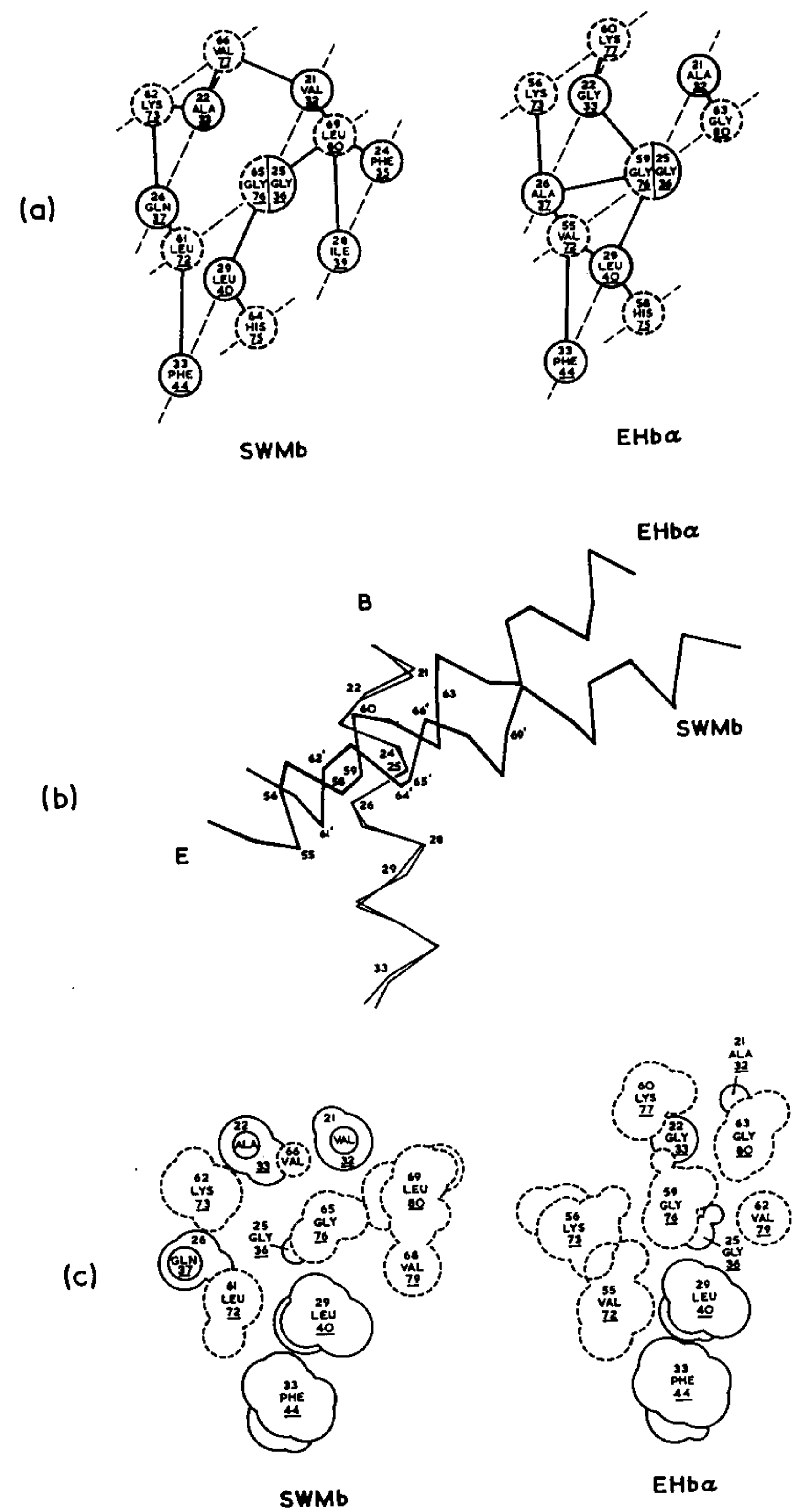


60 LYS
56 LYS
22 GLY
21 ALA
63 GLY
59 GLY
25 GLY
26 ALA
55 VAL
29 LEU
58 HIS
33 PHE

EHbα



29A
84
87
33
32
77
80
83
37
36
76
79
41
40
72
75
44


(a)
SWMb
EHbα
(b)
EHbα
B
SWMb
E
(c)
SWMb
EHbα

Conservation of secondary structure

- globins contain 8/9 helices:
  A, B, C, (D), E, F, G, H
- C helix is a 3_{10} helix
- lengths of helices vary somewhat
- in mammalian globins F helix broken (F′-F)
- create a pocket for haem
- iron linked to residues in F, G helices

