

plex reveals a novel mechanism for DNA recognition by a protein containing a Zn<sub>2</sub>Cys<sub>6</sub> binuclear cluster. *Nature Struct. Biol.* **4**, 751–759.

- Travers, A. A. (1995). DNA bending by sequence and proteins. In *DNA-Protein: Structural Interactions* (Lilley, D. M. J., ed.), pp. 49–75, IRL Press, Oxford.
- Yanagi, K., Privé, G. G. & Dickerson, R. E. (1991). Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J. Mol. Biol.* **217**, 201–214.

## Appendix: Calculation of Major- and Minor-groove Widths

As stated in the main text, we have measured major- and minor-groove widths as distances between phosphate groups on the two anti-parallel strands. There is no universally agreed way of measuring groove width (e.g. for variants, see Suzuki & Yagi, 1996; Y. Kim *et al.*, 1993); but our scheme seems to be adequate for the purposes of the present study. Our algorithm provides consistent values for the two groove widths at a given dinucleotide step, provided the step is sufficiently far from the ends of the molecule.

Figure A1(a) illustrates the layout of base-pairs, backbones and phosphate groups for an eight-base-pair piece of DNA. On strand I, the phosphate groups are identified as  $P_i$ , where  $i$  increases in the 5'-to-3' direction, while on strand II they are identified as  $p_j$ , with  $j$  increasing in the 3'-to-5' direction. Thus two phosphate groups carry the same index number as in the corresponding dinucleotide step. The diagram shows a "skew ladder", as if the base-pairs were viewed from the minor-groove side in an "unrolled" version of the helix.

If distances are measured from a particular phosphate  $P_i$  on strand I to  $p_{i+m}$  on strand II for  $-5 \leq m \leq 0$ , then it is generally found that there is a minimum distance at around  $m = -3$ , as shown in Figure A1(b). Two such distances are shown by broken lines in Figure A1(a); and this is, broadly, how we define the width of the minor-groove, using an "offset" of three phosphate groups. Such distances are clearly not the shortest lengths as measured geometrically across the schematic two-dimensional skew-ladder in Figure A1(a); but no plane diagram, of course, can adequately represent the actual three-dimensional geometry of the double helix.

The major-groove width is not quite so straightforward to describe, because a plot of the distance from  $P_i$  to  $p_{i+m}$  for  $0 \leq m \leq 5$  sometimes shows a minimum at around  $m = 4$ , but sometimes only a point of inflection there, as shown in Figure A1(b): cf. Suzuki & Yagi (1996), Figure 1. (In this connection we should note that a minimum distance, which may occasionally be found as far away as  $m = 7$ , would correspond to the close approach of a remote phosphate in strongly bent DNA rather than a groove width.) In general, we shall define the major-groove width by using an offset  $m = 4$ , as shown in Figure A1(a) by the chain-dotted line

