Structural Basis of DNA Bridging by Barrier-to-Autointegration Factor‡

Timothy C. Umland, Shui-Qing Wei,§ Robert Craigie, and David R. Davies*

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

Received March 13, 2000; Revised Manuscript Received May 17, 2000

ABSTRACT: Barrier-to-autointegration factor (BAF) is a host cell protein that plays a crucial role in retroviral integration. Preintegration complexes (PICs) stripped of BAF lose their normal integration activity, which can be restored by incubation with purified BAF. BAF bridges double-stranded DNA both intra- and intermolecularly in a non-sequence-specific manner, leading to the formation of a nucleoprotein network. BAF also binds to the nuclear protein lamina-associated polypeptide 2 (LAP2), and is localized with chromatin during interphase and mitosis. The crystal structure of homodimeric human BAF has been determined to 1.9 Å resolution. The fold of the BAF monomer resembles that of the second domain of RuvA. This comparison revealed the presence of the helix−hairpin−helix (HhH) nonspecific DNA binding motif within BAF. A novel feature of BAF's HhH motif is the occupation of the metal binding site by the ϵ-amino group of Lys 6, providing an alternative means of sequestering positive charge. Mutational analysis corroborates the HhH motif's prominent role in DNA binding and argues against a previously proposed helix−turn−helix (HTH) binding site located in another region of the monomer. A model of BAF bridging DNA via the HhH motif is proposed.

A crucial process in the life cycle of retroviruses is the integration of a DNA copy of the viral genome into the host cell's genome (*1*). Integration is mediated in vivo by the preintegration complex (PIC)[1] (*2*, *3*). This is a large nucleoprotein complex derived in part from the core of the infecting virion. It has not been thoroughly characterized, but it is known to contain viral proteins, including integrase, viral DNA, and cellular proteins (*4*−*7*). PICs isolated from cells infected by Moloney murine leukemia virus (Mo-MLV) (*8*) or by human immunodeficiency virus type 1 (HIV-1) (*9*, *10*) exhibit a strong preference for intermolecular integration of linear viral DNA into a target DNA molecule, and not for suicidal intramolecular integration, termed autointegration, of viral DNA into itself. The host cell protein BAF was first identified by its ability to protect Mo-MLV PICs against autointegration (*7*, *11*), and has been shown to play a similar role in promoting intermolecular integration activity of HIV-1 PICs (*12*).

Human BAF is a homodimeric protein, with each 10.1 kDa subunit composed of 89 amino acid residues. BAF does not exhibit any significant degree of sequence similarity to other known proteins. However, many species express a transcript that can encode a protein that is highly similar (*11*). BAF binds double-stranded but not single-stranded DNA in a highly base-nonspecific manner (*11*; R. Zheng and R. Craigie, personal communication). It is present in both cytoplasmic and nuclear fractions of cell extracts. BAF possesses the remarkable property of being able to bridge together multiple segments of DNA. This bridging activity has been proposed to be responsible for BAF's ability to protect viral DNA from autointegration within the preintegration complex (*11*). The endogenous function of BAF within uninfected cells is unknown. However, it appears to be vital. Inhibitory RNA experiments in *Caenorhabditis elegans* have shown that embryonic development is arrested in the absence of BAF (M. S. Lee, R. Cragie, and M. Krause, personal communication). Additionally, BAF binds lamina-associated polypeptide (LAP) 2, a lamin and chromatin binding nuclear protein, and has been suggested as a mediator of LAP2−chromosome interaction at the end of mitosis (*13*, *14*). Immunofluorescence staining demonstrated that BAF is preferentially localized to the nucleus in interphase cells and to chromosomes during mitosis (*13*).

The crystal structure of BAF, determined at 1.9 Å resolution using phases determined by a multiwavelength anomalous dispersion (MAD) experiment, is reported here. This structure strongly suggests a model by which BAF nonspecifically binds double-stranded DNA using a previously unrecognized helix−hairpin−helix (HhH) motif. This is a nonspecific DNA binding motif in which binding occurs through interactions between amide groups of the peptide chain and phosphate groups of the DNA backbone (*15*), and at least in certain cases a metal-mediated interaction (*16*). This motif has been predicted, on the basis of sequence, to exist in many proteins (*15*, *17*−*19*). It has been unambiguously observed in the crystal structures of the three DNA glycosylases, endonuclease III (*15*), AlkA (*20*, *21*), and MutY (*22*), as well as in the Holliday junction binding protein RuvA

‡ Accession numbers for the coordinates and structure factors deposited in the Protein Data Bank are 1ci4 and r1ci4sf, respectively.

* Corresponding author. Telephone: (301) 496-4295. Fax: (301) 496-0201. E-mail: drd@vger.niddk.nih.gov.

§ Present address: Schering-Plough Research Institute, Kenilworth, NJ 07033.

[1] Abbreviations: PIC, preintegration complex; BAF, barrier-to-autointegration factor; Mo-MLV, Moloney murine leukemia virus; HIV-1, human immunodeficiency virus type 1; HhH, helix−hairpin−helix; HTH, helix−turn−helix; MAD, multiwavelength anomalous dispersion; LAP2, lamina-associated polypeptide 2; rmsd, root-mean-square deviation; ncs, noncrystallographic symmetry.

10.1021/bi000572w
Published on Web 07/08/2000

(*23*−*26*) and the DNA gap-filling enzyme polymerase β (*18*, *27*, *28*). However, except for this motif, the overall structure of dimeric BAF is dissimilar from that of these five proteins. Both the crystal structure and mutational analysis strongly suggest a model of BAF's mode of binding and bridging double-stranded DNA, and this is presented here.

Recently, the solution structure of BAF was reported (*29*). While the BAF crystal and solution structures are similar, particularly on the level of the main chain fold of the monomer, differences do exist which may be significant. For example, the dimer interfaces are not identical, with the crystal structure exhibiting a tighter and more extensive interface. In addition, the positions of several potential important side chains at the dimer interface and at the putative DNA binding site differ between the two structures. Furthermore, the putative DNA binding site predicted in this work and in the NMR study (*29*) involves distinctly different regions of the protein.