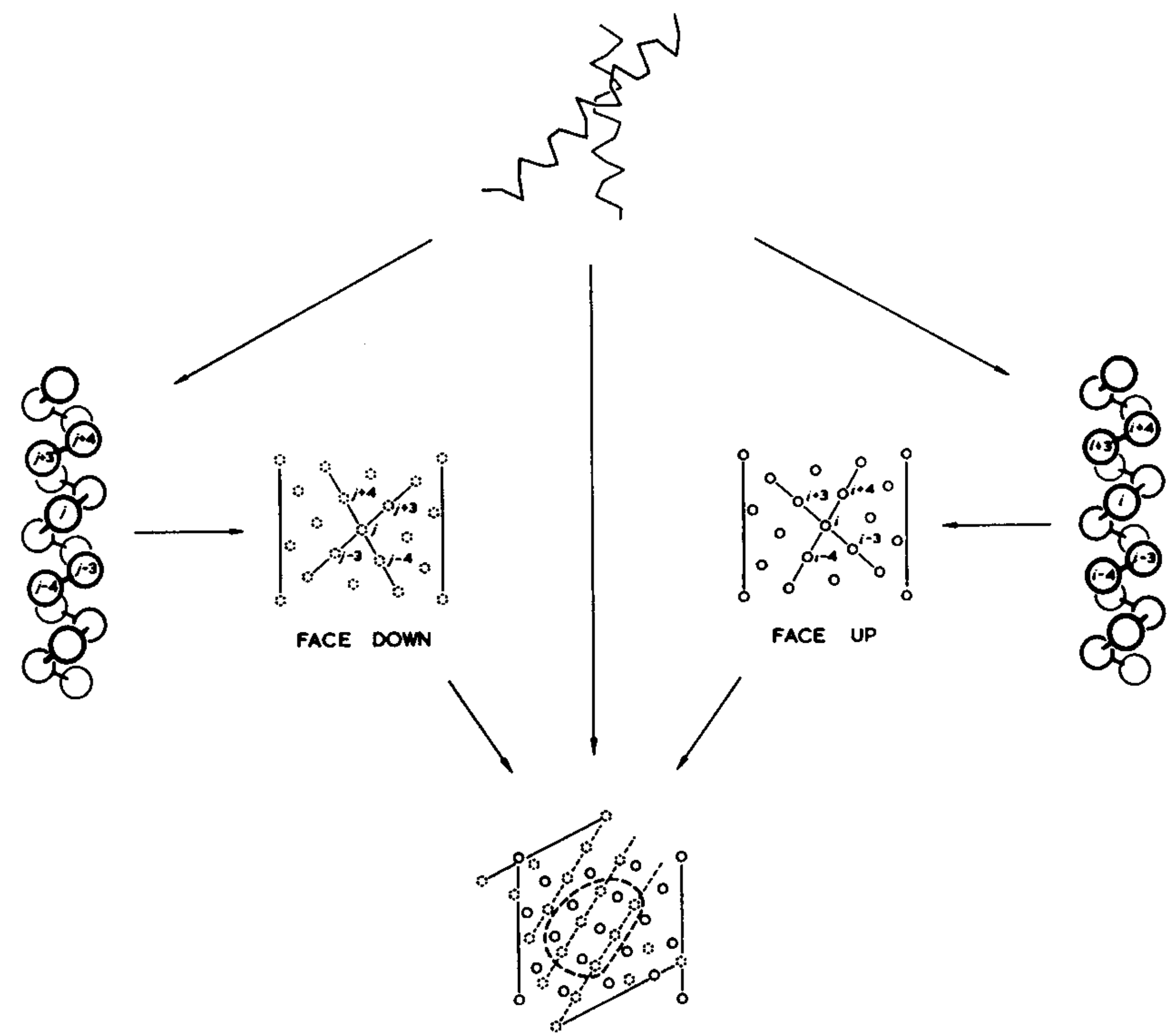
Conservation of tertiary structure

- common folding pattern
- globins stabilized by packings at helix-helix interfaces
- five helix packings common: A/E, B/E, B/G, F/H, G/H
- other packings in some but not all

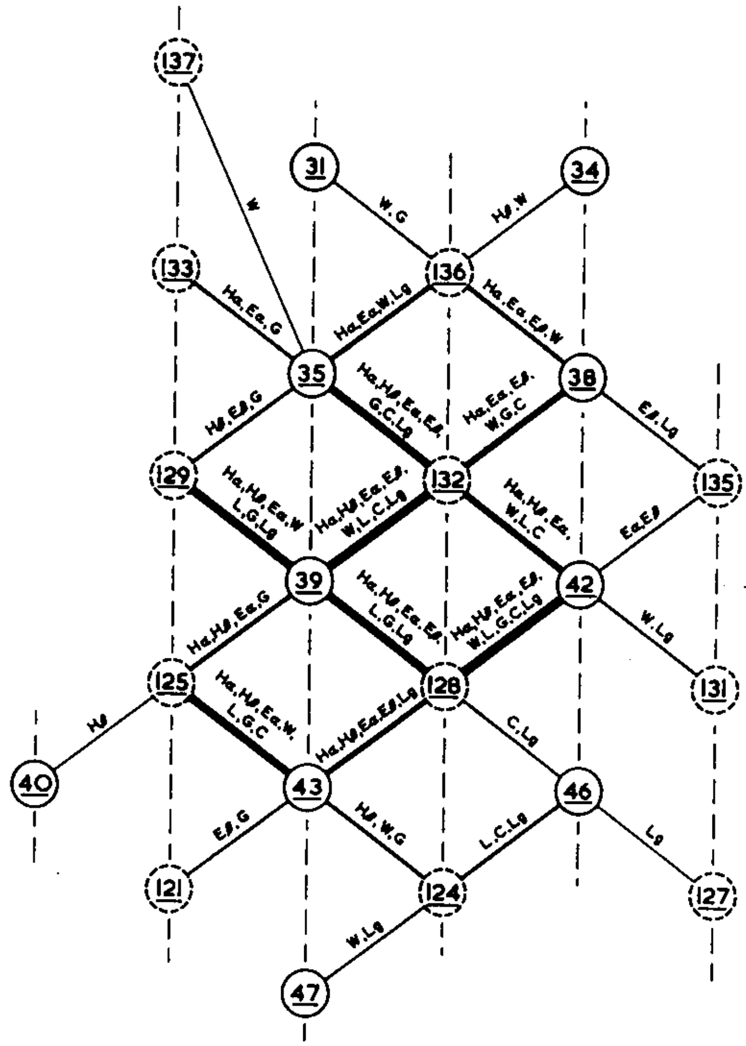
Structure of helix interfaces

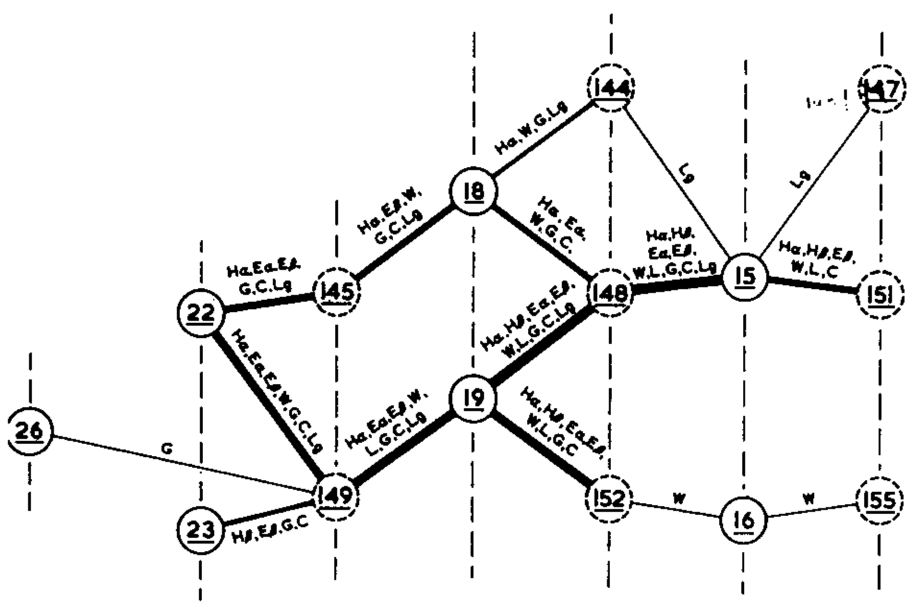
- ridges-into-grooves model
- sidechains on helices create ridges on helix surfaces
- grooves between parallel ridges
- ridges from one helix packs into grooves between ridges on another

