The EMBO Journal vol.14 no.7 pp.1571–1580, 1995

A novel type of interaction between cruciform DNA and a cruciform binding protein from HeLa cells

Christopher E.Pearson[1], Maria Zannis-Hadjopoulos[2], Gerald B.Price and Haralabos Zorbas[3]

McGill Cancer Centre, McGill University, 3655 Drummond Street, Montreal, Quebec, Canada H3G 1Y6 and [3]Institut für Biochemie der Ludwig-Maximilians-Universität München, 82152 Martinsried, Germany
[1]Present address: Institute of Biosciences and Technology, Texas A & M University, Houston, TX 77030, USA

[2]Corresponding author

M.Zannis-Hadjopoulos and H.Zorbas contributed equally to this work

Communicated by E.Winnacker

We recently identified and enriched a protein (CBP) from HeLa cells with binding specificity for cruciform-containing DNA. We have now studied the interaction of CBP with stable cruciform DNA molecules containing the 27 bp palindrome of SV40 on one strand and an unrelated 26 bp palindrome on the other strand by hydroxyl radical footprinting. The CBP–DNA interaction is localized to the four-way junction at the base of the cruciforms. CBP appears to interact with the elbows of the junctions in an asymmetric fashion. Upon CBP binding, structural distortions were observed in the cruciform stems and in a DNA region adjacent to the junction. These features distinguish CBP from other cruciform binding proteins, which bind symmetrically and display exclusively either contacts with the DNA backbone or structural alterations in the DNA.

Key words: four-way DNA junction/hydroxyl radical footprinting/protein–DNA interaction/stem–loop/structure-specific recognition

Introduction

Soon after the discovery of the double-stranded helical structure of DNA (Watson and Crick, 1953) the formation of cruciform DNA was proposed (Platt, 1955) as a means to 'untwist' the two strands for DNA replication. Inverted repeat (IR) DNA sequences have the potential to form cruciform structures through intra-strand base pairing. IRs are present and distributed in a non-random fashion in the chromosomal DNA of many eucaryotes (Klein and Welch, 1980, and references therein). Cruciforms have been shown to form in procaryotic (Zheng *et al.*, 1991) and mammalian (Ward *et al.*, 1990, 1991) cells.

IRs are a common feature of many procaryotic, viral and eucaryotic origins of DNA replication (Muller and Fitch, 1982; Zannis-Hadjopoulos *et al.*, 1984). Cruciforms have been implicated in the initiation of DNA replication in procaryotic plasmids (Lin and Meyer, 1987; Noirot *et al.*, 1990), eucaryotic viruses (Pogue and Hall, 1992) and are thought to be involved in the initiation of DNA replication in mammalian cells (Hand, 1978). We had previously associated cruciform structures with mammalian DNA replication (Zannis-Hadjopoulos *et al.*, 1988; Ward *et al.*, 1990; Bell *et al.*, 1991). The formation of cruciforms and their role in the regulation of DNA transcription and replication may be controlled by cruciform-specific binding proteins (Einck and Bustin, 1985; Zannis-Hadjopoulos *et al.*, 1988; Noirot *et al.*, 1990; Waga *et al.*, 1990).

Using cruciform-containing heteroduplexes (Nobile and Martin, 1986; Frappier *et al.*, 1987, 1989), we have identified and enriched a novel DNA binding protein (CBP) from HeLa cell nuclear extracts with specificity for the cruciform-containing heteroduplex DNA molecule (Pearson *et al.*, 1995). In this study we have characterized the binding specificity of the CBP for cruciform DNA using hydroxyl radical footprinting analysis. The footprints reveal a novel mode of interaction of a protein with cruciforms.

Results

Properties of the heteroduplex cruciforms

Previous enzymatic studies (Nobile and Martin, 1986; Frappier *et al.*, 1989) have demonstrated that both heteroduplex molecules AD and CB (Figure 1I) are susceptible at the tips of the loops to cleavage by S1 and mung bean nucleases, are naturally resistant to DNase I cleavage at the base of the cruciform and are recognized and restricted by T7 endonuclease I, which is specific for four-way DNA junctions (Parsons and West, 1990). Thus the structures are stereochemically equivalent to a cruciform, except that they lack the symmetry of a true cruciform in that the stems are not complementary. Cruciform molecules AD and CB (Figure 1I) migrate in polyacrylamide as a single sharp band (Figure 1II, lane 2; Figure 1III, lanes 1 and 3). Co-migration of the two possible cruciform-containing molecules occurs over a wide range of NaCl and/or Mg^{2+} concentrations during electrophoresis (data not shown), suggesting that they possess very similar conformations.

Cruciform binding activity in HeLa cell extracts

CBP has novel cruciform binding activity, is enriched from HeLa nuclear extracts and sediments in glycerol at 66 kDa; it is devoid of any detectable nuclease activity. By the criteria of molecular weight, migration rates in retention gels and immunochemical properties, this new activity is distinct from that reported for HMG1 (Pearson *et al.*, 1995). Competition experiments indicate that the binding is specific for cruciform structures, regardless of sequence (Pearson *et al.*, 1995). CBP does not bind to linear homoduplexes (Figure 1II, lane 4 versus 3) or to

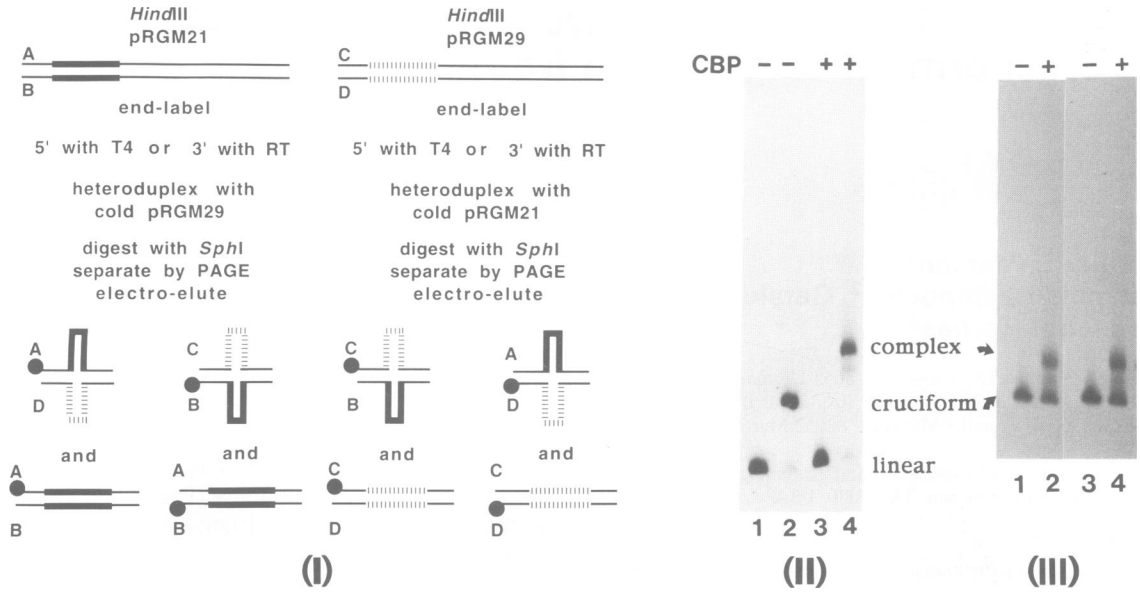

Fig. 1. Schematic of labeling/cruciform isolation and cruciform-specific binding activity. (**I**) The bold and hatched lines represent the different inverted repeats; the thin lines represent regions of homology between plasmids pRGM21 (strands A and B) and pRGM29 (strands C and D). Only ^{32}P-labeled molecules are shown in the figure; the labeled 3831 bp pBR322 *Sph*I–*Hin*dIII fragment is not shown. ^{32}P-Labeling is indicated by a dot. For details, see Materials and methods and Results. (**II**) A mixture of isolated ^{32}P-labeled linear homoduplexes (AB and CD) (lanes 1 and 3) or of isolated labeled cruciform heteroduplexes (AD and CB) (lanes 2 and 4) were used as binding substrates with CBP in a band shift assay, as indicated in the figure. (**III**) Isolated ^{32}P-labeled cruciform heteroduplex AD (lanes 1 and 2) or CB (lanes 3 and 4) were used separately as binding substrates with CBP in a band shift assay, as indicated in the figure. + indicates addition of CBP. For details see Materials and methods and Results.