MATERIALS AND METHODS

Protein Purification. Selenomethionine human BAF containing a hexahistidine tag was synthesized in the *Escherichia coli* methionine autotroph B834(DE3)LysS (Novagen) using a previously described construct (*11*). Cells were grown at 37 °C using a defined medium containing 1% dextrose (w/v) as the carbon source and with 40 mg/mL selenomethionine. BAF was purified as described previously (*11*), with modification. Following the nickel chelate affinity column run under denaturing conditions, BAF was refolded by dialysis against 50 mM K_2PO_4 (pH 6.5), 200 mM NaCl, 10 mM EDTA, and 15 mM 2-mercaptoethanol. The histidine tag was cleaved using thrombin (20 units/mg of BAF for 75 min at room temperature). The thrombin was removed using a benzamidine Sepharose column. BAF was further purified using a Superdex 200 (20/60) column (Pharmacia), with a buffer composed of 20 mM Tris-HCl (pH 7.0), 10% (w/v) glycerol, 150 mM NaCl, 10 mM DTT, and 0.1 mM EDTA. The selenomethionine BAF was then concentrated to 8.3 mg/mL using Centriprep 10 concentrators (Amicon) at room temperature.

Crystallization. BAF was crystallized by microdialysis against 20 mM imidazole (pH 6.5), 80 mM NaCl, and 10 mM DTT at room temperature. Crystals were serially transferred into a final cryoprotectant solution of 30% (w/v) glycerol, 20 mM imidazole (pH 6.5), 50 mM NaCl, and 10 mM DTT, with increases in the glycerol concentration by increments of 5% (w/v). Cryoprotected crystals were plunged into liquid propane to flash-freeze them.

Data Collection. MAD data were collected at Beamline X-9B of the National Synchrotron Light Source (NSLS) at 95 K, and recorded on a MAR 345 imaging plate. Data were collected employing monochromatic X-rays at three separate wavelengths experimentally selected to optimize the differences in both the real and imaginary parts of the Se scattering factor (Table 1). The crystal was tetragonal, in space group $P4_32_12$, with the following cell parameters: $a = b = 41.8$ Å and $c = 214.6$ Å. The asymmetric unit was estimated to contain two BAF monomers, using a Matthews coefficient of 2.30 Å^3/Da.

Data Processing and Refinement. Data were processed using HKL (*30*). Three of the four Se atoms in the asymmetric unit were located using SOLVE (*31*). The Se positions were refined using maximum-likelihood phase refinement, and the resultant phases were modified using solvent flattening. These pseudo-SIRAS calculations were undertaken in PHASES-95 (*32*). Table 1 contains details of the data processing and refinement results. This yielded an electron density map of excellent quality (Figure 1). The protein model was built into the density using O (*33*). Refinement was conducted using CNS version 0.5 (*34*), employing torsion angle dynamics simulated annealing with a maximum likelihood target function. Bulk solvent correction and an anisotropic temperature factor correction were employed. All data where $I/\sigma \geq 0$ were utilized. Prior to refinement, 10% of the data were randomly selected as a *test* data set, to be used for calculating R_{free}, while the model was refined against the remaining 90% of the data. R_{free} was used throughout the refinement to optimize the refinement scheme and to prevent overfitting of the data. The final model included the use of weak noncrystallographic symmetry (ncs) restraints (ncs-related main chain atoms, 20 kcal mol^{-1} Å^{-2}; ncs-related side chain atoms, 10 kcal mol^{-1} Å^{-2}; non-ncs-related main chain atoms, 5 kcal mol^{-1} Å^{-2}; and no restraints on side chain atoms judged not related by ncs) and restrained individual temperature factors. Water molecules were added to the model at the later stages of refinement. The final

Table 1: Data Collection and Refinement Statistics

	λ_1	λ_2	λ_3
wavelength (Å)	0.9793	0.9789	0.9686
resolution (Å)	40.0−1.9	40.0−1.9	40.0−1.9
no. of observations	215877	216434	219206
no. of unique reflections	15938	15976	16011
completeness (%)[a]	99.1(94.0)	99.4(97.3)	99.5(100.0)
R_{sym} (%)[a,b]	6.8(24.0)	7.3(26.9)	6.1(18.4)
phasing ($I/\sigma(I) \geq 1.0$)			
R_{Cullis} (%)[c]	-	-	54.8
R_{Kraut} (%)[d]	-	2.0	1.9
phasing power[e]	-	4.20	1.97
⟨FOM⟩	0.61 for 15 311 phased reflections		

Refinement Statistics	
no. of reflections ($I/\sigma(I) \geq 0.0$)	
working	14 112
test	1588
R (R_{free})[f]	0.214 (0.265)
no. of protein atoms	1389
no. of waters	228
⟨B-factor⟩ (Å^2)	
monomer A	23.1
monomer B	21.5
water	33.6
rms deviation from ideality	
bond lengths (Å)	0.0099
bond angles (deg)	1.37°

[a] Values in parentheses refer to statistics for data in the 1.94−1.90 Å resolution shell. [b] $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$. [c] $R_{Cullis} = \Sigma||FPH_O| \pm |FP_O|| - |FH_C||/\Sigma||FPH_O| \pm |FP_O||$ for centric reflections. [d] $R_{Kraut} = \Sigma||FPH_O| - |FPH_C||/\Sigma|FPH_O|$ for acentric reflections, isomorphous case. $R_{Kraut} = \Sigma||FPH_O^+| - |FPH_C^+|| + ||FPH_O^-| - |FPH_C^-||/\Sigma(|FPH_O^+| + |FPH_O^-|)$ for acentric reflections, anomalous case, where FP is the protein, FPH the derivative, and FH the heavy atom structure factor, respectively. FPH^+ and FPH^- denote Bijvoet mates. [e] Phasing power, FH_C/E for the isomorphous case and $2FH_{C''}/E$ for the anomalous case, where E is the rms lack of closure. [f] $R = \Sigma|FP_O - FP_C|/\Sigma FP_O$. R_{free} is the same as R but computed using a randomly selected 10% of the data which was excluded throughout the refinement.