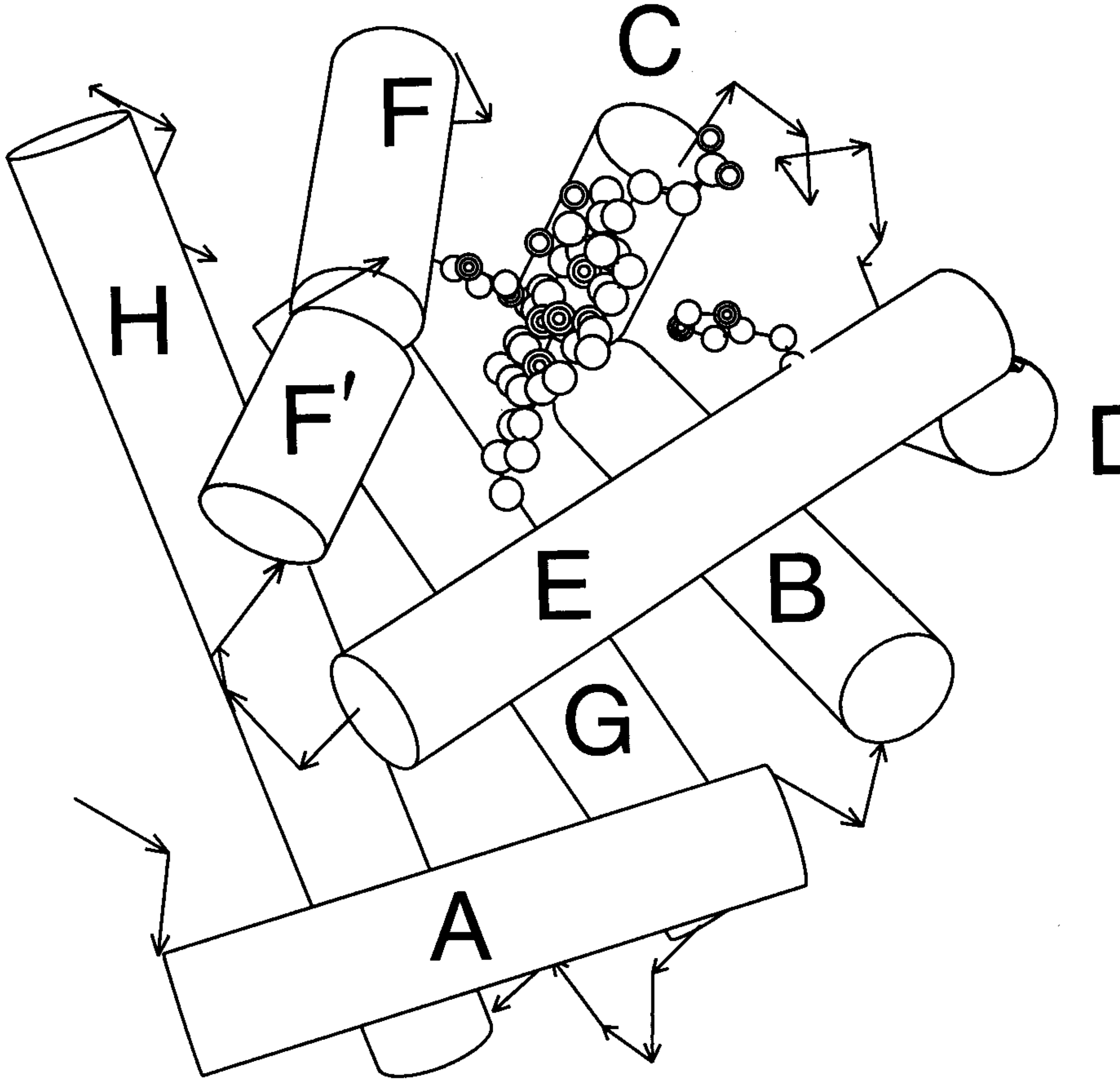
- ridge structure: i+4/i+4 most common
- globins rich in exceptions: 'crossed-ridge' structure

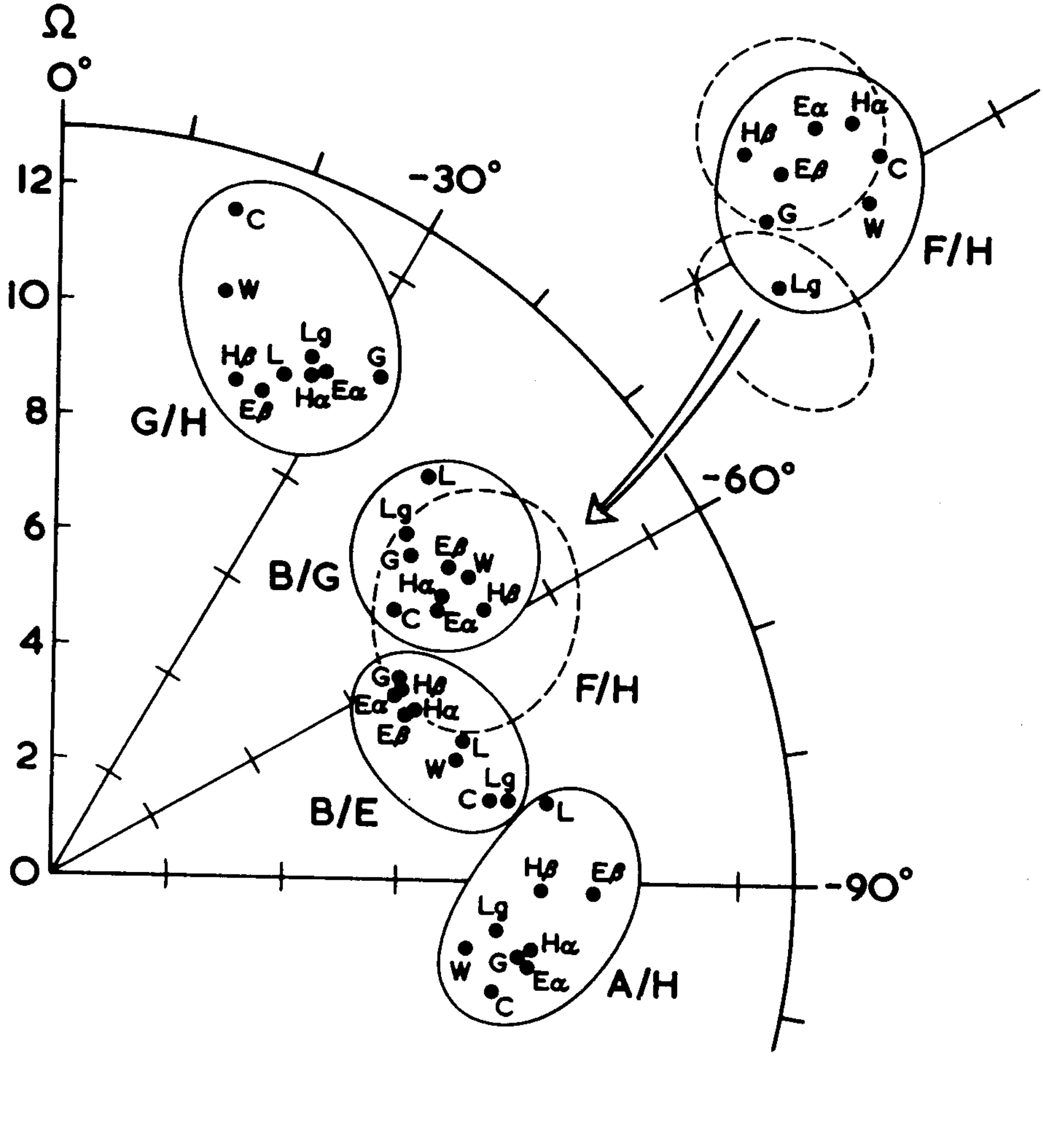


FACE DOWN
FACE UP


137
31
34
W
W,G
Hβ,W
133
136
Hα,Eα,G
Hα,Eα,W,Lg
Hα,Eα,Eβ,W
35
38
Hβ,Eβ,G
Hα,Hβ,Eα,Eβ, G,C,Lg
Hα,Eα,Eβ, W,G,C
Eβ,Lg
129
132
135
Hα,Hβ,Eα,W L,G,Lg
Hα,Hβ,Eα,Eβ, W,L,C,Lg
Hα,Hβ,Eα, W,L,C
Eα,Eβ
39
42
Hα,Hβ,Eα,G
Hα,Hβ,Eα,Eβ, L,G,Lg
Hα,Hβ,Eα,Eβ, W,L,G,C,Lg
W,Lg
125
128
131
Hβ
Hα,Hβ,Eα,W, L,G,C
Hα,Hβ,Eα,Eβ,Lg
C,Lg
40
43
46
Eβ,G
Hβ,W,G
L,C,Lg
Lg
121
124
127
W,Lg
47

