European Journal of Medicinal Chemistry 91 (2015) 100–108

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Towards understanding the molecular recognition process in prokaryotic zinc-finger domain

Luigi Russo [a], Maddalena Palmieri [a], Jolanda Valentina Caso [a], Gianluca D' Abrosca [a], Donatella Diana [b], Gaetano Malgieri [a], Ilaria Baglivo [a], Carla Isernia [a, c], Paolo V. Pedone [a], Roberto Fattorusso [a, c, *]

[a] Department of Environmental, Biological and Pharmaceutical Sciences and Technology, Via Vivaldi 43, 81100 Caserta, Italy
[b] Institute of Biostructures and Bioimaging -CNR, Via Mezzocannone 16, 80134 Naples, Italy
[c] Interuniversity Centre for Research on Bioactive Peptides (CIRPEB), University of Naples Federico II, Via Mezzocannone 16, 80134 Naples, Italy

ARTICLE INFO

Article history:
Received 10 June 2014
Received in revised form
5 September 2014
Accepted 11 September 2014
Available online 16 September 2014

Keywords:
Prokaryotic zinc finger
Molecular Dynamics simulations
NMR
Docking
Protein-DNA complex

ABSTRACT

Eukaryotic Cys_2His_2 zinc finger domain is one of the most common and important structural motifs involved in protein-DNA interaction. The recognition motif is characterized by the tetrahedral coordination of a zinc ion by conserved cysteine and histidine residues. We have characterized the prokaryotic Cys_2His_2 zinc finger motif, included in the DNA binding region (Ros87) of Ros protein from *Agrobacterium tumefaciens*, demonstrating that, although possessing a similar zinc coordination sphere, this domain presents significant differences from its eukaryotic counterpart. Furthermore, basic residues flanking the zinc binding region on either side have been demonstrated, by Electrophoretic Mobility Shift Assay (EMSA) experiments, to be essential for Ros DNA binding. In spite of this wealth of knowledge, the structural details of the mechanism through which the prokaryotic zinc fingers recognize their target genes are still unclear. Here, to gain insights into the molecular DNA recognition process of prokaryotic zinc finger domains we applied a strategy in which we performed molecular docking studies using a combination of Nuclear Magnetic Resonance (NMR) and Molecular Dynamics (MD) simulations data. The results demonstrate that the MD ensemble provides a reasonable picture of Ros87 backbone dynamics in solution. The Ros87-DNA model indicates that the interaction involves the first two residue of the first α-helix, and several residues located in the basic regions flanking the zinc finger domain. Interestingly, the prokaryotic zinc finger domain, mainly with the C-terminal tail that is wrapped around the DNA, binds a more extended recognition site than the eukaryotic counterpart. Our analysis demonstrates that the introduction of the protein flexibility in docking studies can improve, in terms of accuracy, the quality of the obtained models and could be particularly useful for protein showing high conformational heterogeneity as well as for computational drug design applications.

1. Introduction

Specific DNA recognition by proteins is fundamental to many regulatory process that control the flow of genetic information. Transcription, DNA replication and repair are key dynamic cellular processes that require extremely tight regulation by DNA-binding proteins. Zinc fingers of the Cys_2His_2 class (or classical), first discovered in Transcription factor IIIA (TFIIIA) [1–3], constitute one of the most abundant and important structural motifs involved in protein-DNA interactions and is also known to be involved in binding of RNA, lipids, and proteins [4–7]. The classical Cys_2His_2 zinc finger is a small domain that typically contains a repeated 28–30 amino acid sequence in which a zinc ion, crucial for its stability, is tetrahedrally coordinated by two cysteines and two histidine residues. This motif folds into a compact 3D structure, stabilized by the zinc ion and a small hydrophobic core and consists in an antiparallel β-sheet faced by an α-helix (ββα fold) [5]. The α-helix is constituted of three turns including the two coordinating histidines on two successive turns at the C-terminus of the finger, whereas the β-sheet is located at the N-terminal part and contains

* Corresponding author. Department of Environmental, Biological and Pharmaceutical Sciences and Technology, Via Vivaldi 43, 81100 Caserta, Italy.
E-mail address: roberto.fattorusso@unina2.it (R. Fattorusso).

http://dx.doi.org/10.1016/j.ejmech.2014.09.040
0223-5234/

the two cysteines. Structural studies on classical zinc finger protein-DNA complexes have demonstrated that sequence specific molecular recognition is achieved by the interaction of the DNA bases with side chains from the surface of the N-terminus part of the α-helix [5]. A single zinc finger domain in itself is not sufficient for high-affinity binding to a specific DNA target sequence. In fact, proteins with multiple zinc finger domains require a minimum of two zinc fingers for high affinity DNA binding [4,8]. Nevertheless, the single zinc finger domain present in the *Drosophila* GAGA transcription factor [9,10], as well as the single zinc finger domain of the *Arabidopsis thaliana* SUPERMAN protein [11,12], are capable of sequence specific DNA binding when flanked by basic regions. Recently, we have characterized the prokaryotic Cys_2His_2 zinc finger motif, included in the DNA binding region (Ros87) of Ros protein from *Agrobacterium tumefaciens*, demonstrating that, although possessing a similar zinc coordination sphere, this domain presents significant differences from its eukaryotic counterpart. In particular, the prokaryotic Cys_2His_2 zinc finger domain adopts a globular fold, consisting of 58 amino acids, arranged in a ββββαα topology and stabilized by an extended hydrophobic core [13]. This domain can also overcome the metal requirement to properly fold and function [14,15]. Furthermore, basic residues flanking the zinc binding region on either side have been demonstrated, by EMSA experiments, to be essential for Ros87 DNA binding [16]. In spite of this wealth of knowledge, the structural details of the mechanism through which the prokaryotic zinc fingers recognize their target genes are still unclear. Molecular Docking represents an efficient tool to describe the mechanism underlying specific DNA recognition, giving detailed structural information on DNA-protein complexes. A pre-condition for molecular docking is the availability of a three-dimensional structure. To date, two main experimental techniques are commonly used for protein structure determination, *i.e.* X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy. X-ray structures normally have higher precision with respect to NMR structures. On the other hand, NMR structures are determined in solution and therefore are often thought to be more biologically relevant than X-ray structures [17]. Furthermore, NMR spectroscopy is a powerful tool for describing protein dynamic motions [18–24] and the solved NMR ensemble reflects protein conformational variations. Interestingly, docking studies performed using the NMR ensemble demonstrated that a single X-ray or average NMR structure may not be suitable for docking calculations [25]. In this context, Molecular Dynamic (MD) simulations give a detailed description of proteins dynamical process at the picoseconds to milliseconds timescale representing, thereby, a valuable tool complementary to experimental data obtained from X-ray and NMR [26–28].

