Research Journal of Pharmaceutical, Biological and Chemical Sciences

On The Conformation Of Lys79 And Lys98 Residues In The Sperm While Myoglobin Structure.

Mahmud A Basharov*.

Laboratory of Functional Biophysics of Proteins, Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Institutskaya 3, Pushchino, Moscow Region, 142290, Russia.

ABSTRACT

The myoglobin structure determined by John Kendrew and associates by X-ray crystallography is the first protein structure described in atomic detail. Plenty of myoglobin structures are available in the Protein Data Bank now. A visible feature of the structures is that conformation of Lys79 and Lys98 of nonglycine residues has a positive value of the backbone dihedral angle ϕ. Here I suspect the conformation of both the residues to be wrong and persuade that angle ϕ of the residues should be negative. Careful rebuilding of the structure model of EF and FG loops including Lys79 and Lys98 residues and anew refinement of the myoglobin structure by X-ray crystallography or other methods of structural biology is required. Verification of the suggestion is important for structural biology and bioinformatics. This would lead to further refinement of the existing many Mb structures and stimulate the studies on the improvement of other protein structures.
Keywords: Myoglobin, Protein structure, Conformation, X-ray crystallography, Amino acid.

*Corresponding author

INTRODUCTION

Proteins are polypeptide chains of the amino acid residues sequentially joined by peptide bonds, generated by ribosomes, folded into a unique conformation – three-dimensional (3D) structure to fulfill some particular biological functions in the living cells. The conformation of a protein is described by a set of the values of dihedral (torsion) angles ϕ, ψ and ω of the main chain (backbone) and χ_i of the side group of the consecutive amino acid residues (Fig. 1). Backbone angles describe rotations about the backbone bonds, ϕ and ψ about the NC^α and $C^\alpha C$ bonds and ω about the peptide bond between the residues. Peptide bonds are usually trans-planar (values of ω are around ± 180°); cis- peptide (ω is around ± 0°) bonds occur quite rarely. With this in mind, conformation of the residues is described by (ϕ, ψ) angles. The set of (ϕ,ψ) values defines the backbone conformation of the polypeptide chain, the overall folding pattern or the secondary structure of the protein.

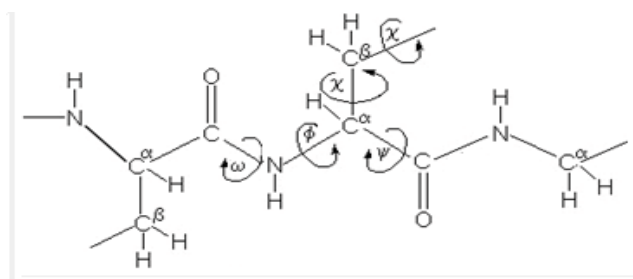**Figure 1: Scheme of a polypeptide chain segment with backbone atoms, valence bonds and backbone ϕ, ψ and ω dihedral angles of a residue labeled.**

Restricted values of ϕ and ψ angles are accessible to amino acid residues in peptides because of the steric hindrances and unfavorable interactions of the nonbonded atoms during rotations about NC^α and $C^\alpha C$ bonds by ϕ and ψ conformational angles [1,2]. The accessible values demarcate a 2D contour graph in the (ϕ,ψ) space, the (ϕ,ψ) map of the regions of allowed conformations. Figure 2 displays U (ϕ,ψ) conformational energy map versus ϕ and ψ angles for alanine residue obtained by calculations of alanine dipeptide with the use of the van der Waals potentials [3]. The map, popularly known as (classic) Ramachandran map, is typical for all the common nonglycine (non-Gly) residues and represents the conformations accessible to non-Gly residues in proteins [2-7]. For a protein, the values of (ϕ,ψ) are calculated from the atomic coordinates of the protein determined experimentally (by X-ray crystallography, neutron diffraction, NMR, etc.). The values represent a 2D scatter plot, the plot of the distribution of (ϕ,ψ) angles (similar to that of shown in Fig. 3) known as the Ramachandran plot of the protein. Protein Data Bank (PDB) [8,9] holds the data on the determined protein structures.

X-ray crystallography is the most powerful tool for exploring the 3D structure of proteins. Solving of a protein structure by this tool requires several procedures to be carried out, such as determination of the amplitudes and phases of the diffraction reflexes from the protein crystals, building and interpretation of the electron-density (e.d.) map, detection and fitting of the atomic coordinates in the e.d. by building an adequate initial structure model of the protein molecule, refinements of the coordinates, and building of the final structure model [10-12]. Introduction of an error into the protein structure model is unavoidable during almost all the stages of structure solution; the presence of local errors is habitual even in the well determined protein structures [10-17].

The sperm while (SW) myoglobin (Mb) (the met form, metMb) structure determined by John Kendrew and associates by X-ray crystallography [18,19] is the first protein structure described in the atomic detail. The structure of the protein of a single polypeptide chain of 153 amino acid residues and one heme represents a compact globule consisting of eight α-helices, labeled A through H, connected by short loops and a C-terminal short tail [19]. The atomic coordinates of the protein were published by Watson [20]. In the PDB, the entry code of the structure is 1MBN. The structure was refined using the atomic coordinates by Watson [20], the original phase angles by Kendrew et al. [18] and the newly collected intensity data [21]. It is available in the PDB with code 4MBN now. In the PDB, there are a great deal of structure models of SW Mb in different functional, mutant, ligand and modified forms solved at different experimental conditions and resolutions,

including the highest 1 Å resolution X-ray crystallographic structure 1A6M [22]. The 3D structure of Mb highly conserved across Mb species with many invariant residues [21,23]. The highest resolution structure of SW 1A6M is considered here as the reference structure.

