The EMBO Journal vol.13 no.13 pp.2994-3002, 1994

Crystal structure of the adenovirus DNA binding protein reveals a hook-on model for cooperative DNA binding

Paul A.Tucker, Demetrius Tsernoglou[1], Alec D.Tucker[2], Frank E.J.Coenjaerts[3], Henk Leenders[3] and Peter C.van der Vliet[3]

European Molecular Biology Laboratory, D-69012 Heidelberg, Germany and [3]Laboratory for Physiological Chemistry, Utrecht University, PO Box 80042, 3508 TA Utrecht, The Netherlands

[1]Present address: Biochemistry Department, Universita 'La Sapienza', Piazzale Aldo Moro 5, I-00185 Rome, Italy

[2]Present address: Zeneca Pharmaceuticals, Alderley Park, Macclesfield, Cheshire SK10 4TF, UK

Communicated by P.C.van der Vliet

The adenovirus single-stranded DNA binding protein (Ad DBP) is a multifunctional protein required, amongst other things, for DNA replication and transcription control. It binds to single- and double-stranded DNA, as well as to RNA, in a sequence-independent manner. Like other single-stranded DNA binding proteins, it binds ssDNA cooperatively. We report the crystal structure, at 2.6 Å resolution, of the nucleic acid binding domain. This domain is active in DNA replication. The protein contains two zinc atoms in different, novel coordinations. The zinc atoms appear to be required for the stability of the protein fold rather than being involved in direct contacts with the DNA. The crystal structure shows that the protein contains a 17 amino acid C-terminal extension which hooks onto a second molecule, thereby forming a protein chain. Deletion of this C-terminal arm reduces cooperativity in DNA binding, suggesting a hook-on model for cooperativity. Based on this structural work and mutant studies, we propose that DBP forms a protein core around which the single-stranded DNA winds.

Key words: adenovirus/cooperativity/crystal structure/DNA binding protein/SSB

Introduction

Single-stranded DNA binding proteins (SSBs) are a class of polypeptides that bind ssDNA without apparent sequence specificity and are usually multifunctional. The prototype of the class is the T4 gene 32 protein (Alberts and Frey, 1970), and many others have been discovered in both prokaryotes and eukaryotes (for a review see Chase and Williams, 1986). In contrast to sequence-specific dsDNA binding proteins, common domains or motifs required for interaction with DNA are not evident. Despite this, there are some similarities between SSBs in that single-stranded DNA is often bound in a cooperative fashion and in many cases the proteins bind to RNA and to double-stranded DNA. In addition, there is evidence that similar structural changes can be induced in DNA by complex formation with various SSBs (Van Amerongen *et al.*, 1987), and functional exchange is sometimes possible. As an example, during SV40 DNA replication the role of replication protein A in T antigen-mediated unwinding can also be performed by *Escherichia coli* SSB and adenovirus single-stranded DNA binding protein (Ad DBP) (Wold and Kelly, 1988). However, despite intensive investigation of their functions and their nucleic acid binding properties, structural information on SSBs is still largely lacking.

Ad DBP belongs to this class of proteins. The product of the viral E2A gene, it is a nuclear phosphoprotein that accumulates to high levels, up to 2×10^7 molecules/cell, in infected cells (Van der Vliet and Levine, 1973). The adenovirus 5 (Ad5) protein contains 529 amino acids (Kruijer *et al.*, 1982) and can, by limited chymotrypsin digestion (Tsernoglou *et al.*, 1985), be divided into two domains. The N-terminal domain (1-173) is only weakly conserved and is extensively phosphorylated (Klein *et al.*, 1979; Linne and Philipson, 1980). It contains the nuclear location signal (Morin *et al.*, 1989) and is involved in determining the host range (Klessig and Grodzicker, 1979). The non-phosphorylated C-terminal domain (174-529) is highly conserved among the various serotypes (Neale and Kitchingman, 1990) and contains the nucleic acid binding properties. This domain is functional in DNA replication in a manner similar to the intact molecule (Tsernoglou *et al.*, 1985). As shown by a ^{65}Zn blotting technique, this domain contains at least one zinc atom in the region encompassing amino acids 273-286 (Eagle and Klessig, 1992), whereas zinc was also shown to be required during *in vitro* synthesis of DBP to obtain functional protein capable of DNA binding (Vos *et al.*, 1988a).

Analysis of mutants, as well as *in vitro* studies, have revealed the multifunctional character of Ad DBP. The protein is involved in DNA replication, transcription control, transformation, virus assembly and possibly recombination, whilst it also helps the replication of adeno-associated viruses. Its function in DNA replication has been studied in detail using a reconstitution system. Ad DBP is absolutely required during chain elongation. It protects the single strand which originates from displacement synthesis from nuclease attack. In addition, it increases the rate of synthesis and processivity of the viral DNA polymerase and modifies its sensitivity to nucleotide analogues (Lindenbaum *et al.*, 1986; Mul *et al.*, 1989). DBP also has helix destabilizing properties that could be important for DNA unwinding during chain elongation (Zijderveld and Van der Vliet, 1994). Ad DBP also enhances initiation of replication by lowering the K_m of formation of a pTP-dCMP initiation complex (Mul and Van der Vliet, 1993) and by increasing the binding of NFI to its recognition site in the auxiliary origin (Cleat and Hay, 1989; Stuiver and Van der Vliet, 1990). Finally, Ad DBP enhances the rate of renaturation of complementary strands originating from displacement synthesis (Zijderveld *et al.*, 1993).

Table I. Data collection and phasing summary

	Native	$KAu(CN)_2$	$Pb(NO_3)_2$	K_2PtCl_4
λ (Å)	1.009	1.009	1.5419	1.5419
Maximum resolution (Å)	2.6	2.6	2.7	3.2
Observations	31 209	32 905	18 942	17 703
Unique reflections	10 432	10 686	10 473	5022
Percentage completion (%)	84	85	91	93
R (%)	6.5	5.1	7.6	5.6
Heavy-atom sites		1	1	1
Phasing power		3.02	0.98 (3.5 Å)	1.40
Cullis R factor (%)		51.2	74.3 (3.5 Å)	69.0

Merging $R = \Sigma_{h,k,l} \sum_{i=1}^{n} |I(h,k,l)_i - \langle I(h,k,l) \rangle| / \Sigma_{h,k,l} \sum_{i=1}^{n} I(h,k,l)_i$. Cullis $R = \Sigma_{h,k,l} |F_{h,o}(h,k,l) - F_{h,c}(h,k,l)| / \Sigma_{h,k,l} F_{h,o}(h,k,l)$ for centric reflections only. Overall phasing power $= [\Sigma_{h,k,l} F_h^2(h,k,l) / \Sigma_{h,k,l} e_h^2(h,k,l)]^{1/2}$. h, k and l are the Miller indices defining crystal planes.

