Frontiers in Bioscience Landmark

Front. Biosci. (Landmark Ed) **2024**; 29(11): 379
https://doi.org/10.31083/j.fbl2911379

Opinion

Formation of the Native Topology of a Protein is due to the "Conserved but Non-Functional" Residues: A Case of Apomyoglobin Folding

Valentina E. Bychkova[1,*,†], Dmitry A. Dolgikh[2,3], Vitalii A. Balobanov[1], Alexei V. Finkelstein[1,3,*,†]

[1]Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia
[2]Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117871 Moscow, Russia
[3]Biology Department, Lomonosov Moscow State University, 119192 Moscow, Russia
*Correspondence: vbychkova@mail.ru (Valentina E. Bychkova); afinkel@vega.protres.ru (Alexei V. Finkelstein)
[†]These authors contributed equally.

Academic Editor: Yudong Cai
Submitted: 17 July 2024 Revised: 9 September 2024 Accepted: 12 September 2024 Published: 8 November 2024

Abstract

This paper is dedicated to the memory of Oleg B. Ptitsyn (1929-1999) and presents an answer to his question: "What is the role of conserved non-functional residues in protein folding?". This answer follows from the experimental works of three labs. The role of non-functional but conserved residues of apomyoglobin (apoMb) in the formation of the native protein fold in the molten globule state has been experimentally revealed. This research proves that the non-functional but conserved residues of apoMb are necessary for the formation and maintenance of the correct topological arrangement of the main elements in the apoMb secondary structure already in the early folding intermediate.

Keywords: globin; apomyoglobin; conserved residues; non-functional residues; protein stability; protein folding; folding intermediate; native fold topology

1. Introduction

The discovery of conserved but non-functional residues in c-type cytochromes and globins posed the issue regarding their possible role in protein folding [1,2]. It has been hypothesized [1–3] that such conserved residues play a key role in the early stages of the process of protein folding. In this review, we present verification of this hypothesis using the experiments on apomyoglobin (apoMb) folding.

It is well known that most of the native protein structures are characterized by a fixed native topology of the protein chain and rigid native positions of all or nearly all amino acid residues in this chain [4–8]. The native protein structure can be experimentally determined by X-ray crystallography or high-resolution electron cryo-microscopy of protein crystals, or the nuclear magnetic resonance (NMR) spectroscopy of proteins in solution; all these methods show the positions of all or nearly all protein atoms.

The protein topology is determined by the native arrangement of the main elements of the secondary structure of the protein chain. The main elements of the hologlobin secondary structure are A, B, E, F, G, and H helices (Fig. 1A, Ref. [8,9]), though helix F is disordered in apo-globins, i.e., in the absence of the heme [9]. These helices are connected by non-covalent contacts. The most conserved close contacts (except those directly involved in the binding of the heme, the functional group of a globin) exist between A, G, and H helices of globins [1,2,10]; these helices form a π-like structure (Fig. 1B,C, Ref. [1, 10]), and the contacts between them glue together the N-ends and C-ends of the globin chain. Thus, these conserved (or virtually—by $\geq$70% [1]—conserved, see Fig. 1D (Ref. [1])) hydrophobic—aliphatic or aromatic—residues and contacts between them play a very important role in protein folding, but at *what* stage of the folding process they come into play was unclear until recently.

2. Conserved Residues and a Pathway of ApoMb Folding

Shakhnovich, Abkevich, and Ptitsyn put forward a seminal idea about the special role performed by conserved but non-functional residues in the formation of protein structure and especially its folding nucleus [3]. Later, this idea was developed and expanded in the theoretical works by Shakhnovich's group *et al*. [11–15] and Ptitsyn & Ting [1]. The latter study reported that in each of 728 globin sequences, there are 6 conserved non-functional residues localized in a complex of A, G, H helices. Then it was already known that these helices form at the early stage of apoMb folding and are rather stable in the equilibrium molten globule of apoMb [16,17], which allowed the assumption that in this state these three helices are tightly packed with one another. However, the exact folding steps where they form, and pack together were elucidated only over the next two decades (see below).