unrelated linear competitors. CBP binding to a mixture or to either one of the cruciform-containing molecules also yields single co-migrating bands (Figure 1II, lane 4; Figure 1III, lanes 2 and 4). This indicates that CBP does not distinguish between the two cruciforms and suggests a stable and similar conformation for both complexes. We observed efficient binding in the presence of EDTA. Because Mg^{2+} can affect the structure of four-way junctions (Duckett *et al.*, 1990), we performed binding of CBP in the presence of this metal ion, in the absence of a chelator; the results indicate that Mg^{2+} impairs the interaction (data not shown). Therefore, to maximize complex formation, all binding reactions in this study were performed in the absence of Mg^{2+}. In this context it is important to note that other cruciform binding proteins have been also observed and their interactions were analyzed in the absence of Mg^{2+} (Parsons and West, 1990; Parsons *et al.*, 1990; Bennett *et al.*, 1993; Varga-Weisz *et al.*, 1993).

Hydroxyl radical analysis of the cruciform molecules

To obtain quantitative information about every single residue in the DNA, we performed an analysis of the CBP interaction with cruciform-containing molecules AD and CB (Figure 1I) using free hydroxyl radicals, which are sensitive both to alteration of DNA structure (Price and Tullius, 1992) and to protein–DNA contacts (Dixon *et al.*, 1991).

In the absence of protein, hydroxyl radical attack on either of the four strands in the linear duplex forms approximated an even ladder of bands (Figure 2I, lanes 1 and 7; Figure 3I, lanes 3 and 8). For each of the four strands in the cruciform molecules, in addition to similar, evenly distributed ladders of bands, there were also reproducible and quantitative differences in intensity at particular residues specific to each strand (Figure 2I, strand A, lanes 2 or 3 versus 1; strand D, lane 6 versus 7; Figure 3I, strand C, lane 4 versus 3; strand B, lane 7 versus 8). These differences in intensity are manifested as both decreases and increases compared with their linear counterparts. Decreases in intensities to varying extents are apparent in all four strands, with the maximum decrease centered mainly at the junction region (Figures 2I and II and 3I and II), except in strand D (Figure 2I, lane 6 versus 7), in which the 3′-decrease is slightly shifted away from the junction in the 3′ direction. Reduced strand cleavage could be a consequence of self-protection of the sugar backbone, because of the difficulty of accessibility to the cleft at a bent junction. Alternatively, the residues in these regions may have become single-stranded or unstacked; it has been previously shown (Prigodich and Martin, 1990) that the bases in single-stranded DNA can scavenge radicals, thus reducing local concentrations and consequently their effect on strand cleavage. However, since the single-stranded loop residues (Nobile and Martin, 1986; Frappier *et al.*, 1989) do not show extensive decreases in intensity (Figure 2I, strand A, lanes 2 or 3; strand D, lane 6; Figure 3I, strand C, lane 4; strand B, lane 7), the decrease at the junctions may not be due solely to single-strandedness of the DNA in these regions. It is likely that the structure at the junction deviates from 'normal' doubled-stranded DNA in both linearity and stacking.

We also observed changes in the cleavage pattern on one loop of each cruciform; a strong increase for a cytosine nucleoside at the 3′ site of the loop of strand A (Figure 2I, lanes 2 or 3 versus 1; Figure 2II, filled symbol) and a

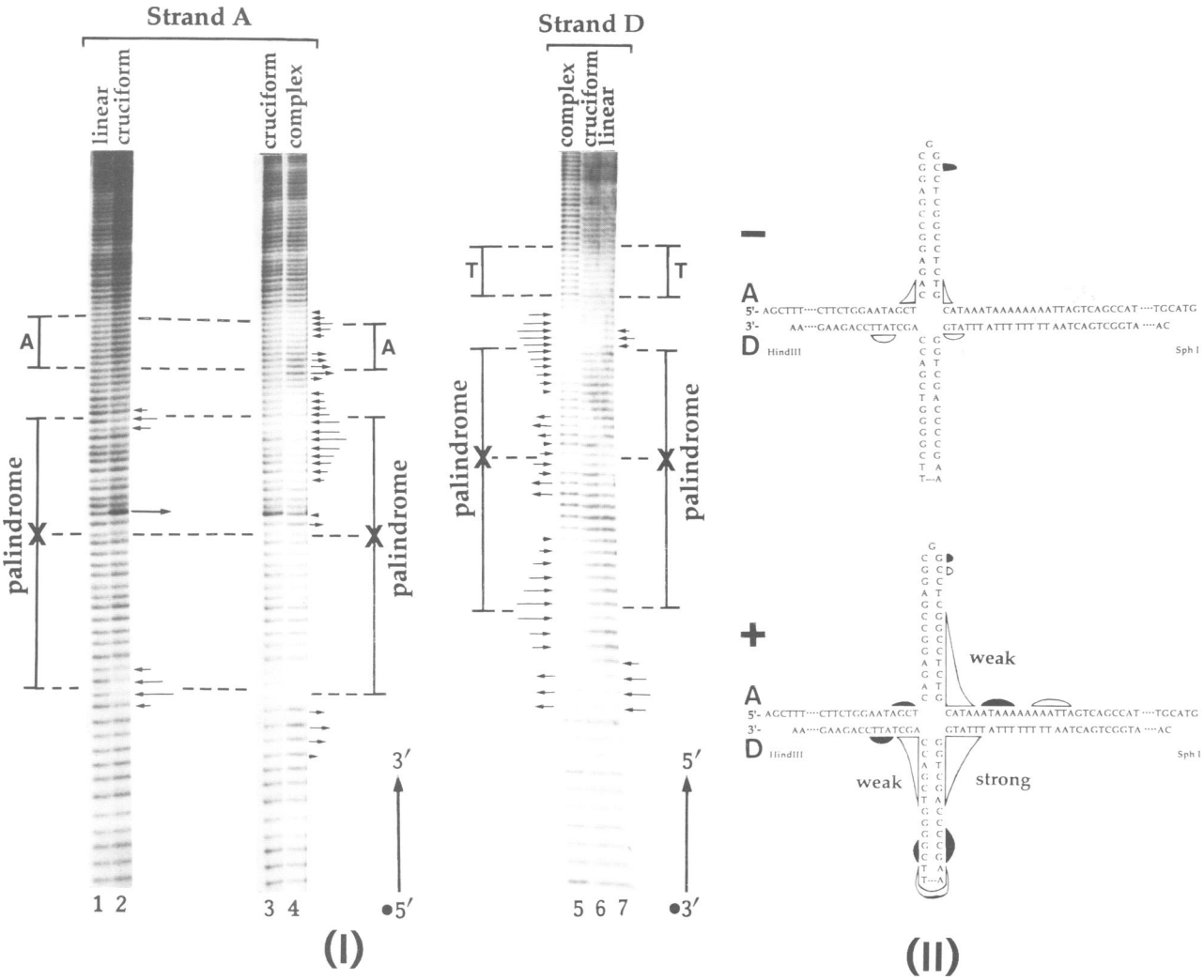

Fig. 2. Structural analysis of cruciform AD and the CBP–cruciform complex. (**I**) Hydroxyl radical analysis (as described in Materials and methods) of each of the two different strands A or D (underlining indicates the ^{32}P-labeled strand) in either the linear homoduplex conformation (AB, lane 1; CD, lane 7) or the cruciform conformation in the absence (AD, lanes 2 or 3; AD, lane 6) or presence of CBP (AD, lane 4; AD, lane 5). ^{32}P-Labeling is as indicated in the figure. The A-rich region (A) of strand A, the T-rich region (T) of strand D and the palindromes that form the cruciform are indicated. Arrows pointing towards or away from the lane(s) indicate decreased or increased chemical reactivity respectively of the bound versus free cruciform or cruciform versus linear homoduplexes. Arrow length is proportional to the relative amount of decrease or increase. Assignment of reactivity to specific bases was done according to Maxam and Gilbert (1980) sequence lanes, as indicated in Figure 3 (not shown here). (**II**) Schematic map of the sites of protection and increased chemical reactivity are represented by hollow and filled sites respectively; the size of the sites is proportional to the altered reactivity. Free and CBP-bound DNA are indicated by − and + respectively.