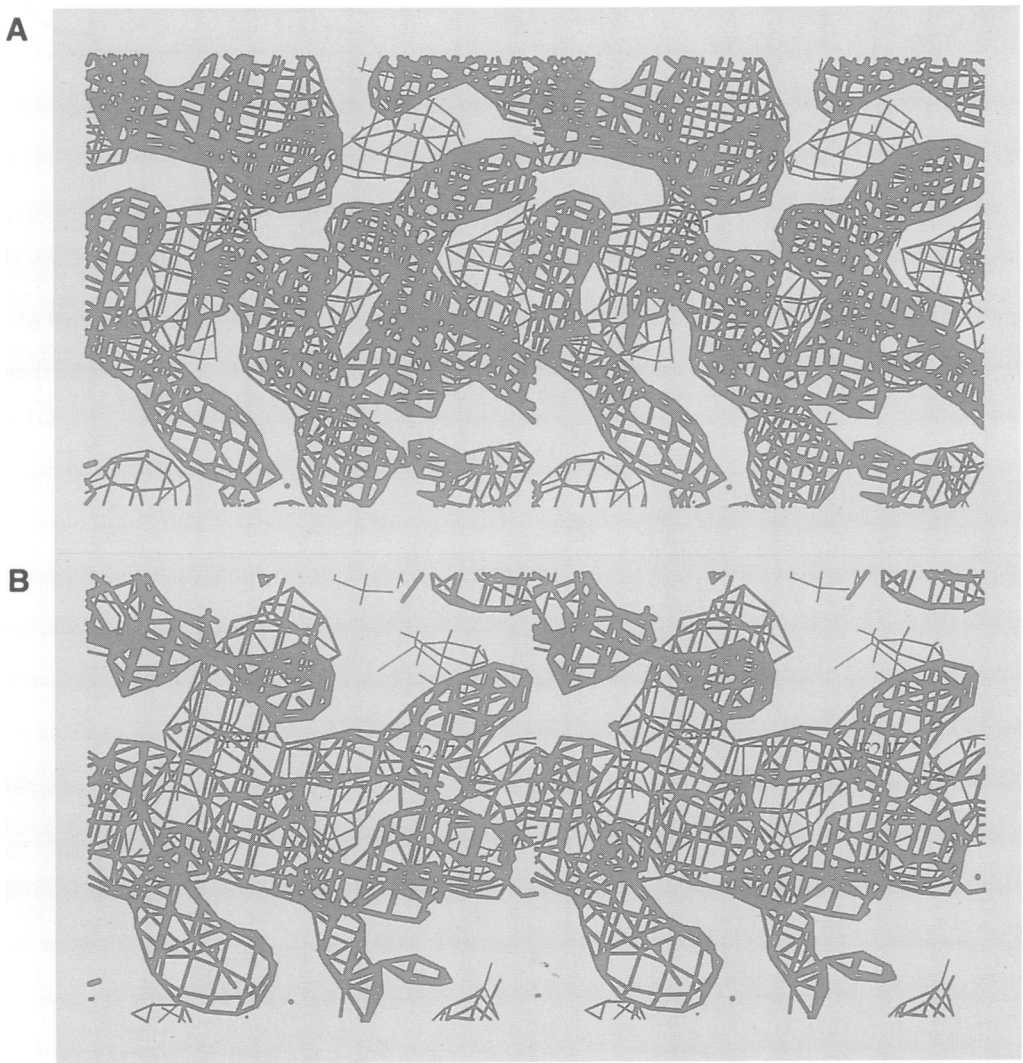

Fig. 1. Stereo diagrams produced using the program 'O' (Jones *et al.*, 1991) of (**A**) the current $2F_o - F_c$ electron density map in the region of helix D and (**B**) the initial solvent flattened map in the same region. Contours for the maps are at 1.2 times the r.m.s. electron density level.

It is likely that these various functions are influenced by the nucleic acid structure induced upon binding to Ad DBP. The structure of Ad DBP in complex with single-stranded DNA has been studied by electron microscopy (EM), sedimentation analysis, CD and optical density measurements (Van der Vliet *et al.*, 1978; Van Amerongen *et al.*, 1987). The Ad DBP site has been determined to be 10–15 nt and the complex is reported to have a regular structure in which the DNA is thought to be extended (Van Amerongen *et al.*, 1987). A cooperativity constant $\omega = 20-30$ was calculated from binding to poly(rA) (Kuil *et al.*, 1989), and binding to single-stranded DNA is also cooperative (Van der Vliet

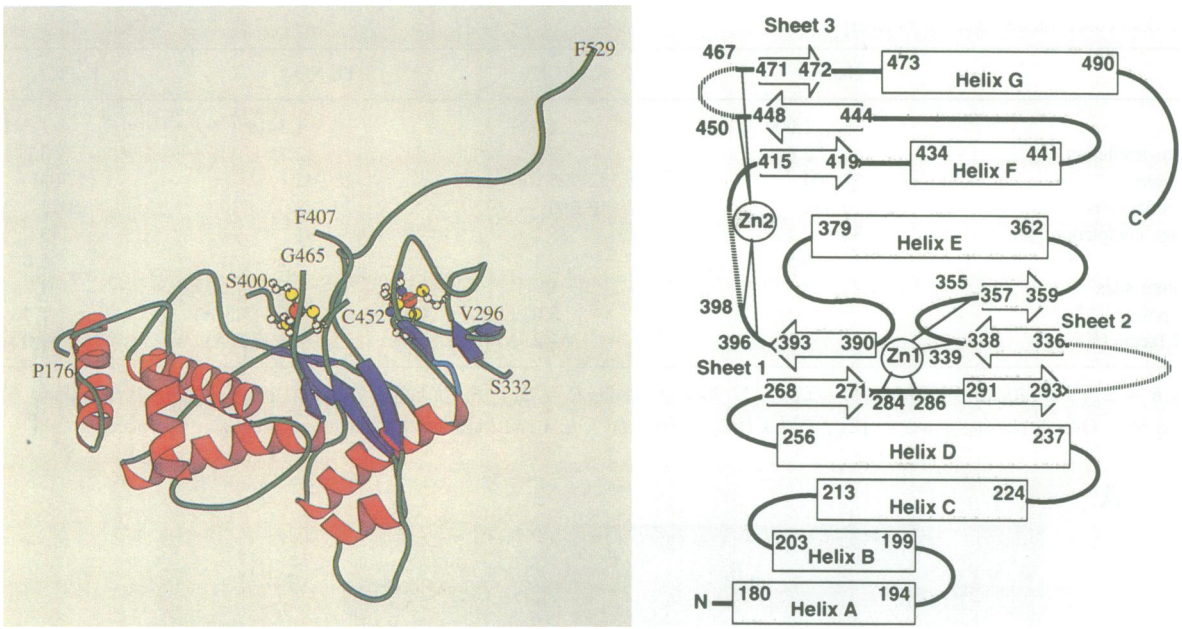

Fig. 2. (**A**) A representation of the Ad5 DBP nucleic acid binding domain (176–529) using the program Molscript (Kraulis, 1991). The two zinc coordination sites are shown separately, the CHCC site to the right and the CCCC site to the left. Helices A–D and G are on the left and helices E and F to the right. To the front is the β-sheet 3 that seems certain to be involved in interaction with DNA. The cleft between helices G and E/F is where the straight stretch of the C-terminal arm of the adjacent molecule docks and the cavity into which F527 fits is at the back. To the top right is an amino acid stretch for which no electron density is visible. The other two loops with poor electron density, which are omitted from the current model, are at the top left. The N-terminus of the domain is to the left and at the back. (**B**) A topological diagram of the structure showing the position in the sequence of the secondary structure elements together with the residues coordinating zinc. Regions of the polypeptide chain that we are not currently able to model are shown by dashed lines. The topology shows the importance of the zinc ions for structural stability.