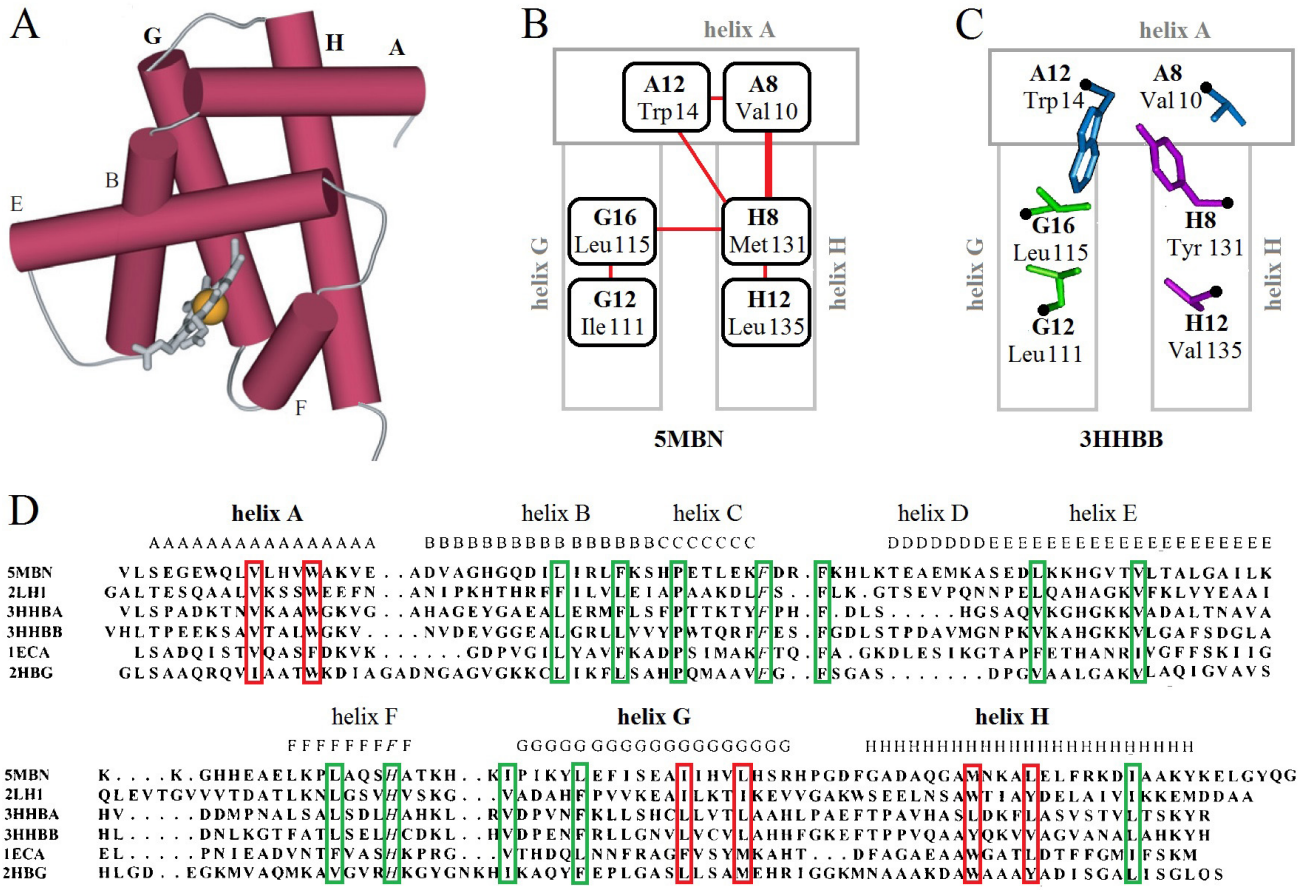

Fig. 1. Conserved amino acid residues in globins. (A) The structure of a globin consisting of two crossed layers of three α-helices each. The A, E, F helices (lettered by their positions in the protein chain) form the upper layer, and H, G, B helices form the lower layer. The small C, D helices (of 1–2 turns each) are not shown because they are present not in all globins. Bold letters indicate helices A, G, and H, which have intriguing conserved but "non-functional" contacts with each other. A cavity in the bottom of the upper layer of helices houses the heme (Adapted with permission from Finkelstein AV, Protein Physics. A Course of Lectures. 2nd edn.; published by Academic Press, An Imprint of Elsevier Science, 2016 [8]; helix F is disordered in the absence of heme [9]). (B) A cluster of the most conserved "non-functional" residues and contacts [1,10] is located at the interface of A, G, H helices (while the conserved "functional" residues [1] in the heme-containing region are not shown here); the helices A, G, H form a π-like topology. In this picture, the amino acid residue positions are numbered following their places in A, G, H helices; the residue names and numbers correspond to those in the sperm whale apomyoglobin (apoMb) structure 5MBN presented in the Protein Data Bank (PDB). The thick red bar shows the strongest (20 atom-atom contacts) and the most conserved residue-residue contact between helices A and H [1]; the thin red bars show the weaker (8–15 atom-atom contacts) interactions between the conserved residues of helices A, G, and H [1,10]. Adapted with permission from Ptitsyn OB, Non-functional conserved residues in globins and their possible role as a folding nucleus; published by Elsevier, 1999 [1]. (C) Analogous cluster of the most conserved or virtually conserved "non-functional" side chains and contacts in human hemoglobin B (PDB code: 3HHBB) [1] on the background od schematically presented helices A, G, H; the drawing from the laboratory archive of O.B. Ptitsyn. Adapted with permission from Ptitsyn OB, Non-functional conserved residues in globins and their possible role as a folding nucleus; published by Elsevier, 1999 [1]. (D) Alignment of six representative globin sequences (out of 728 ones examined in [1]). The PDB code of protein structure is in the first column. The conserved or virtually conserved *non*-functional residues singled out by Ptitsyn [1] are encircled in red; the singled out [1] conserved or virtually conserved *functional* residues are encircled in green; and two completely conserved *functional* residues (CD1 and F8) are in italics. Adapted with permission from Ptitsyn OB, Non-functional conserved residues in globins and their possible role as a folding nucleus; published by Elsevier, 1999 [1].

It should be noted that similar conserved clusters of non-functional residues were found also in c-type cytochromes [1,18], where their interactions likewise glue together N- and C-terminal α-helices. Similarly to A, G, H helices in globins, in cytochrome c, the N- and C-terminal helices form at the earliest stage of folding of this protein,

remain rather stable in the molten globule state [19–21], and are closely packed with one another when they are in the molten globule.