slight decrease for the central cytosine nucleoside of the loop of strand B (Figure 3I, lane 8 versus 7; Figure 3II, hollow symbol). Neither change is simply attributable to the single-strand state of the loops, because of the opposite effects (increase versus decrease) and because neither the neighboring residues nor the loops of strands C and D, which are also single-stranded, show any alterations. It is more likely that the particular conformations of loops A and B are similar in that, in each, some deoxyriboses alter their positioning in a way that renders them differentially accessible to hydroxyl radicals, as opposed to the loop residues of strands C and D. The opposite effects of hydroxyl radical susceptibility in the loops of strands A and B (i.e. increase versus decrease) might result from the single nucleotide difference in the sequence at the tips of the loops, affecting their exact local structure.

Hydroxyl radical fine mapping of the CBP–cruciform complex

Binding of CBP to the cruciforms caused pronounced alterations in the chemical cleavage pattern (Figures 2I and 3I, complex). We obtained both decreases and increases in band intensities compared with the naked cruciform molecules. Although increases in intensity clearly indicate an altered structure in the corresponding region, the decreases are more difficult to assign unequivocally as direct protein contacts or as induced structural changes due to protein binding. Major decreases in intensity were obtained mainly at the junctions of the cruciforms (Figure 2I, strand A, lane 4 versus 2 or 3; strand D, lane 5 versus 6; Figure 3I, strand C, lane 5 versus 4; strand B, lane 6 versus 7). In the cruciform AD complex, strand A is protected at the 3′ elbow only (Figure 2I, lane 4 versus 2

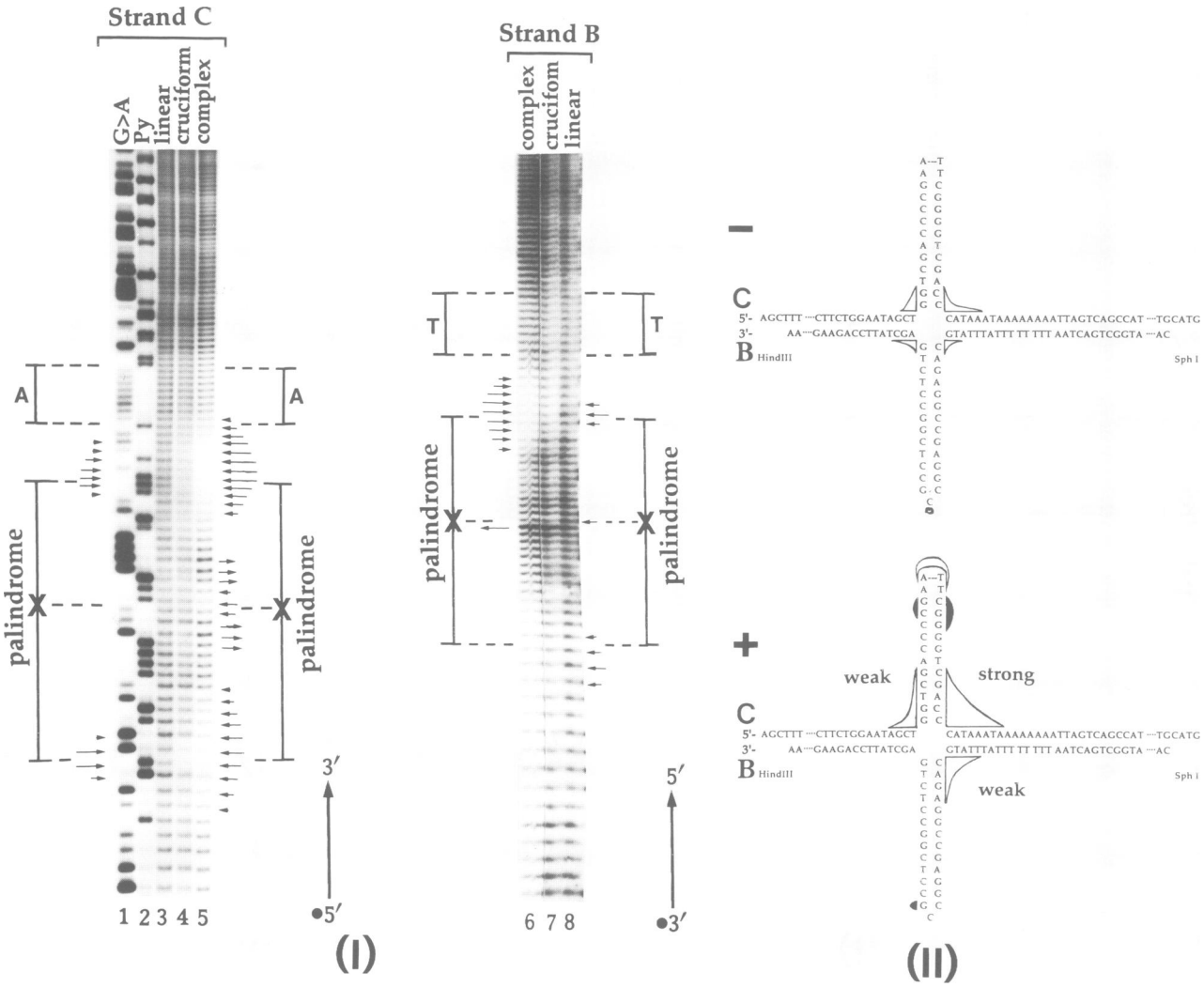

Fig. 3. Structural analysis of cruciform CB and the CBP−cruciform complex. **(I)** Hydroxyl radical analysis (as described in Materials and methods) of each of the two different strands C or B (underlining indicates the ^{32}P-labeled strand) in either the linear homoduplex conformation (CD, lane 3; AB, lane 8) or the cruciform conformation in the absence (CB, lane 4; CB, lane 7) or presence of CBP (CB, lane 5; CB, lane 6). ^{32}P-Labeling is as indicated in the figure. **(II)** Maxam and Gilbert sequence ladders (1980) for strand C are shown (G > A, lane 1; pyrmidines, Py; lane 2). The A-rich region (A) of strand C, the T-rich region (T) of strand B and the palindromes that form the cruciform are indicated in the figure. Other symbols are as described in Figure 2.

or 3; Figure 2II+), whereas strand D displays protection at both elbows (Figure 2I, lane 5 versus 6; Figure 2II+). In the cruciform CB complex, it is the C strand which is protected at both elbows (Figure 3I, lane 5 versus 4; Figure 3II+), while strand B shows clear protection only at the 5′ elbow (Figure 3I, lane 6 versus 7; Figure 3II+). In both cruciform complexes, a 'strongly' protected elbow, i.e. one in which the band intensities approach zero, is flanked by two more 'weakly' protected elbows (compare Figures 2II+ and 3II+). In the manner drawn, which preserves the positioning and helical pitch of the homologous parts of the linear counterparts, the pattern of protection at the junctions is similar after a 180° rotation around the axis of the branch helices. These results suggest that the DNA at the junctions of both complexes must be folded in an equivalent way.