Figure 3 displays (ϕ, ψ) Ramachandran plot for the reference structure 1A6M of SW Mb. A pronounced feature of the plot is that the two residues Lys79 and Lys98 are in the conformation with a positive value of the backbone dihedral angle ϕ of 139 non-terminal non-Gly residues ((ϕ,ψ) values are (50, 51) for Lys79 and (57, 60) for Lys98). This is also true for the pioneering structure 1MBN ((ϕ,ψ) values are (38, 48) for Lys79 and (65, 37) for Lys98). Both the Lys79 and Lys98 residues are invariant residues across Mb species [21,23] and located in the exterior loops on the surface crevices of the protein globule, Lys79 in the EF loop between helices E and F and Lys98 in the FG loop between helices F and G [18-20].

In this study, I suspect that the conformation of both Lys79 and Lys98 residues is wrong in both 1MBN and 1A6M structures of SW Mb, due to local errors made during determination of the structures, and persuade that the backbone angle ϕ of the residues should be negative. Highly likely, the errors were introduced yet in determining the pioneering structure [18], in the course of detecting and fitting the coordinates of the backbone atoms forming the angle ϕ of both the residues in the e.d. by building an initial model. The statement is based on the results obtained by conformational analysis of the model peptides [24-26] and revisions of the 3D structure models of proteins from the PDB [26,27]. A partial summary of the theoretical basis and the experimental data vindicating the statement is given below.

MATERIALS AND METHODS

The Mb structures determined at a resolution 2.0 Å or better were selected arbitrary for analysis from the PDB at the Research Collaboratory for Structural Bioinformatics (RCSB) (http://www.rcsb.org/pdb/). The structures were 106 of SW and 42 of different species. PDB codes of the SW Mb structures: 1F6H, the powder diffraction structure; 2BW9, the Laue structure; 1MYF, the NMR structure; 1L2K, 1MBD and 2MB5, the neutron diffraction structures; X-ray crystallographic structures: 1A6K, 1A6M, 1ABS, 1AJG, 1BVC, 1BVD, 1BZ6, 1BZP, 1BZR, 1CH7, 1CH9, 1CIK, 1CIO, 1DTI, 1DTM, 1DUK, 1DXD, 1FCS, 1IRC, 1J3F, 1JDO, 1JW8, 1H1X, 1LTW, 1MBC, 1MBI, 1MBN, 1MBO, 1MBS, 1MCY, 1MGN, 1MLF, 1MLS, 1MOA, 1MTI, 1MTK, 1MYM, 1MYT, 1NAZ, 1SPE, 1SWM, 1TES, 1UFJ, 1UFP, 1V9Q, 1VXA, 1VXB, 1VXC, 1VXF, 1YOH, 1YOG, 1YOI, 2BLH, 2BLI, 2CMM, 2EB8, 2EB9, 2EF2, 2EKT, 2EVK, 2FRF, 2G12, 2JHO, 2MBN, 2MBW, 2MGB, 2MGC, 2MGD, 2MGE, 2MGF, 2MGG, 2MGH, 2MGI, 2MGJ, 2MGK, 2MGL, 2MGM, 2MYA, 2MYE, 2SPL, 2SPM, 2SPN, 2SPO, 2ZSN, 2ZSQ, 2ZSS, 2ZSZ, 2ZT2, 3M3A, 3O89, 3RJ6, 3U3E, 4H07, 4LPI, 4MBN, 4OF9, 4OOD, 4PG6, 5MBN, 5HAV; PDB codes of other species Mb: 3RGK, human; 1DWT, 3WFT, 1HRM, 1NPF, 1NPG, 1WLA, 2FRF, 2FRJ, 2NSS, 2V1K, 3RJ6, 3VM9, 4NS2, 4TWU, 4TWV, 5AZR, and 5AZQ, pig; 1M6M and 1MYG, horse; 1LHS and 1LHT, loggerhead sea turtle; 1MBA, 1DM1, 2FAL, 2FAM, 3MBA, 4MBA, and 5MBA, slug sea hare; 2NRL, 2NRM, 2NX0, 3QM5, 3QM6, 3QM7, 3QM8, and 3QMA, blackfin tuna; 1MYT, yellowfin tuna; 1EMY, Asian elephant; 1MBS, harbor seal (resolution 2.5 Å).
Computations and data processing were performed by own programs in Fortran. Plots were drawn by SigmaPlot, figures designed by Adobe Photoshop.

