Chang, H.K., Yeh, S.H., and Shieh, R.C. (2003). J. Physiol. *553*, 101–112.

Domene, C., Doyle, D.A., and Venien-Bryan, C. (2005). Biophys. J. *89*, L01–L03.

Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., and MacKinnon, R. (1998). Science *280*, 69–77.

Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B.T., and MacKinnon, R. (2002). Nature *417*, 523–526.

Jogini, V., and Roux, B. (2005). J. Mol. Biol., in press.

John, S.A., Xie, L.H., and Weiss, J.N. (2004). Mechanism of inward rectification in Kir channels. J. Gen. Physiol. *123*, 623–625.

Kuo, A., Gulbis, J.M., Antcliff, J.F., Rahman, T., Lowe, E.D., Zimmer, J., Cuthbertson, J., Ashcroft, F.M., Ezaki, T., and Doyle, D.A. (2003). Science *300*, 1922–1926.

Kuo, A., Domene, C., Johnson, L.N., Doyle, D.A., and Vénien-Bryan, C. (2005). Structure *13*, this issue, 1463–1472.

Shin, H.G., and Lu, Z. (2005). J. Gen. Physiol. *125*, 413–426.

Structure, Vol. 13, October, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.09.002

Bending in the Right Direction

Rohs et al. (2005) report the success of a new Monte Carlo algorithm in describing sequence-specific DNA bending. The approach offers the possibility of improved treatments of "indirect readout" effects in the prediction of transcription factor binding sites.

What distinguishes a small number of transcription factor (TF) binding sites from the large number of other nucleotide sequences that could potentially bind to the same protein? For those sites that are recognized, what are the structural and energetic origins of the subtle differences in binding affinity that are often observed? In addition to direct readout mechanisms, which involve interactions between contacting groups on the protein and DNA, it has become increasingly clear that indirect readout mechanisms such as sequence-specific DNA bending are often operative in tuning these affinities.

An interesting example involves targets of the human papillomavirus E2 protein (HPV-18 E2), which are the subject of a computational study published in this issue of *Structure* by Rohs et al. (2005). A high-affinity site, ACCGAATTCGGT, and a low-affinity site, ACCGACGTCGGT, differ only in the spacer sequence (underlined) that connects the two flanking half sites (Hegde, 2002). There are no contacts observed between the spacer sequence and the bound protein (HPV-18 E2 for the AATT sequence and bovine papillomavirus E2 [BPV-1 E2] for the ACGT sequence), and indeed the shape of the DNA is very similar for both sequences (Hegde, 2002). In contrast, crystallographic studies of the free DNA molecules show that the sequence with the AATT spacer is bent toward the minor groove in a manner similar to that seen in the complex while the dodecamer with the ACGT spacer is essentially straight (Hizver et al., 2001). Rohs et al. (2005) employed a novel Monte Carlo (MC) approach to simulate the structure and dynamics of these same two DNA dodecamers in solution. The bending patterns observed in the MC simulations are in good agreement with those observed for the free DNA molecules.

The simulation technology, which combines an analytical chain closure method carried out in torsion and bond angle space with internal- and collective-variable MC moves, is in itself a major technical achievement (Rohs et al., 2005). Used with an all-atom force-field and a simple dielectric screening model, the approach enables effective sampling and fast conformational equilibration. Remarkably, the MC simulations with the simple solvent model are in better agreement with the X-ray results than those obtained from MD simulations that use explicit solvent (Byun and Beveridge, 2004; Djuranovic et al., 2004). This appears to be due, in part, to more efficient sampling in the MC simulations. The success of the simple dielectric screening model implies that the MC simulations accurately reflect the internal conformational energetics of the DNA and that solvation effects may be of only secondary importance.