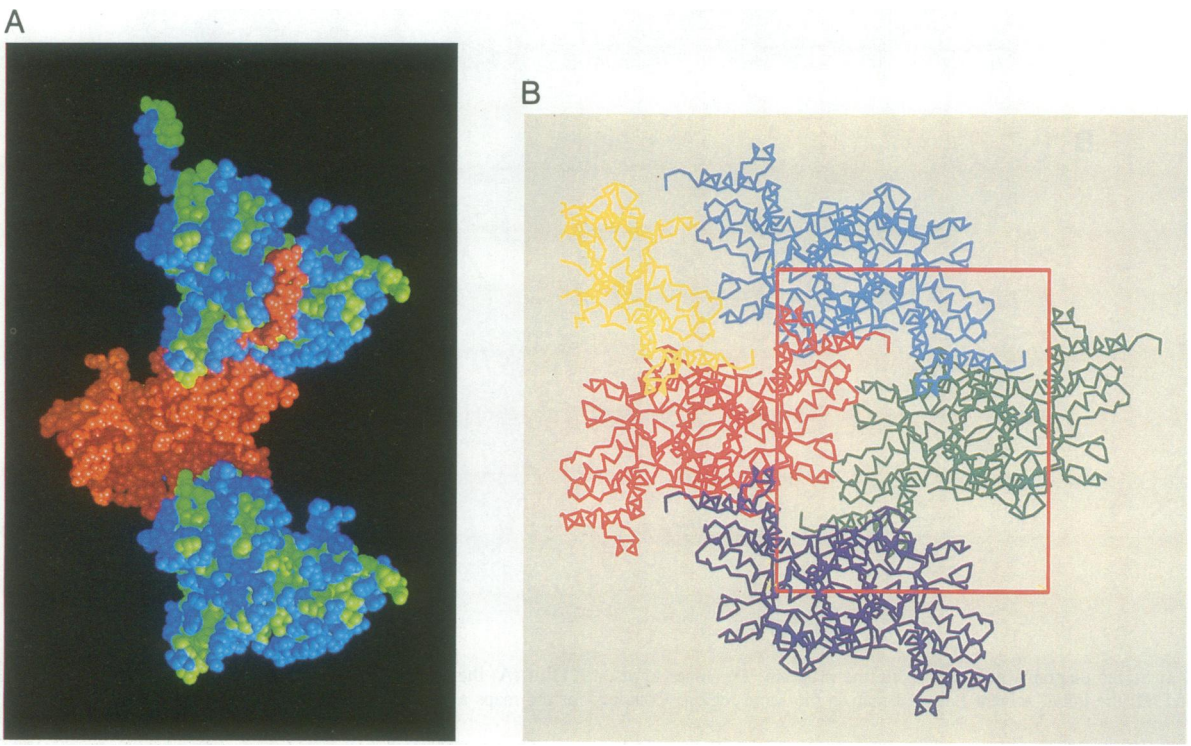

Fig. 3. (**A**) The packing of three molecules along the crystallographic *y*-axis. The central molecule is coloured red and is related to the other two by a crystallographic 2_1 screw axis. The colour code is green for hydrophobic and aromatic residues, and blue otherwise. The C-terminal arm of the upper molecule is at the left. The C-terminal arm of the central molecule crosses the base of the adjacent upper molecule and covers a largely hydrophobic patch. The view is rotated ~180° about a vertical axis relative to that in Figure 2A. (**B**) The packing of the protein chains looking along the crystallographic *y*-axis.

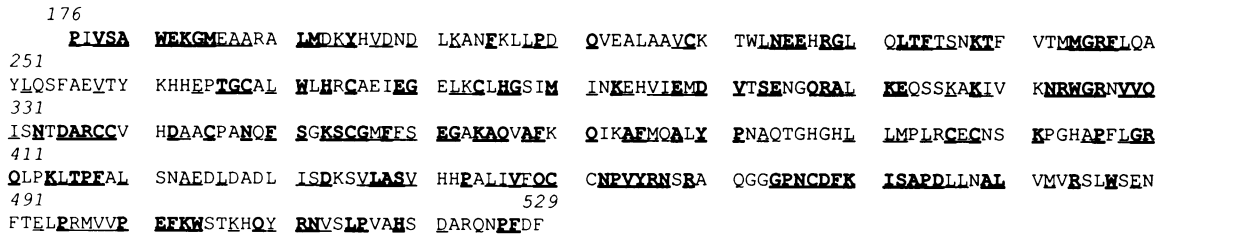

Fig. 4. Amino acid sequence of the crystallized C-terminal Ad5 DBP fragment. In underlined bold are the residues absolutely conserved in the seven adenovirus serotypes sequenced so far (types 2, 4, 5, 7, 12, 40 and 41) (Kitchingman, 1985; Vos *et al.*, 1988b). Underlined are conservative replacements S = T, I = V = L = M = A, F = Y, D = E, K = R and Q = N.

et al., 1978). DBP binds dsDNA in a non-cooperative manner and forms a dynamic multiprotein–DNA complex in which subtle structural changes in DNA can be observed (Stuiver *et al.*, 1992)

We have reported previously the crystallization of the C-terminal fragment which is active in DNA replication and binds ssDNA cooperatively (Tsernoglou *et al.*, 1984). Here we present the 3-D structure and propose a model that explains the cooperative binding.

Results

Solving the crystal structure

X-ray diffraction data were collected and processed as described in Materials and methods and the results are summarized in Table I. The structure of the C-terminal domain of Ad5 DBP was determined by isomorphous replacement and the initial solvent flattened map (Figure 1), although of poor quality, was mostly interpretable. The model was improved and refined as described in Materials and methods.

Structure of the protein

The molecule forms a single domain of approximate size 51×42×27 Å, with a C-terminal extension ~40 Å long (Figure 2A). The topology is shown in Figure 2B. Residues 176–260 form a predominantly α-helical region (helices A–D) against which helix G packs. Above the largely α-helical base (Figure 2A) there are three small regions of β-sheet. The polypeptide chain around the C-terminal extension has no well-defined secondary structure and is apparently dependent for structural stability on two zinc atoms (Figure 2). Indeed, in this region of the molecule there are three sections of polypeptide chain that are either invisible or poorly defined in the electron density map (I297–I331, F401–P406 and P453–G464; Figure 2A).