In the bound complexes two other regions of altered susceptibility to chemical attack are apparent in the stem−loops of strands D (Figure 2I, lane 5 versus 6) and C (Figure 3I, lane 5 versus 4): slight but clear decreases in band intensity appeared at the tips of these loops, flanked by some increases in intensity in the stem residues. This pattern is compatible with an interaction of CBP with these loops, which forces bending of the respective stem−loops towards the protein. Bending might have caused a widening of a DNA groove of the stems (Bennett *et al.*, 1993, and references therein), thus exposing the C4′ atoms of the deoxyriboses to increased radical attack (Tullius, 1987). Alternatively, the altered pattern in these two stem−loops might reflect structural changes alone, resulting from protein interaction only at the bases of the hairpins. The loops in strands A (Figure 2I, lane 4 versus 2 or 3) and B (Figure 3I, lane 6 versus 7) show only very weak changes in the cleavage pattern, which, due to their confinement to single residues, we consider represent minor structural changes rather than direct protein contacts (cf. analysis of the uncomplexed cruciforms above). The patterns of the pairwise homologous stem−loops (of strands C and D or A and B) are very similar to each other (see Figures 2II+ and 3II+), which suggests that

they are recognized by the protein as equivalent structures. Interestingly, the pattern of stem–loop alterations in each cruciform are again related to each other by a 180° rotation around the axis of the branch helices (cf. above). Since several elements (stem–loops, folding at the junctions) seem to adopt a comparable structure in the two complexed cruciform molecules, the overall conformations of the complexes may resemble each other to a high degree; therefore, the complexes might share a common basic structure. This is in accordance with CBP interactions being of indistinguishable specificity and affinity for both cruciforms and resulting in complexes that display the same migration rate in retention gels (cf. Figure 1II, lane 4; Figure 1III, lanes 2 and 4).

Another decrease in intensity occurs at the 3′ end of the A tract in the 3′ region of strand A (Figure 2I, lane 4 versus 2 or 3). This is a common phenomenon in some A tracts probed with hydroxyl radicals (Burkhoff and Tullius, 1988) and has been observed in all A tracts of intrinsically curved DNA (Burkhoff and Tullius, 1987), isolated single A tracts (Burkhoff and Tullius, 1987) and at induced bends in A tracts without direct, but neighboring, protein contact (Zorbas *et al.*, 1989). It might be due to a narrowing of the minor groove in the AT stretch (Burkhoff and Tullius, 1987), giving rise to a bend, indirectly induced at this site. Although the complementary T-rich strand would also be expected to show some protection (Burkhoff and Tullius, 1987; Zorbas *et al.*, 1989), this is not apparent here (Figure 2I, strand D, lane 5 versus 6). This may be due to the very small degree of this bend. A mild reduction in radical cleavage in this region compared with flanking sequences is also seen in all strands of the uncomplexed heteroduplex and homoduplex DNA (Figures 2I and 3I, all lanes). This might be an indication that this region is smoothly curved before protein binding and, in cruciform AD, becomes more bent as a result of binding. If this decrease were due to direct protein contacts rather than to an indirectly induced bend, we would also expect a decrease at the analogous position in the CB cruciform, which is not the case (Figure 3I, lane 5 versus 4).

The 3′ decrease in the A tract of strand A is accompanied at the 5′ border by an increase in intensity (Figure 2I, lane 4 versus 2 or 3), again suggesting a widening of the minor groove at this site (see discussion of the stem–loops above). The maxima of these two changes are about a half helix turn apart and thus on opposite sides of the branch. The increase lies some 10 bp from the maximal protection at the neighboring junction. In fact, this is the pattern one would expect if the branch were bent towards the protein occupying a major groove at the junction; this would narrow the minor grooves on the same side of the helix and widen the intervening minor groove on the opposite side.

Finally, we observed very faint increases at the 5′ side of the junction of strand A (Figure 2I, lane 4 versus 2 or 3; Figure 2II+) and at the 3′ side of the junction of strand D (Figure 2I, lane 5 versus 6; Figure 2II+). Such increases were absent in cruciform CB. These may represent minor structural changes in cruciform AD, which together with the described bend (above) comprise a certain heterogeneity in the fine structure of the complexes.

Footprinting of the CBP–cruciforms with other methods

We tried to obtain information about essential contacts of CBP with particular residues of the cruciforms by performing methylation interference and missing contact analysis of hydroxyl radical-treated DNA (Hayes and Tullius, 1989). Neither method revealed any residues required for stable CBP–cruciform interaction (data not shown). This suggests that there is in fact no required direct interaction of CBP with any specific base(s), in accordance with the apparent absolute insensitivity of CBP to competition with the linear AB or CD homoduplex molecules and with the conclusion that the interaction is purely structural (Pearson *et al.*, 1994).

A model for the CBP–cruciform complex

As stated above, the footprinting patterns at the junctions and at the stem–loops are similar for the two cruciforms and related to each other by the axis of the branch helices. Additionally, as shown by retention gel analysis (this study and Pearson *et al.*, 1994), the two complexes migrate indistinguishably, suggesting a similar shape. Therefore, the overall structure of the cruciforms and the manner of CBP interaction with the cruciforms, assuming a homogeneous CBP composition, may be very similar in both complexes; thus the complexes probably share a common basic stucture. A common, three-dimensional model for the cruciform molecules AD and CB complexed with CBP must account for the fact that complementary (A and B, C and D) and not homologous (A and C, B and D) strands display a similar footprinting pattern. This means the following: if the spatial arrangement of the homologous strands in each of the two cruciforms were to be equivalent, CBP would have to interact in a different manner with the two cruciforms, by recognizing different surfaces, to give these footprints. This, of course, is in contradiction to an assumed similar interaction. Therefore, it is the complementary strands which must have the same spatial arrangement in each of the two cruciforms. Additionally, in the model, the structure and orientation of homologous stem–loops (A and B, C and D) and their interaction with the protein must also be pairwise equivalent. A model should also enable us to infer the small but apparent differences in interaction between the two cruciform molecules, especially the apparent lack of bending of the AT tract in CB versus AD. The common basic structure may display only few elements of symmetry: these are reflected in the cleavage patterns of the strands of the two opposite elbows in each cruciform, both of which show a similar 'weak' protection [cruciform AD, strands A and D at the downstream (3′) elbow; cruciform CB, strands C and B at the upstream (5′) elbow]; therefore, the strands at these elbows may all interact with CBP pairwise in a similar manner (Figures 2II+ and 3II+). In contrast, in each complex the other two junction elbows are affected in a different manner, which suggests that the surfaces defined by these two sets of junction elbows are not equivalent. Also, the two stem–loops on each cruciform are recognized differently by the protein, and so they must differ in orientation and/or shape. In other words, the interaction is asymmetric overall. One possible type of asymmetric interaction could derive from a sequence-dictated preference of CBP for a

