Multimedia in Biochemistry and Molecular Biology Education

Proteopedia: 3D Visualization and Annotation of Transcription Factor–DNA Readout Modes

Ana Carolina Dantas Machado‡, Skyler B. Saleebyan§, Bailey T. Holmes§, Maria Karelina§, Julia Tam§, Sharon Y. Kim§, Keziah H. Kim¶, Iris Dror‡, Eran Hodis‡‡, Eric Martz§§, Patricia A. Compeau§, and Remo Rohs‡‡

From the ‡Molecular and Computational Biology Program, Departments of Biological Sciences, Chemistry, and Physics and Astronomy, University of Southern California, 1050 Childs Way, Los Angeles, California 90089 §La Cañada High School, 4463 Oak Grove Drive, La Cañada Flintridge, California 91011 ¶Flintridge Preparatory School, 4543 Crown Avenue, La Cañada Flintridge, California 91011 ‡‡Harvard/MIT MD-PhD Program, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115 §§Department of Microbiology, University of Massachusetts, 639 North Pleasant Street, Amherst, Massachusetts 01003

3D visualization assists in identifying diverse mechanisms of protein-DNA recognition that can be observed for transcription factors and other DNA binding proteins. We used Proteopedia to illustrate transcription factor-DNA readout modes with a focus on DNA shape, which can be a function of either nucleotide sequence (Hox proteins) or base pairing geometry (p53).

Keywords: Transcription factors, protein-DNA recognition, Hox, p53, DNA shape.

Proteopedia is an online encyclopedia designed to illustrate the three-dimensional (3D) structures of proteins, nucleic acids, and their complexes [1, 2]. To better describe protein–DNA binding at the molecular level, we created two Proteopedia pages (http://proteopedia.org/w/Hox-DNA_Recognition and http://proteopedia.org/w/P53-DNA_Recognition), where we annotated and visualized key interactions at protein–protein and protein–DNA interfaces within protein–DNA complexes for the Hox and p53 transcription factors (Fig. 1). The p53 tumor suppressor plays an instrumental role in the body's defense against stress and cancer by promoting distinct biological responses, while Hox proteins are crucial transcription factors that regulate the development of distinct segments of the embryo.

Protein–DNA readout modes have been traditionally classified into "direct and indirect readout," terms coined when the very first crystal structures of protein–DNA complexes were solved and analyzed. Direct readout originally referred to hydrogen bonds between amino acids and bases, and indirect readout to water-mediated hydrogen bonds [3]. Since then, the 3D structures of thousands of protein–DNA complexes have become available, shedding light on many variations of protein–DNA recognition. Different protein families bind to DNA with various levels of specificity, from non-specific to
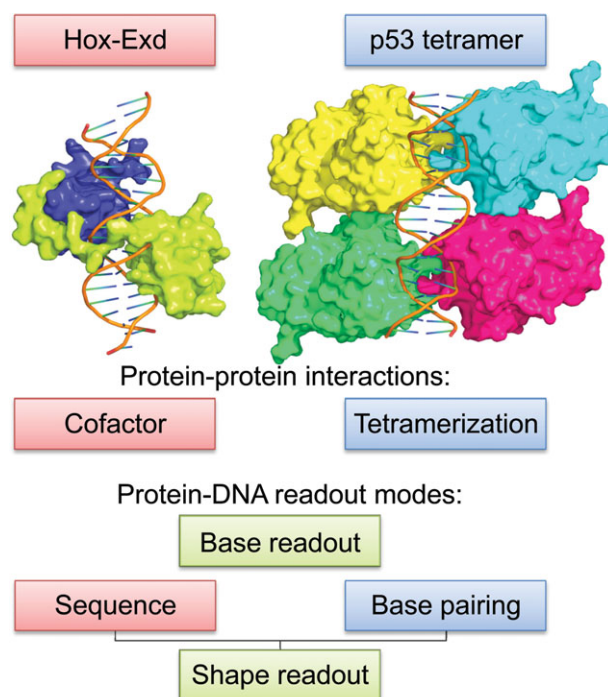

FIG. 1. The crystal structures of a Hox-Exd-DNA ternary complex (PDB ID 2R5Z) and a complex of a p53 DNA-binding-domain tetramer and its DNA target (PDB ID 3KZ8) form distinct protein–protein and protein–DNA interactions. Hox proteins bind DNA with a cofactor (Exd) while p53 forms a tetramer. Whereas both systems use the mechanisms of base and shape readout, DNA shape is a sequence-dependent feature in Hox binding sites and is modulated through the alteration of the base pairing geometry in a p53 consensus site.

highly specific, which can be achieved through a variety of means [3]. Some proteins bind alone, while others recruit cofactors or bind as dimers or tetramers. Proteins can contact the DNA either in the major or minor groove, or bind to both grooves at once. Some proteins leave the intrinsic DNA structure essentially intact, while others drastically deform the DNA. The term direct readout is not restricted to sequence-specific hydrogen bonds or hydrophobic contacts with functional groups of bases. Indirect readout, in turn, has become a catch-all term that is poorly defined. Thus, a classification into the two categories, direct and indirect readout, is simply not extensive enough to reflect our current knowledge of thousands of structures.