RESULTS AND DISCUSSION

As is evident from Fig. 2, the allowed regions of the Ramachandran map for non-Gly residues consist of two insulated regions each of which embraces the allowed conformations with negative and positive values of angle ϕ separately. The preferable kinetic path for possible conformational interconversions between both the regions was investigated by computations of the U (ϕ, ψ, χ) potential energy surface map of the model peptides [25]. It was revealed that the path runs along the profile of the lowest potential energy depending on ϕ angle in the U (ϕ,ψ) map, or along the U (ϕ) profile [25,26]. Figures 4 and 5 illustrates the U (ϕ, ψ) map and the U (ϕ) profile for alanine residue. Three extremes on the U (ϕ) profile present the values of the potential energy in the three prominent conformations in the U (ϕ,ψ) map: 0 kcal/mol (ϕ ~ −80°) of the lowest energy in the allowed region with negative values of ϕ (at nearly (−80°, 80°)), 6 kcal/mol (ϕ ~ 60°) of the local minimum in the allowed region with positive values of ϕ (at nearly (60°, 50°)), and 16 kcal/mol (ϕ ~ 0°) of the transition state saddle point (at nearly (0°, 90°)). Accordingly, for non-Gly residues, the lowest energy conformation with negative ϕ angle is more favorable than that with positive ϕ by about 6 kcal/mol, the transition barrier between the conformations being about 16 kcal/mol.
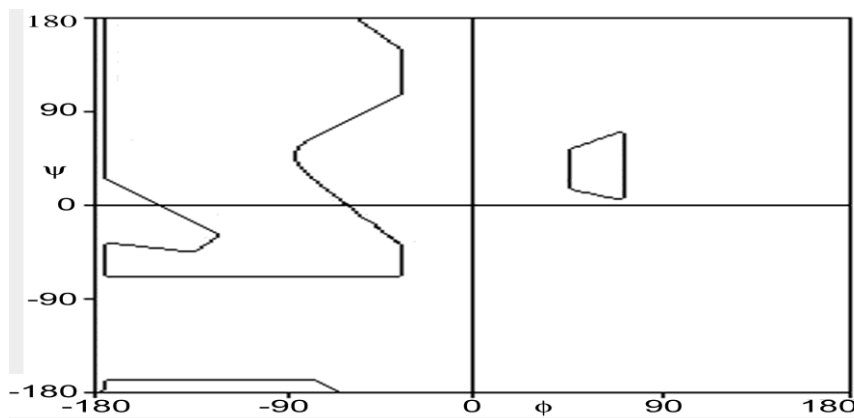
Figure 2: Alanine dipeptide classic U (φ,ψ) Ramachandran map for non-Gly residues (adapted from [3]).
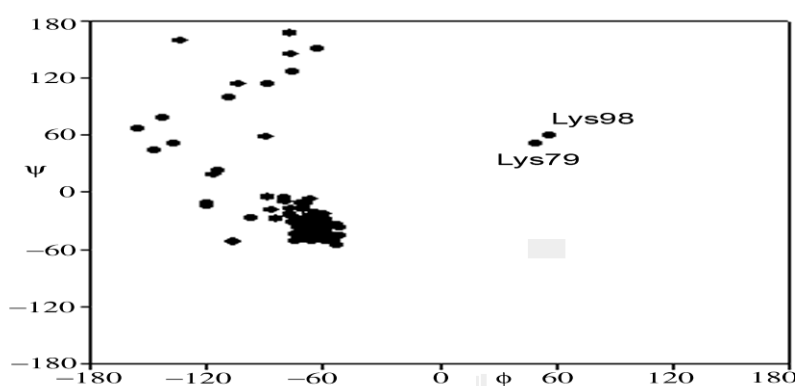

Figure 3: Ramachandran plot for non-Gly residues of the highest 1.0 Å resolution structure 1A6M of SW Mb.
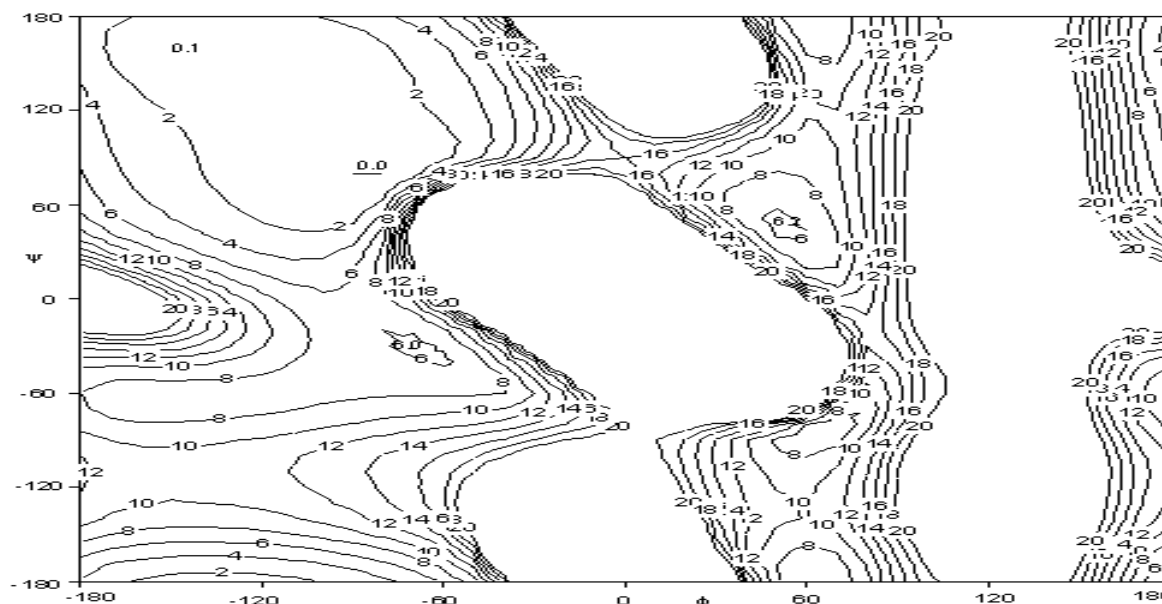