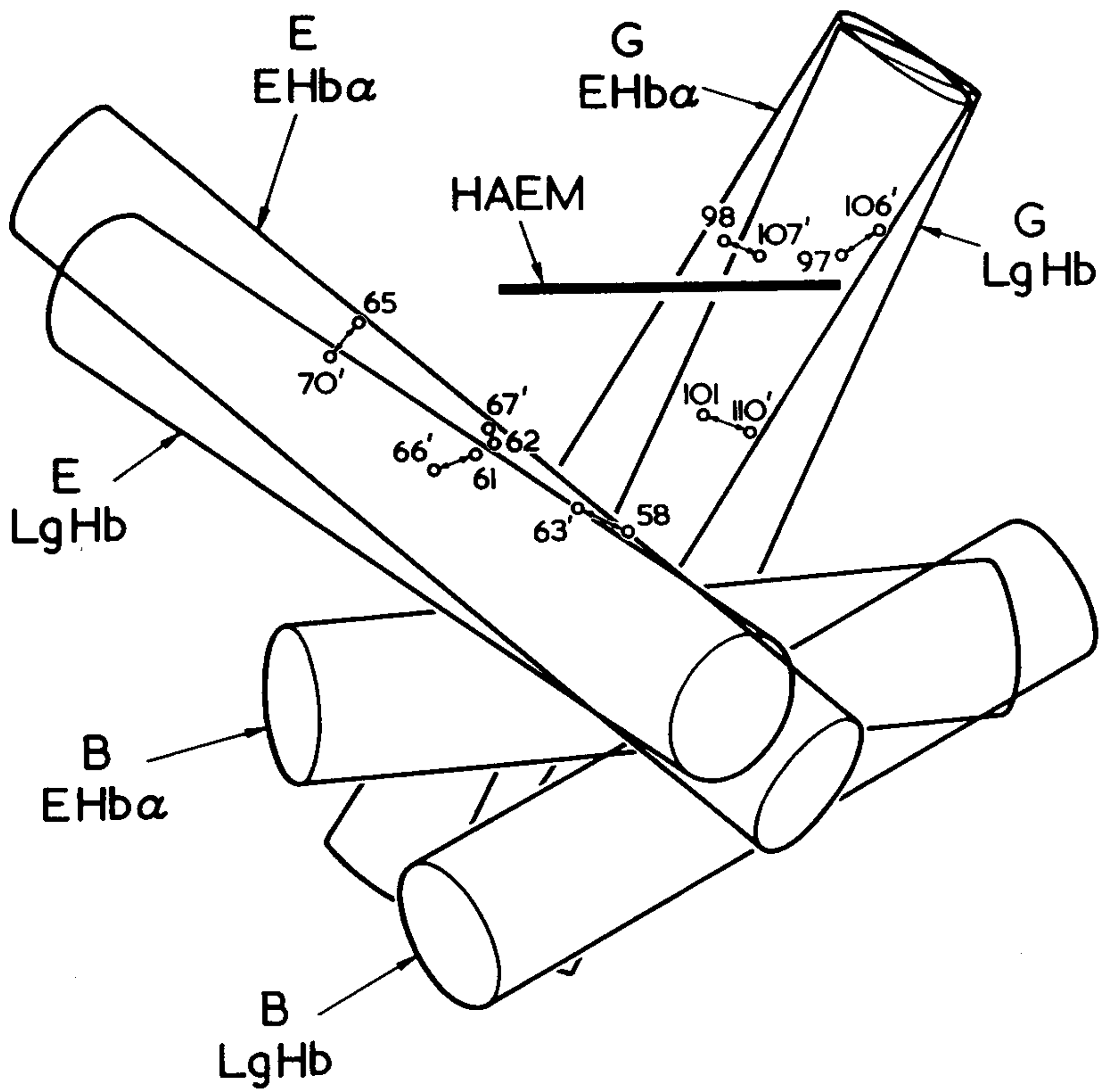

144
147
Hα,W,G,Lg
Lg
Lg
18
Hα,Eβ,W, G,C,Lg
Hα,Eα, W,G,C
Hα,Hβ, Eα,Eβ, W,L,G,C,Lg
Hα,Hβ,Eβ, W,L,C
Hα,Eα,Eβ, G,C,Lg
145
148
15
151
22
Hα,Eα,Eβ,W,G,C,Lg
Hα,Hβ,Eα,Eβ, W,L,G,C,Lg
19
Hα,Eα,Eβ,W, L,G,C,Lg
Hα,Hβ,Eα,Eβ, W,L,G,C
26
G
149
152
W
16
W
155
23
Hβ,Eβ,G,C


C
F
H
F′
D
E
B
G
A


Ω
0°
-30°
-60°
-90°
r (Å)
12
10
8
6
4
2
0
G/H
C
W
Lg
Hβ
L
G
Hα
Eα
Eβ
F/H
Eα
Hα
Hβ
Eβ
C
G
W
Lg
B/G
L
Lg
Eβ
G
W
Hα
Hβ
C
Eα
F/H
B/E
G
Hβ
Eα
Hα
Eβ
L
W
Lg
C
A/H
L
Hβ
Eβ
Lg
Hα
W
G
Eα
C

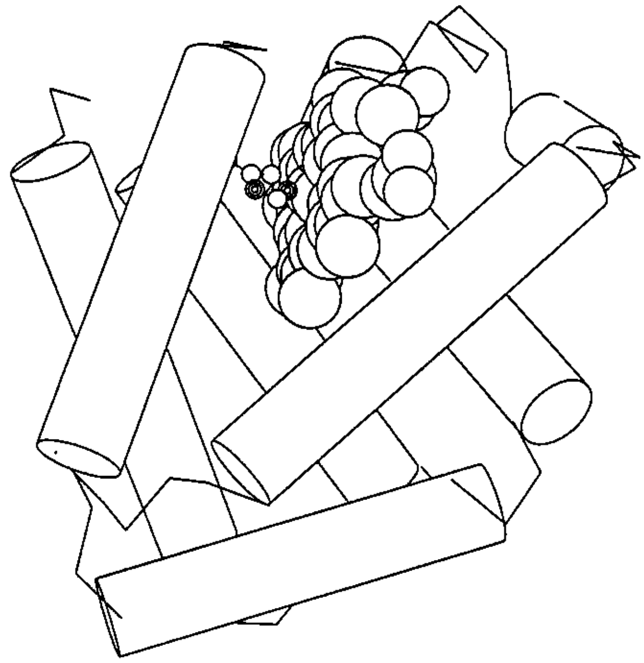
Helix interface structures conserved

- Same ridge/groove pattern
- 59 residues involved in helix interfaces, conserved hydrophobic
- template based on these residues does good job of discriminating globins
- Mutations cause shifts and rotations of packed helices – up to 7 Å / 30 °
- Shifts coupled to preserve binding site
- Reticulation of residues conserved