To assist researchers and educators, we have recently proposed a new classification scheme for protein–DNA readout modes into the more descriptive categories of "base and shape readout," which allow subsequent sub-classifications such as local and global shape readout, and major and minor groove base readout [3, 4]. We illustrated that *Drosophila* Hox proteins recruit a cofactor, Exd (Extradenticle), to achieve *in vivo* specificity, a mechanism that we call latent specificity [5]. In comparison to Hox proteins, human p53 binds to DNA as a tetramer, and thereby engages in cooperative protein–protein interactions [6]. Both the Hox and p53 transcription factors engage in hydrogen bonding between protein side chains and major groove edges of the base pairs. In particular, bidentate hydrogen bonds between arginine side chains and guanine bases contribute to highly specific sequence readout. Similarly, both transcription factors use shape readout of minor groove geometry and electrostatic potential to enhance binding specificity [7]. However, the mechanisms through which the shapes of their binding sites are modulated differ (Fig. 1). Hox binding sites are recognized based on their sequence-dependent minor groove shape [8]. In contrast, it has been observed that the DNA targets of p53 alter their shape through a transition of some base pairs from Watson–Crick to Hoogsteen geometry [6]. It still remains to be seen, however, if the Hoogsteen geometry is recognized by p53 as an intrinsic feature of DNA or if it is induced through the formation of the complex [9].

Protein structures have been classified in much detail over the years, while the same cannot be said for protein–DNA recognition. Under the new classification scheme, which was already mentioned in a textbook [10], we are now able to analyze protein–DNA readout modes in more detail and integrate them using 3D visualization to better understand molecular interactions. Proteopedia provides the framework for utilizing intricate details of protein–DNA recognition in both research and education.

REFERENCES

[1] E. Hodis, J. Prilusky, E. Martz, I. Silman, J. Moult, J. L. Sussman (2008) Proteopedia—A scientific 'wiki' bridging the rift between three-dimsional structure and function of biomacromolecules, *Genome Biol.* **9**, R121

[2] J. Prilusky, E. Hodis, D. Canner, W. A. Decatur, K. Oberholser, E. Martz, A. Berchanski, M. Harel, J. L. Sussman (2011) Proteopedia: A status report on the collaborative, web-encyclopedia of proteins and other biomolecules, *J. Struct. Biol.* **175**, 244–252.

[3] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, R. S. Mann (2010) Origins of specificity in protein–DNA recognition, *Annu. Rev. Biochem.* **79**, 233–269.

[4] R. Harris, T. Mackoy, A. C. Dantas Machado, D. Xu, R. Rohs, M. O. Fenley, in T. Schlick, Ed. (2012) Opposites attract: Shape and electrostatic complementarity in protein–DNA complexes, Chapter 3, Volume 2. Innovations in Biomolecular Modeling and Simulation. Biomolecular Sciences series, Royal Society of Chemistry, RCS Publishing, London, UK, pp 53–80.

[5] M. Slattery, T. Riley, P. Liu, N. Abe, P. Gomez-Alcala, I. Dror, T. Zhou, R. Rohs, B. Honig, H. J. Bussemaker, R. S. Mann (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins, *Cell* **147**, 1270–1282.

[6] M. Kitayner, H. Rozenberg, R. Rohs, O. Suad, D. Rabinovich, B. Honig, Z. Shakked (2010) Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs, *Nat. Struct. Mol. Biol.* **17**, 423–429.

[7] R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann, B. Honig (2009) The role of DNA shape in protein-DNA recognition, *Nature* **461**, 1248–1253.

[8] R. Joshi, J. M. Passner, R. Rohs, R. Jain, A. Sosinsky, M. A. Crickmore, V. Jacob, A. K. Aggarwal, B. Honig, R. S. Mann (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure, *Cell* **131**, 530–543.

[9] B. Honig, R. Rohs (2011) Biophysics: Flipping Watson and Crick, *Nature* **470**, 472–473.

[10] J. Kuriyan, B. Konforti, D. Wemmer (2012) The Molecules of Life: Physical and Chemical Principles, Garland Science, Taylor & Francis Group, New York, United States.