

The effects of mutations, deletions and insertions of single amino acids on the three-dimensional structure of globins

N.A. Kolchanov, V.V. Soloviov and A.A. Zharkikh

Institute of Cytology and Genetics of the USSR Academy of Sciences, Siberian Department, Novosibirsk 630090, USSR

Received 29 June 1983

Using the method of calculation of low energy packings of α -helices, we studied the effects of mutations, deletions and insertions of amino acids on the structure of sperm whale myoglobin. It was found that these events can cause complete breakdown of this structure.

Mutation Deletion Insertion Protein structure

1. INTRODUCTION

The traditional approach to analysis of the effects of amino acid replacements is based on testing their sterical permissibility in a folded 3-dimensional structure of globular proteins [1,2]. Most sterical and energetical strains might have been eliminated by conformational accommodations of a new side chain in a preformed sterical environment, and this led to the conclusion that the structure is not strongly affected by replacements [1].

The overall spatial structure of globular proteins is the result of the sequential folding of polypeptide chain fragments [3] and, hence, it is affected only by replacements interfering with any of these steps. Clearly, estimates obtained by an approach disregarding these facts rather reflect chemical modifications of amino acids in a ready structure than the real conformational effects of mutations. A broader basis is provided when replacement effects on the main steps of the self-organization of globular protein are taken into consideration [4].

We here report the results obtained for the effects of mutations, deletions and insertions of amino acids on one of the steps of the folding of the spatial structure of sperm whale myoglobin. These effects were analysed by calculating low energy packings of the α -helices [5–7] in the par-

ticular case of the folding of the G- and H-helices. The data presented demonstrate that mutations, deletions and insertions can cause complete breakdown of the normal 3-dimensional structure of globular proteins. There is good agreement between these data and those for fixation of mutations, deletions and insertions derived from analysis of the evolutionary tree of the globin family.

2. METHODS

The effects of all single amino acid substitutions in each position of the G and in several of the H, which are possible within the framework of the genetic code, on the lowest energy packings of two-helix system GH of sperm whale myoglobin were analysed. (To specify, amino acid substitutions resulting from replacement of single nucleotide in a codon were considered.) The effects of deletions of single amino acids and insertion of tryptophan in each position of the G were also analysed. In each case, the lowest energy conformation of the GH was calculated and compared with that in the normal GH. The calculations were based on a method (for details see [5–7]) in which the side chains of the amino acids are represented by hard spheres whose centres are on points corresponding to the C_{β} -atoms and the backbone of the α -helix is

represented by a cylinder with a radius of 4 Å (fig.1,2). The interactions between the two α -helices are expressed by the following energy function:

$$E = \frac{1}{4} \sum_{ij} [(E_i^s + E_i^p) + (E_j^s + E_j^p)] + \sum_i [\frac{1}{4}(E_i^s + E_i^p) - F] \\ R_{ij} < R_i + R_j + 2.8 \text{ Å} \quad P_i < R_i + R + 2.8 \text{ Å} \\ + \sum_j [\frac{1}{4}(E_j^s + E_j^p) - F] + \sum_{ij} E_{ij}^e \\ P_j < R_j + R + 2.8 \text{ Å} \quad R_{ij} < R_i + R_j + 2.5 \text{ Å}$$

In this energy function, the first sum includes the interactions between the side chains of the α -helices, the second and the third sums include the interactions between the side chains and the backbones, and the fourth term is the sum of the energies of the electrostatic contacts between the charged side chains of amino acids. The designations are:

R_{ij} , distance between the i -th and j -th C_β -atoms in the first and second α -helices, respectively;

R_i , radius of hydration of the i -th side chain;

E_i^s and E_i^p , surface and polar contributions to the energy of hydrophobic interactions of the i -th side chain;

P_i , distance between the i -th side chain of one helix and the backbone of the other helix;

R , radius of the cylinder representing the backbone;

F , energy of dehydration of a solvent contact area of the backbone (that area of the backbone of one of the helices which comes into contact with the side chain of the other);

E_{ij}^e , energy of electrostatic contact between the i -th and j -th amino acids. For parameter values see [5-7].