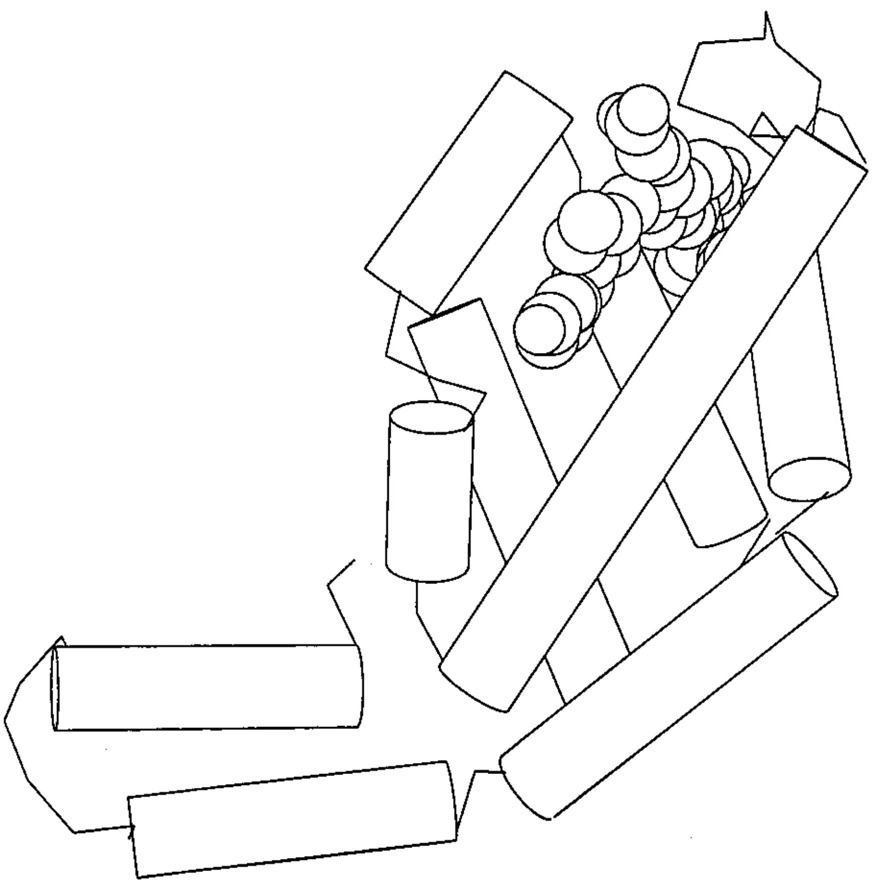

E
EHbα
G
EHbα
HAEM
98
107′
106′
97
G
LgHb
65
70′
67′
62
66′
61
101
110′
E
LgHb
63′
58
B
EHbα
B
LgHb

Phycocyanin

- Antenna pigment in photosynthesis
- No sequence similarity to globins
- When structure solved – globin fold
- Why is this:
  distant relationship or convergence?



sea hare myoglobin



Phycocyanin

Are globins and phycocyanins related?

- How to answer?

  look for similarities in details

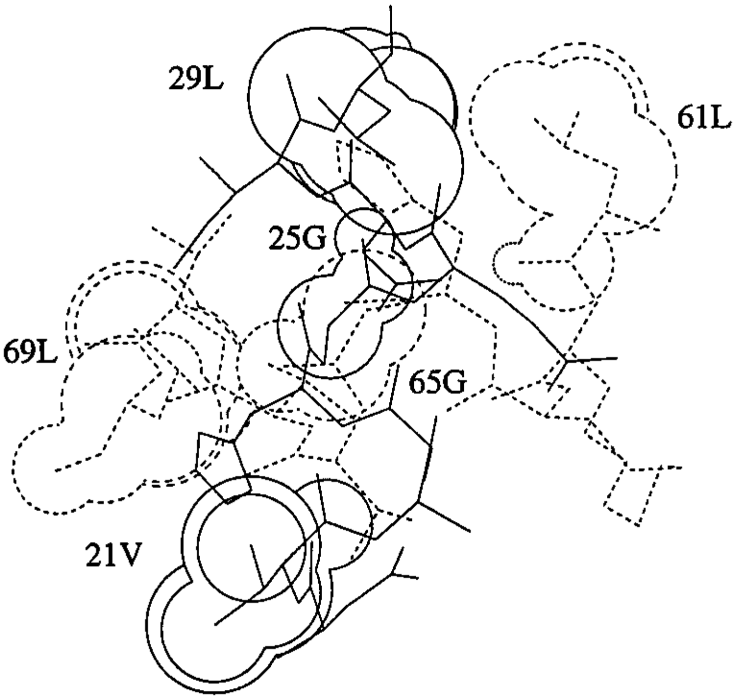
  not required for structure or function

  (But, how would you prove that?)
- Analysis of structural details:

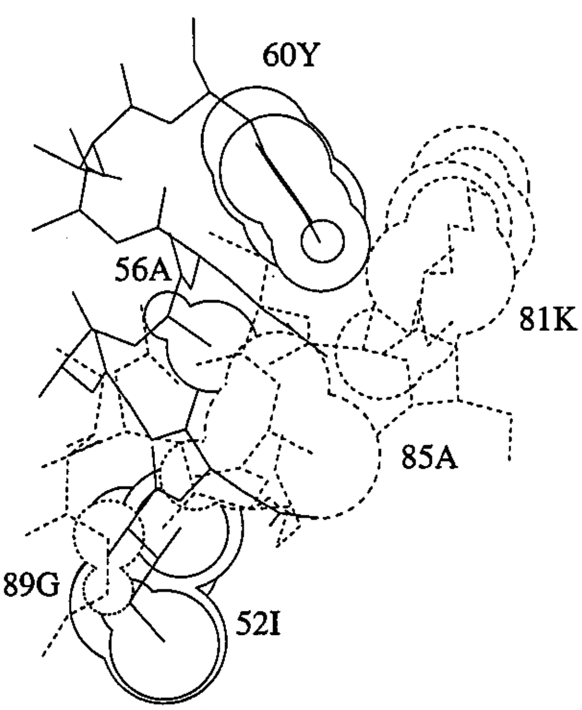
  C helix in phycocyanin a 3_{10} helix

  ridge/groove packing patterns similar

  including exceptional crossed-ridge pattern



myoglobin B/E contact



phycocyanin B/E contact

Everything looked good

- Then truncated globins discovered
- Prokaryotes, also eukaryotes (Paramecium)
- as few as 109 residues