The individual molecules form an infinite chain along the crystallographic *y*-axis, in which the C-terminal extension of one molecule binds in the cleft between helices G and E of the next molecule in the chain. Adjacent molecules are related by a crystallographic 2_1 screw axis. The molecular organization of the chain is shown in Figure 3A. Importantly, the N-terminus of the fragment points out perpendicular to the direction of the protein chain, as would be necessary to accommodate the N-terminal domain of the intact protein. The packing of these protein chains in the crystal is shown in Figure 3B.

The interactions between the C-terminal arm (residues 513–529) with the adjacent molecule are primarily hydrophobic in nature. Residue F527, conserved within the seven known adenovirus serotypes sequenced (Figure 4), sits in a hydrophobic depression (Figure 5) formed by L270, M376, L391, V499, L477 and L480, much like a hook. Conserved residue P516 packs close to conserved residues P474 and F375, and conserved residue P526 packs against conserved residue Y380. Further hydrophobic contacts are between V513 and L515 and hydrophobic residues L420, A423, L426, V436 and V440. All these residues are conserved or conservatively replaced in other serotypes (Figure 4). There are few hydrogen bonding interactions between the C-terminal arm and the adjacent molecule. The most important are from the conserved R484 to the main chain carbonyl of P526 and the side chain of D521. Others are between R496 and the C-terminal carboxylate, and between the main-chain nitrogen of Q384 and N525. The interactions between molecules involve the C-terminal arm alone. There are no other apparent interactions between the molecules and indeed the residues in loops at the molecule–molecule boundary do not, on average, have lower temperature factors than in other loop regions.

The crystal structure shows the presence of two zinc atoms. One is coordinated by four conserved cysteine residues located at positions 396, 398, 450 and 467 (Figure 4). This zinc atom is loosely bound and may be removed by EDTA soaking of the crystals. Indeed, it is substituted by gold and is the site of the gold atom in the most useful heavy-atom derivative (see Materials and methods). The second zinc atom is coordinated by three cysteine residues (C284, C339 and C355) and one histidine residue (H286), again all conserved. This zinc atom is not readily exchangeable, at least in the crystal.

Consideration of the 10 conserved glycine residues in visible portions of the polypeptide chain is instructive. Only two (G280 and G362) are involved in tight β-turns. Of the remainder, three are adjacent to residues bonding zinc (G267, G287 and G356), and five adjacent to positively charged residues (G184, G229, G245, G325 and G409). At least some of this group of positively charged residues are implicated in interactions with DNA (*vide infra*), and it is tempting to postulate that the adjacent glycine is necessary to allow good interaction with DNA.

Cooperativity of DNA binding of a C-terminal deletion mutant

The structure suggests that the cooperative binding to single-stranded DNA observed previously (Van der Vliet *et al.*, 1978; Kuil *et al.*, 1989) is mainly caused by the C-terminal arm. To test this hypothesis, we expressed a fusion gene between glutathione-S-transferase (GST) and the coding sequences of either the intact C-terminal domain (174–529)

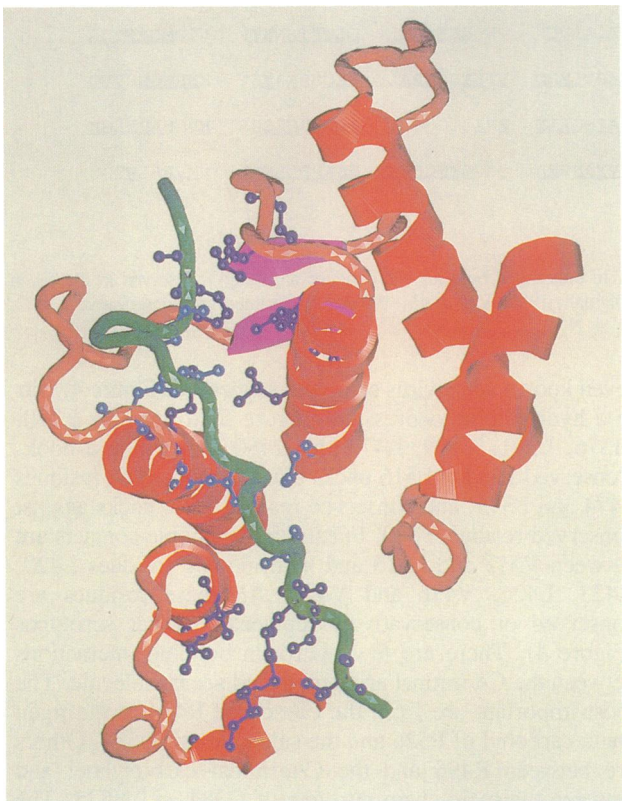

Fig. 5. The interaction region of the C-terminal hook of one molecule, in green, with the base of an adjacent molecule, showing the hook passing between helices E (on the left) and G (on the right). At the top of the figure in purple are F527 of one molecule (the barb of the hook) and L270, M376, L391, L477, L480 and V499 of the second molecule (residues form the hydrophobic depression in which F527 sits). At the bottom of the figure in purple are the residues V513 and L515 of the hook, together with L420, A423, L426, V436 and V440 of the adjacent molecule. Other hydrophobic interactions are shown in light blue. The figure was produced using 'O' (Jones *et al.*, 1991).

or a 17 aa deletion mutant (174–512) lacking the C-terminal arm completely. The fusion proteins were purified by glutathione agarose affinity chromatography and analysed for binding to a 114 base single-stranded DNA probe. At increasing concentrations of the intact C-domain (Figure 6, lanes 2–7), a saturated protein–DNA complex is formed consisting of six to seven DBP molecules per 114 bases, without intermediate complexes, indicating cooperative binding. Such a result was also observed with native DBP (Stuiver and Van der Vliet, 1990). In contrast, with the deletion mutant intermediate complexes are observed consisting of one or several proteins bound to DNA (Figure 6, lanes 8–16). This indicates that the cooperativity of binding is mostly lost, without a reduction of the intrinsic DNA binding capacity to isolated binding sites. Similar results were obtained over a wide range of salt concentrations (not shown).

Discussion

On the basis of the structure and the loss of cooperativity of DNA binding of the C-terminal deletion mutant, it is clear that the protein–protein interaction, involving the C-terminal arm of one molecule and the base of an adjacent molecule, is responsible for the cooperativity of DNA binding. The

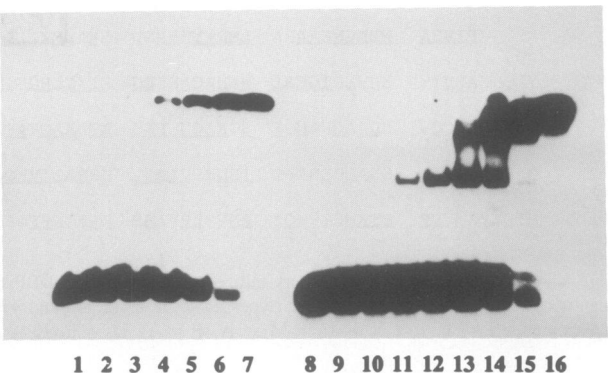

Fig. 6. The C-terminal domain of Ad DBP (lanes 1–7) and the 17 aa C-terminal deletion (lanes 8–16) were incubated for 30 min at 0°C at increasing concentrations with a labelled 114 base single-stranded probe (0.1 ng) containing the Ad2 origin of replication in 20 mM HEPES (pH 7.5), 1 mM EDTA, 1 mM DTT, 100 mM KCl, 0.01% NP-40, 4% Ficoll and 1 μg BSA, and analysed by gel retardation as described previously (Mul *et al.*, 1989). The amounts of DBP were 0, 10, 20, 40, 80, 160 and 320 ng for 173ΔN (lanes 1–7) and 0, 5, 10, 20, 40, 80, 160, 320 and 640 ng for 173ΔNΔC17 (lanes 8–16).