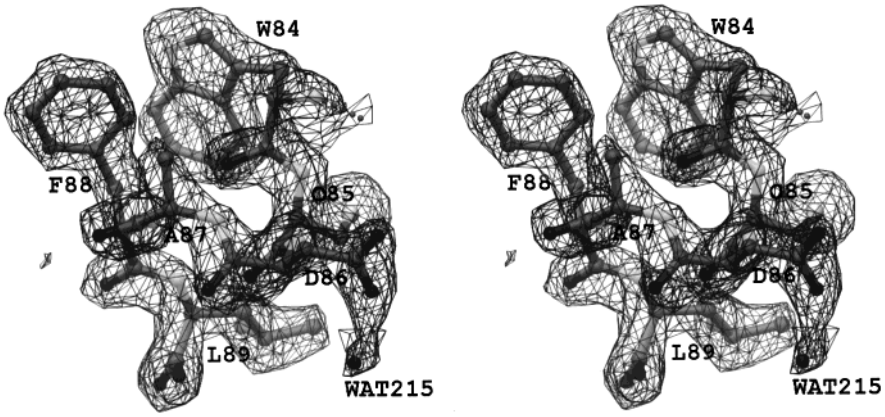

FIGURE 1: Experimental electron density map, calculated using phases derived from MAD data to 1.9 Å resolution followed by solvent flattening. The map is contoured at 1.2σ. Residues Trp 85−Leu 89 of monomer B of the final refined structure are displayed.

protein model was evaluated using PROCHECK (*35*). Figures were created using RIBBONS (*36*) and GRASP (*37*).

Site-Directed Mutagenesis. Residues of interest in human BAF were selectively mutated using the QuikChange Site-Directed Mutagenesis Kit (Strategene). Mutant proteins were synthesized in *E. coli* and purified in a manner similar to that described above, but on a reduced scale. The BAF mutants were analyzed for DNA binding using a gel-shift assay.

RESULTS AND DISCUSSION

Description of the Monomer Structure. The crystal structure of selenomethionine BAF was determined as described in Materials and Methods. The 1.9 Å resolution MAD data produced an experimentally phased electron density map of exceptional quality (Figure 1). BAF is a homodimer within the crystallographic asymmetric unit (Figure 2A), in agreement with the observation of the dimeric state in solution from NMR (*29*), gel filtration, and dynamic light scattering experiments (data not shown). The two monomers composing this dimer were restrained by weak noncrystallographic symmetry restraints during refinement. The monomers will be termed monomer A and monomer B when features pertinent to only one are being discussed. To denote that a particular residue belongs to a second monomer, a prime (′) will be added to its name. Each monomer contains five α-helices (α_1, Gln 5−Val 11; α_2, Glu 28−Glu 36; α_3, Ala 42−Val 51; α_4, Glu 56−Cys 67; and α_5, Ala 71−Phe 88) along with a single turn of 3_{10}-helix (Val 20−Ser 22). The monomer itself forms a compact globular structure with the helices packed about a hydrophobic core. The exceptions to this are the first two residues proximal to the N-terminus, which are highly solvent exposed. No electron density was observed for the N-terminal selenomethionine of monomer B. The BAF monomer contains four cysteine residues, all of which are buried within the hydrophobic core. It is impossible for any of these cysteines to interact with one another to form disulfide bonds, either intra- or intermolecularly, without significant unfolding of the protein.

Description of the Dimer Structure. BAF displays an extensive dimer interface (Figure 2B). The core of this interface is composed of hydrophobic residues belonging to helix α_3. This helix contains a central notch created by the inclusion of a glycine residue (Gly 47), which allows the α_3 helices from each of the two opposing monomers to cross while simultaneously packing tightly together. Also contributing to the dimer interface are residues belonging to the loops immediately preceding and following helix α_3 and to the loop preceding the single turn of 3_{10}-helix and the C-terminal residue Leu 89. Approximately 725 Å^2 per monomer is buried upon dimerization. Surrounding the hydrophobic interface core are polar interactions extending across the interface, including four salt bridges. These salt bridges are between Glu 17 and Lys 54′, Asp 40 and Lys 53′, and their mates related by noncrystallographic 2-fold symmetry. Lys 53 is also positioned to hydrogen bond to the main chain carbonyl group of Gly 38′. The polypeptide's terminal carboxyl group on Leu 89 forms hydrogen bonds across the dimer interface, interacting with the main chain amide groups of Met 15′ and Gly 16′.

The dimer exhibits a saddle-shaped hydrophobic surface patch centered upon the solvent-exposed faces of helices α_3 and α_3′ (Figure 3). It is composed of aromatic and aliphatic residues (Phe 39, Val 44, Val 51, Leu 52, Leu 58, and Trp 62) from both constituents of the dimer. The only polar residue within this patch is Gln 48. Positively and negatively charged residues immediately surround this hydrophobic patch. This region has not been implicated in any type of functionality. However, it is an atypical protein surface formed by highly conserved residues and is a candidate for the LAP2 binding site.

Comparison to Solution Structure. The structure of BAF has been determined using high-quality data obtained by crystallographic methods (MAD phasing to high resolution) and also by NMR methods (*29*) [including the assessment of dipolar couplings which provide long-range structural information (*38*)]. The main chain structures of the monomers determined from both methods compare reasonably well, yielding root-mean-square deviations (rmsd) of 0.73 and 1.3 Å for the superposition of Cα atoms and all atoms, respectively. The coordinates of the BAF solution structure were obtained from the Protein Data Bank, with the identifying code 2EZX. For comparison, the rmsds are 0.27 and 0.59 Å for the superposition of the Cα atoms and all atoms, respectively, between the two monomers within the crystallographic asymmetric unit. Weak noncrystallographic restraints had been employed during refinement of the crystal structure, with their weights being optimized through the use of R_{free}. There was no evidence that the use of the ncs restraints caused the model to significantly disagree with the electron density maps.