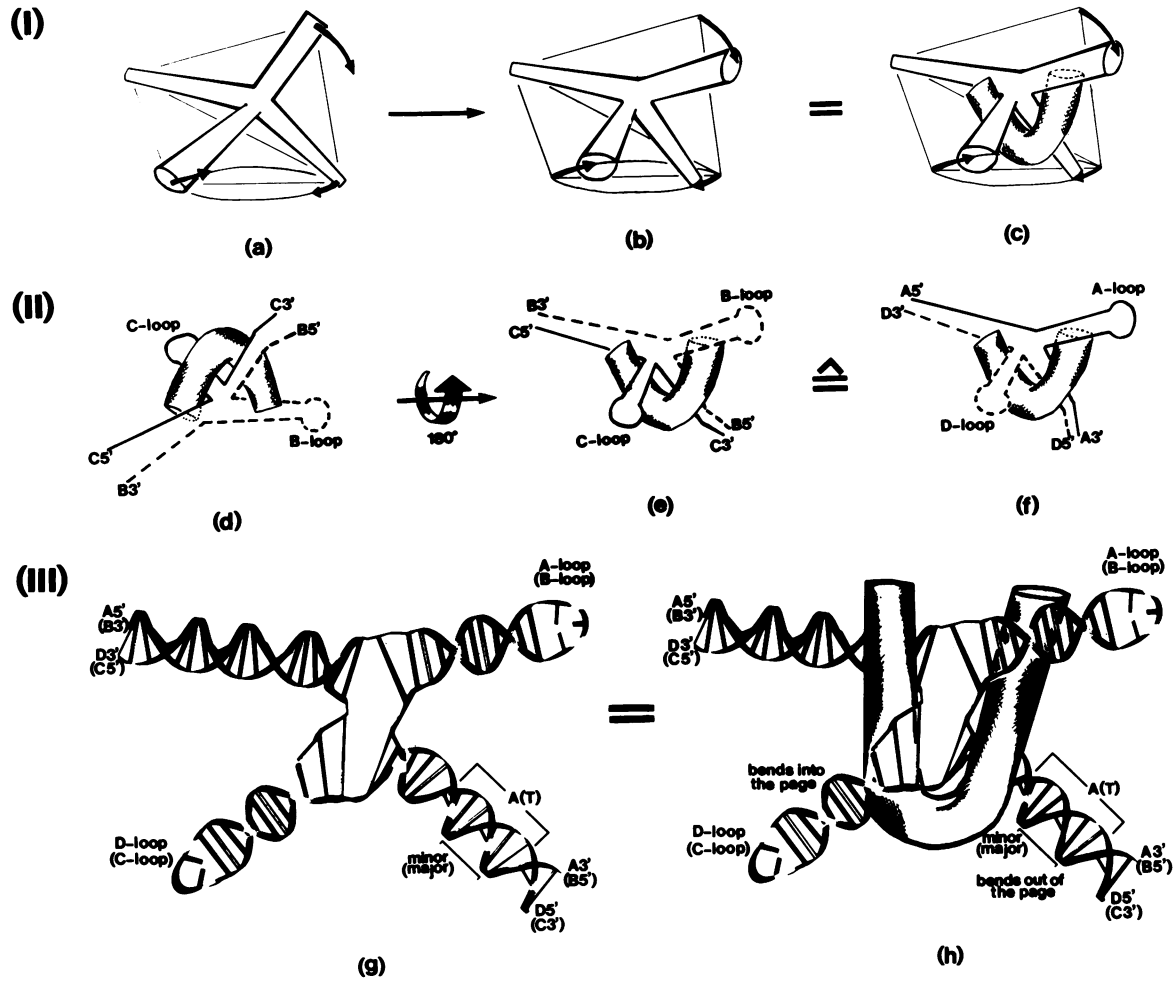

Fig. 4. A model for the CBP–cruciform complex. (**I**) Derivation of the distorted tetrahedron. (a) Cruciform shape with helices symbolized by straight cylinders, adjusted to a regular tetrahedron, indicated by thin lines. Arrows (on thin ellipsoids) indicate the direction of the movement of arms to yield a distorted tetrahedron. (b) Result of the movement of three arms. (c) As (b), with bound protein, symbolized as a bent cylinder; since there is no experimental information regarding the protein folding, one alternative is presented which accounts for the contacts between the protein and the cruciform. (**II**) Schematic drawing of cruciforms CB (d and e) and AD (f) in the complex with CBP protein. The paths of individual strands as well as the smooth inherent DNA curvature on branch B5′/C3′ (A3′/D5′) are indicated. The equivalent strands A and C are depicted as continous; the equivalent strands B and D are depicted as broken lines. (**III**) Helical representation of cruciform AD (CB) complexed with CBP. In (g) the protein has been omitted for clarity, in (h) it is bound to the cruciform(s). The DNA double helix is shown without emphasis of major and minor grooves to account for the fact that these are reversed in the two cruciforms, as indicated in the figure. The adenosine (A) and thymine [(T)] components of cruciforms AD and (CB) respectively are indicated in the figure. In the complex, branch A3′/D5′ bends out of the page and stem–loop D (C) bends into the page, as indicated.

particular helix (strand) of a symmetric cruciform. However, this is not likely, because we have largely excluded a direct influence of sequence on the CBP–cruciform interaction (this study and Pearson *et al.*, 1994). Consequently, the protein should be able to recognize a different arrangement of the helices. Thus, the cruciform structure *per se* in the complexes appears to be asymmetric, and therefore we may exclude a planar–tetragonal, symmetrical X-shaped or symmetrical tetrahedral–pyramidal geometry.

Based on the above, we propose a model for the protein-bound cruciforms, best described as a distorted tetrahedron, with the following general characteristics (Figure 4I–III). (i) Double-stranded DNA helices with average sequence composition are in the B-DNA conformation. (ii) The structure deviates from a regular tetrahedron (Figure 4Ia) in the following manner: (a) the stem–loop A (B) approaches co-linearity with branch arm A5′/D3′ (B3′/C5′) (Figure 4Ib, see arrows, and 4II); (b) stem–loop D (C) is bent towards stem–loop A (B) (Figure 4Ib, see arrows, and 4II); (c) branch arm A3′/D5′ (B5′/C3′) bends to a similar extent towards branch arm A5′/D3′ (B3′/C5′) (Figure 4Ib, see arrows, and 4II). In this arrangement the opposite angles subtended by stem–loop D and branch arm A5′/D3′ (stem–loop C and branch arm B3′/C5′) and by stem–loop A and branch arm A3′/D5′ (stem–loop B and branch arm B5′/C3′) are obtuse and pairwise equal; in contrast, branch arm A5′/D3′ and stem–loop A (branch arm B3′/C5′ and stem–loop B) and stem–loop D and branch arm D5′/A3′ (stem–loop C and branch arm C3′/B5′) subtend unequal flattened obtuse angles. (iii) CBP straddles the four-way junction from the region between stem–loop D and branch arm A3′/D5′ (stem–loop C and arm B5′/C3′) (Figure 4Ic, II and IIIh), causing areas of protection on three of the four junction strands, with the greatest protection at the 5′ elbow of strand D (3′ elbow

of strand C) (compare 'strong' protection in Figures 2II+ and 3II+). The continuous, quasi-linear region of strand A (B) spanning the junction from the opposite side is not affected.

Preserving the pitches of the strands of corresponding helices in each of the two cruciforms, e.g. branch arm A5′/D3′ and branch arm C5′/B3′ or branch arm D5′/A3′ and branch arm B5′/C3′ would give rise to 'reciprocal' structures, as depicted in Figure 4IIf and d. By rotating cruciform complex CB (Figure 4IId) by 180° (Figure 4II, as indicated by the arrow), the overall structures of the two cruciform complexes, CB and AD, (Figure 4e and f, respectively) become superimposable. This rotation brings the footprinting patterns to congruency, as recognized in the diagrammatic presentations of the hydroxyl radical data (see previous section and Figures 2II+ and 3II+). However, as an immediate consequence of the rotation, in the superimposed structures, where cruciform complex CB presents a major groove, cruciform complex AD presents a minor groove (Figure 4IIIg and h, see reversal of A with T and of minor with major). Although the overall structure of the two cruciforms is superimposable, the equivalent strands A and C or D and B follow different paths, particularly in the loop regions, with respect to the helical axes. Thus they provide a different footprinting pattern. On the other hand, the homologous stem–loops (A and B or D and C) have the same orientation, as required. Since we suggested that all four helices in one cruciform have a different orientation, stacking at the junctions of the complex might be imperfect, resembling that of the naked cruciform molecules. This is reflected in the cleavage pattern of the junction residues of strands A or B, which in the model are assumed not to be contacted by CBP.

