

Exposing the secrets of sex determination

Remo Rohs, Ana Carolina Dantas Machado & Lin Yang

Sex-determining transcription factors recognize their genomic target sites through mechanisms of DNA base-and-shape readout in combination with cooperative binding. Murphy *et al.* reveal that for one such transcription factor, DMRT1, the DNA sequence-and-shape features of its binding sites determine whether it binds DNA as a dimer, trimer or tetramer; they also characterize protein-DNA contacts that affect gender phenotypes in flies and humans.

The genes encoding the Doublesex (Dsx) and Male abnormal-3 (Mab-3)-related transcription (DMRT) factors play a key part in the sexual development of metazoans. One member of this family of transcription factors in humans, DMRT1, regulates genes involved in sex determination by influencing gonadal differentiation and testis development¹. Although DMRT1 has important biological functions, mechanistic understanding of how it binds DNA had been lacking. A new report by Murphy *et al.*², in this issue, sheds light on the structural mechanisms used by DMRT1 to recognize its genomic target sites.

The DNA-binding motif (DM) domain of the closely related *Drosophila* Dsx protein had been previously characterized by NMR spectroscopy as a zinc-binding module consisting of two intertwined Zn²⁺-binding sites formed by α -helices, with the C-terminal tail forming an α -helix that functions as the DNA-recognition element³. Because chemical modifications in the major groove had been shown not to affect Dsx binding, the DM domain was previously thought to bind in only the minor groove. The degeneracy of base-specific contacts in the minor groove⁴, however, precluded an understanding of the DNA binding specificity of DMRT proteins.

The cocrystal structure of three DMRT1 DM domains bound to DNA, reported now by Murphy *et al.*², has solved this conundrum and has revealed a very unusual architecture of the quaternary complex (Fig. 1). DM domains of three DMRT1 proteins bind to DNA by inserting their C-terminal α -helices into the major groove and anchoring the zinc-binding module into the minor groove. Remarkably, two of the α -helices insert into the same region of the major groove, thus leading to an unusual deformation of that groove due to the spatial requirement of the two antiparallel α -helices. Although the relatively low

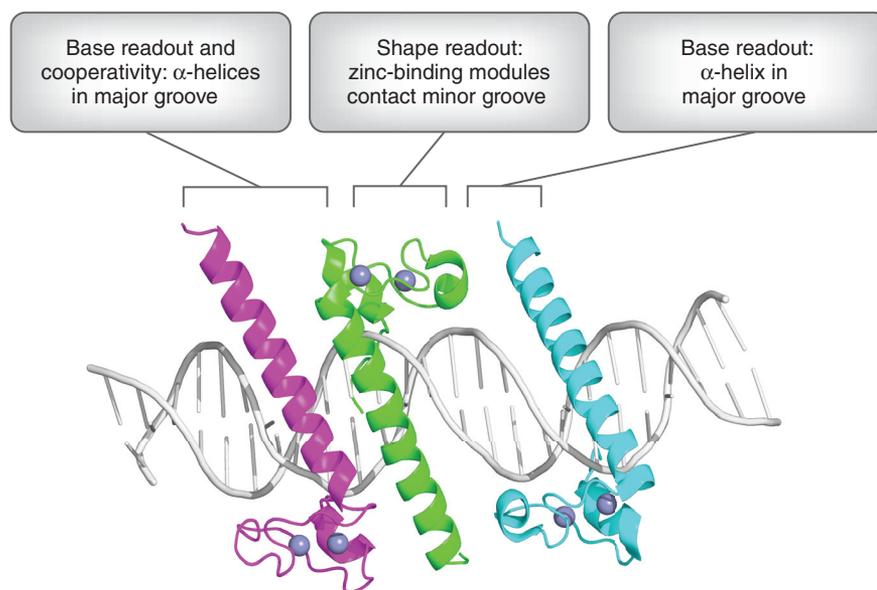


Figure 1 Cocrystal structure of the quaternary DMRT1–DNA complex. Three α -helices intrude into the major groove (base readout), two of them (purple and green) in the same region and the third α -helix (cyan) a helical turn away. The zinc-binding modules bind the minor groove (shape readout). In addition, protein-protein contacts are formed between the three DM domains (cooperativity).

resolution of the cocrystal structure merits future studies, the exposed binding geometry has not been previously observed for any DNA-binding protein. The recognition helix of the third DMRT1 protein inserts into the major groove one helical turn away, so that all three α -helices bind DNA on the same side of the double helix through base readout in the major groove.