To exclude sterically non-permissive conformations, we introduced a matrix of permissive minimum distances between the C_β -atoms of one of the helices and C_β -atoms of the other. In search for the lowest energy packings, the relative orientation of the helices G and H is given by 6 parameters (fig.1). The values of these parameters were chosen randomly within their variation ranges [5-7]. In this search, 20000 different relative orientations of the G and H were examined, and the packing with the lowest energy was chosen.

Search for deletions and insertions, which have been fixed in the primary structure of globins during evolution, was carried out as in [8]. Search for

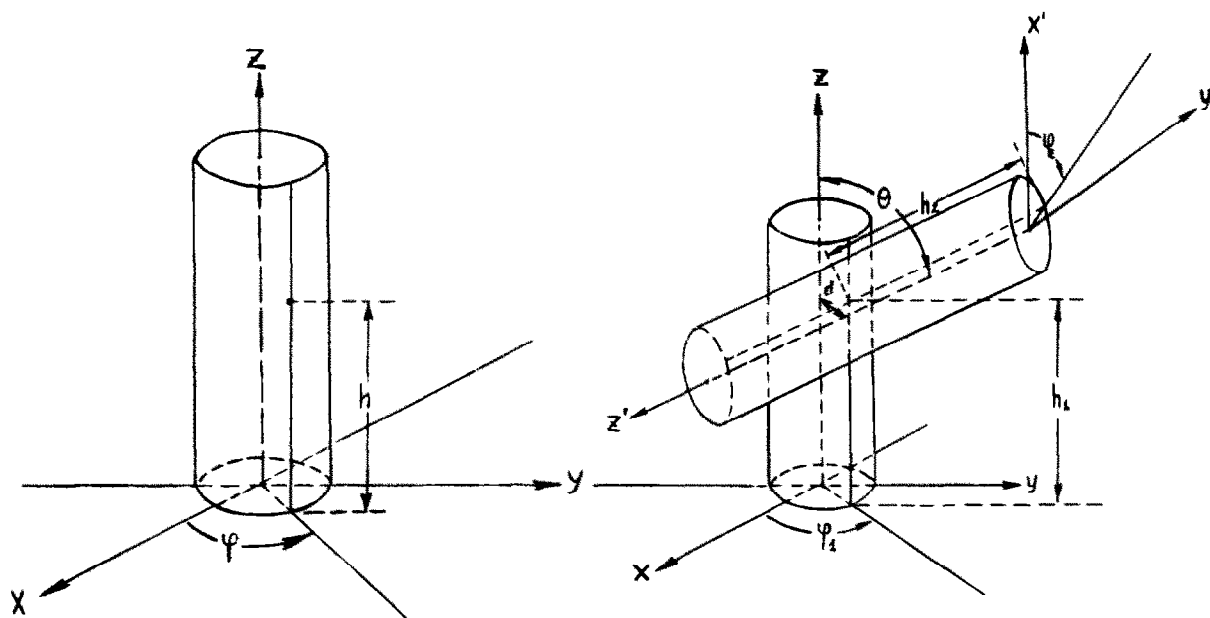


Fig.1. Parameters giving the spatial orientation of two α -helices. The cylinder represents the backbone of the α -helices. ψ_1, ψ_2 are the angles giving the contact line on the first and second helices, respectively; h_1, h_2 denote the position of the points of contact on the first and second helices; d and θ , the distance and the angle between the axes of the helices, respectively.

mutations was performed on the evolutionary tree derived from 40 known primary structures of globins of taxonomically distant groups (for method, see [9]).

3. RESULTS AND DISCUSSION

Our calculations enabled us to establish that certain mutations can produce complete breakdown of the normal spatial structure of globins. Fig.2 compares 3 projections of the spatial structure of the GH without any mutations with ones having a single substitution Arg \rightarrow Gly in position 16 of the H. As seen in fig.2, this mutation distorts drastically the normal structure. The G and H make contacts with those side chains of their surfaces which have not been earlier involved in in-

terhelical interactions. As compared with the orientation in the normal structure, there occurs a rotation of the helices about their own axes, namely a rotation of the contact line of the G and that of the H by 50°. As a result, out of a total of 23 normal contacts between the side chains of the G and H, only 5 are retained by the distorted structure and 12 new ones arise (table 1).