Figure 4 U (φ,ψ) map and U (φ) profile for Ala residue obtained from the computation of Ala tripeptide at 10° intervals. The values are for U in kcal/mol and for (φ,ψ) in (°). The equipotential energy contours with the potential energy 20 kcal/mol and less are drawn at 2 kcal/mol intervals. The map is similar to that of for Ala dipeptide [25,26] and is proposed for diversity and for comparison with the map in Fig. 2.
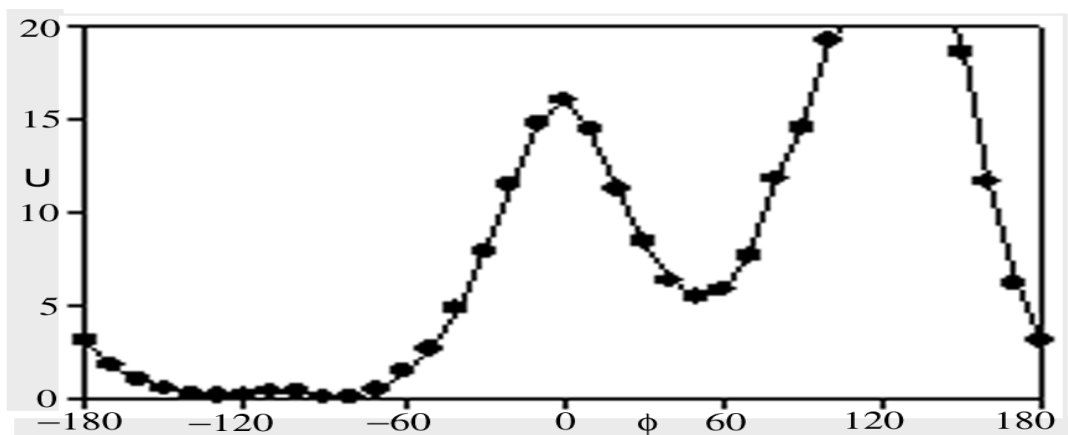
Figure 5 U (ϕ) profile for Ala residue plotted based on the U (ϕ⯑⯑ψ) map in Fig.3. The values are for U in kcal/mol and for ϕ in (°).

Using the estimated values of the energy, the equilibrium and the rate constants and the residence times of the allowed conformations with negative and positive values of angle ϕ can be estimated by famous Arrhenius formulas $K = \exp(-\Delta H°/RT)$ and $\tau = \tau_o*K$ (K is the equilibrium constant, τ is the rate constant, $\Delta H°$ is the energy difference of two conformations, RT = 0.5961 kcal/mol at 27°C, and $\tau_o \sim 10^{-12}$ s). The parameters obtained are $2.35*10^4$ for the rate constant and 0.454 s for the residence time of the lowest energy conformation and 0.454 s for the rate constant and $1.93*10^{-5}$ s for the residence time for the local minimum. From these, the probability of a non-Gly residue to reside in allowed conformations with positive ϕ will be $4.25*10^{-5}$ ($1.93*10^{-5}/0.454$).

The above results strongly prove that it is unlikely for non-Gly residues to get and to reside in the conformations with a positive value of ϕ angle from both kinetic and thermodynamic points of view. A fundamental hypothesis can be advanced that non-Gly residues in proteins should have an allowed conformation with a negative value of ϕ angle. Accordingly, any conformation of non-Gly residues with a positive value of ϕ angle is dubious and can be a result of errors introduced during the structure solution; such residues should be reinvestigated thoroughly for coordinate errors. The deduction is the basis why both Lys79 and Lys98 residues in the SW Mb structure should have the conformation with a negative value of ϕ torsion angle. The experimental data and discussions vindicating the suggestion are below.