To validate the importance of these major-groove contacts for the DNA binding specificity of DMRT1, Murphy *et al.*² replaced G-C base pairs by 2-aminopurine–uracil base pairs, which resulted in different patterns of functional groups in the major groove but not the minor groove. This substitution strongly

reduced binding. In a parallel experiment, replacing guanine by inosine kept the major groove intact and modified the minor groove. This substitution, in contrast, had no effect on binding. Confirming this observation, mutation of the Arg111 residue that was observed to form hydrogen bonds in the major groove abolished DMRT1 binding. On the basis of exome sequencing for a person with sex reversal, Murphy *et al.*² also associated the mutation of Arg111 with gonadal dysgenesis. When taken together, the results strongly suggest that major-groove contacts are one key component for DMRT1 binding specificity.

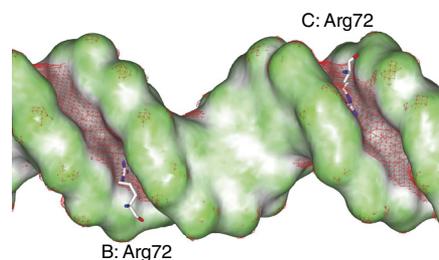


Figure 2 Biophysics of minor-groove contacts in the DMRT1–DNA cocrystal structure. The negative electrostatic potential, derived from a Poisson-Boltzmann calculation under physiologic ionic conditions in the absence of the protein⁵ and shown by a red mesh representing an isopotential surface at $-5 k_B T/e$, attracts the positively charged Arg72 residues of DM domains B and C to the minor groove.

Remo Rohs, Ana Carolina Dantas Machado and Lin Yang are at the Molecular and Computational Biology Program, Department of Biological Sciences, University of Southern California, Los Angeles, California, USA. e-mail: rohs@usc.edu

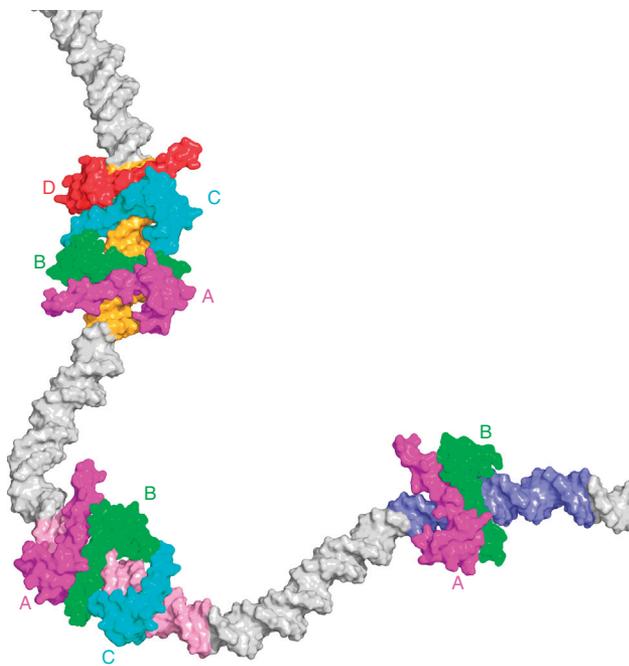


Figure 3 Schematic representation of DMRT1-DNA binding stoichiometry in the genome. Depending on the DNA sequence-and-shape characteristics of putative binding sites, DMRT1 was found to bind DNA both *in vitro* and *in vivo* as a dimer (A–B), trimer (A–B–C) or tetramer (A–B–C–D).

The mutation of a different arginine residue, Arg72, likewise substantially reduced binding. This residue of the zinc-binding module inserts deeply into the minor groove (Fig. 2). Whereas hydrogen bonds in the minor groove lack base-pair specificity⁴, basic side chains are attracted by the negative electrostatic potential in the minor groove⁵. This observation that DMRT1 uses an arginine as an electrostatic anchor provides biophysical evidence that sequence-dependent DNA-shape features probably contribute to the binding specificities of DMRT1 proteins. This form of shape readout has previously been observed for many other transcription factors⁶.

Apart from the unexpected DNA binding architecture, another key finding of the current study is that DMRT1 binds DNA in multimeric form, more specifically either as a dimer, trimer or tetramer (Fig. 3). Different modes of cooperativity were previously observed for other transcription factors^{7,8} but not for DMRT1. DNase I-hypersensitivity profiles revealed that DMRT1-footprint size depended on the DNA sequence of the putative binding sites and their flanking regions. Whereas the cocrystal structure included a DMRT1 trimer, gel shift experiments indicated the formation of dimers and trimers within the same sequence context. Complexes that included sites with larger DNase I footprints migrated more slowly through the gel, thus indicating the additional formation of tetramers. Further changes to the DNA sequence within the region where the third

DMRT1 protein binds DNA led to complexes that moved more quickly through the gel, thus suggesting the exclusive formation of dimers.