importance of an intact C-terminus is also underscored by the observation that an insertion of 22 aa at the C-terminus destroys viral replication (Vos *et al.*, 1989), whereas a deletion of 59 aa reduces DNA binding considerably (Cleghon and Klessig, 1992). In addition, a 5 aa insertion at R484, which is part of the helix G that the C-terminal extension of the next molecule is packed against, results in a non-viable virus (Vos *et al.*, 1989). The rotational and translational symmetry relating the molecules in the crystal need not be exact in solution because, as stated above, the interactions between the two molecules largely involve the C-terminal arm and there are few other interactions at the protein–protein interface. This would allow a degree of flexibility in the relative orientation of adjacent molecules. In support of protein–protein interactions in the absence of DNA and in solution, we note that dynamic light scattering shows the C-terminal domain in solution to be polymeric.

Zinc ions are required during *in vitro* synthesis of DBP to obtain a protein capable of DNA binding (Vos *et al.*, 1988a), and zinc also protects DBP temperature-sensitive mutants against thermal denaturation (Tsuji and Kitchingman, 1992). The separation in the sequence between the conserved cysteine residues that bind the first zinc ion does not conform to any of the known zinc binding motifs for DNA binding proteins (Berg, 1992). No point mutations have been described for any of these cysteine residues, which makes it difficult to establish a function for zinc binding in this region. Since DNA binding is insensitive to zinc chelators, such as EDTA and copper phenanthroline, we assume that the area around this zinc atom is not directly involved in interaction with DNA, and indeed the zinc atom may be absent in the Ad DBP–ssDNA complex. Using a zinc blotting technique, it was shown that ^{65}Zn binds to the intact Ad DBP but not to a mutant with amino acids 273–286 deleted. This region contains two of the conserved residues that bind the second zinc atom. The deletion also leads to reduced DNA binding (Eagle and Klessig, 1992). Although CHCC motifs for zinc binding have been identified in some retroviral nucleocapsid proteins (Summers *et al.*, 1990; Vallee *et al.*, 1991), the $CXHX_{52}CX_{15}C$ motif found

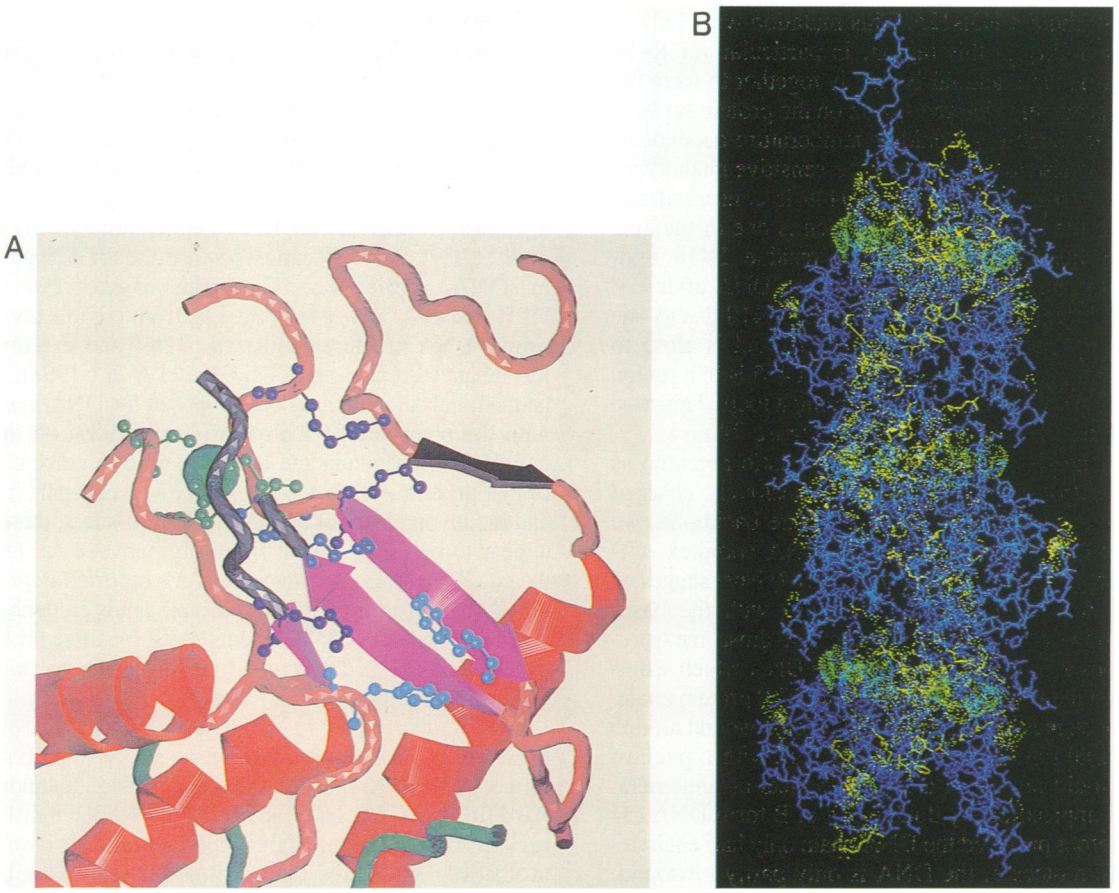

Fig. 7. (**A**) The proposed DNA binding region drawn using the program 'O' (Jones *et al.*, 1991). In purple running top to bottom at the approximate centre of the picture are the conserved residues R410, K414 and K470. The conserved surface residues F234 and F418 (in blue) are to the bottom right. Also shown in green is the CCCC zinc coordination. Towards the top left between the ends of the black rattler is the omitted region P453–G464 which contains Y455. To the top right is the disordered region I297–I331, which is also implicated in interactions with ssDNA. (**B**) The surface of a chain of three molecules coloured yellow at points with the highest positive electrostatic potential. Also coloured in yellow are the conserved positively charged residues, as well as the two conserved surface phenylalanines. In green are shown the ends of the polypeptide chain between which are regions that, although not visible in the electron density, are implicated in interactions with ssDNA. The path of the DNA thus mapped passes from bottom to top across the front of the bottom molecule diagonally from left to right, then around the back of the middle molecule diagonally right to left and finally again across the front of the top molecule diagonally from left to right. The electrostatic potential was calculated and displayed with the MIDAS program (Ferrin *et al.*, 1988) using the default charges file. The regions of highest positive potential are the surface around R410/K414 and K470, with the region between these points being, on average, at higher positive potential than other parts of the molecule.