In the normal GH structure, one surface is almost entirely polar while the other is hydrophobic [5-6]. It is this hydrophobic surface of the GH which interacts with the hydrophobic one of the three-helix system ABE during the formation of the globular nucleus of the normal structure of globins [7]. In the distorted GH structure, this interaction is impossible because both surfaces become partially polar [4].

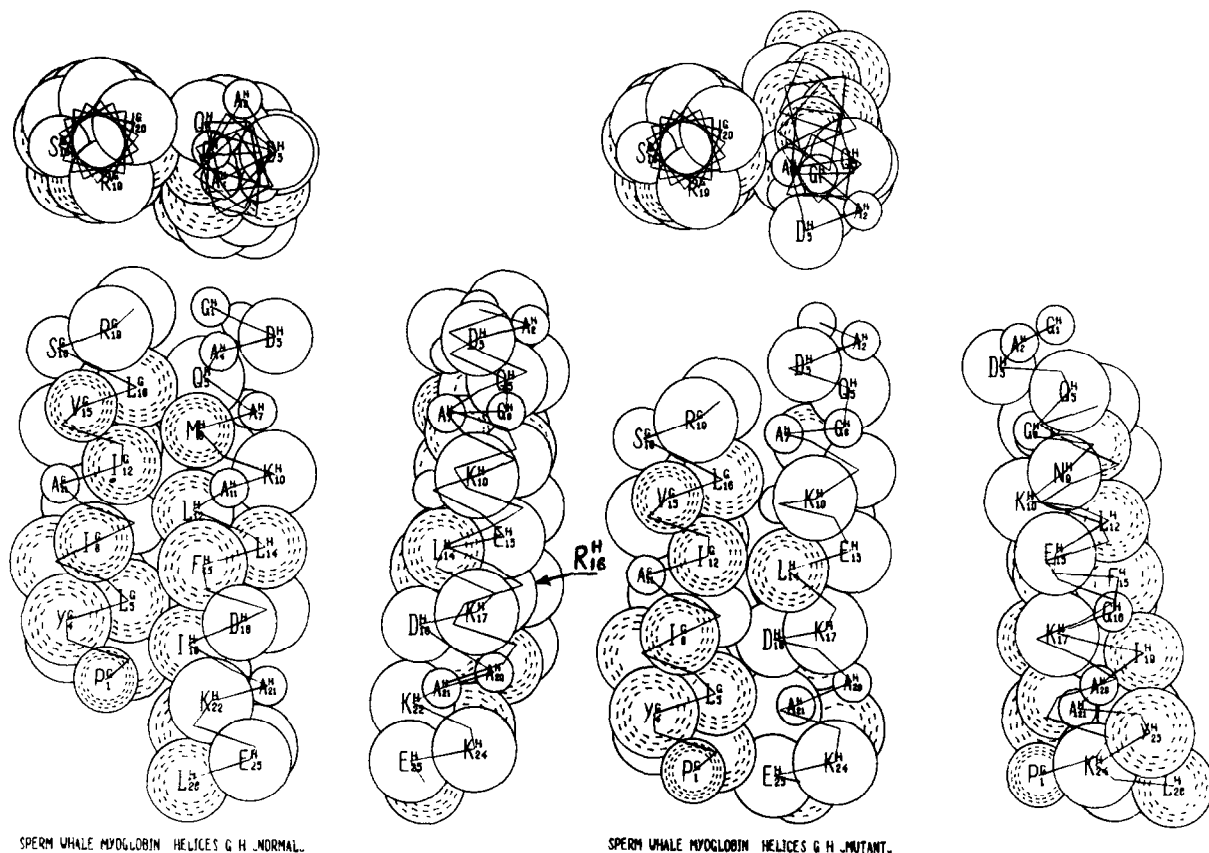


Fig.2. Effect of mutation Arg \rightarrow Gly in position 16 of the H helix on the lowest energy conformation of the two-helix system GH in sperm whale myoglobin. Right - 3 projections of the GH without mutations. Left - 3 projections of the GH with a mutation. The side chains of the amino acids are represented by a hard sphere. The hydrophobic side chains are hatched. Mutation position is indicated by an arrow (not shown in the left projection).

