CellPress

Indirect DNA Sequence Readout by LAGLIDADG Homing Endonucleases

Matthias Bochtler[1,2,*]
[1]International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland
[2]Institute of Biochemistry and Biophysics PAS, Pawinskiego 5a, 02-106 Warsaw, Poland
*Correspondence: mbochtler@iimcb.gov.pl
http://dx.doi.org/10.1016/j.str.2016.05.008

In this issue of *Structure*, Lambert et al. (2016) describe extensive structural and functional work on meganucleases, the group of homing endonucleases most commonly adapted to genome engineering applications. The data are of interest to structural biologists, evolutionary biologists, protein designers, and genome engineers.

Homing endonuclease (HE) genes are selfish genetic elements that insert into other genes (Burt and Trivers, 2006). HEs skew inheritance in their own favor by cleaving alleles not containing the HE gene (Dujon, 1989). Target sequences of HEs are long (14–40 nucleotides) and therefore nearly unique in entire genomes. They occur in highly conserved regions and are recognized imprecisely to protect the genetic drive against mutations or polymorphisms. Presumably to minimize deleterious effects on the host, HE genes are embedded in introns (of type I or type II) (e.g., I-SceI, also known as the ω of *S. cerevisiae*) or in inteins (e.g., PI-SceI) (Belfort and Roberts, 1997). In organelles, cleavage of non-HE-containing alleles probably eliminates them. In the nucleus, preferential inheritance of HE containing alleles is due to gene conversion, a by-product of double-strand break repair by homologous recombination (Stoddard, 2005).

LAGLIDADG HEs (named for a pattern of conserved amino acids), also called meganucleases (MNs), represent the most thoroughly studied HE family (several other families have also evolved independently). Meganucleases are predominantly found in the fully or partially bi-parentally inherited organelles of lower eukaryotes (especially algae and fungi) and come in different varieties: some are dimeric and cleave palindromic or nearly palindromic target sequences (e.g., I-CreI) (Jurica et al., 1998), others are dimeric but cleave strongly asymmetric targets (e.g., I-CeuI) (Spiegel et al., 2006), and some are pseudodimeric (e.g., I-AniI) (Bolduc et al., 2003). Some come with an intein domain (e.g., PI-SceI) (Moure et al., 2002) or double as maturases (Bolduc et al., 2003). For all the prototypes, crystal structures have been determined. Altogether, ~20 structures of different meganucleases are already known. So what remains to be learned?

In their work in this issue of *Structure*, Lambert et al. (2016) identify a group of I-OnuI related, previously uncharacterized MNs from fungi (mostly Ascomycetes colonizing plants and arthropods). The new MNs reside in mitochondrial genes for ribosomal RNA and for respiratory chain components. Importantly, target sequences could be predicted based on a comparison of closely related alleles with and without HE insertion. In an amazing tour de force, the group expressed the candidate MNs on the yeast surface and succeeded in the majority of cases (22 out of 34).

Cleavage activity against the predicted target sequences was confirmed for all but two expressed proteins, validating the target prediction strategy. After expression in *E. coli* and additional biophysical characterization, an astonishing nine MNs could be crystallized together with their targets. Structures were then solved and refined to resolutions of 3.2 Å and better. In a single strike, this work increases by roughly 50% the number of genuinely different MN structures since the first report of a meganuclease structure in complex with DNA in 1998 (I-CreI, also by the Stoddard group) (Jurica et al., 1998). Three crystal structures were described separately (I-OnuI, I-LtrI, and I-SmaMI); the crystal structures of I-AabMI, I-CpaMI, I-GpeMI, I-GzeII, I-LtrWI, and I-PanMI are presented for the first time in the report in *Structure*. All enzymes belong to the group of dimeric meganucleases and are 34%–48% identical to each other. Despite the dimeric nature of the new MNs, their target sequences are at most distantly related to palindromes, and also quite different from each other. Based on the analysis of the crystal structures, Stoddard and colleagues then generate the currently most comprehensive maps of amino acids in the MN fold that can contact nucleobases in a DNA target.

Meganucleases require A-T or T-A base pairs at the centers of their target sequences, even though the enzymes make barely any contacts with these base pairs. The "indirect" readout is of course highly reminiscent of the indirect readout of the central base pairs (A-T and T-A) by EcoRV, which has been traced at least in part to energetic penalties associated with DNA bending and base pair unstacking (Martin et al., 1999). The work of Lambert and colleagues is more than "just similar work for other enzymes," because the authors compare indirect readout in a large set of related enzymes. They show that all tested MNs bend DNA at the center. Most enzymes tolerate substitutions in the central base pairs at the level of binding, but not at the level of DNA cleavage. Interestingly, one of the MNs, I-PanMI, is more tolerant to changes in the target sequence than the other enzymes. Nevertheless, it bends DNA to a similar extent as the other MNs. Comparison of the crystal structures suggests an explanation for this paradox. In the I-PanMI DNA complexes, the helical rise between the central base pairs is much lower than in the other complexes. Lambert and colleagues suggest that this

difference may mitigate the extra penalty for bending G-C or C-G pair-containing DNA over the penalty for bending DNA with A-T or T-A pairs.