The cocrystal structure of the DMRT1–DNA complex (Fig. 1) reveals how this form of modular cooperativity can be achieved. The α -helices of two neighboring DMRT1 proteins that insert into the same region of the major groove interact extensively with each other. Hydrogen bonds between one of these DM domains and a third DMRT1 protein, which form between their recognition helices and zinc-binding modules, establish protein-protein interactions that enable the formation of higher-order complexes.

Using high-throughput DNA-shape analysis, Murphy *et al.*² predicted minor-groove width patterns of sequences bound by DMRT1 *in vivo*. They grouped target sites derived from chromatin immunoprecipitation with exonuclease treatment (ChIP-exo) experiments⁹ by predicted minor-groove topographies. In one group, sites with asymmetric minor-groove width profiles showed unilateral narrowing in the 5' flanking region. For these sites, the average minor-groove profile almost exactly reflected the DNA conformation observed in the crystal structure, and the ChIP-exo cross-linking pattern indicated binding by dimers and trimers. In a second group, sites with symmetric minor-groove width profiles were characterized by a bilateral narrowing in the 5' and 3' flanking regions. The cross-linking pattern for these sites predicted binding by tetramers.

This result led to the hypothesis that the narrow minor groove in the 3' flanking region would enable the formation of DMRT1 tetramers. The binding of the fourth C-terminal α -helix would probably widen the major groove in the region where, in the cocrystal structure, one recognition helix alone inserts into the major groove and thus narrows the minor groove in the 3' flanking region. An intrinsic minor-groove narrowing in this region might provide an initial anchoring point for the Arg72 residue of the fourth DMRT1 protein. Such narrowing of the minor groove due to the sequence environment of putative binding sites in the genome would enhance its negative electrostatic potential and thereby probably stabilize tetramer formation.

The importance of DNA shape in regions flanking transcription factor-binding sites for binding specificity was previously shown^{10,11}. Another recent study showed that DNA-shape features of sequences derived from an improved ChIP-exo approach were indicative of transcription-factor binding preferences¹². The finding by Murphy *et al.*² that minor-groove width correlates with binding stoichiometry suggests an additional role of DNA shape in the formation of larger complexes. Thus, the modeling of transcription-factor binding^{13,14} should include the interplay between DNA-readout modes and cooperative binding. By combining knowledge derived from X-ray crystallography and ChIP-exo experiments, the study by Murphy *et al.*² also supports a recent review concluding that the integration of structural biology and genomics furthers the understanding of gene regulation¹⁵.

ACKNOWLEDGMENTS

This work was supported by grant GM106056 from the US National Institutes of Health (to R.R.).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Raymond, C.S., Murphy, M.W., O'Sullivan, M.G., Bardwell, V.J. & Zarkower, D. *Genes Dev.* **14**, 2587–2595 (2000).
2. Murphy, M.W. *et al. Nat. Struct. Mol. Biol.* **22**, 442–451 (2015).
3. Zhu, L. *et al. Genes Dev.* **14**, 1750–1764 (2000).
4. Seeman, N.C., Rosenberg, J.M. & Rich, A. *Proc. Natl. Acad. Sci. USA* **73**, 804–808 (1976).
5. Rohs, R. *et al. Nature* **461**, 1248–1253 (2009).
6. Abe, N. *et al. Cell* **161**, 307–318 (2015).
7. Kitayner, M. *et al. Mol. Cell* **22**, 741–753 (2006).
8. Panne, D., Maniatis, T. & Harrison, S.C. *Cell* **129**, 1111–1123 (2007).
9. Rhee, H.S. & Pugh, B.F. *Cell* **147**, 1408–1419 (2011).
10. Gordán, R. *et al. Cell Reports* **3**, 1093–1104 (2013).
11. Levo, M. *et al. Genome Res.* doi:10.1101/gr.185033.114 (11 March 2015).
12. He, Q., Johnston, J. & Zeitlinger, J. *Nat. Biotechnol.* **33**, 395–401 (2015).
13. Stormo, G.D. *Quant. Biol.* **1**, 115–130 (2013).
14. Zhou, T. *et al. Proc. Natl. Acad. Sci. USA* **112**, 4654–4659 (2015).
15. Slattery, M. *et al. Trends Biochem. Sci.* **39**, 381–399 (2014).