Truncated globins

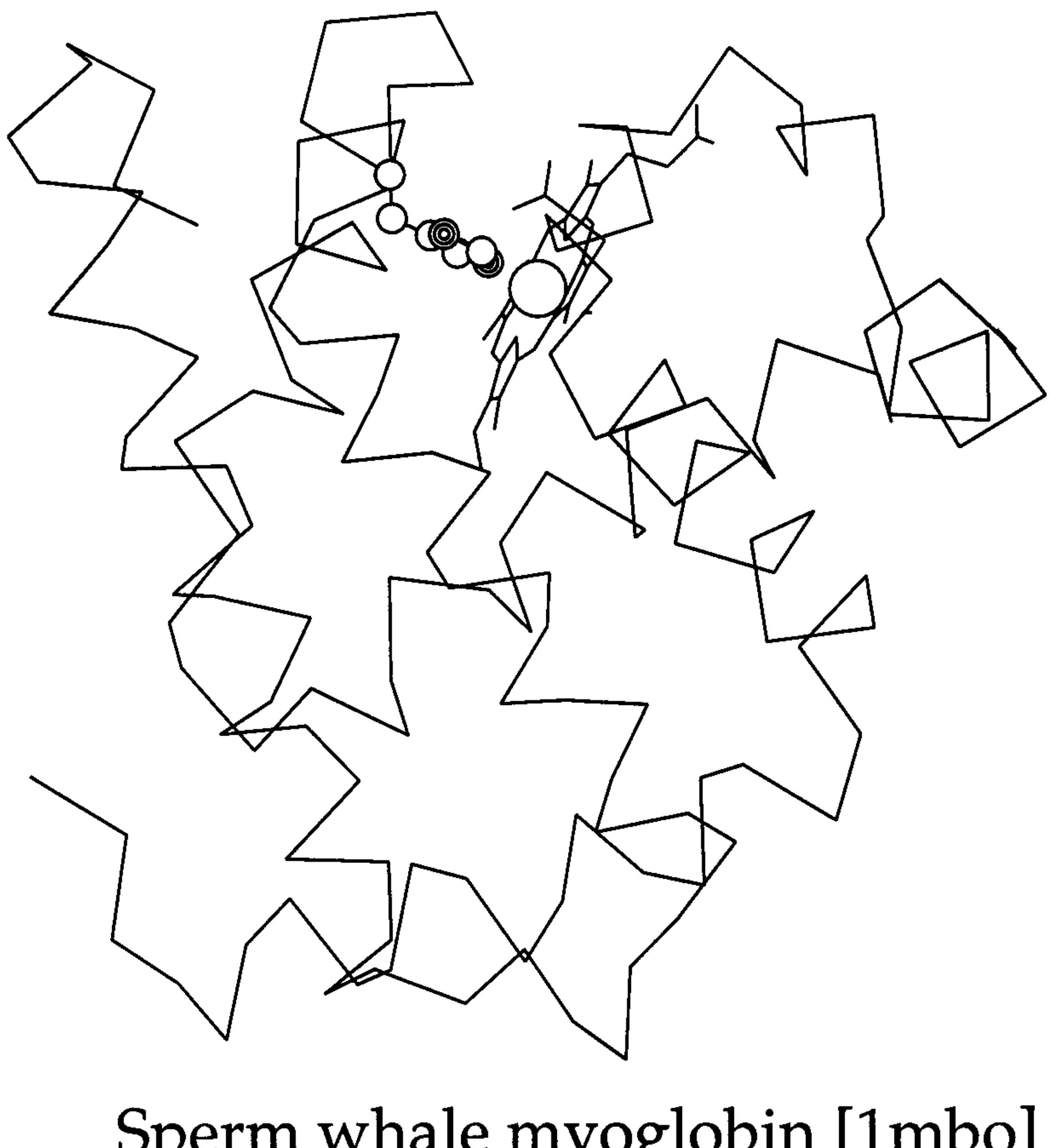
- as few as 109 residues
- deletions not only at termini
- change in 'conserved' structural details

☞ reopens a lot of questions

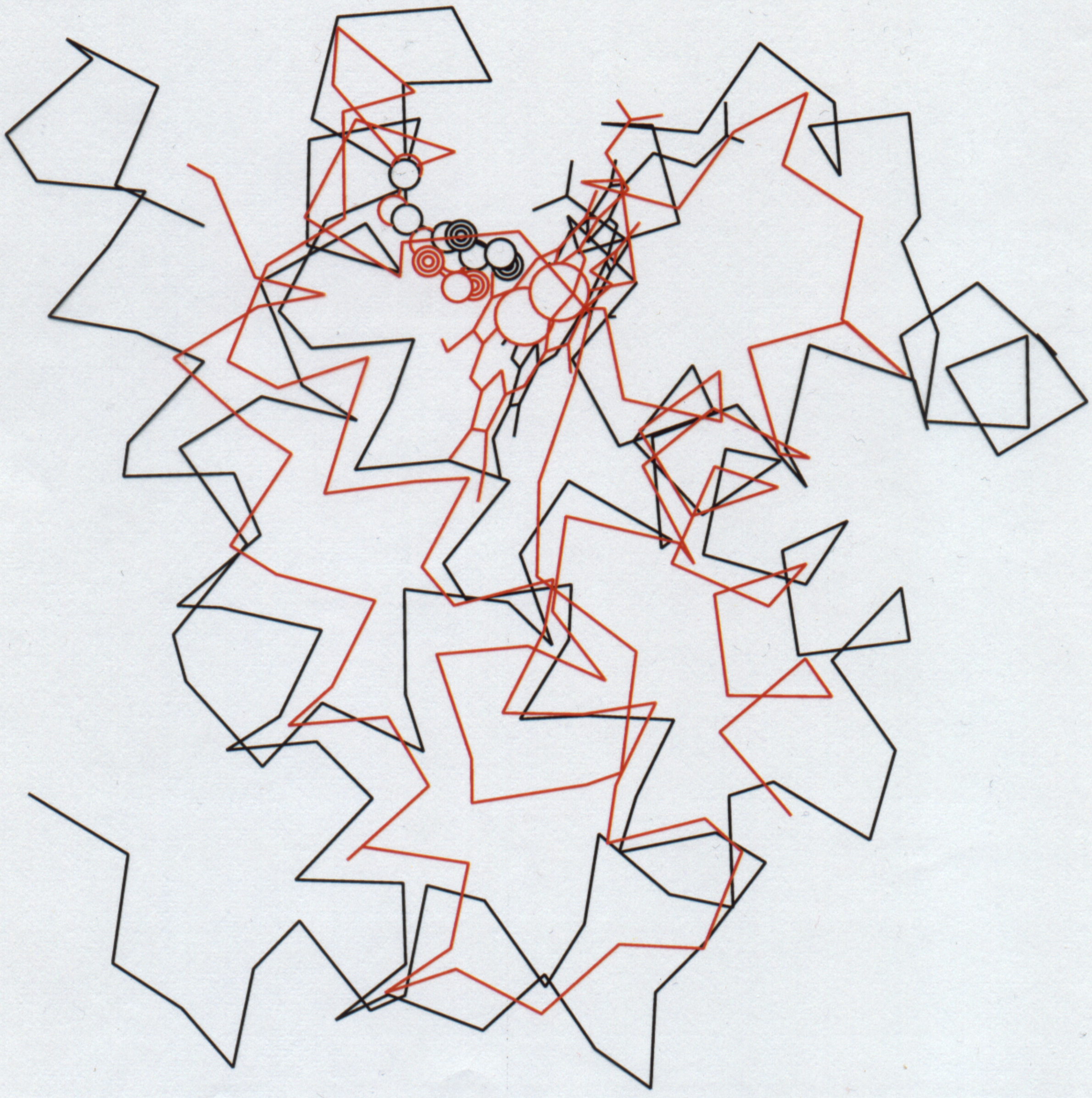
Unusual features of truncated globins

- Short
- F-helix unfolded
- crossed-ridge B/E interface shows normal i+4/i+4 packing

