Proc. Natl. Acad. Sci. USA
Vol. 89, p. 11653, December 1992

Commentary

The importance of being flexible

Alan D. Frankel

Department of Biochemistry and Biophysics, and Gladstone Institute of Virology and Immunology, University of California, San Francisco, P.O. Box 419100, San Francisco, CA 94141

Although the precise three-dimensional structure of a protein is an essential determinant of its function, it is sometimes desirable that parts of its structure not be too rigid. For example, when induced fit occurs in an antibody–antigen, receptor–ligand, or enzyme–substrate interaction, the protein and/or its partner must undergo a conformational change; structural flexibility can help minimize the energetic cost of the conformational transition. A paper by Qian *et al.* (1) illustrates a case in which a region of the *Drosophila* Antennapedia (Antp) protein thought to be important for its function may be highly flexible, and indeed unstructured, in the absence of its partner.

The structure of a 60-amino acid Antp homeodomain–DNA complex has been determined by NMR (2). The homeodomain consists of three α-helices; the third helix binds in the DNA major groove and is considered to be the recognition helix. Additional contacts are made in the minor groove by a flexible N-terminal arm, which is disordered in the absence of DNA. In the intact protein, the Antp homeodomain is located near the C terminus and is connected via its arm to a region containing a Tyr-Pro-Trp-Met (YPWM) tetrapeptide that is conserved among several homeotic proteins. The YPWM region does not appear to directly contact the DNA but may help determine the specificity of homeotic function *in vivo*, perhaps through protein–protein interactions. Qian *et al.* (1) have reexamined the structure of the homeodomain, now with a 14-amino acid N-terminal extension including the YPWM motif, and find that the DNA-binding arm and the N-terminal extension are both disordered in the absence of DNA. Although it cannot be ruled out that other parts of the Antp protein may help to stabilize the structure of this region, the suggestion is that the N-terminal arm of the homeodomain and the adjacent YPWM region are part of a flexible linker that specifies interactions with DNA and other proteins.

The cocrystal structures of the *Drosophila* engrailed (3) and the yeast α2 (4) homeodomain–DNA complexes show strikingly similar arrangements to the Antp complex, suggesting that the homeodomain structure and its flexible N-terminal DNA-binding arm are highly conserved features of this class of proteins (4). However, even in the crystals, the first few residues of the arm are disordered. Such unstructured flexible arms have been seen in other helix–turn–helix DNA-binding proteins and can contribute significantly to DNA recognition. In the case of λ repressor, an N-terminal arm wraps around the DNA to make sequence-specific contacts in the major groove, and even subtle mutations in the arm can dramatically reduce DNA binding (5).

The yeast α2 homeodomain protein provides a clear example of how an unstructured flexible region can enhance DNA-binding specificity through protein–protein interactions. The α2 protein is organized into two domains, a C-terminal homeodomain and an N-terminal dimerization domain that are connected by a protease sensitive, and presumably flexible, linker of 30–40 amino acids (6). The flexibility of the connection is needed to allow α2 dimers to recognize DNA operator half-sites that are separated by various numbers of base pairs. To achieve recognition of a specific operator, a second protein, MCM1, binds between the two α2 subunits, thus setting the proper spacing and orientation of the subunits (7). Recent experiments demonstrate that the flexible linker of α2 interacts directly with MCM1 and allows two α2 monomers to bind cooperatively (8). Furthermore, when the linker is grafted onto the engrailed homeodomain, MCM1-dependent cooperative binding to an engrailed binding site is observed, demonstrating that the flexible linker is sufficient to specify the interaction with MCM1. As with Antp, NMR data indicate that the flexible linker region of α2 is unstructured in the absence of MCM1 (8), although α2 does not contain the YPWM motif.

There are other examples among transcription factors in which unstructured or flexible segments of proteins can mediate specific interactions with nucleic acids or other proteins (for review, see ref. 9). Recent NMR data with a peptide–RNA complex from the human immunodeficiency virus (HIV) Tat–TAR system (10) and with an acidic activation domain from the herpes simplex virus VP16 protein (11) indicate that these domains are unstructured. Other examples of unstructured domains are likely to be found, and as the current studies already seem to indicate, the roles of these flexible regions may be as flexible as the structures themselves.

1. Qian, Y. Q., Otting, G., Furukubo-Tokunaga, K., Affolter, M., Gehring, W. J. & Wuthrich, K. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10738–10742.
2. Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J. & Wuthrich, K. (1990) *EMBO J.* **9**, 3085–3092.
3. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990) *Cell* **63**, 579–590.
4. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. (1991) *Cell* **67**, 517–528.
5. Clarke, N. D., Beamer, L. J., Goldberg, H. R., Berkower, C. & Pabo, C. O. (1991) *Science* **254**, 267–270.
6. Sauer, R. T., Smith, D. L. & Johnson, A. D. (1988) *Genes Dev.* **2**, 807–816.
7. Smith, D. L. & Johnson, A. D. (1992) *Cell* **68**, 133–142.
8. Vershon, A. K. & Johnson, A. D. (1992) *Cell*, in press.
9. Frankel, A. D. & Kim, P. S. (1991) *Cell* **65**, 717–719.
10. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992) *Science* **257**, 76–80.
11. O'Hare, P. & Williams, G. (1992) *Biochemistry* **31**, 4150–4156.