It is noteworthy that the results obtained by Musto *et al.* [10] using site-directed mutants of *Aplysia limachine* apoMb (distantly related to mammalian myoglobins) agree with results obtained for the sperm whale apoMb, and they clearly show that the conserved but non-functional position H8 is involved in stabilization of the main folding intermediate in both these proteins. The residue occupying this H8 position thus plays a very important role: it is evolutionarily conserved, as a large hydrophobic residue, in all the globin family, from invertebrates to fish and mammals. These results strongly support the statement that the cluster formed by the A, G, H helices is an important element for the maintenance of the proper folding pathway of different globins.

The papers [22,23] from Wright's laboratory at The Scripps Research Institute, USA, describe, for a set of mutants, the hydrogen exchange pulse-labeling kinetic experiments with the NMR detection. They have shown that helix H forms contacts with residues of helices G and A already at the stage of formation of a molten globule, when other helices (B, C, D, E, F) apparently have not yet formed, at least at acidic pH [24–26]. This stage yields an ensemble of metastable compact intermediates with a native-like arrangement of helices A, B, E, G, H. Though, there is evidence for a small shift of helix H in these burst-phase intermediates, as compared to its location in the native protein or the equilibrium molten globule. This shift, by approximately one turn of the helix H towards its N-end, maximizes hydrophobic contact of helices H and G. The folding, in experiments, was caused by a very fast pH jump (pH2.2 $\rightarrow$ pH6.0) monitored by stopped-flow circular dichroism and fluorescence measurements.

It was only the technique of kinetic measurements with continuous-flow capillary mixing, developed in Roder's lab at Fox Chase Cancer Center, USA, that allowed investigating very rapid, microsecond-long folding events [24–27] occurring in apoMB, some other proteins, and their mutants. It made clear how fast the A-G-H complex of helices forms and the contacts of which residues connect these helices. It has been found that the initial free energy barrier is overcome during the protein chain compaction, and the resulting loosely packed folding intermediate accumulates before the main rate-limiting step in the formation of the native tertiary structure. Molecular compaction and helix formation (seemingly associated with a conformational search for the correct tertiary contacts [28]) occur within the first few hundred microseconds of initiating the apoMb folding.

Kinetic studies of the molten globule state formation in apoMb and its mutants shed light on the early stages of protein folding [27]. It was confirmed that conserved "non-functional" residues, theoretically discovered by Ptitsyn, play an essential role mainly at the stage of formation of the apoMb molten globule, albeit they certainly contribute to the apoMb native structure stability too. The contacts between these conserved non-functional residues keep the correct mutual position of A, G, H helices already during the formation of the molten globule state. Later, their interaction provides the fixation of the correct (native) topology of apoMb [27,29].