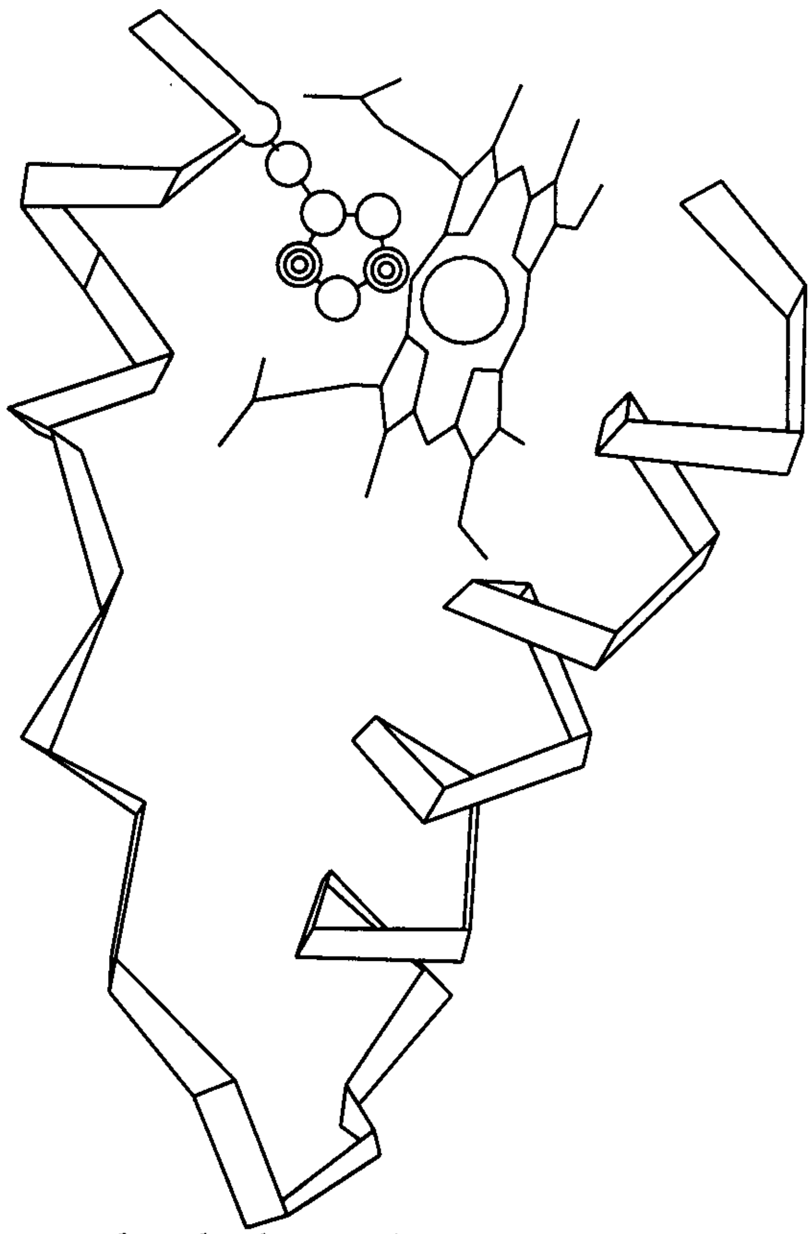
truncated globin *(P. caudatum)* [1dlw]



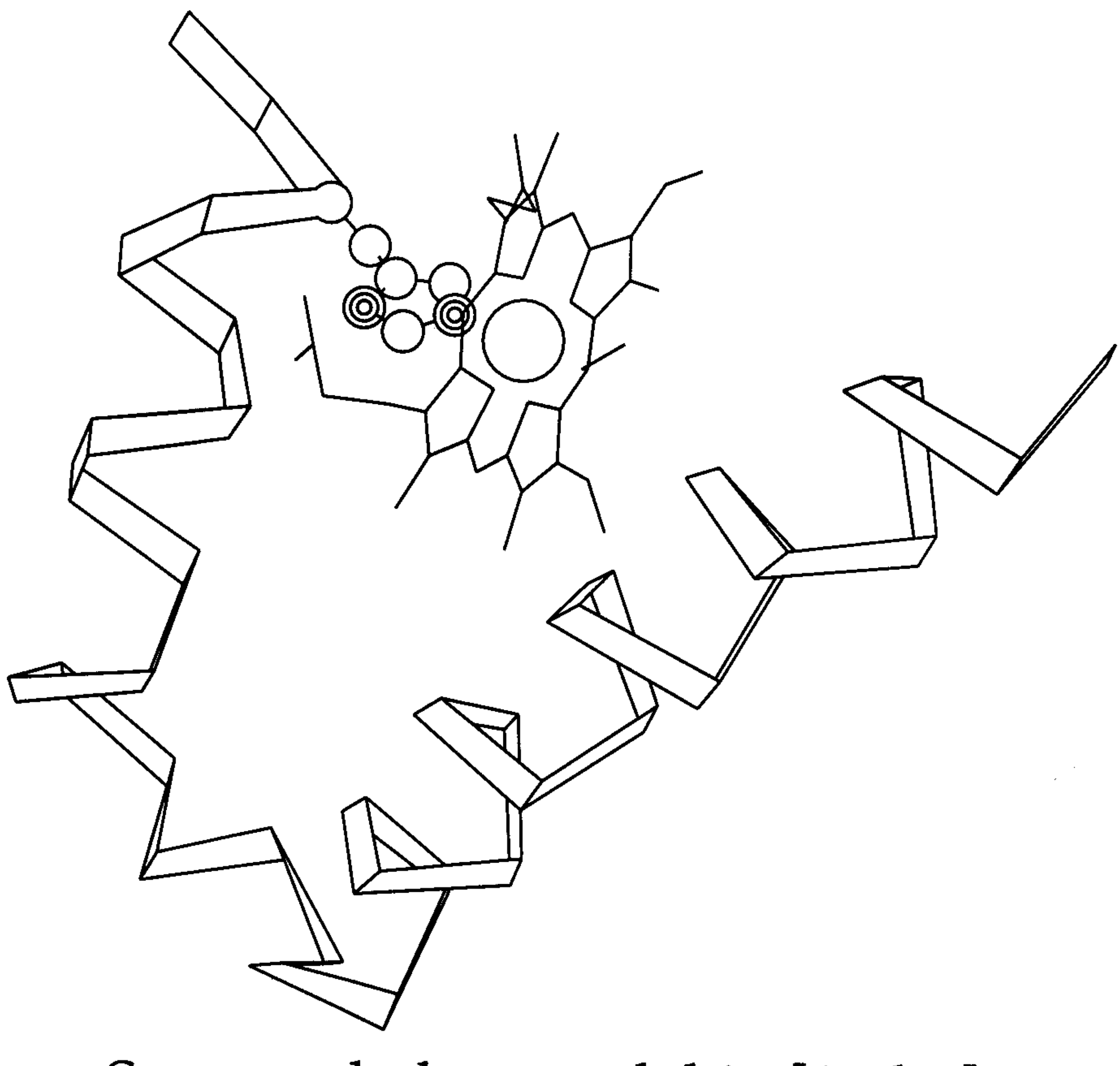
Sperm whale myoglobin [1mbo]