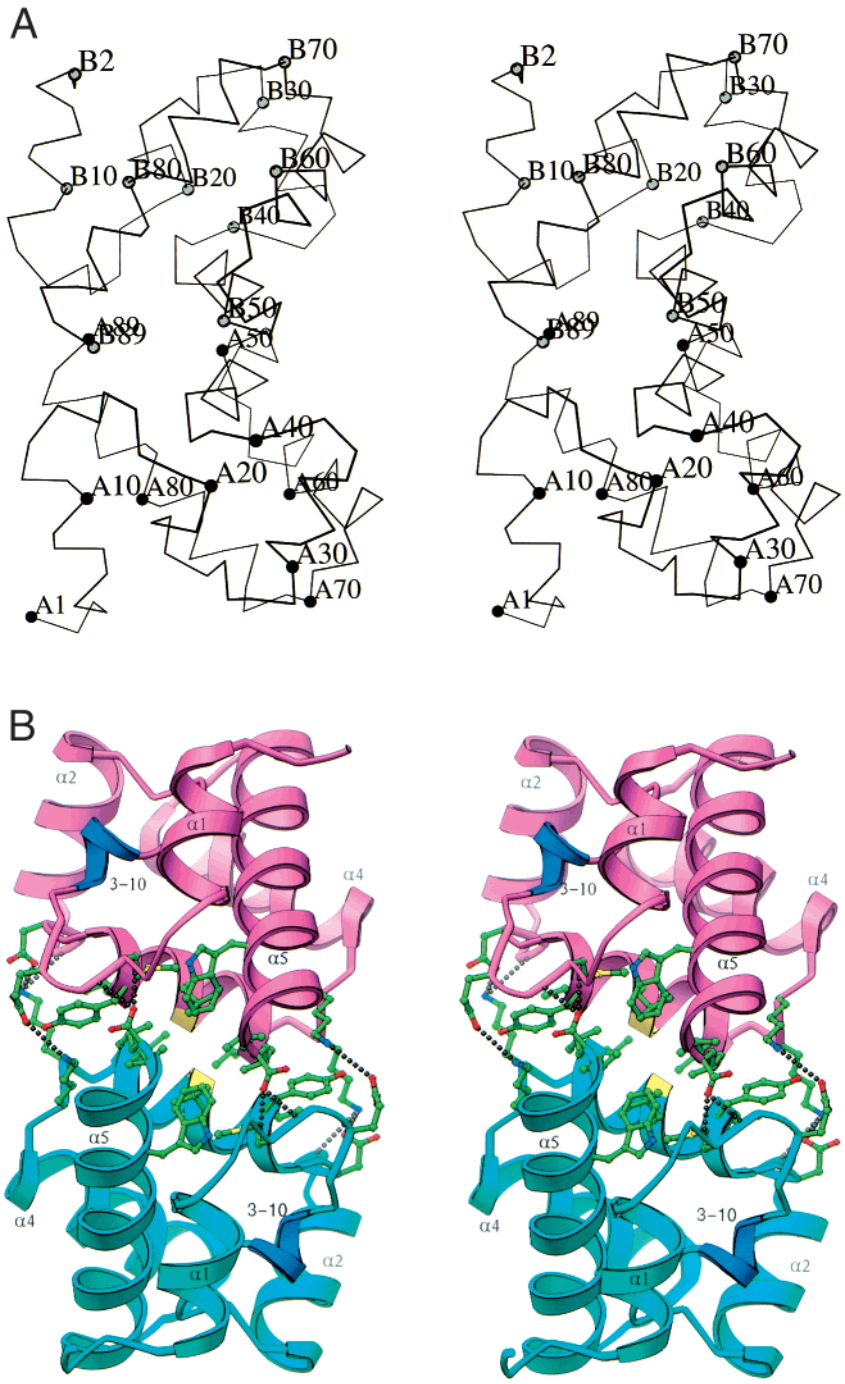

FIGURE 2: BAF crystal structure. (A) Stereoview of the Cα trace of the dimer. (B) Stereoview of a ribbon representation of BAF down the dimer axis. Monomer A is cyan, and monomer B is magenta. The individual helices are labeled, except for helix α_3 which is a fundamental element of the dimer interface. Gly 47 is located in the middle of helix α_3 and highlighted in yellow. The side chains of the residues contributing to the dimer interface are displayed as ball-and-stick representations (oxygen, red; nitrogen, blue; carbon, green; and sulfur, yellow). The 10 polar interactions across the interface are denoted by black spheres. The single 3_{10}-helix in each monomer is colored dark blue.

However, comparison of the intact dimers results in significantly poorer agreement. If the dimers are aligned on the basis of the transformation matrix derived from the superposition of the Cα atoms from a single monomer of each structure, then the rmsd is 1.8 Å for the Cα atoms of the second monomers (Figure 4). In the rmsd calculations presented above, residues 1−3 were excluded as their positions were poorly defined in the NMR study and were greatly influenced by packing contacts in the crystal structure. This large rmsd primarily reflects a disparity between the dimer interfaces observed in the two structures. In this superposition of structures, the second monomer of the NMR structure is related to the corresponding monomer of the crystal structure by a 6.9° rotation and a 1.1 Å translation. This results in the α_3 helices of opposing monomers crossing at angles of approximately 141° and 148° in the crystal and NMR structures, respectively. Additionally, the NMR structure possesses fewer contacts across the dimer interface, including fewer polar interactions. The surface area buried at the dimer interface of the NMR structure is approximately