It is known that the reason for the conservation of specific protein features—such as specific amino acid residues or their interacting pairs, residue conformations, *etc.*—in a certain position within the given protein structure is their strong stabilization of this structure precisely by existing in this position [8,30,31]; this results from the natural selection of sequences that stabilize the given protein structure [8,31]. Thus, equivalent parts of the structure of similar protein folds must consist of the same or similar amino acid residues, which leads to the conservation of even "non-functional" amino acid residues and their interacting pairs. It is not out of place mentioning that Fig. 1B presents a good illustration of the principle of consistency [32–34] of the short-range (intra-helical) and long-range (inter-helical) interactions, which leads to a significant stabilization of protein structure.

It is worth noting that the modern deep learning-based approaches to predicting the protein structures from amino acid sequences of their chains (AlphaFold, etc.) [35–38] are largely based on identifying similar contacts of amino acid residues both in the protein whose spatial structure is to be predicted and in already known protein structures—in other words, these approaches amply use the conservation of non-functional amino acid residues.

3. Conclusions

Studying the involvement of non-functional but conserved amino acid residues of apoMb in the process of its folding has led to a deeper understanding of how the topology of the native protein is formed.

Resulting from the joint work of three laboratories: ours, at the Institute of Protein Research, Russia; Wright's, at The Scripps Research Institute, USA; and Roder's, at the Fox Chase Cancer Center, USA, the pathway of formation of the native topology in apoMb has been first demonstrated, with the elucidation of the key role of "conserved non-functional residues" in this process [27,29].

Abbreviations

ApoMb, apomyoglobin; NMR, nuclear magnetic resonance spectroscopy.

Availability of Data and Materials

All data reported in this paper are publicly available.

Author Contributions

This manuscript was written in tight collaboration of all authors of this paper (VEB, DAD, VAB, AVF): DAD and

VAB collected and analyzed the literature data, VEB and DAD prepared the initial draft, VAB and AVF corrected and extended it, and passed to VEB, who collected and added the additional references and edited the final text herself, and AVF has composed the Figure. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

We are grateful to professor P. Wright of the Scripps Research Institute and professor H. Roder of the Institute for Cancer Research, Fox Chase Cancer Center, for their contribution in discussion of this research, and to our coworkers who participated in this prolonged investigation.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Ptitsyn OB, Ting KL. Non-functional conserved residues in globins and their possible role as a folding nucleus. Journal of Molecular Biology. 1999; 291: 671–682.

[2] Ptitsyn OB. Protein evolution and protein folding: non-functional conserved residues and their probable role. Pacific Symposium on Biocomputing. 1999; 494–504.

[3] Shakhnovich E, Abkevich V, Ptitsyn O. Conserved residues and the mechanism of protein folding. Nature. 1996; 379: 96–98.

[4] Perutz MF. Protein Structure. W.H. Freeman & Co.: New York, USA. 1992.

[5] Branden C, Tooze J. Introduction to Protein Structure. 2nd edn. Garland Science: New York, USA. 1999.

[6] Frauenfelder H. The Physics of Proteins. An Introduction to Biological Physics and Molecular Biophysics. Springer: New York, USA. 2010.

[7] Lesk AM. Introduction to Protein Science: Architecture, Function, and Genomics. 2nd edn. Oxford University Press: Oxford, New York. 2010.

[8] Finkelstein AV, Ptitsyn OB. Protein Physics. A Course of Lectures. 2nd edn. Academic Press, An Imprint of Elsevier Science: Amsterdam • Boston • Heidelberg • London •New York • Oxford • Paris • San Diego •San Francisco • Singapore • Sydney • Tokyo. 2016.

[9] Eliezer D, Wright PE. Is apomyoglobin a molten globule? Structural characterization by NMR. Journal of Molecular Biology. 1996; 263: 531–538.

[10] Musto R, Bigotti MG, Travaglini-Allocatelli C, Brunori M, Cutruzzolà F. Folding of Aplysia limacina apomyoglobin involves an intermediate in common with other evolutionarily distant globins. Biochemistry. 2004; 43: 230–236.

[11] Michnick SW, Shakhnovich E. A strategy for detecting the conservation of folding-nucleus residues in protein superfamilies. Folding & Design. 1998; 3: 239–251.