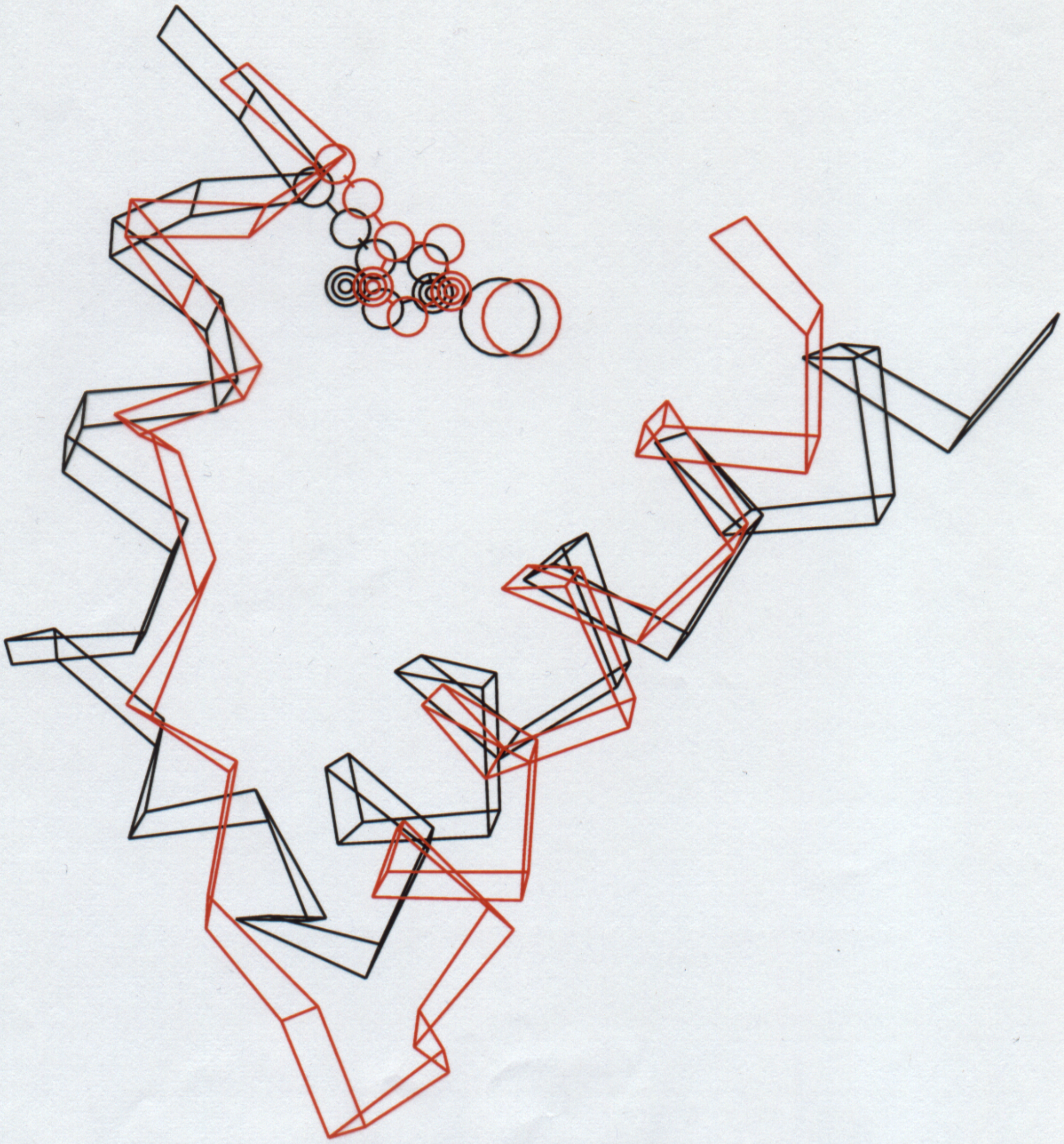
superposition



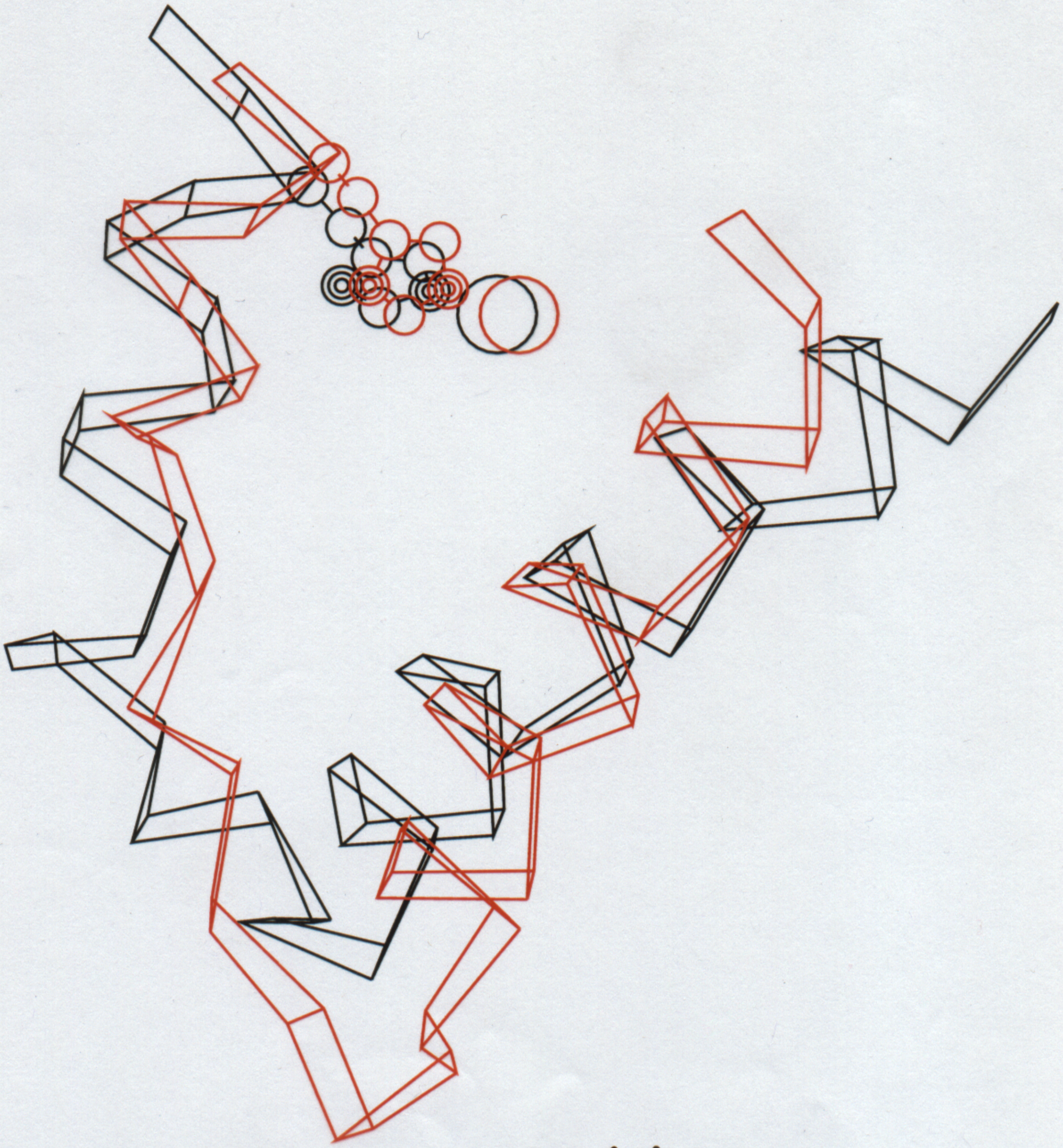
truncated globin *(P. caudatum)* [1dlw]



Sperm whale myoglobin [1mbo]



superposition



superposition

Possible to align sequences

- Based on structural superposition
- Truncated globins largely similar
- Chain termini differ
- E-F turn varies

Topics not treated

- In full-length globins, binding site blocked
- Molecule must 'breathe' to let oxygen in and out
- Truncated globins seem to have a tunnel connecting binding site to outside

Conclusion

- To understand protein structure, watch evolution
- To understand protein structure better, look for very distant relatives

Colleagues

- Cyrus Chothia (Cambridge)
- Martino Bolognesi (Genova)
- Annalisa Pastore (London)