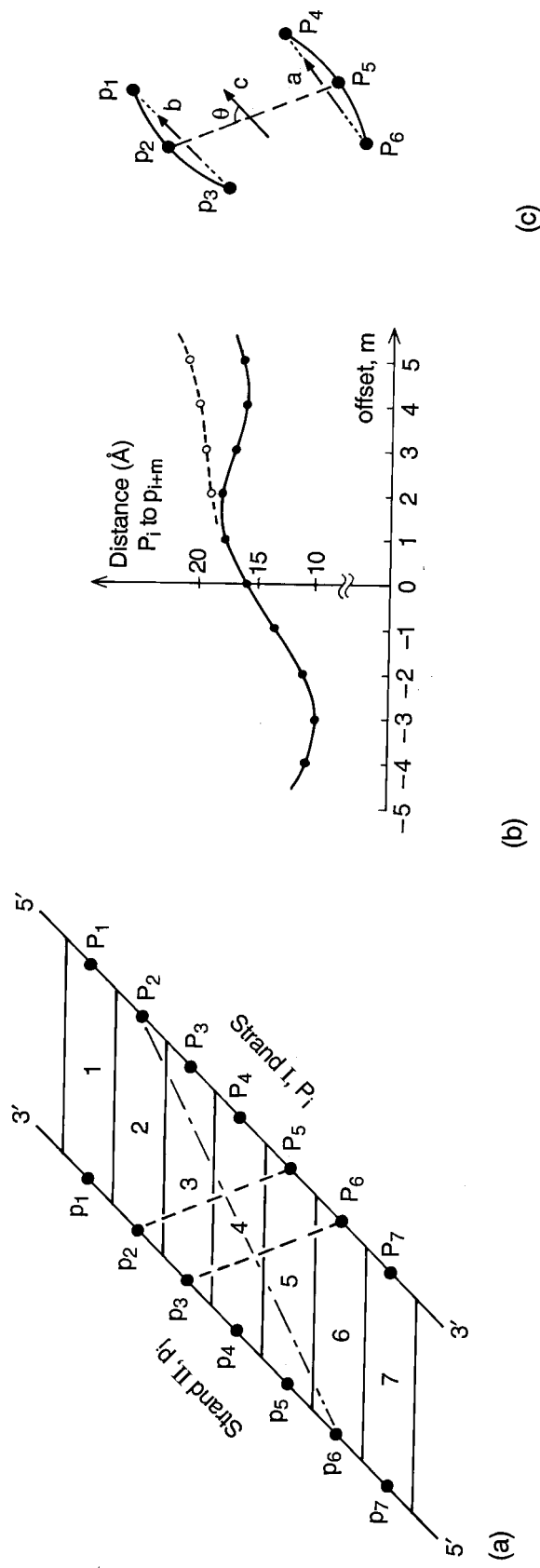
$P_2$  to  $p_6$ . Now on the plane diagram that appears to be a very large distance; but if the ladder were to be wrapped into something approaching its true three-dimensional cylindrical form, then the two points would obviously move much closer together. The fact that the major-groove width does not always correspond to a true minimum in the cross-strand phosphate-to-phosphate distance is not a significant disadvantage. Thus we are interested in "groove widths" because we know that protein moieties such as  $\alpha$ -helices and  $\beta$ -sheets fit into the major groove, with hydrogen bond and other links to backbone phosphate groups. Where phosphate groups on both strands are bound in this way, there is often an "offset" of about 4: for a collection of examples see, e.g. Calladine & Drew (1977), chapter 8.

As a rule, then, we shall measure minor-groove and major-groove widths from  $P_i$  to  $p_{i-3}$  and from  $P_i$  to  $p_{i+4}$ , respectively. There are, however, three practical complications in using such a scheme, as follows.

First, if we wish to associate a major-groove width with a given dinucleotide step  $k$ , we can measure the distance between  $P_{k-2}$  and  $p_{k+2}$ . But in relation to the minor-groove width, the odd-numbered offset of 3 is awkward. Accordingly, we define the minor-groove width for a step such as 4 in Figure A1(a) as the mean of the two marked distances; or in general:

$$\frac{1}{2}((P_{k+1} - p_{k-2}) + (P_{k+2} - p_{k-1})) \quad (\text{A1})$$

Second, while offsets  $m = -3$  and  $m = +4$  often locate the shortest distances between particular cross-chain phosphate groups, the minimum distances for a smooth curve connecting the measured points in Figure A1(a) generally lie to one side or the other of these groups, depending on the degree of distortion from the classical B form. In our calculations we have taken account of this situation in the following way. Suppose we wish to compute the shortest cross-chain length in the vicinity of the distance  $P_5$  to  $p_2$  in Figure A1(a). First we imagine that a smooth curve is drawn through the phosphate groups along each strand, as shown schematically in perspective for a short piece of the molecule in Figure A1(c). We calculate a unit vector **a** tangential to this curve at  $P_5$  on strand I by assuming it to be parallel with the vector  $P_6 \rightarrow P_4$ . Similarly, we calculate the unit vector **b** at  $p_2$  on the other strand, parallel with  $p_3 \rightarrow p_1$ . Finally, we calculate a unit vector **c**, for a hypothetical "mid-strand" curve by taking the mean of directions **a** and **b**: thus  $\mathbf{c} = (\mathbf{a} + \mathbf{b})/|\mathbf{a} + \mathbf{b}|$ . Now if vector **c** happens to be perpendicular to the cross-strand line  $P_5 - p_2$ , that line is indeed close to the shortest cross-strand distance. The geometrical device of a third strand, mid-way between the actual strands I and II, and with local tangent parallel with **c**, enables us to assess whether  $P_5 - p_2$  is practically perpendicular to both strands, as it would be if it