Elements of symmetry, i.e. equal angles giving the opposite pairwise 'weak' protections in each cruciform, in the model are retained. Thus, in both cruciforms CBP faces the same structure with quasi-2-fold symmetry of the junction angles, therefore interacting with similar affinity and specificity with it, as expected from previous analyses (Figure 1II and III; Pearson *et al.*, 1994). However, within one molecule, the interacting surface is unique, because of the unique identity and angles of neighboring helices, giving rise to asymmetric interaction, complex and footprints. The same points may be the cause for the different hydroxyl radical cleavage patterns of the two stem–loops within one cruciform molecule. Stem–loop D (C), being next to the protein may bend smoothly towards it, thus widening the groove of the stem at the opposite surface next to the loop. In contrast, stem–loop A (B) remains largely unaffected. The same would be true for the corresponding region of arm A5′/D3′ (B3′/C5′). This arrangement predicts different structures for the two stem–loops in the same cruciform complex and a similar structure for homologous stem–loops (A and B or D and C) of the two different cruciform complexes, as required.

We suggested above that arm A3′/D5′ also bends towards CBP, causing the observed alterations in the cleavage pattern around the AT region (see Figures 2 and 4). A possible reason for the observed difference in bending in the analogous branch B5′/C3′ of cruciform CB might be the different geometry of the unbound molecule for branch B5′/C3′ as compared with branch A3′/D5′, i.e. in cruciform AD, strand A is A-rich, whereas in cruciform CB, strand C is A-rich (Figure 4IIIg and h, see reversal of A with T and of minor with major). However, according to the model, strand C follows the path of strand D, not A (Figure 4IIIg and h). Thus in the superimposed structures, the A-rich and the T-rich strands in cruciform CB are reversed, with the helix being curved in the opposite direction as compared with that of cruciform AD (Figure 4II, compare e and f; Figure 4IIIg and h). This might constitute an orientation that would not allow CBP to attract this branch towards itself. The assumption that neither branch A3′/D5′ nor B5′/C3′ are contacted, thereby stabilizing the interaction differentially, is in accordance with the apparently similar binding affinity of CBP for the two cruciforms (Pearson *et al.*, 1994).

Discussion

The model described above is a first approximation and contains several implicit assumptions which can be tested by additional, independent experiments. However, most of our experimental results on CBP–cruciform interaction (this study and Pearson *et al.*, 1994) can be explained by this model. How probable is the suggested structure? The three-dimensional structure of four-way DNA junctions has been a field of intense investigation. Most of the structural analysis of four-way DNA junctions has been modeled on small stable junctions composed of four synthetic oligonucleotides (reviewed in Lilley and Clegg, 1993). The structure of these stable junctions depends critically on the DNA sequence, mainly at the junction point directly, which will determine the distribution of the stereoisomers (Chen *et al.*, 1988; Duckett and Lilley, 1991), and on the type and amount of counterion used in the solutions, determining the geometry of the helices (Duckett *et al.*, 1990). Briefly, in the absence of salt, the junction is in an extended conformation, probably planar–tetragonal with 4-fold symmetry, with unstacked bases at the junction (Duckett *et al.*, 1988). Micromolar concentrations of Mg^{2+} enable the four-way junction to adopt a more compact, X-shaped structure with 2-fold symmetry, with pairwise coaxial stacking of helices and apparently no unpaired bases (Duckett *et al.*, 1988). In the presence of Na^+ ($\geq$50 mM, as in this study), the structure is similar to that in the presence of Mg^{2+} in that it is also compact (X-shaped) (Clegg *et al.*, 1992; Lilley and Clegg, 1993), however, the structure has only imperfect 2-fold symmetry (Duckett *et al.*, 1990) and the junction bases are still unstacked (Duckett *et al.*, 1988, 1990). The cruciform substrates used in this study differ from four-way DNA junctions generated from short oligonucleotides in that they contain loops and are significantly longer (200 bp). In spite of these differences, the characteristics of the complexed cruciforms in the proposed model show remarkable similarity to protein-free four-way junctions in Na^+-containing buffers as determined by others (Duckett *et al.*, 1990; Clegg *et al.*, 1992; Lilley and Clegg, 1993): h pair of opposite helices (for example branch A5′/D3′, stem–loop A) crosses the other one (for example branch A3′/D5′, stem–loop D) in a non-orthogonal way, thus rendering the molecule compact, much like an X, with equal opposite but unequal neighboring angles. However,

the axes of opposite helices are claimed to be imperfectly co-linear, resulting in non-regular base pairing at the junctions, which was implied from the cleavage patterns of the junction residues of strands A and B. Finally, the two crossing helix pairs (cf. above) deviate from co-linearity to different degrees, creating a partial asymmetry in the folding of the molecules. It is also noteworthy that cruciform helices with the same sequence in the common basic structure assume the same spatial orientation: compare, for example, in Figure 4IIe and f, branch arm B3′/C5′ with equivalent branch arm A5′/D3′ or stem–loop A with the homologous stem–loop B, which has an identical sequence but for the central base. This reflects a particular mutual orientation of the base pairs at the junction points and reveals that in both cruciforms the identity of adjacent base pairs of quasi-co-linear helices directly at the junctions are entirely preserved. Therefore, although base pair stacking at the junctions might not contribute much to stabilization of the structure under our experimental conditions (see above), to a certain degree the sequence may dictate the 'partner choice' among the four helices, thus determining the bias of the possible isomers. This has been exactly found with free cruciforms not only in the presence of Mg^{2+}, but also in the presence of (enough) NaCl (Chen *et al.*, 1988; Duckett *et al.*, 1988, 1991; Murchie *et al.*, 1989, 1991; Clegg *et al.*, 1992; Lilley and Clegg, 1993). Taken together, it is most likely that the basic structure of the cruciform substrates in the complex with CBP, which we have inferred from our data, is very similar in crucial aspects to that described in other model systems analyzed by different methods under comparable conditions, and does not differ significantly from the free form. Thus the likelihood of the model is substantiated by its similarity to known structures.

The reactivity pattern of the residues in the loop regions are also in agreement with known properties of hairpin/loop architecture in other systems. For example, it has been reported that the details of loop folding conformation depend on the base sequence of the loop and the stem (Germann *et al.*, 1990; Blommers *et al.*, 1991). In accordance, we observed the same cleavage patterns for entirely homologous loops (of strands D and C), but different patterns for the loops of strands A and B that differ in the central base (G versus C). Generally, loop formation may be accompanied by unusual deoxyribose structures (Blommers *et al.*, 1991). Hairpin loops probably involve base stacking (Blommers *et al.*, 1991), non-Watson–Crick base pairing (Blommers *et al.*, 1991) and possibly extraloop (unpaired and unstacked) bases (Zhou and Vogel, 1993). Such unusual structures could account for varying degrees of cleavage at residues within the loops of strands A and B.

The smoothly curved DNA in the AT tract in branch arms A3′/D5′ and B5′/C3′ is not a consequence of cruciform formation, since the same cleavage pattern was also found in the linear counterparts. In fact, this region has been previously shown to be curved (Deb *et al.*, 1986). Binding of CBP to cruciform AD induced increased bending at this site. Such bending may have biological significance for facilitating the interaction of other factor(s) present in HeLa cell nuclei with the cruciform substrates (Pearson *et al.*, 1994). Previous work (Borowiec and Hurwitz, 1988) has shown that the AT tract is structurally 'distorted' following the binding of SV40 T-antigen to site II of the SV40 *ori* (corresponding here to the linear inverted repeat of homoduplex AB), without apparent essential protein contacts. Possibly, this region is prone to structural change upon binding of proteins nearby. Thus, within a suitable sequence and/or structural context, CBP, upon binding to the junction of the cruciform, may exert certain 'distant effects' on both the branch arms and stem–loops which may prove important for its function.