Conformations of both Lys79 and Lys98 were inspected in a set of Mb structures from the PDB (PDB codes of the structures can be found in section Methods), to ascertain whether these residues are in the conformation with negative value of ϕ in some Mb structures. In the structures of SW Mb, Lys79 has a negative value of ϕ angle in two structures 1VXB and 1MYF. In 1MYF, the nuclear magnetic resonance (NMR) structure, Lys79 has a negative value of angle ϕ in the 7-th, 9-th and 12-th subunits out of all 12 subunits. Lys98 is with a negative value of ϕ in two structures 2EB8 and 1DTI. Lys98 (also Lys96) is invisible in 2MBN (Table 9; in [21]). The value of ϕ of Lys98 is uncertain in three structures 1J3F, 2EB9 and 2EF2 because the preceding residue His97 is not located in these structures in experiments, as the respective PDB files remarks. In the structures of other Mb species, Lys79 has a negative value of ϕ in both subunits of horse Mb dimeric structure 3VM9. The homologue of Lys79 of SW Mb is Lys75 in blackfin tuna Mb. Value of ϕ of Lys75 is negative in the three structures 2NRL, 2NRM and 3QMA of eight blackfin tuna Mb structures inspected. Ala78 of slug sea hare Mb is the homologue of the SW Mb Lys79. Ala78 has a negative value of ϕ in the all the inspected seven structures of slug sea hare Mb. As for Lys98, it has negative value of ϕ in the structures 4NS2 of horse heart Mb and 1MBS of harbor seal Mb (resolution 2.5 Å). The homologue of SW Mb Lys98 is Phe98 of slug sea hare Mb. The value of ϕ of Phe98 is negative in the all seven structures of slug sea hare Mb inspected here. The examples vindicate that Lys79 and Lys98 and their homologies do occur in the conformation with negative value of ϕ dihedral angle in a number of Mb structures.

Table 1 provides (ϕ, ψ) values of the EF and FG loops residues in some selected Mb structures, mainly those in which both or either of Lys79, Lys98 or the homologous residue is with a negative value of ϕ angle. As

is seen from the Table, the negative value of angle ϕ of Lys79, Lys98 or the homologous residue differs from the positive value of ϕ angle of the corresponding homologous residue (Lys79 or Lys98) in the reference structure 1A6M by about 180°. Furthermore, the value of ψ angle of the preceding residue, either Lys78 or His97 or the homologue, differs from the value of ψ angle of the homologous residue in the reference structure (Lys78 or His97) by about 180° as well. Thus the conformations of the residues in the Lys78Lys79 pair with negative ϕ angle of Lys79 and in the His97Lys98 pair with negative ϕ angle of Lys98 (and the corresponding homologous residues) interrelated with the conformations of the homologous residues in the Lys78Lys79 or His97Lys98 pair in the reference structure 1A6M by imaginary flipping of the peptide plane between the pairs. In this connection, it should be reminded that a peptide plane flip ('pep-flip') is known as a rotation of the peptide plane between the residues i-1 and i that alters the ψ_{i-1} dihedral angle of residue i-1 and the ϕ_i angle of residue i by about 150-180° with a little or no displacement of the side chains [28]. Also it should be noted that utilizing of the pep-flip tool is in common use in protein structure solution by X-ray crystallography, especially in determination of the backbone atomic coordinates of the residues in loop regions of proteins [15,28-30].

The backbone geometry of the EF and FG loops was inspected in the selected Mb structures. A considerable amount of the backbone angles deviating from the standard values were revealed in almost all the structures. In the 1 Å resolution reference structure 1A6M, the values of the following backbone bond angles in (º) can be cited for the residues in Lys78Lys79 and His97Lys98 pairs compared with the standard values given in the parentheses: $C^\alpha CN_{i+1}$ = 122 (117) and $CN_{i+1}C^\alpha_{i+1}$ = 125 (121) of Lys78 ($_{i+1}$ is Lys79), $C^\alpha CN_{i+1}$ = 113 (117) of Lys79 ($_{i+1}$ is Gly80), $CN_{i+1}C^\alpha_{i+1}$ = 126 (121) of His97 ($_{i+1}$ is Lys98), and $C^\alpha CN_{i+1}$ = 111 (117) of Lys98 ($_{i+1}$ is Ile99). These and other noncited unusual values for the backbone bond angles in short EF and FG loops indicate that the backbone of the loops is strained enough. Deviations of the backbone bond angles of an unstructured chain segment detached from the globule and exposed to solvent from the standard values seemingly doubtful, especially if these angles are adjacent to the conjugated bonds, as above. The angles should be weighted and relaxed around the standard values.

An analysis of the literature on early crystallographic studies of SW Mb structure revealed the following. The structure of metMb 2MBN (replaced by 4MBN later) was refined using the coordinates of the pioneering structure 1MBN by Watson [20], the original phase angles by Kendrew et al. [18] and the newly collected intensity data [21]. The structure of oxyMb 3MBN (replaced by 5MBN later) was refined using the atomic coordinates and phase angles of the structure 2MBN and newly collected intensity data [31]. The metMb structure was solved and refined starting with structure 2MBN and new intensity data [32]. The oxyMb structure 1MBO was solved and refined using the structure 2MBN and new intensity data [33]. Structures 1MBD [34] and 2MYE [35] were solved and refined starting with structure 1MBO [33]. Structure 1MBC was solved and refined starting with the metMb structure [32] against the X-ray data [36]. Structure 1A6M was solved using structure 1MBC and 1.0 Å resolution intensity data [22]. Structures 1A6G, 1A6K, and 1A6N [22] were solved using structure 1MBC. Structures 1BVD and 1BVC were solved using structure 4MBN [37] (4MBN is replacement for 2MBN). The examples remind that the structures of many forms of SW Mb were solved starting with 2MBN, the refined (by Takano [21]) release of the pioneering SW metMb structure 1MBN solved by Kendrew et al. [18] by means of homology modeling in fact. In this template structure 2MBN residue Lys98 (also Lys96) is invisible (Table 9 in [22]).