Table 1

A list of contacts between the side chains of the G- and H-helices in a normal lowest energy packing and after mutation in Arg→Gly in position 16 of the H-helix

Side chains involved in contacts of the G- and H-helices in the normal G-H structure		Side chains involved in contacts of the G- and H-helices in the G-H structure after mutation	
H	G	H	G
I ₁₉	P ₁	I ₁₉	I ₂ ^a
A ₂₀	P ₁	K ₂₂	I ₂
R ₁₆	I ₂	Y ₂₃	I ₂ ^a
I ₁₉	I ₂	L ₂₆	I ₂
Y ₂₃	I ₂	K ₂₂	L ₅
F ₁₅	L ₅	F ₁₅	E ₆
R ₁₆	L ₅	I ₁₉	E ₆
I ₁₉	L ₅	K ₂₂	E ₆
R ₁₆	E ₆	F ₁₅	S ₉ ^a
R ₁₆	I ₈	M ₈	I ₁₃ ^a
L ₁₂	S ₉	A ₁₁	I ₁₃
F ₁₅	S ₉	L ₁₂	I ₁₃ ^a
R ₁₆	S ₉	F ₁₅	I ₁₃
M ₈	I ₁₂	M ₈	H ₁₇
L ₁₂	I ₁₂	A ₄	H ₂₀
Q ₅	I ₁₃	M ₈	H ₂₀
M ₈	I ₁₃	F ₁₅	E ₁₀
N ₉	I ₁₃		
L ₁₂	I ₁₃		
Q ₅	L ₁₆		
M ₈	L ₁₆		
Q ₅	H ₁₇		
Q ₅	H ₂₀		

^a Contacts retained after mutation

Index refers to the position of the side chain in a helix

Table 2

Effects of mutations in different positions of the G-helix on the packing of the two-helix system GH in sperm whale myoglobin

Position in the G-helix	Amino acid substitution	Number of contacts ^a retained between the side chains of the G- and H-helices
2	I → R	2
3	K → I	23
4	Y → D	23
6	E → V	3
6	E → Q	2
8	I → N	23
9	S → W	14
11	A → E	0
12	I → R	1
12	I → F	23
16	L → Q	23

^a In the two-helix system GH without mutations the number of contacts between the side chains of G and H is 23

The calculations we performed indicate that the spatial structure of the GH is altered by mutations which change polarity or charge of the amino acids (table 2). Furthermore, the effects of mutations are not strongly dependent on position (central or terminal) in the α -helix. Mutations in the N- or C-end equally affect the GH structure. Of 161 mutations possible in the G, only 15 (10%) can completely break down the normal GH structure, and the re-

Table 3

The effects of deletions of single amino acids and insertions of tryptophan in all positions of the G-helix on the packing of the two-helix system GH in sperm whale myoglobin

Number of normal contacts retained between the side chains of the G- and H-helices after	Position of deletion or insertion in the G-helix																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Deletions	21	20	5	5	14	6	7	0	0	0	0	17	17	15	16	19	15	21	21	21
Insertions	—	19	17	2	2	1	0	0	0	2	15	15	0	20	19	19	20	21	21	15

maintaining mutations weakly, if at all, distort the structure.

Table 3 summarizes the data computed for the effects of deletions of single amino acids and insertions of tryptophan in all positions of the G on the lowest energy conformation of the GH. It is immediately apparent that deletions and insertions cause complete breakdown of the structure much more frequently than mutations (40% of deletions and 45% of mutations distort so drastically the GH structure that there remain no more than 10 normal contacts between the side chains). The

strongest distorting effects are in the centre of the G, and the weakest at its ends. Thus, the extent to which deletions or insertions distort the normal conformation is related to their position in the α -helix.