At this stage, the MC simulations have confirmed insights as to the source of bending that had been noticed through the analysis of crystal structures (Hegde, 2002; Hizver et al., 2001). In the cases studied here, the AATT-containing dodecamer sequence is bent because there is bending at the CG steps in the flanker regions that is reinforced by bending at the AT step in the spacer. The ACGT dodecamer sequence is mostly straight because bending at the CG step in the spacer opposes bending in the two flanking sequences. There are other examples to look at in this and other systems, but if the MC simulations continue to be successful in reproducing sequence-specific tendencies for DNA bending, we will have a powerful tool that can be applied to the study of protein-DNA binding. Moreover, the results obtained so far offer the promise of being able to predict how DNA will bend when bound to different proteins, although this will require further methodological developments.

Accounting for the energetic cost of deforming DNA so as to bind to a given protein is clearly an essential aspect of understanding sequence-specific DNA recognition. The driving force for this deformation is provided by attractive interactions between the protein and the DNA, some of which are highly specific. Indeed there are often strong similarities between the interfaces formed by cognate and noncognate complexes involving the same proteins even if this requires significant distortion of one or both DNA strands (Siggers et al., 2005). For the case of papillomavirus E2 proteins, the crystallographic evidence, and now the MC simulations, suggest that the cost of deforming the AATT dodecamer is less than that for the ACGT dodecamer because the former is already bent in the right direction

in the free DNA while the latter is essentially straight, thus requiring a higher deformation energy. However there may be more to the story. For example the two dodecamers form complexes with the BPV-1 E2 protein which are structurally quite similar to those formed with HPV-18. However, in contrast to binding to HPV-18, the binding affinities of the two BPV-1 complexes are very similar to one another. It has been suggested that these differences in relative affinities are associated with changes in the width of the minor groove between the BPV-1 and HPV18 complexes (Hizver et al., 2001), but accounting for subtleties of this type clearly poses further theoretical challenges.

The use of structural information to understand protein-DNA binding specificity and to identify TF binding sites will undoubtedly see rapid progress in the coming years. The paper of Rohs et al. (2005) is likely to be an important step in the development of approaches of this type because it offers a means to describe efficiently the conformation of the unbound DNA molecule in atomistic terms and hence to improve treatments of the energetic effects of indirect readout. It will be necessary to develop methods to calculate strain energies for the conformations that are generated. In addition to the direct use of all-atom force-fields for this purpose (Paillard and Lavery, 2004), a number of papers have used the structure-derived parameters of Olson et al. (1998) to calculate deformation energies. The structure-based prediction of TF binding sites must account for direct readout effects as well, and the energy-based analysis of known TF binding sites is an essential step in this direction. In addition, much has already been learned from experimental and statistical studies of known TF sites (Bulyk, 2003), and the integration of this information with physical-chemical insights promises to become a new and exciting research area (see e.g., Sarai and Kono, 2005). Indeed we will need to learn the lessons of many diverse approaches if we are to fully understand the subtle differences in protein-DNA binding specificity that underlie transcriptional regulation.

Trevor Siggers, Tonya Silkov, and Barry Honig
Howard Hughes Medical Institute
Center for Computational Biology
and Bioinformatics and
Department of Biochemistry and Molecular
Biophysics
Columbia University
1130 St. Nicholas Avenue
ICRB
Mail Box 200
New York, New York 10032

Selected Reading

Bulyk, M.L. (2003). Genome Biol. *5*, 201.

Byun, K.S., and Beveridge, D.L. (2004). Biopolymers *73*, 369–379.

Djuranovic, D., Oguey, C., and Hartmann, B. (2004). J. Mol. Biol. *339*, 785–796.

Hegde, R.S. (2002). Annu. Rev. Biophys. Biomol. Struct. *31*, 343–360.

Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D., and Shakked, Z. (2001). Proc. Natl. Acad. Sci. USA *98*, 8490–8495.

Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M., and Zhurkin, V.B. (1998). Proc. Natl. Acad. Sci. USA *95*, 11163–11168.

Paillard, G., and Lavery, R. (2004). Structure *12*, 113–122.

Rohs, R., Sklenar, H., and Shakked, Z. (2005). Structure *13*, this issue, 1499–1509.

Sarai, A., and Kono, H. (2005). Annu. Rev. Biophys. Biomol. Struct. *43*, 379–398.

Siggers, T.W., Silkov, A., and Honig, B. (2005). J. Mol. Biol. *345*, 1027–1045.