Meganucleases recognize their long target sequences imprecisely. The contributions of individual base pairs to binding are relatively small, and in some cases, mutations in the target interfere only with DNA cleavage, but not DNA binding. The authors take advantage of this situation to structurally characterize I-SmaI bound to DNA duplexes harboring cleavage-blocking mutations at the center of the target sequence (5′-TTGT-3′ and 5′-TTCT-3′ instead of 5′-TTAT). The effects of the changes on the mode of binding are subtle, but a (marginally significant) slip of the altered base pairs compared to the base pair in the substrate can be observed. This slight structural change propagates to the active site, where it leads to the loss of a catalytic metal ion, explaining the loss of activity of the enzyme. Crystallography is frequently useful to explain how small molecular changes alter binding affinities. Lambert and colleagues have hit on a rare case in which crystallography can convincingly explain changes in catalytic rates!

The study by Lambert and colleagues is unusually exhaustive, but because so much new information is made available, more analysis can be done. Protein designers will be interested to see whether the new data help to better understand the MN "cipher," which is so much more complicated than the "cipher" for TALENs or ZFNs (zinc finger nucleases). Evolutionary biologists can test correlations between degenerate sequence recognition and conservation of targets, and protein engineers aiming to make useful MNs now have a much larger repertoire of well-characterized MNs to start from.

ACKNOWLEDGMENTS

The author's laboratory is supported by NCN grant UMO-2011/02/A/NZ1/00052.

REFERENCES

Belfort, M., and Roberts, R.J. (1997). Nucleic Acids Res. *25*, 3379–3388.

Bolduc, J.M., Spiegel, P.C., Chatterjee, P., Brady, K.L., Downing, M.E., Caprara, M.G., Waring, R.B., and Stoddard, B.L. (2003). Genes Dev. *17*, 2875–2888.

Burt, A., and Trivers, R. (2006). Genes in Conflict: the Biology of Selfish Genetic Elements (Cambridge, MA: Belknap Press of Harvard University Press).

Dujon, B. (1989). Gene *82*, 91–114.

Jurica, M.S., Monnat, R.J., Jr., and Stoddard, B.L. (1998). Mol. Cell *2*, 469–476.

Lambert, A.R., Hallinan, J.P., Shen, B.W., Chik, J.K., Bolduc, J.M., Kulshina, N., Robins, L.I., Kaiser, B.K., Jarjour, J., Havens, K., et al. (2016). Structure *24*, this issue, 862–873.

Martin, A.M., Horton, N.C., Lusetti, S., Reich, N.O., and Perona, J.J. (1999). Biochemistry *38*, 8430–8439.

Moure, C.M., Gimble, F.S., and Quiocho, F.A. (2002). Nat. Struct. Biol. *9*, 764–770.

Spiegel, P.C., Chevalier, B., Sussman, D., Turmel, M., Lemieux, C., and Stoddard, B.L. (2006). Structure *14*, 869–880.

Stoddard, B.L. (2005). Q. Rev. Biophys. *38*, 49–95.

JAK1 Takes a FERM Hold of Type II Cytokine Receptors

Jeffrey J. Babon,[1,*] Nicholas P.D. Liau,[1] and Nadia J. Kershaw[1]
[1]Department of Medical Biology, The Walter and Eliza Hall Institute of Medical Research, University of Melbourne, Parkville, VIC 3052, Australia
*Correspondence: babon@wehi.edu.au
http://dx.doi.org/10.1016/j.str.2016.05.007

Janus kinases (JAKs) initiate the intracellular signaling cascade triggered by exposure of cells to cytokines and interferons. In order to achieve this, JAKs are bound to the intracellular domain of specific cytokine receptors immediately adjacent to the cell membrane. In this issue of *Structure*, Ferrao et al. (2016) provide structural details of such an interaction and in doing so, identify for the first time the motif used by type II cytokine receptors to recruit JAK1.

Cytokines and interferons are critical mediators of hematopoiesis and the immune response. They are small, soluble, secreted glycoproteins that act as intercellular messengers, instructing their target cells to differentiate, proliferate or apoptose (Nicola and Hilton, 1998). Hematopoietic cytokines (such as EPO, TPO, and many interleukins) and interferons stimulate the JAK/STAT pathway inside their target cells to effect the desired biological response. The JAK/STAT pathway is one of the major classes of signaling pathways that use cell surface receptors to drive gene transcription and development of the organism. It is also one of the most direct, requiring only four key components: cytokine, receptor, kinase, and transcription factor. The crucial elements of this pathway were determined more than two decades ago by Stark, Darnell, and Kerr (Darnell et al., 1994; Darnell, 1997; Bromberg et al., 1999). In the simplest systems, a cytokine receptor consists of two individual transmembrane protein chains, each of which binds a JAK molecule in their intracellular region, close to the cell membrane. Under basal conditions, the kinase domains of these JAKs are inactive but become