In the studies of Mb structure the main aims were to determine accurately relative dispositions of the helices, iron atom, heme, and ligands, to study structure, function, stability and dynamics of the protein with emphasis on the residues surrounding the functional center, to explore the structure of the mutants and the protein in a variety of experimental conditions, the mechanism of ligand binding, the role of distal histidine, etc. Almost no special attention was given to the exact definition of the backbone atomic coordinates of the residues in the EF and FG loops including Lys79 and Lys98 residues. The reason is probably that both the loops are located in the poorly defined exterior surface crevices and exposed to solvent, besides, the loops backbone do not play appreciable role in the protein functioning.

A critical step in the protein structure solution is building of an initial adequate model of the protein molecule from the e.d. map. More difficult modeling case is modeling of loops. In such cases, many residues are manually adjusted to the best fit in the model and then refined. An error introduced at this early stage may persist throughout the result [10-17,28,29]. This is the case of both EF and FG loops of Mb involving Lys79 and

Lys98 residues, as noted in section Results. An error was probably introduced as early as at the stage of building of an initial model of the pioneering structure.

Table 1. Dihedral angles (ϕ, ψ) of the EF and FG loops residues in some Mb structures
_____

EF loop:
SW: -αLeu76(α)Lys(-70, -13)Lys(-79, -10)**Lys79(50, 51)**Gly(94, -23)His(-89, 58)His82(-114, 23)α-; 1A6M
SW: -αLeu76(-66, -33)Lys(-77, -5)Lys(-100, -4)**Lys79(45, 60)**Gly(89, -22)His(-94, 63)His82(-115, 16)α-; 4MBN
SW: -αLeu76(α)Lys(-39, -23)Lys(-84, 3)**Lys79(38, 48)**Gly(96, -25)His(-91, 58)His82(-94, 4)α-; 1MBN
SW: -αLeu76(α)Lys(-40, -33)Lys(-87, 21)**Lys79(-19, 107)**Gly(82, -11)His(-125, 86)His82(-129, 20)α-; 1VXB
SW: -αLeu76(α)Lys(-59, -43)Lys(-78, 83)**Lys79(-59, 43)**Gly(145, -32)His(-117, 77)His82(-126, 177)α-; 1MYF (7s)
SW: -αLeu76(α)Lys(-54, -43)Lys(-83, 64)**Lys79(-38, 74)**Gly(88, 0)His(-135, 66)His82(-93, 64)α-; 1MYF (9s)
HH: -αLeu76(α)Lys(-71, -17)Lys(-84, -1)**Lys79(52, -127)**Gly(-80,-8)His(-111,57)His82(-100,10)α-; 4NS2
HU: -αLeu76(α)Lys(-70, -13)Lys(-92, 2)**Lys79(51, -127)**Gly(-78, -11)His(-114, 70)His82(-119, 20)α-; 3RGK
LT: -αLeu76(α)Lys(-64, -20)Gln(-87, 2)**Lys79(55, -125)**Asn(-78, -24)Asn(-104, 87)His82(-126, 28)α-; 1LHS
HH: -αLeu76(α)Lys(-63, -45)Lys(-68, -35)**Lys79(-62, -45)**Gly(-58,-44)His(-63,-42)His82(-67,-46)α-; (I) 3VM9
HH: -αLeu76(α)Lys(-63, -46)Lys78(-68, -34)**Lys79(-60, -46)**Gly(-59,-42)His(-62,-44)His82(-64,-49)α-; (II) 3VM9
BT: -αLeu72(α)Lys(-88, -3)Ala(-61, 130)**Lys75(-97, -5)**Gly(-114, -149)Ser(-66, 123)His78(-71, -40)α-; 2NRL
BT: -αLeu72(α)Lys(-83, -12)Ala(-57, 133)**Lys75(-100, -23)**Gly(-91, -162)Ser(-54, 122)His78(-69, -35)α-; 2NRM
BT: -αLeu72(α)Lys(-99, -1)Ala(-68, 136)**Lys75(-104, -7)**Gly(-110, -155)Ser(-63, 114)His78(-68, -40)α-; 3QMA
BT: -αLeu72(α)Lys(-68, -24)Ala(-75, -17)**Lys75(56, 42)**Gly(91, 2)Ser(-119, 90)His78(-142, -19)α-; 2NX0
SH: -αVal75(α)Asn(-63, -38)Asn(-105, 8)**Ala78(-52, -29)**Ala(-92, 7)Asn(-119, 95)Ala(-49, -33)Gly82(α)α-; 1MBA
SH: Ala78(ϕ, ψ): **(-62,-32)** in 1DM1, **(-52, -29)** in 2fal, **(-52, -24)** in 2FAM, **(-47, -36)** in 3MBA,
          (-96, -21) in 4MBA, and **(-46, -35)** in 5MBA.