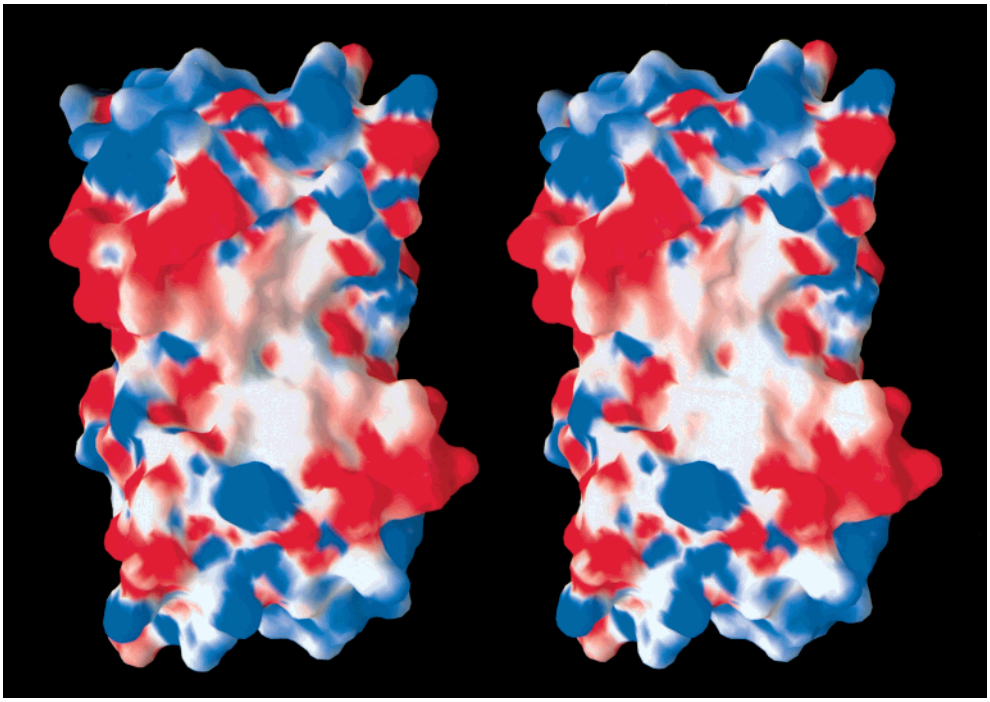

FIGURE 3: Stereoview of the electrostatic surface potential of the saddle-shaped hydrophobic patch on the BAF dimer. The displayed gradient is from −5 (red) to 5 (blue) kT/e$^-$. This view is down the dimer axis.

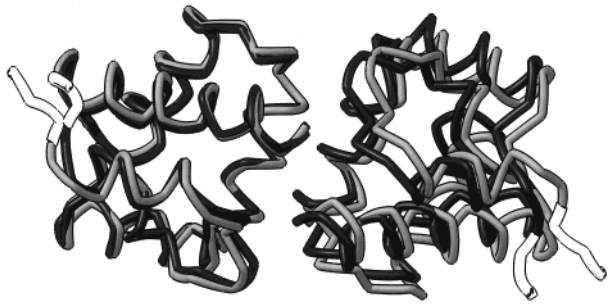

FIGURE 4: Superposition of the BAF dimer crystal and solution structures. Monomers on the left were superimposed, and then the two dimers were aligned according to this transformation. The crystal structure is colored black, and the solution structure is colored gray. The regions colored white (Met 1−Ser 3) were excluded from rmsd calculations.

590 Å^2 per monomer, or 135 Å^2 less than that found in the crystal structure. Also, within a monomer, fewer stabilizing polar interactions are observed in the NMR structure than in the crystal structure. The differences between the BAF structures determined via the two methods may arise, in part, due to the observation that NMR structures tend to be expanded when compared to crystal structures (*39*). This expansion of the structure can result as an artifact of the nature of the experimental NMR data and the computational methods used in refining solution structures and may not represent a meaningful difference between the two structures.

Comparison to Structures in the PDB. The PDB was searched for proteins that were structurally similar to BAF (*40*). No proteins having a fold highly similar to that of BAFs were detected. Several proteins in the database were identified as weakly similar to BAF, with rmsds on Cα atoms in the range of 3−4 Å over 50−76 residues distributed across multiple segments. It was found that the helices between residues Glu 17 and Phe 88 of the BAF monomer were arranged like those of domain II of RuvA (*23*−*26*). However, the overall tertiary and quaternary structures of BAF and RuvA are dissimilar. Detailed examination of the superposition of domain II of RuvA and BAF revealed that the BAF monomer exhibits both the consensus sequence hxxhxGhGxxxsxxhh [h being a hydrophobic residue (V, I, L, M, W, F, Y, or A), s a small residue (A, G, C, or S), and x any residue], and the three-dimensional structure of the helix−hairpin−helix (HhH) DNA-binding motif (*15*, *17*). The existence of this motif within BAF has not been previously reported. It resides within BAF between residues Val 20 and Glu 35 (VGSLAGIGEVLGKKLE), with the only disagreement with the consensus sequence being at the final position, Glu 35. This final residue of the motif is located away from the DNA binding site, and the first HhH motif of polymerase β (*18*, *28*) also contains an acidic residue at this position. The rmsds between the Cα atoms within the HhH motif of BAF and the first and second such motifs in polymerase β are 0.56 and 0.64 Å, respectively. This is similar to what has been reported for comparisons between other HhH motifs (*17*). Domain II of RuvA contains a second example of the HhH motif. The corresponding region in BAF is between residues Leu 63 and Phe 78 (LKDTCGANAKQSRDCF) which resembles the HhH motif in both sequence and structure, but is sufficiently different that it cannot be classified as such. The most significant deviation is the presence of an asparagine residue (Asn 70) in place of the second conserved glycine of the motif.