There are only three other examples of proteins interacting with four-way junctions which have been analyzed by hydroxyl radical footprinting: the bacteriophage-encoded resolvases T4 endonuclease VII (Parsons *et al.*, 1990) and T7 endonuclease I (Parsons and West, 1990) and the *Escherichia coli* RuvC resolvase (Bennett *et al.*, 1993). Irrespective of the exact structure of the cruciform substrates in the complex with CBP, it is apparent from the footprinting pattern of these proteins that their modes of interaction differ significantly from that of CBP. (i) CBP interacts with the junction giving clear areas of protection and simultaneously introducing several changes in the fine structure of the cruciform. In contrast, the T7 endonuclease I (Parsons and West, 1990) and the T4 endonuclease VII (Parsons *et al.*, 1990) contact the DNA backbone without detectably altering its structure. On the other hand, the RuvC resolvase (Bennett *et al.*, 1993), surprisingly, reveals only structural alterations, apparent as increases in radical cleavage intensity, but no protection footprint. Since a sequence dependence of the cleavage reaction of the RuvC resolvase was demonstrated (Bennett *et al.*, 1993), this protein might not interact directly with the DNA backbone (Bennett *et al.*, 1993). (ii) Most importantly, in the hydroxyl radical cleavage patterns of the bacteriophage endonuclease–cruciform complexes only two diametrically opposed (T4; Parsons *et al.*, 1990) or all four junction strands (T7; Parsons and West, 1990) are protected, revealing a radically different structural association of these proteins with the substrates than that of CBP. In addition, the fact that CBP protects both junction strands of stem–loops D and C as well as both strands of branch arms A3′/D5′ and C3′/B5′ indicates that CBP apparently interacts at both major and minor groove faces of the cruciform junctions. In contrast T4, yeast and calf thymus resolving enzymes interact with the minor groove face of junctions (Bhattacharyya *et al.*, 1991). The high mobility group protein HMG1 binds to four-way junctions (Bianchi *et al.*, 1989; Pearson *et al.*, 1994), amongst other non-B-DNA structures (Hamada and Bustin, 1985). It has been reported that HMG1 can protect the single-stranded tips of cruciform stems from S1 nuclease digestion (Waga *et al.*, 1990), but not from T4 endonuclease VII (Bhattacharyya *et al.*, 1991), which interacts at the junction, suggesting that HMG1 binds in a different manner to CBP. Apparently, CBP provides a novel type of cruciform DNA–protein interaction in that there are firm contacts with the sugar–phosphate backbone (protection) as well as structural alterations of the cruciform substrate, both reflecting a putative cruciform stabilizing function of CBP in the cell. The ability to structurally alter the DNA by binding of CBP provides a putative role for it in the preparation of DNA for the processes of replication, transcription or recombination. The asymmetric binding further predicts that there may well

be a specific orientation of CBP required for interaction with other proteins at the functional DNA element.

Materials and methods

DNA substrates

Plasmids pRGM21 and pRGM29 (Nobile and Martin, 1986; Frappier *et al.*, 1987, 1989) were used for construction of the heteroduplexes. Plasmid pRGM21 contains the *Hin*dIII–*Sph*I fragment (200 bp; strands AB, Figure 1I) of the wild-type SV40 origin of replication, cloned into pBR322. pRGM29 (strands CD, Fig 1I) is identical to pRGM21 except that the wild-type SV40 27 bp palindrome has been replaced by an unrelated 26 bp palindromic sequence. Heteroduplex formation between linearized pRGM21 and pRGM29 results in molecules with strands AD and CB (Figure 1I), each containing a stable cruciform at the inverted repeat sequences on opposite strands. The cruciform structures are stable (non-mobile), as the symmetry is such that no alternative base pairing can occur through branch migration.

Plasmids were linearized by digestion with *Hin*dIII and end-labeled with either T4 kinase (BRL) or with AMV reverse transcriptase (Boehringer) using [γ-^{32}P]ATP or [α-^{32}P]dATP respectively (Figure 1I, T4 and RT). The labeled pRGM21 or pRGM29 was then heteroduplexed (Frappier *et al.*, 1989) with an equal amount of unlabeled *Hin*dIII digest of pRGM29 or pRGM21 respectively. The heteroduplex mixtures were then digested with *Sph*I. To purify the uniquely labeled cruciform (heteroduplex) from the uniquely labeled linear (homoduplex) *Hin*dIII–*Sph*I molecules (200 bp), the digestion products were separated by electrophoresis on 4% polyacrylamide and the appropriate DNA fragments were localized by wet autoradiographic exposure and eluted from the gel by isotachophoresis (Ofverstedt *et al.*, 1984).

Extract preparation

HeLa S3 cell extract preparation and enrichment of cruciform binding activity was as described (Pearson *et al.*, 1991, 1994). The CBP-enriched protein fraction (2.65 mg/ml) from the heparin column, in 0.01 M K_2HPO_4, pH 7.4, 0.15 M NaCl, 2.5 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, was free of any detectable non-specific or sequence-specific DNA binding activities. In all experiments this fraction was diluted to 1 mg/ml with 25 mM HEPES–KOH, pH 7.5, 0.15 M NaCl.

Assay for DNA binding

DNA binding assays were performed in 20 mM Tris–HCl, pH 7.5, 1 mM dithiothreitol, 1 mM EDTA, 100 ng/ml double-stranded poly(dI–dC) (Pharmacia), in a final volume of 20 ml. Cruciform DNA was used at a concentration of 65–75 fmol/reaction; the protein fraction containing CBP was added to 0.5 mg/ml protein. Addition of the CBP protein fraction resulted in a final concentration of 75 mM NaCl; binding activity was evident over a range of NaCl concentrations (20–150 mM). After a 30 min incubation on ice, loading buffer (25% Ficoll, 25 mM EDTA, 0.2% bromophenol blue, 0.2% xylene cyanol) was added and the samples were subjected to electrophoresis on 4% polyacrylamide in 1× TBE at 12–13 V/cm for 1.5–2 h at room temperature. The gels were dried and exposed for autoradiography.

Footprinting

Hydroxyl radical modification was performed on binding reactions as previously described (Zorbas *et al.*, 1989). Binding reactions were performed as described above. Reactions were incubated on ice for 30 min; then, at room temperature, sodium ascorbate (Sigma), ammonium iron sulfate (Merck), EDTA (Boehringer) and hydrogen peroxide (Sigma) were added, yielding final concentrations of 10 mM, 10 and 20 μM and 0.03% respectively. Reactions were stopped by the addition of thiourea and EDTA to final concentrations of 10 and 2 mM. Loading dye was added and reactions were electrophoresed on 4% polyacrylamide gels in 1× TBE at 4°C to separate the bound species from free cruciforms. After electrophoresis, a wet exposure was made to locate the bound and unbound species. The bound species was excised and eluted by isotachophoresis (Ofverstedt *et al.*, 1984). The eluate was extracted with phenol and precipitated with ethanol in the presence of linear polyacrylamide (Zorbas *et al.*, 1992). Protein-free reactions were treated exactly the same (75 mM NaCl), except that the final concentration of sodium ascorbate was 1 mM and immediately after stopping the reaction the samples were precipitated with ethanol, without gel separation. Samples and sequence ladders (Maxam and Gilbert, 1980) were loaded onto a denaturing 8% polyacrylamide gel, electrophoresed, dried and exposed for autoradiography as previously described (Zorbas *et al.*, 1989; Schreck *et al.*, 1990).