FG loop:
SW: -α-Thr95(-97, -27)Lys(-106, -52)His(-86, -19)**Lys98(57, 60)**Ile(-103, 114)Pro100(-63, 150)α-; 1A6M
SW: -α-Thr95(-81, -31)Lys(-101, -54)His(-83, -23)**Lys98(61, 56)**Ile(-100, 113)Pro100(-54, 138)α-; 4MBN
SW: -α-Thr95(-75, -14)Lys(-138, -63)His(-53, -39)**Lys98(65, 37)**Ile(-90, 109)Pro100(-61, 135)α-; 1MBN
SW: -α-Thr95(-135, 144) Lys(-88, -14)His(-94, 155)**Lys98(-80, 128)**Ile99(-79, 102)Pro100(-61, 148)α-; 2EB8
SW: -α-His93(-100, 2)Ala94(-121, -20)-
     -Thr95(-123, -55)Lys(-51, -63)Asp(-120, 154)**Lys98(-114, 94)**Ile(-106, 113)Pro100(-64, 154)α-; 1DTI
SW: -α-Thr95(-132, unc.)Lys96(n.l.)His(n.l.)**Lys98(**unc., **37)**Ile(-105, 102)Pro100(-65, 158)α-; 1J3F
SW: -α-Thr95(-67, -30)Lys96(-164, unc.)His(n.l.)**Lys98(**unc., **37)**Ile(-85, 115)Pro100(-64, 146)α-; 2EF2
SW: -α-Thr95(n.l.)Lys96(n.l.)His(n.l.)**Lys98(**unc.**, 37)**Ile(-105, 102)Pro100(-65, 158)α-; 2EB9
HU: -α-His93(-78, -29)Ala(-73, -58)-
     -Thr95(-86, -34)Lys(-102, -52)Lys(-85, -21)**Lys98(61, 57)**Ile(-100, 114)Pro100(-68, 154)α-; 3RGK
HH: -αHis93(-88, 6)Ala(-114, -6)-
     -Thr95(-129, -77)Lys(-59, -43)His(-130, 104)**Lys98(-69, 97)**Ile(-108, 112)Pro100(-63, 150)α-; 4NS2
HH: -α- His93(-73, -30)Ala(-76, -47)-
      -Thr95(-86, -47)Lys(-87, -54)His(-85, -18)**Lys98(63, 47)**Ile(-98, 116)Pro100(-57, 144)α-; 5AZR
HS: -α-Thr95(-66, -56)Lys96(-107, -77)His(-90, 53)**Lys98(-19, 149)**Ile(-79, 68)Pro(-58, 124)Ile101(-9, -139)α-; 1MBS
SH: -α-His95(-59, -44)Val(-60, -39)Gly(-60, -29)**Phe98(-90, 3)**Gly(92, 6)Val(-88, 132)Gly(-126, -174) Ser102(α)α-; 1MBA
SH: Phe98: **(-102, 11)**, in 1DM1; **(-94, 2)**, in 2FAL; **(-95, 2)**, in 2FAM; **(-93, 4)**, in 3MBA;
          (-93, 8), in 4MBA; **(-93, 4)**, in 5MBA.
_____

SW – sperm while: 1A6M, the highest 1 Å resolution structure of metMb[22], the reference structure; 1MBN, the pioneering structure [18-20]; 4MBN, the refined pioneering structure[21]; 1MYF, NMR structure (7s and 9s designate the 7th and 9th subunit). HH – horse heart (I, II the 1-st and 2-d subunits of 3VM9), HU – human, LT – loggerhead sea turtle, BT – blackfin tuna, SH – slug sea hare, HS – sea harbor seal 1MBS. α, designates α-helix. Lys79, Lys98 and their homologies shown in bold and those of these residues are with negative value of ϕ also underlined; n.l., residue not located in experiment; unc., value of the angle uncertain as the preceding residue not defined in experiment.

CONCLUSIONS

The results of this study evince that the coordinates of all or some backbone atoms at least of the Lys78Lys79 and His97Lys98 pairs residues of the EF and FG loops are to be refined in SW Mb reference structure 1A6M (also in other Mb structures inspected). There is a need for the studies aimed at rerefinement of the backbone atomic coordinates of both EF and FG loops in the structures of Mb species by means of careful rebuilding of the structure models of the loops and refinement by X-ray protein crystallography or by

other appropriate experimental methods of structural biology. It is reasonably safe therewith to suggest that the coordinates to be looked for should be such that to provide a negative value of angle ϕ for residues Lys79 and Lys98, in conformity with both theoretical finding and a few experimental observations enlightened in section Results.

Seemingly there is nothing out of the ordinary if the backbone conformation of the two distant individual residues in a protein Mb structure is fault, especially since these residues do not play a noticeable role in the protein function. But hard experimental confirmation that the backbone ϕ dihedral angle of both Lys79 and Lys98 residues has a negative value in the pioneering protein structure of Mb has fundamental importance in structural biology and bioinformatics. The confirmation will lead to further improvement of the existing many structures of Mb species and also vindicate the eligibility of the theoretical finding enlightened in this study. In turn, this will stimulate the studies on the further trial of the hypothesis to use it as a simple means for both solution of the new and further improvement of the existing protein structures.