in Ad DBP is different from any zinc binding domains described previously. This region is structurally important and stabilizes at least three other regions in the molecule. The first two residues, C284 and H286, are located in a conserved region for which several mutations have been described. A substitution of the first C-residue (C284S) shows impaired zinc binding and a decreased affinity for single-stranded DNA (Eagle and Klessig, 1992). Two mutants temperature-sensitive for DNA binding and DNA replication (Tsuji *et al.*, 1991) are located in close proximity. H5ts111 contains a G280V substitution (Prelich and Stillman, 1986). Such a mutation will inhibit the formation of a turn at that position and thus destroy the zinc binding potential. In another mutant Ad2ND2ts23, L282 is substituted by a phenylalanine residue (Kruijer *et al.*, 1982) which would cause bad contacts at this position in the structure and should destabilize the protein. Both mutants show temperature-dependent conformational changes, as indicated by a changed proteolytic breakdown pattern (Tsuji and Kitchingman, 1992). We propose that the zinc atom at this position is essential for the overall structural conformation of DBP. It is unlikely that the region directly contacts DNA, as has been suggested earlier (Eagle and Klessig, 1992).

A detailed description of the way in which single-stranded DNA binds to DBP cannot be given because suitable crystals of the protein–DNA complex have not yet been obtained. However, evidence that the upper edge of β-sheet 3 (the front face of the molecule shown in Figure 2A) is involved in interactions with DNA is compelling, both from structural and biochemical studies. 5 aa insertion mutations at I471 and L415 (Vos *et al.*, 1989) result in non-viable viruses. The region is shown in more detail in Figure 7A. Calculation of the electrostatic potential (Figure 7B) reveals a region of high positive potential that might be expected to interact with the phosphate backbone, approximately at right angles to the β-strand direction. Residues R410, K414 and K470 are all conserved. Point mutation K470T has been shown to have lower DNA binding affinity (Neale and Kitchingman, 1989) and the point mutation F469I reduces adeno-associated virus (AAV) helper function 10-fold (Quinn and Kitchingman, 1986). Also in this region lies P413, which is mutated to serine (Kruijer *et al.*, 1982) in the prototype of the DNA-

negative DBP mutant H5ts125. This mutation would allow increased flexibility in this region, in particular for R410 and K414, which are held fairly rigidly together to form the most highly positively charged region on the protein surface. This increased flexibility would be temperature-dependent, possibly giving rise to the temperature-sensitive phenotype. On the central strand, the residue Q449 is conserved and is adjacent to the conserved C450 which is one of the zinc coordinating ligands. The aromatic residue F418 and, possibly, Y455 are thought to be bound to DNA after UV crosslinking (Cleghon and Klessig, 1992). F418 lies to one side of the line between K414 and K470 and is close to another conserved phenylalanine, F234. Y455 is in a region of poorly defined electron density forming a loop. This loop lies to the other side of the line (see Figure 7A).

Assuming the ridge of positive electrostatic potential as well as the majority of the conserved positively charged residues outlines the trace of the phosphate backbone, we may attempt to describe the protein–DNA complex. CD measurements (Van Amerongen *et al.*, 1987) have suggested that the DNA has an extended structure with the bases unstacked, however, the interpretation of those measurements assumes a regular structure for the DNA which seems unlikely given the irregularities of the Ad DBP protein chain. The structure suggests that the DNA would be wound around the protein chain in a right-handed fashion with a pitch of 76 Å, but would not adopt a regular helical arrangement. This pitch is approximately double that of B form DNA (34 Å), and therefore provided the DNA chain only half encircles one Ad DBP molecule, the DNA is only partly unwound. A shortening of DNA on complexation with DBP by 13% has been estimated from EM measurements (Van der Vliet *et al.*, 1978). A coverage value of 13 bases per Ad DBP molecule would give a shrinkage of 14%, this value being within the range of measurements derived from other techniques. The EM measurements also put a width on the protein–DNA chain of ~200 Å. A DNA–protein complex based on the structure reported here would have a smaller diameter than this, ~110 Å, but this excludes the N-terminal domain that extends outwards from the protein core. We cannot yet be certain as to whether the DNA runs 3′ to 5′ from the base to the C-terminus of the Ad DBP molecule; however we assume that to be the case because to either side of the base of the molecule there are many negatively charged residues that could interact with proteins at the replication fork, whereas the presence of the long, somewhat hydrophobic, C-terminal arm seems less likely to interact favourably with the replication machinery.

Electron density for the region I297–I331 is not visible and that part of the structure may well be flexible. Interestingly, this area is highly conserved. Mutations (R323L and W324L) are located here that show much reduced adeno-associated virus (AAV) helper function (Quinn and Kitchingman, 1986) and reduced affinity for DNA binding (Neale and Kitchingman, 1989). In addition, M299 becomes crosslinked to DNA after UV irradiation (Cleghon and Klessig, 1992). This suggests that this region may fold after DNA binding. It is even possible that the whole domain from I278 to A363 changes conformation on binding to DNA because the majority of contacts between this region and the rest of the molecule are solvent-mediated.

It is tempting to speculate that the formation of a protein chain around which DNA is wound is a general property of SSBs. SSBs are involved in several different processes, such as DNA replication, recombination and repair, and change the DNA structure (Chase and Williams, 1986). Moreover, Ad DBP can substitute for other SSBs under certain circumstances. Ad DBP, *E.coli* SSB and T4 GP32 are all functional in SV40 antigen-mediated unwinding and allow DNA polymerase δ-mediated leading-strand synthesis in a reconstituted SV40 DNA replication system (Kenny *et al.*, 1989). Ad DBP can also substitute for φ29 p5 protein in DNA replication (Gutierrez *et al.*, 1991). The DBP–ssDNA and T4 GP32–ssDNA complexes have a similar DNA structure (Kuil *et al.*, 1989) and are compatible for enhanced renaturation (Zijderveld *et al.*, 1993). A (left-handed) helical wind has been proposed for DNA in complex with the phage fd GP5 protein (Gray, 1989) on the basis of EM measurements and has been used in current models of that protein–ssDNA complex. However, apart from this and the involvement of a three-stranded β-sheet, there would appear to be no structural similarity between Ad DBP and the fd GP5 protein. Neither the work reported here nor the NMR structure for the fd GP5 protein lends credence in any detail to a general ssDNA binding motif spanning prokaryotic and eukaryotic ssDNA binding proteins, as has been proposed (Gutierrez *et al.*, 1991).

ssRNA also binds to the C-terminal domain of DBP, presumably in a similar fashion to ssDNA. Although we propose a three-stranded β-sheet as a nucleotide binding site, Ad DBP does not have an RNP1 or RNP2 motif in this region. Although some of the important residues (two surface aromatics, glutamine and arginine) seem similar, the topology is not, and there is no similarity in structure to the RNA binding domains of small nuclear ribonucleoproteins (Nagai *et al.*, 1990) or cold-shock proteins (Schindelin *et al.*, 1993).