Mutational Analysis of Residues That Are Important to DNA Binding. Engineered mutants of BAF were assayed for their DNA binding activity to identify residues that potentially constitute the DNA binding site. The BAF mutants behaved like wild-type BAF during purification, including the elution profiles during the gel filtration step. Thus, there was no indication that any of the mutations caused a gross misfolding of the protein. All of the lysine and the arginine

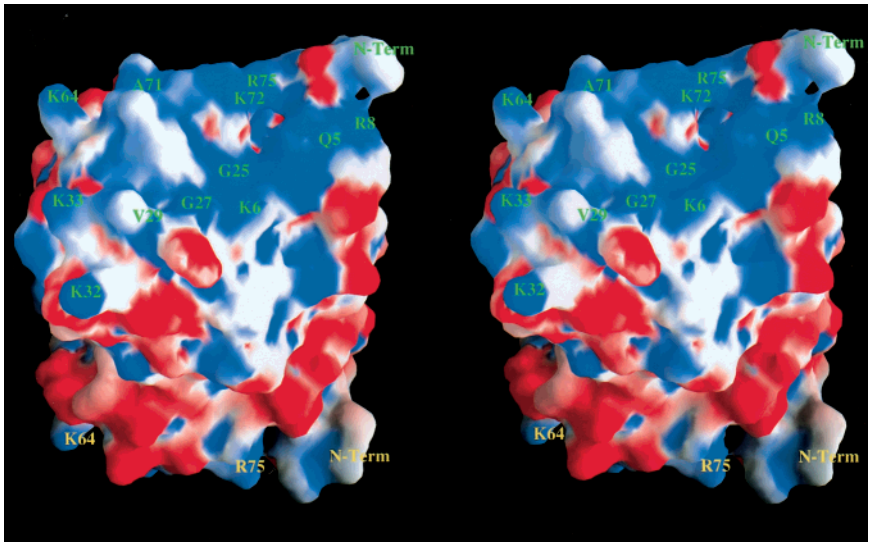

FIGURE 5: Stereoview of the electrostatic surface potential of the BAF dimer. This view emphasizes the region about one HhH motif contained within the dimer. This motif is located in the upper portion of the figure and participates in forming an extensive surface of positive electrostatic potential. Important residues of the proposed DNA binding site are labeled in green, including several residues of the HhH motif (Gly 25, Gly 27, and Val 29). Basic residues that may interact with bound DNA are also labeled, except Arg 60 which is obscured. The labels for basic residues are positioned near the termini of their side chains, and near Cα atoms for other residues. The dimer's second DNA binding site is mostly obscured in this view, with only two visible basic residues, labeled in yellow. The N-termini of both monomers are also labeled. The displayed gradient is from −5 (red) to 5 (blue) kT/e$^-$.

residues were screened by individually mutating each to a glutamate. The two conserved glycine residues of the HhH motif were mutated to both glutamate and glutamine. Single mutations which resulted in undetectable DNA binding activity were K6E, K18E, G25E, G27E, G27Q, K33E, K54E, R60E, K64E, K72E, and R75E. Mutations which significantly decreased but did not completely eliminate BAF's DNA binding activity were R8E, G25Q, K32E, and K53E. Mutations that caused no discernible change in DNA binding activity were R37E, K41E, and R82E. Most of the residues thus identified as potentially belonging to the DNA binding site were located near BAF's HhH motif. The three exceptions participate either in the dimer interface (Lys 53 and Lys 54) or in a salt bridge (Lys 18) which aids in the stabilization of the large loop between helix α_1 and the 3_{10}-helix.

The detection of the HhH motif within the human BAF monomer is highly suggestive of BAF's mode of binding DNA. This motif has been implicated in the nonspecific binding of DNA in a number of proteins (*15*, *17*). It has been observed to be directly interacting with DNA in the crystal structures of polymerase β (*18*), RuvA (*23*, *24*), and AlkA (*16*), and in an NMR study mapping the binding site of endonuclease III (*41*). This motif exists in proteins which do not share a common overall fold, and BAF is a further example of this. It should also be noted that unlike these proteins for which structures of DNA complexes have been reported, BAF has not been shown to recognize any particular feature of DNA structure, such as a gap, a Holliday junction, or a damaged base.

The importance of the HhH motif for the DNA binding activity of BAF is supported by additional data. The signature sequence of the HhH motif is conserved within BAF from four distinct species (human, mouse, zebrafish, and *C. elegans*) (*11*). Moreover, electrostatic potential calculations (*37*) demonstrate that this motif lies within the only large positively charged region on BAF's surface (Figure 5). This positively charged surface patch would interact favorably with the negatively charged phosphate backbone of DNA. As previously mentioned, this region contains the majority of the basic residues that negatively influence DNA binding upon mutation. The exceptions possess other structurally important roles.

The HhH motif is defined as containing two helices connected by a type II reverse turn accompanied by a characteristic sequence. Nonspecific binding to DNA occurs via hydrogen bonds formed between two adjacent phosphate groups of DNA's backbone and main chain amide groups on an extended surface presented by the reverse turn and the N-terminus of the second helix of the motif. To further test the hypothesis that the HhH motif composes BAF's DNA binding site, the two conserved glycine residues of the motif were mutated individually to glutamate or to glutamine. The first conserved glycine, corresponding to Gly 25 in BAF, is required to allow the formation of the type II reverse turn, as the required main chain torsion angles are normally inaccessible to any residue possessing a side chain. The main chain amide of this glycine is then positioned so that it can form a hydrogen bond to a backbone phosphate group of DNA. In polymerase β (*42*) and AlkA (*16*), this reverse turn also contributed to the formation of a metal binding site which further mediated DNA binding. The G25E mutant of BAF lost its DNA binding ability, whereas the G25Q mutant displayed a reduced level of binding. Both mutations would be expected to distort the main chain geometry of this reverse turn. This distortion may account for the reduced binding activity of the G25Q mutant. Elimination of binding by the G25E mutant may be due to a combination of a disruption of the structure of the turn and an introduction of a negative charge near the binding site, creating unfavorable electrostatic interactions with the DNA phosphate groups. The second conserved glycine, corresponding to Gly 27 in BAF, is also critical for DNA binding. The lack of a side chain on this residue allows the main chain amide groups of the residues