The results indicate that mutations, deletions and insertions, which interfere with the main steps of folding, can cause complete breakdown of the protein structure. Based on these results, a mutation is considered structurally permissible; i.e.,

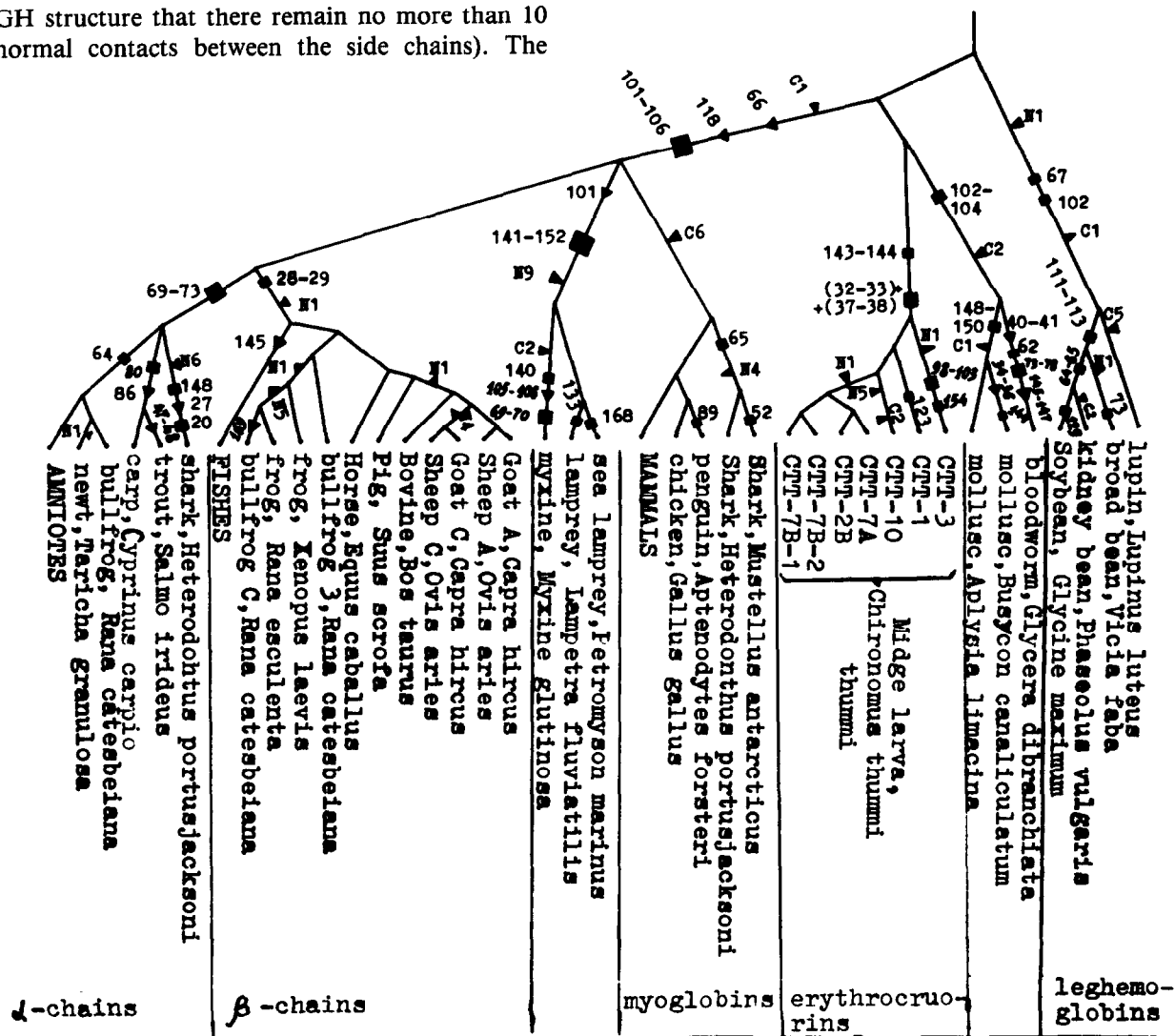


Fig.3. Evolutionary tree of globin family. The tree was derived from primary structure data concerning 40 globins of distantly related taxonomic groups. Internal deletions are indicated by squares and internal insertions by triangles on the branches. The pendant triangles with apexes pointing to the branches represent insertions, and those with apexes pointing away from the branch represent deletions at the ends of the globin chain. The numbers indicate the length and position of deletions and insertions in the primary structure.

drastically not effecting the 3-dimensional structure and limited to the vicinity of the mutation position, when all the main intermediate steps of folding are unaltered. Only under these conditions, structural distortion is local and occurs at the last steps of folding when the intermediate globular structure becomes native through fine stereochemical adjustments.