**Figure A1.** (a) Schematic, "unrolled" view of eight base-pairs and seven phosphate groups on each DNA strand, in order to explain the measurement of major and minor-groove widths. Shortest distances across the minor groove are marked by broken lines, while a corresponding major-groove distance (which appears rather long in this "unrolled" view) is marked by a chain-dotted line. The nomenclature for phosphate groups and dinucleotide steps is explained in the main text. (b) Schematic plot of cross-strand distance from  $P_i$  on strand I to  $p_j$  on strand II, where  $j = i + m$ , showing that the minor-groove width always corresponds to a minimum (but not necessarily at  $m = -3$ ), while the major-groove width usually corresponds to a minimum near  $m = 4$  (filled points); but there may instead be a point of inflection (open points). (c) Schematic, three-dimensional view of a short piece of the molecule, together with three unit vectors that are used in a more refined calculation of minor-groove width.

were indeed the shortest distance between them. In general, however, vector  $\mathbf{c}$  will lie at angle  $\theta$  to  $\mathbf{P}_5 - \mathbf{p}_2$ , where  $\theta \neq 90^\circ$ . In this case, we take the minimum cross-strand distance to be  $(\mathbf{P}_5 - \mathbf{p}_2)\sin\theta$ ; that is, we take the projection of  $(\mathbf{P}_5 - \mathbf{p}_2)$  perpendicular to  $\mathbf{c}$ . In order to complete the calculation of the minor-groove width corresponding to step 4, we repeat the above calculation for  $(\mathbf{P}_6 - \mathbf{p}_3)$ , and then take the average of the two. A similar calculation is done for major-groove width. In this case, the mid-strand tangential direction for step 4 is taken as the mean of the tangents to the two strands at points  $\mathbf{P}_2$  and  $\mathbf{p}_6$ . In practice, the value of  $\sin\theta$  is not much different from 1 in most of these calculations; and so the correction to the simple calculation, equation (A1), and its counterpart for major-groove width is generally small. These algorithms have been applied uniformly to all of our samples.

A third complication in the practical calculation of major and minor-groove widths is that if they are to be found for a step such as 4 in Figure A1(a), then data on phosphate positions are obviously required for several steps on either side. Examination of Figure A1(a) shows that, in order for the

calculations described above for step 4 to be completed, data on phosphate positions are required for three steps on either side, i.e. for all of the steps shown in the diagram. (But a total of five steps would be sufficient if the  $\sin\theta$  correction factor were to be omitted.) This means that we compute only one pair of groove widths for a DNA octamer, three for a decamer and five for a dodecamer. In the case of co-crystals, it turns out that our data sets are such that all bridging-region groove widths are available; and with very few exceptions in two of the 434 repressor co-crystals, all contacted-region groove widths may be found.

## References

- Calladine, C. R. & Drew, H. R. (1997). *Understanding DNA*, 2nd edit., Academic Press, London.
- Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
- Suzuki, M. & Yagi, N. (1996). An in-the-groove view of DNA structures in complexes with proteins. *J. Mol. Biol.* **255**, 677–687.

*Edited by J. Karn*

*(Received 28 January 1998; received in revised form 5 June 1998; accepted 12 June 1998)*