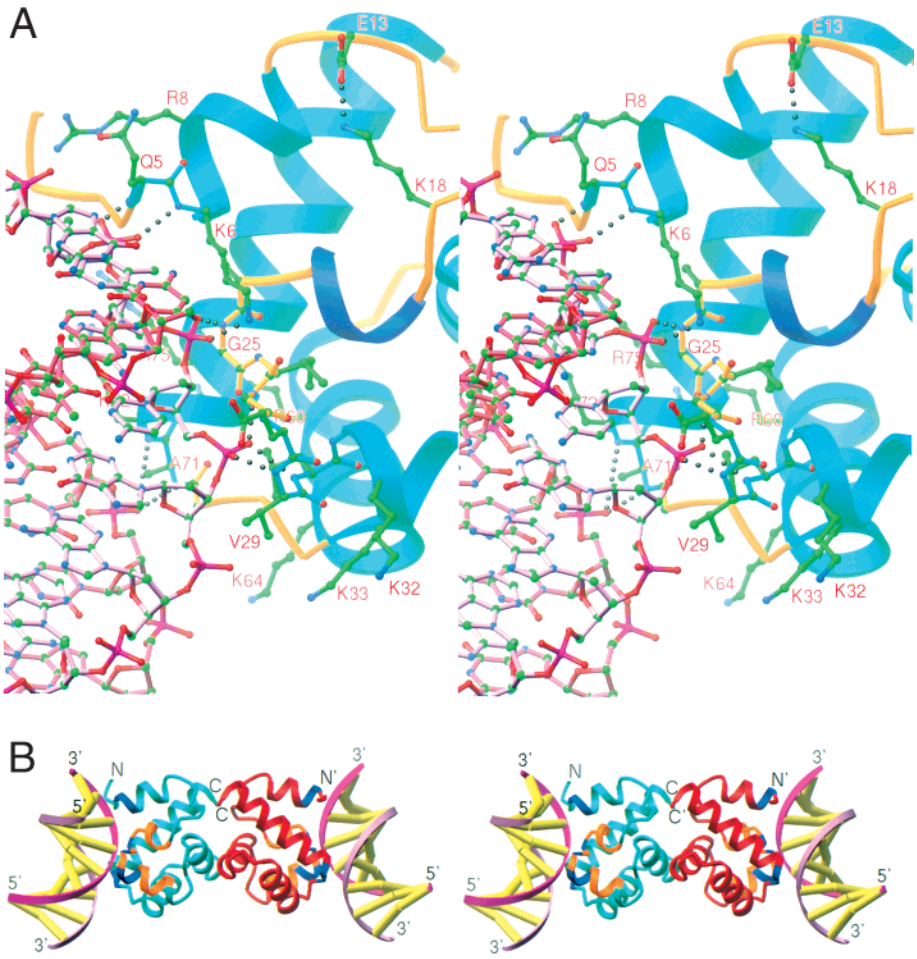

FIGURE 6: Model of B-DNA bound to the BAF dimer. (A) A stereoview of a detailed model of B-DNA bound to one of BAF's HhH motifs and neighboring residues. Hydrogen bonds and salt bridges are indicated by black spheres. Residues that interact with DNA in this model are shown as ball-and-stick representations. The color scheme for the spheres representing atoms is as follows: phosphorus, magenta; oxygen, red; nitrogen, blue; and carbon, green. Additionally, Glu 13 and Lys 18 are displayed. These residues form a salt bridge within the loop connecting helix α_1 and the 3_{10}-helix. Within the protein, cyan indicates α-helix, dark blue indicates 3_{10}-helix, and orange indicates nonhelical regions. (B) A stereoview of a ribbon representation of the BAF dimer modeled bridging two segments of B-DNA. BAF's orientation is approximately the same as in panel A. The individual monomers are colored red and cyan, respectively. The HhH motif within each monomer is colored orange. Protein residues which are proposed to form hydrogen bonds between their main chain amide groups and the DNA phosphate backbone are colored in dark blue.

at the N-terminus of the second helix of the motif (Val 29 and Leu 30 in BAF) to be exposed. Thus, they are available to form hydrogen bonds to a DNA backbone phosphate group. The addition of a residue containing a side chain at this position would sterically block the formation of these hydrogen bonds, but would not necessitate a change in main chain conformation. Both the G27E and G27Q BAF mutants exhibited no DNA binding activity.

A Model of DNA Bound to BAF. An idealized B-DNA decamer was modeled bound to BAF (Figure 6A). Interactions between DNA and the second HhH motif within polymerase β (*18*) formed the basis for this modeling. In addition, the reported interactions between DNA and the HhH motifs in RuvA (*23*, *24*) were used as a guide. Both the B-DNA decamer and BAF were treated as rigid bodies. It should be noted that preliminary data indicate that a large BAF−DNA complex may form, containing more that a single BAF dimer and involving either multiple DNA oligomers or multiple contacts with a single large DNA molecule (R. Zheng and R. Craigie, personal communication). Hence, the model presented here probably represents a subunit of this larger complex. This model revealed the expected hydrogen bonding scheme involving exposed main chain amide groups within the HhH motif, those of Gly 27, Val 29, and Leu 30, to the phosphate group of nucleotide *n*, and that of Gly 25 to the phosphate group of nucleotide $n + 1$. Several other favorable interactions were also present in this model. The aliphatic side chain of Val 29, which is solvent-exposed in the crystal structure, forms a hydrophobic interaction with the deoxyribose group of nucleotide $n - 1$. The side chain of Lys 6 is packed against the reverse turn of the HhH motif. Its ϵ-amino group may form hydrogen bonds to the main chain carbonyl groups of Gly 21 and Ile 26, and less ideally to Leu 23. This serves to sequester additional positive charge directly at the DNA binding site. In the proposed model, the side chain amino group of Lys 6 is then positioned to interact with the phosphate group of nucleotide $n + 1$. Interestingly, this ϵ-amino group occupies the metal binding site observed within each of the two HhH motifs of the polymerase β−DNA complex (*18*) and within the single HhH motif of the AlkA−DNA complex (*16*). It is unknown whether under certain conditions a metal ion may displace the side chain of Lys 6 and bind to BAF's HhH motif. This metal ion has only been observed in the

DNA complexes of polymerase β and AlkA (*16*, *42*), and possibly in the RuvA−Holliday junction complex (*43*), and not in the crystal structures of AlkA (*20*, *21*), endonuclease III (*15*), and MutY (*22*) determined in the absence of DNA. However, DNA binding is abolished in the BAF K6E mutant. In the solution structure, the side chain of Lys 6 is not associated with this reverse turn. However, this may be due to the geometry of the residues near the N-terminus being poorly defined in the NMR experiment (*29*).