Here, to gain insights into the molecular DNA recognition process of prokaryotic zinc finger domains we applied a strategy in which we performed molecular docking studies using a combination of Nuclear Magnetic Resonance (NMR) and Molecular Dynamics (MD) simulations data. Since in protein-DNA recognition, adaptation and plasticity appear to be two major properties that confer specific binding capability, we generated a set of MD ensembles to better describe the conformational heterogeneity. Successively, we evaluated the quality of each ensemble by using experimental chemical shifts. Finally, the obtained MD ensembles and the cognate DNA motif have been used to drive the docking. Our results indicate that the DNA recognition mechanism in the prokaryotic zinc finger domain is mediated by the first α-helix that docks the major groove of the DNA. Moreover, the protein-DNA interaction is stabilized by additional residues located around the zinc binding region, making the protein-DNA binding interface more extended with respect to the eukaryotic Cys_2His_2 domains.

2. Results and discussion

2.1. Detection of Ros87 protein dynamics using chemical shifts and NMR ensemble

Isotropic Chemical Shifts, which are averaged over manifold of conformations on the millisecond timescale, have been used as probes of conformational dynamics. Chemical shifts can be utilized to predict S^2 order parameter [29,30], demonstrating that are sensitive reporters of fluctuations in the ps-ns timescale. They are also used to predict backbone and side-chain dihedral angles, reflecting their sensitivity to conformational variations on the millisecond timescale [31,32]. So far, to explore the dynamic behavior of Ros87 in solution we used chemical shift data (^{1}H, ^{15}N and ^{13}C) and the determined NMR ensemble [13]. At first, in order to check the quality of the chemical shifts prediction tool we fitted all experimental data to each conformer of the NMR ensemble using 4DSPOT software [33,34]. All backbone chemical shifts (^{13}Cα, ^{15}N, ^1H_N and ^{1}Hα) provided good quality fits as reflected by Q factor values in the range of 0.032–0.111 for ^{15}N, 0.051–0.109 for ^1H_N, 0.024–0.027 for ^{13}Cα, 0.062–0.069 for ^{1}Hα (Table SI1). Then, in order to assess the protein motions, describing the conformational space sampled by Ros87 in solution, we analyzed the backbone chemical shifts against the NMR ensemble as reported in *Materials and Methods*. As expected, by using all 20 NMR conformers, we obtained a small but significant improvement of Q factors to values of 0.027 for ^{15}N, 0.050 for ^1H_N, 0.023 for ^{13}Cα, 0.061 for ^{1}Hα (Fig. 1). These results fully support the concept that single NMR conformer cannot be considered a fully realistic model of protein in solution. Overall, the data indicate that Ros87 adopts a compact fold in the region Pro9-Tyr66 showing moderate backbone motions. In particular, Q factors of residues located in the N-terminal and C-terminal tails show a larger improvement with respect to the rest of the protein, suggesting that in these regions the protein samples a larger conformational space. Our analysis demonstrates that the NMR ensemble, including all the conformers, is a better representation of the internal motions than any of them separately. All together our results are in a good agreement with the protein dynamics characterization of Ros87 based on NMR relaxation parameters, as reported in a previous publication [13] but the NMR ensemble, containing twenty conformers, only partially describes the experimental data. In particular, the main deviations are related to those residues that show high conformational heterogeneity: N-terminal and C-terminal regions. The residues within the region Pro9-Tyr66, with the exception of loop regions, sample a conformational space that appears more restricted. In conclusion, our data indicate that Ros87 samples a limited conformational space in the region Pro9-Tyr66, while in the two extended N-terminal and C-terminal tails explores a larger conformational space.

2.2. Dynamic averaging of chemical shifts using Molecular Dynamics (MD) simulations data

In order to consider the protein flexibility and include this piece of information in molecular docking of Ros87-DNA interaction, we generated different MD ensembles. In fact, Molecular Dynamics simulations are widely used as a tool to study the biomolecules dynamics, providing information on the motions of all atoms [24,35,36]. Moreover, as mentioned before, chemical shifts are sensitive probes of conformational dynamics. Therefore, the combination of MD ensembles with experimental chemical shifts may provide a more realistic picture of the conformational space sampled by Ros87 in solution. At first, we ran MD simulations of varied length, ranging from 1 (Fig. 2A) to 10 ns using the Amber99SB force fields in TIP3P water. Then, we used the 4DSPOT