Materials and methods

Crystallization

The C-terminal domain of Ad5 DBP obtained by chymotryptic cleavage of the native protein was prepared and crystallized as described previously (Tsernoglou *et al.*, 1984). The crystals used in the work described here were first transferred to an acetate buffer (pH 5.8) containing 2 M $NaNO_3$.

X-ray data collection

The crystals belong to space group P $2_12_12_1$ with unit cell dimensions a = 79.5, b = 76.0 and c = 67.4 Å. X-ray intensities of the native protein and the $KAu(CN)_2$ derivative (soak conditions 0.4 mM for 22 h) were collected to 2.6 Å resolution at the EMBL Outstation at DESY, Hamburg, on an image plate detector using a wavelength of 1.009 Å and processed with the program MOSFLM. The 31 209 observations yielded 10 432 unique reflections, with a merging R of 0.065 (on intensities). For the Au derivative, 32 905 observations yielded 10 686 unique reflections with a merging R of 0.051, with Bijvoet pairs treated independently (0.079 with Bijvoet pairs merged). $Pb(NO_3)_2$ (soak conditions 0.45 mM for 18 h) and K_2PtCl_4 (soak conditions 0.19 mM for 18 h) derivatives were collected on a Siemens Xentronics area detector with a wavelength of 1.5419 Å and processed with the XENGEN (Howard *et al.*, 1987) package and XDS program (Kabsch, 1988), respectively. The number of observations, unique reflections, merging R and maximum resolution were 18 942, 10 473, 0.076 and 2.7 Å, respectively, for the Pb derivative and 17 703, 5022, 0.056 and 3.2 Å, respectively, for the Pt derivative. For the native data set, image plate observations were combined with 44 969 observations measured on a Siemens Xentronics area detector and were reduced to 12 956 unique reflections with a merging R factor of 0.073 and a completeness of >95% over the resolution range 50.0–2.6 Å.

Solving the crystal structure

Heavy-atom binding sites (each derivative having one site) were located from inspection of anomalous difference Patterson (Au derivative), difference

Patterson and difference Fourier maps and refined using the program PHARE. The Cullis *R* factors were 0.512 (to 3.0 Å), 0.743 (to 3.5 Å) and 0.690 (to 3.2 Å) for the Au, Pb and Pt derivatives, respectively. The phasing analysis gave a mean figure of merit of 0.56 for reflections in the range 50–3 Å. The phasing powers, to the resolution limits given above, of the Au, Pb and Pt derivatives were 3.02, 0.98 and 1.40, respectively, and therefore the initial maps were predominantly phased on the Au derivative using the anomalous signal. Solvent flattening (Wang, 1985; Leslie, 1989) did not significantly improve the quality of the map and indeed the solvent content is low (~40%). Map interpretation was carried out using the program FRODO (Jones, 1978) running on an Evans and Sutherland PS390. Iterations by simulated annealing refinement (X-PLOR) (Brünger, 1992a) and rebuilding against electron density maps (coefficients $2F_o-F_c$ and F_o-F_c) and introduction of data to higher resolution resulted in substantially better maps. Residues I297–I331 were not visible in the electron density maps and are assumed to be disordered. The electron density for residues K401–P406 and P453–G464 is very poor and these residues are omitted from the current model. The model was checked and decisions about whether or not to include a residue were made using free *R* value refinement (Brünger, 1992b) and simulated annealing omit maps. Sixty water molecules were added to the model. Refinement with X-PLOR was terminated with a crystallographic *R* factor of 0.189 for the resolution range 6.0–2.6 Å. The final refinement cycles were performed using TNT (Tronrud *et al.*, 1987). The crystallographic *R* factor for all 12 956 X-ray data in the resolution range 50.0–2.6 Å was 0.197. The r.m.s. deviations from stereochemical ideality are 0.022 Å for bond distances, 3.3° for bond angles, 0.022 Å for planar groups and 19° for torsion angles. Apart from glycine residues, no amino acids other than E294 (close to one of the zinc binding sites) have disallowed backbone conformations according to the Ramachandran analysis (Ramakrishnan and Ramachandran, 1965). With the exception of the refinement and graphics programs, all calculations in the structure determination were made using the CCP4 suite of programs (SERC, Daresbury). A more detailed description of the structure determination will be given elsewhere after further refinement. The current coordinates will be deposited with the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).

Construction of deletion mutant

The intact C-terminal domain of DBP containing amino acids 174–529 (173ΔN) or deletion mutant protein containing amino acids 174–512 (173ΔNΔC17) were expressed as fusions with GST (Smith and Johnson, 1988). For this purpose, the fragments were amplified by PCR using Ad5 DNA as template and appropriate oligonucleotide primers containing *Hin*dIII and *Bam*HI restriction enzyme recognition sites. The primers used were: GGCAC**GGATCC**AGTGTGCCGATCGTG (I) and GGCAC**AAGCTT**T-CAAAAATCAAAGGGG for 173ΔN, and (I) and GGCAC**AAGCTT**T-CAGTTGCGATACTGGTG for 173ΔNΔC17. Amplification was for 30 cycles using Vent DNA polymerase which contains a proofreading function. The fragments were purified by PAGE and cloned into the expression plasmid pRP265 (Verrijzer *et al.*, 1992). Overnight BL21 (DE3) cultures grown at 37°C in the presence of 1 mM $ZnCl_2$ were diluted 1:20, and after 90 min growth at 37°C isopropylthiogalactopyranoside (IPTG) was added to 0.4 mM. Further incubation was at 37°C for 3 h. Bacteria were harvested by centrifugation and resuspended in 20 ml lysis buffer per litre culture. Lysis buffer contained 50 mM Tris–HCl (pH 8.0), 250 mM NaCl, 1 mM EDTA, 1 mM DTT and 0.35 mg of lysozyme per ml. After 20 min at 20°C, cells were lysed by two freeze–thaw steps and ultrasound sonification in the presence of 0.1% NP-40. Insoluble material was removed by centrifugation. The supernatant, containing ~0.1% of the expressed protein, was purified by DEAE ion exchange chromatography followed by affinity chromatography on a 1 ml glutathione–agarose column (Sigma). The column was washed with 150 mM NaCl, 16 mM Na_2HPO_4, 4 mM NaH_2PO_4, 1 mM DTT, 1% Triton X-100 and 10% glycerol and the proteins were eluted with 50 mM Tris–HCl pH 8.0, 10% glycerol, 1 mM DTT, 0.02% NP-40 and 5 mM glutathione. Further purification was by fast flow Q ion exchange chromatography. The final products were >90% pure, as determined by gel electrophoresis and silver-staining.

Acknowledgements

We thank L.Philipson for his continued interest in this project. Thanks are also due to M.H.Stuiver for preparations of Ad5 DBP, to C.P.Verrijzer for advice on bacterial expression, and to K.Wilson and Z.Dauter of the EMBL Hamburg Outstation for their advice and hospitality during the synchrotron experiments. This work was supported in part by The Netherlands Foundation for Chemical Research (SON) with financial support from The Netherlands Organization for Scientific Research (NWO).