[12] Mirny LA, Shakhnovich EI. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. Journal of Molecular Biology. 1999; 291: 177–196.

[13] Dokholyan NV, Shakhnovich EI. Understanding hierarchical protein evolution from first principles. Journal of Molecular Biology. 2001; 312: 289–307.

[14] Donald JE, Hubner IA, Rotemberg VM, Shakhnovich EI, Mirny LA. CoC: a database of universally conserved residues in protein folds. Bioinformatics. 2005; 21: 2539–2540.

[15] Shakhnovich E. Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. Chemical Reviews. 2006; 106: 1559–1588.

[16] Jennings PA, Wright PE. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science. 1993; 262: 892–896.

[17] Tcherkasskaya O, Ptitsyn OB. Direct energy transfer to study the 3D structure of non-native proteins: AGH complex in molten globule state of apomyoglobin. Protein Engineering. 1999; 12: 485–490.

[18] Ptitsyn OB. Protein folding and protein evolution: common folding nucleus in different subfamilies of c-type cytochromes? Journal of Molecular Biology. 1998; 278: 655–666.

[19] Jeng MF, Englander SW, Elöve GA, Wand AJ, Roder H. Structural description of acid-denatured cytochrome c by hydrogen exchange and 2D NMR. Biochemistry. 1990; 29: 10433–10437.

[20] Marmorino JL, Pielak GJ. A native tertiary interaction stabilizes the A state of cytochrome c. Biochemistry. 1995; 34: 3140–3143.

[21] Roder H, Elöve GA, Englander SW. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. Nature. 1988; 335: 700–704.

[22] Nishimura C, Dyson HJ, Wright PE. Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. Journal of Molecular Biology. 2006; 355: 139–156.

[23] Aoto PC, Nishimura C, Dyson HJ, Wright PE. Probing the non-native H helix translocation in apomyoglobin folding intermediates. Biochemistry. 2014; 53: 3767–3780.

[24] Shastry MC, Roder H. Evidence for barrier-limited protein folding kinetics on the microsecond time scale. Nature Structural Biology. 1998; 5: 385–392.

[25] Roder H, Maki K, Cheng H. Early events in protein folding explored by rapid mixing methods. Chemical Reviews. 2006; 106: 1836–1861.

[26] Xu M, Beresneva O, Rosario R, Roder H. Microsecond folding dynamics of apomyoglobin at acidic pH. The Journal of Physical Chemistry. B. 2012; 116: 7014–7025.

[27] Mizukami T, Xu M, Fazlieva R, Bychkova VE, Roder H. Complex Folding Landscape of Apomyoglobin at Acidic pH Revealed by Ultrafast Kinetic Analysis of Core Mutants. The Journal of Physical Chemistry. B. 2018; 122: 11228–11239.

[28] Uzawa T, Akiyama S, Kimura T, Takahashi S, Ishimori K, Morishima I, *et al*. Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of alpha-helical content and compactness. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101: 1171–1176.

[29] Bychkova VE, Dolgikh DA, Balobanov VA. Function of the Conserved Non-Functional Residues in Apomyoglobin - to Determine and to Preserve Correct Topology of the Protein. Biochemistry. Biokhimiia. 2023; 88: 1905–1909.

[30] Pohl FM. Empirical protein energy maps. Nature: New Biology. 1971; 234: 277–279.

[31] Finkelstein AV, Badretdinov AYa, Gutin AM. Why do protein architectures have Boltzmann-like statistics? Proteins. 1995; 23: 142–150.

[32] Ptitsyn OB, Finkel'shteĭn AV. Predicting the spiral portions of globular proteins from their primary structure. Doklady Akademii Nauk SSSR. 1970; 195: 221–224.
[33] Ptitsyn OB, Finkel'shteĭn AV. Relation of the secondary structure of globular proteins to their primary structure. Biofizika. 1970; 15: 757–768.
[34] Go N. Theoretical studies of protein folding. Annual Review of Biophysics and Bioengineering. 1983; 12: 183–210.
[35] Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, *et al*. Improved protein structure prediction using potentials from deep learning. Nature. 2020; 577: 706–710.
[36] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al*. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596: 583–589.
[37] Ahdritz G, Bouatta N, Floristean C, Kadyan S, Xia Q, Gerecke W, *et al*. OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. Nature Methods. 2024; 21: 1514–1524.
[38] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, *et al*. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024; 630: 493–500.