REFERENCES

[1]   Sasisekharan V. Stereochemical criteria for polypeptide and protein structures. In: Collagen: Ramanathan N, ed., Wiley & Sons, Madras, 1962, pp. 39–78.
[2]   Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 1963; 7: 95–99.
[3]   Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. Adv. Prot. Chem. 1968; 23: 283–437.
[4]   Brant DA, Schimmel PR. Analysis of the skeletal configuration of crystalline hen egg-white lysozyme. Proc. Natl. Acad. Sci. USA 1967; 58: 428–435.
[5]   Scheraga HA. Calculations of conformations of polypeptides. Adv. Phys. Org. Chem. 1968; 6: 103–184.
[6]   Momany FA, McGuire RF, Burgess AW, Scheraga HA. Energy parameters in polypeptides, VII: geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J. Phys. Chem. 1975; 79: 2361–2380.
[7]   Vasquez M, Nemethy G, Scheraga H. Conformational energy calculations of polypeptides and proteins. Chem. Rev. 1994; 94: 2183–2239.
[8]   Bernstein FC et al. The protein data bank: A computer based archival file macromolecular structures. J Mol Biol. 1977; 112: 535–542.
[9]   Berman HM, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucl. Acids Res. 2007; 35: D301–D303.
[10]   Blundell TL, Johnson LN. Protein crystallography, Academic Press, New York, 1976.
[11]   Richardson JS. The anatomy and taxonomy of protein structure. Adv. in Prot. Chem. 1981; 34: 167–339.
[12]   Drenth J. Principles of protein X-ray crystallography, Springer, New York, 2007.
[13]   Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and location of errors in these models. Acta Cryst. 1991; A47: 110–119.
[14]   Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. Proteins: Struct. Funct. Genet. 1992; 12: 345–364.
[15]   Kleywegt GJ. Validation of protein crystal structures. Acta Cryst. 2000; D56: 249–265.
[16]   Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Cryst. 2010; D66: 12–21.
[17]   Read RJ et al. A new generation of crystallographic validation tools for the Protein Data Bank. Structure 2011; 19: 1395-1412.
[18]   Kendrew JC et al. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Angstrom resolution, Nature 1960; 185: 422–427.
[19]   Kendrew JC. Myoglobin and the structure of proteins (Nobel Lecture, 1962). Prix Nobel 1963; 103: 676–693.
[20]   Watson HC. The stereochemistry of the protein myoglobin. Prog. Stereochem. 1969; 4: 299–333.
[21]   Takano T. Structure of Myoglobin refined at 2.0 Angstroms resolution I. Crystallographic refinement metmyoglobin from sperm while. J. Mol. Biol. 1977; 110: 537–568.
[22]   Vojtechovsky J, Chu K, Berendzen J, Sweet RM, Schlichting I. Crystal structures of myoglobin-ligand complexes at near-atomic resolution, Biophys. J. 1999; 77: 2153–2174.

[23]   Bashford D, Chothia C, Lesk AM. Determinants of a protein fold. Unique features of the globin amino acid sequences. J. Mol. Biol. 1987; 196: 199–216.

[24]   Basharov MA. Potential energy surfaces of the peptides and the conformations of the amino acid residues in proteins. Preliminary results of treatment of a protein structure bank. Biophysics 1995; 40: 229–241.

[25]   Basharov MA. The internal rotational barriers about NC^{α} and $C^{\alpha}C$ backbone bonds of polypeptides. Eur. Biophys. J. 2012; 41: 53–61.

[26]   Basharov MA. Nonglycine residues in proteins should most likely have an allowed conformation with a negative value for backbone torsion angle ϕ. Annual Research & Review in Biology 2017; 17: 1–17.

[27]   Basharov MA. Distribution of the conformations of amino-acid residues encountered in three-dimensional protein structures. Analysis of non-glycine residues in 'positive' conformations. Biophysics 1997; 42: 753–765.

[28]   Kleywegt GJ. Use of non-crystallographic symmetry in protein structure refinement. Acta Cryst. 1996; D52: 842–857.

[29]   Kleywegt GJ, Jones TA. Databases in protein crystallography. Acta Cryst. 1998; D54: 1119–1131.

[30]   Hayward S. Peptide plane flipping in proteins. Protein Sci. 2001; 10: 2219–2227.

[31]   Takano T. Structure of Myoglobin refined at 2.0 Angstroms resolution II. Crystallographic refinement metmyoglobin from sperm while. J. Mol. Biol. 1977; 110: 569–584.

[32]   Frauenfelder H, Petsko GA, Tsernoglou D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. Nature 1979; 280: 558–563.

[33]   Phillips SEV. Structure and refinement of oxymyoglobin at 1.6 Å resolution. J. Mol. Biol. 1980; 142: 531–554.

[34]   Phillips SEV, Schoenborn BP. Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin. Nature 1981; 292: 81–82.

[35]   Johnson KA, Olson JS, Phillips JN, Jr. Structure of myoglobin-ethyl isocyanide. Histidine as a swinging door for ligand entry. J. Mol. Biol. 1989; 207: 459–463.

[36]   Kuriyan J, Wilz S, Karplus M, Petsko GA. X-ray structure and refinement of Carbon-monoxy (FeII)-Myoglobin at 1.5 Å resolution. J. Mol. Biol. 1986; 192: 133–154.

[37]   Wagner UG, Muller N, Schmitzberger W, Falk H, Kratky C. Structure determination of the biliverdin apomyoglobin complex: crystal structure analysis of two crystal forms at 1.4 and 1.5 Å resolution. J. Mol. Biol. 1995; 247: 326–337.