Mutation interference with protein function has been related to local disturbances of protein conformation involving active centres [2]. We obtained evidence for yet another mechanism of functional interference: single amino acid substitutions far from the active site can break down the structure and, as a consequence, the functional centres of the protein molecule. Support comes from experimental results indicating that mutations can deviate from normal the folding of bacteriophage P22 tail proteins and thereby give rise to non-functional proteins [10,11].

The data computed for the damaging effects of mutations, deletions and insertions agree well with those obtained by analysis of the globin evolutionary tree. A schematic representation of the tree is given in fig.3. The set of parameters used to test the uniformity of the distribution of mutations, deletions and insertions along the secondary structure of globins are compiled in table 4.

Apparent from these data is the non-uniformity

Table 4

Distribution of evolutionary deletions, insertions and mutations in the helical and interhelical sections of polypeptide chains of globins

Event location in the chain	Events/position	
	Deletions + insertions	Mutations
Interhelical sections	0.55	23.9
End residues of the helices	0.27	23.3
Internal residues of the helices	0.14	27.0
Total	0.28	25.2
χ^2_a	12.26	2.14

^a The uniformity tested by Chi-square (χ^2) method is obvious for mutations via deletions and insertions

of the distribution of deletions plus insertions and their concentration in the interhelical sections and at the ends of the α -helices. This is in agreement with the results of the conformational analysis according to which deletions or insertions in the α -helix centre always break down the normal structure of the globin molecule.

The fixation frequencies of mutations exceed 100-fold those of deletions plus insertions. In contrast to deletions plus insertions, mutations are uniformly distributed along the secondary structure (table 4). This also agrees with the computed data indicating that:

- (i) Deletions and insertions more frequently damage protein structure than mutations;
- (ii) The damaging effects of mutations are more related to the type of amino acid substitutions than to its location in the α -helix (table 2).

ACKNOWLEDGEMENT

The authors are grateful to A. Fadeeva for translating this paper from Russian into English.

REFERENCES

- [1] Warm, P.K. (1975) *Biochemistry* 14, 3518–3526.
- [2] Perutz, M.M. and Lehmann, H. (1968) *Nature* 219, 902–909.
- [3] Burgess, A.W. and Sheraga, H.A. (1975) *J. Theor. Biol.* 53, 403–423.
- [4] Solovyov, V.V. and Kolchanov, N.A. (1982) *Genetica* 18, 1573–1580.
- [5] Solovyov, V.V. and Kolchanov, N.A. (1981) *Mol. Biol.* 15, 323–335.
- [6] Solovyov, V.V. and Kolchanov, N.A. (1980) in: *Mathematical Models of Evolutionary Genetics* (Ratner, V.A. ed) pp.5–34, Institute of Cytology and Genetics, Novosibirsk.
- [7] Kolchanov, N.A. and Solovyov, V.V. (1980) in: *Mathematical Models of Evolutionary Genetics* (Ratner, V.A. ed) pp.35–48, Institute of Cytology and Genetics, Novosibirsk.
- [8] Needleman, S.B. and Wunsch, C.D. (1970) *J. Mol. Biol.* 48, 443–453.
- [9] Zharkikh, A.A. (1977) in: *Mathematical Models of Evolution and Selection* (Ratner, V.A. ed) pp.5–52, Institute of Cytology and Genetics, Novosibirsk.
- [10] Goldenberg, D.P. and King, J. (1981) *J. Mol. Biol.* 145, 633–652.
- [11] Smith, D.H. and King, J. (1981) *J. Mol. Biol.* 145, 653–676.