References

Alberts,B.M and Frey,L. (1970) *Nature*, **227**, 1313–1315.
Berg,J.M. (1992) *J. Biol. Chem.*, **265**, 6513–6516.
Bernstein,F.C. *et al.* (1977) *J. Mol. Biol.*, **112**, 535–542.
Brünger,A.T. (1992a) X-PLOR, version 3.0. Yale University, New Haven, CT.
Brünger,A.T. (1992b) *Nature*, **355**, 472–475.
Chase,J.W. and Williams,K.R. (1986) *Annu. Rev. Biochem.*, **55**, 103–136.
Cleat,P.H. and Hay,R.T. (1989) *EMBO J.*, **8**, 1841–1848.
Cleghon,V. and Klessig,D.F. (1992) *J. Biol. Chem.*, **267**, 17872–17881.
Eagle,P.A. and Klessig,D.F. (1992) *Virology*, **187**, 777–787.
Ferrin,T.E., Huang,C.C., Jarvis,L.E. and Langridge,R. (1988) *J. Mol. Graph.*, **6**, 13–27.
Gray,C.W. (1989) *J. Mol. Biol.*, **208**, 57–64.
Gutierrez,C., Martin,G., Sogo,J.M. and Salas,M.J. (1991) *Biol. Chem.*, **266**, 2104–2111.
Howard,A.J., Gilliland,G.L., Finzel,B.C., Poulos,T.L., Ohlendorf,D.H. and Salemme,F.R. (1987) *J. Appl. Crystallogr.*, **20**, 383–387.
Jones,T.A. (1978) *J. Appl. Crystallogr.*, **11**, 268–272.
Jones,T.A., Zou,J.-Y, Cowan,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
Kabsch,W. (1988) *J. Appl. Crystallogr.*, **21**, 916–924.
Kenny,M.K., Lee,S.-H. and Hurwitz,J. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9757–9761.
Kitchingman,G.R. (1985) *Virology*, **146**, 90–101.
Klein,H., Maltzman,W. and Levine,A.J. (1979) *J. Biol. Chem.*, **254**, 11051–11060.
Klessig,D.F. and Grodzicker,T. (1979) *Cell*, **17**, 957–966.
Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
Kruijer,W., Van Schaik,F.M.A. and Sussenbach,J.S. (1982) *Nucleic Acids Res.*, **10**, 4493–4500.
Kuil,M.E., Van Amerongen,H., Van der Vliet,P.C. and Van Grondelle,R. (1989) *Biochemistry*, **28**, 9795–9800.
Leslie,A.G.W. (1989) In *Improving Protein Phases*. Proceedings of CCP4 Study Weekend, Daresbury Laboratory, DL/SCI/R26, pp. 25–31.
Lindenbaum,J.O., Field,J. and Hurwitz,J. (1986) *J. Biol. Chem.*, **261**, 10218–10227.
Linne,T. and Philipson,L. (1980) *Eur. J. Biochem.*, **103**, 259–270.
Morin,N., Delsert,C. and Klessig,D.F. (1989) *Mol. Cell. Biol.*, **9**, 4372–4380.
Mul,Y.M. and Van der Vliet,P.C. (1993) *Nucleic Acids Res.*, **21**, 641–647.
Mul,Y.M., Van Miltenburg,R.T., De Clercq,E. and Van der Vliet,P.C. (1989) *Nucleic Acids Res.*, **17**, 8917–8929.
Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) *Nature*, **348**, 515–520.
Neale,G.A.M. and Kitchingman,G.R. (1989) *J. Biol. Chem.*, **264**, 3153–3159.
Neale,G.A.M. and Kitchingman,G.R. (1990) *J.Virol.*, **64**, 630–638.
Prelich,G. and Stillman,B.W. (1986) *J. Virol.*, **57**, 883–892.
Quinn,C.O. and Kitchingman,G.R. (1986) *J. Virol.*, **60**, 653–661.
Ramakrishnan,C. and Ramachandran,G.N. (1965) *Biophys. J.*, **5**, 909–933.
Schindelin,H., Marahiel,M.A. and Heinemann,U. (1993) *Nature*, **364**, 164–168.
Smith,D.B. and Johnson,K.S. (1988) *Gene*, **67**, 31–40.
Stuiver,M.H. and Van der Vliet,P.C. (1990) *J. Virol.*, **64**, 379–386.
Stuiver,M.H., Bergsma,W.G., Arnberg,A.C., Van Amerongen,H., Van Grondelle,R. and Van der Vliet,P.C. (1992) *J. Mol. Biol.*, **225**, 999–1011.
Summers,M.F., South,T.L., Kim,B. and Hare,D.R. (1990) *Biochemistry*, **29**, 329–340.
Tronrud,D.E., Ten Eyck,L.F. and Matthews,B.W. (1987) *Acta Crystallogr.*, **A43**, 489–501.
Tsernoglou,D., Tucker,A.D. and Van der Vliet,P.C. (1984) *J. Mol. Biol.*, **172**, 237–239.
Tsernoglou,D., Tsugita,A., Tucker,A.D. and Van der Vliet,P.C. (1985) *FEBS Lett.*, **188**, 248–252.
Tsuji,M. and Kitchingman,G.R. (1992) *J. Virol.*, **66**, 480–488.
Tsuji,M., Van der Vliet,P.C. and Kitchingman,G.R. (1991) *J. Biol. Chem.*, **266**, 16178–16187.
Vallee,B.L., Coleman,J.E. and Auld,D.S. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 999–1003.
Van Amerongen,H., Van Grondelle,R. and Van der Vliet,P.C. (1987) *Biochemistry*, **26**, 4646–4652.
Van der Vliet,P.C. and Levine,A.J. (1973) *Nature New Biol.*, **246**, 170–174.
Van der Vliet,P.C., Keegstra,W. and Jansz,H.S. (1978) *Eur. J. Biochem.*,

86, 389–398.
Verrijzer,C.P., Van Oosterhout,J.A.W.M. and Van der Vliet,P.C. (1992) *Mol. Cell. Biol.*, **12**, 542–551.
Vos,H.L., Van der Lee,F.M. and Sussenbach,J.S. (1988a) *FEBS Lett.*, **239**, 251–254.
Vos,H.L., Van der Lee,F.M., Reemst,A.M.C.B., Van Loon,A.E. and Sussenbach,J.S. (1988b) *Virology*, **163**, 1–10.
Vos,H.L., Brough,D.E., Van der Lee,F.M., Hoeben,R.C., Verheijden,G.F.M., Dooijes,D., Klessig,D.F. and Sussenbach,J.S. (1989) *Virology*, **172**, 634–642.
Wang,B.C. (1985) *Methods Enzymol.*, **115**, 90–112.
Wold,M.S. and Kelly,T. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 2523–2527.
Zijderveld,D.C. and Van der Vliet,P.C. (1994) *J. Virol.*, **68**, 1158–1164.
Zijderveld,D.C., Stuiver,M.H. and Van der Vliet,P.C. (1993) *Nucleic Acids Res.*, **11**, 2591–2598.

Received on March 3, 1994