Acknowledgements

We acknowledge the expert assistance of Colin Pearson on the 3-D Studio Imaging Graphics. M.Z.-H. is an MRC Scientist. This work was funded by grants from the Cancer Research Society Inc. and the Medical Research Council (MRC) of Canada to M.Z.-H. and a grant from the Deutsche Forschungsgemeinschaft (DFG Zo 59/2-1) to H.Z. C.E.P. is the recipient of a McGill Graduate Faculty Award (Faculty of Medicine) and a Canderel Travel Award. We also particularly acknowledge the generous financial assistance of Dr E.-L.Winnacker.

References

Bell,D., Sabloff,M., Zannis-Hadjopoulos,M. and Price,G.B. (1991) *Biochim. Biophys. Acta*, **1089**, 299–308.
Bennett,R.J., Dunderdale,H.J. and West, S.C. (1993) *Cell*, **74**, 1021–1031.
Bhattacharyya,A., Murchie,A.I.H., von Kitzing,E., Diekmann,S., Kemper,B. and Lilley,D.M.J. (1991) *J. Mol. Biol.*, **221**, 1191–1207.
Bianchi,M.E, Beltrame,M. and Paonessa,G. (1989) *Science*, **243**, 1056–1059.
Blommers,M.J.J., van de Ven,F.J.M., van der Marel,G.A., van Boom,J.H. and Hilbers,C.W. (1991) *Eur. J. Biochem.*, **201**, 33–51.
Borowiec,J.A. and Hurwitz,J. (1988) *EMBO J.*, **7**, 3149–3158.
Burkhoff,A.M. and Tullius,T.D. (1987) *Cell*, **48**, 935–943.
Burkhoff,A.M. and Tullius,T.D. (1988) *Nature*, **331**, 455–456.
Chen,J.-H., Churchill,M.E.A., Tullius,T.D., Kallenbach,N.R. and Seeman,N.R. (1988) *Biochemistry*, **27**, 6032–6038.
Clegg,R.M., Murchie,A.I.H., Zechel,A., Carlberg,C., Diekmann,S. and Lilley,D.M.J. (1992) *Biochemistry*, **31**, 4846–4856.
Deb,S., DeLucia,A.L., Koff,A., Tsui,S. and Tegtmeyer,P. (1986) *Mol. Cell. Biol.*, **6**, 4578–4584.
Dixon,W.J., Hayes,J.J., Levin,J.R., Weidner,M.F., Dombroski,B.A. and Tullius,T.D. (1991) *Methods Enzymol.*, **208**, 380–413.
Duckett,D.R. and Lilley,D.M.J. (1991) *J. Mol. Biol.*, **221**, 147–161.
Duckett,D.R., Murchie,A.I.H., Diekmann,S., von Kitzing,E., Kemper,B. and Lilley,D.M.J. (1988) *Cell*, **55**, 79–89.
Duckett,D.R., Murchie,A.I.H. and Lilley,D.M.J. (1990) *EMBO J.*, **9**, 583–590.
Einck,L. and Bustin,M. (1985) *Exp. Cell Res.*, **156**, 295–310.
Frappier,L., Price,G.B., Martin,R.G. and Zannis-Hadjopoulos,M. (1987) *J. Mol. Biol.*, **193**, 751–758.
Frappier,L., Price,G.B., Martin,R.G. and Zannis-Hadjopoulos,M. (1989) *J. Biol. Chem.*, **264**, 334–341.
Germann,M.W., Kalisch,B.W., Lundberg,P., Vogel,H.J. and van de Sande,J. (1990) *Nucleic Acids Res.*, **18**, 1489–1498
Hamada,H. and Bustin,M. (1985) *Biochemistry*, **24**, 1428–1433.
Hand,R. (1978) *Cell*, **15**, 317–325.
Hayes,J.J. and Tullius,T.D. (1989) *Biochemistry*, **28**, 9521–9527.
Klein,H.L. and Welch,S.K. (1980) *Nucleic Acids Res.*, **8**, 4651–4669.
Lilley,D.M.J. and Clegg,R.M. (1993) *Annu. Rev. Biophys. Biomol. Struct.*, **22**, 299–328.
Lin,L.-S. and Meyer,R.J. (1987) *Nucleic Acids Res.*, **15**, 8319–8330.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Muller,U.R. and Fitch,W.M. (1982) *Nature*, **298**, 582–585.
Murchie,A.I.H., Clegg,R.M., von Kitzing,E., Duckett,D.R., Diekmann, S. and Lilley, D.M.J. (1989) *Nature*, **341**, 763–766.
Murchie,A.I.H., Portugal,J. and Lilley,D.M.J. (1991) *EMBO J.*, **10**, 713–718.
Nobile,C. and Martin,R.G. (1986) *Intervirology* **25**, 158–171.
Noirot,P., Bargonetti,J. and Novick,R.P. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8560–8564.
Ofverstedt,L.-G., Hammarstrom,K., Bagobin,N., Hjeten,S., Pettersson,U. and Chattopadhyaya,J. (1984) *Biochim. Biophys. Acta*, **782**, 120–126.
Parsons,C.A. and West,S. (1990) *Nucleic Acids Res.*, **18**, 4377–4384.
Parsons,C.A., Kemper,B. and West,S.C. (1990) *J. Biol. Chem.*, **265**, 9285–9289.
Pearson,C.E., Frappier,L. and Zannis-Hadjopoulos,M. (1991) *Biochim. Biophys. Acta*, **1090**, 156–166.
Pearson,C.E., Ruiz,M.T., Price,G.B., and Zannis-Hadjopoulos,M. (1994) *Biochemistry*, **33**, 14 185–14 196.
Platt,J.R. (1955) *Proc. Natl Acad. Sci. USA*, **41**, 181–183.
Pogue,G.P. and Hall,T.C. (1992) *J. Virol.*, **66**, 674–684.

Price,M.A. and Tullius,T.D. (1992) *Methods Enzymol.*, **212**, 194–219.
Prigodich,R.V. and Martin,C.T. (1990) *Biochemistry*, **29**, 8017–8019.
Schreck,R., Zorbas,H., Winnacker,E.-L. and Bauerle,P.A. (1990) *Nucleic Acids Res.*, **18**, 6497–6502.
Tullius,T.D. (1987) *Trends Biochem Sci.*, **12**, 297–300.
Varga-Weisz,P., van Holde,K. and Zlatanova,J. (1993) *J. Biol. Chem.*, **268**, 20699–20700.
Waga,S., Mizuno,S. and Yoshida,M. (1990) *J. Biol. Chem.*, **265**, 19424–19428.
Ward,G.K., McKenzie,R., Zannis-Hadjopoulos,M. and Price,G.B. (1990) *Exp. Cell Res.*, **188**, 235–246.
Ward,G.K., Shihab-El-Deen,A., Zannis-Hadjopoulos,M. and Price,G.B. (1991) *Exp. Cell Res.*, **195**, 92–98.
Watson,J.D. and Crick,F. (1953) *Nature*, **171**, 737–738.
Zannis-Hadjopoulos,M., Kaufmann,G. and Martin,R.G. (1984) *J. Mol. Biol.*, **179**, 577–586.
Zannis-Hadjopoulos,M., Frappier,L., Khoury,M. and Price,G.B. (1988) *EMBO J.*, **7**, 1837–1844.
Zheng,G., Kochel,T., Hoepfner,R.W., Timmons,S.E. and Sinden,R.R. (1991) *J. Mol. Biol.*, **221**, 107–129.
Zhou,N. and Vogel,H.J. (1993) *Biochemistry*, **32**, 637–645.
Zorbas,H., Rogge,M., Meisterernst,M. and Winnacker,E.-L. (1989) *Nucleic Acids Res.*, **17**, 7735–7748.
Zorbas,H., Rein,T. and Winnacker,E.-L. (1992) *Nucleic Acids Res.*, **18**, 6160.

Received on August 4, 1994; revised on October 28, 1994