The BAF−DNA model revealed several potential sites of interaction between regions of the protein outside the HhH motif and the DNA phosphate backbone. The exposed main chain amide groups of Gln 5 and Lys 6 are positioned such that they may hydrogen bond to the phosphate group of nucleotide $n + 2$. Additionally, the 11 basic residues surrounding the HhH motif would form stabilizing electrostatic interactions with DNA backbone phosphates. Further favorable electrostatic interactions would occur between the dipoles of helices α_1 and α_2 and the backbone phosphates of nucleotides $n + 2$ and n, respectively.

BAF's pseudo-HhH motif (Leu 63−Phe 78) is also positioned to interact with the backbone of the second complementary DNA strand, with the minor groove facing BAF. However, since this region deviates from the canonical HhH motif, it is likely that any interactions between these residues and DNA differ from the standard HhH motif−DNA hydrogen bonding scheme. Nonetheless, the main chain amide groups of Ala 71 and Lys 72, which are at the N-terminal end of helix α_5, may possibly hydrogen bond to a backbone phosphate group of the second strand. Supplementing this is the interaction between the dipole of helix α_5 and the nucleotide's phosphate. There are also three basic residues (Lys 64, Lys 72, and Arg 75) within this pseudo-HhH motif and a fourth nearby (Arg 60) which may form electrostatic interactions with several additional phosphate groups of the complementary strand. The presence of favorable interactions within this model between BAF and both strands of B-DNA agrees with the observation that BAF only binds to double-stranded DNA.

BAF was previously proposed to bind DNA via a helix−turn−helix (HTH) motif composed of residues Glu 56−Phe 88 (helices α_4 and α_5) (*29*). However, this is unlikely for several reasons. An HTH prediction method (*44*) failed to locate any potential candidates for this motif within BAF's sequence. Functionally, the HTH motif typically binds to a specific nucleotide sequence, whereas BAF exhibits highly non-sequence-specific binding to DNA. Structurally, the recognition helix of the HTH motif protrudes from the remainder of the protein molecule so that it may bind within the major groove of DNA, forming specific interactions with the bases (*45*). However, the proposed recognition helix, α_5, within BAF is packed between helices α_1 and α_4 on one face of the monomer, leaving it unavailable to project into the major groove. Additionally, there are three highly conserved acidic residues (Asp 76, Glu 83, and Asp 86) that are present on the solvent-exposed face of helix α_5. These negatively charged side chains on the recognition helix of the putative HTH motif would be expected to give rise to unfavorable interactions with the phosphate groups of DNA. Furthermore, the mutational data do not fully fit an HTH binding model. Arg 82, which is on the solvent-exposed face of the proposed recognition helix, may be mutated to a glutamate with no loss of DNA binding ability. Several of the basic residues which have been implicated in DNA binding by mutational analysis (Lys 6, Lys 18, Lys 32, and Lys 33) are located away from the proposed HTH motif. The role that these basic residues have in DNA binding, and the negative effect that mutations G25Q, G25E, G27Q, and G27E have upon binding activity, is not adequately explained if DNA binding is assumed to occur via the proposed HTH motif.

Concluding Remarks. A model by which BAF nonspecifically binds double-stranded DNA through a HhH motif has been proposed and tested by both structural and mutational analysis. The HhH motif has been directly observed in only a small number of proteins, with these proteins exhibiting an array of functions and overall folds. The presence of the HhH motif in BAF further extends this array. Furthermore, BAF displays two novel features associated with its HhH motif. A lysine side chain occupies what has been previously observed to be a metal binding site, yielding an alternative method of capturing a positive charge at the DNA binding site. Also, the first helix of the motif is a single turn of 3_{10}-helix instead of an α-helix previously observed in other structures containing this motif. The presence of two distinct binding sites on the BAF dimer correlates with its ability to bridge together discrete segments of double-stranded DNA, and suggests a model of the bridged nucleoprotein complex (Figure 6B).

ACKNOWLEDGMENT

We thank Fred Dyda and Zbigniew Dauter for help during data collection and Myung Soo Lee, Ronglan Zheng, and Mengli Cai for helpful discussions.

REFERENCES

1. Brown, P. O. (1997) in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., Eds.) pp 161−203, Cold Spring Harbor Laboratory Press, Plainview, NY.
2. Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1987) *Cell 49*, 347−356.
3. Bowerman, B., Brown, P. O., Bishop, J. M., and Varmus, H. E. (1989) *Genes Dev. 3*, 469−478.
4. Bukrinsky, M. I., Sharova, N., McDonald, T. L., Pushkarskaya, T., Tarpley, W. G., and Stevenson, M. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6125−6129.
5. Farnet, C. M., and Haseltine, W. A. (1991) *J. Virol. 65*, 1910−1915.
6. Farnet, C. M., and Bushman, F. D. (1997) *Cell 88*, 483−492.
7. Lee, M. S., and Craigie, R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9823−9827.
8. Fujiwara, T., and Mizuuchi, K. (1988) *Cell 54*, 497−504.
9. Ellison, V., Abrams, H., Roe, T., Lifson, J., and Brown, P. (1990) *J. Virol. 64*, 2711−2715.
10. Farnet, C. M., and Haseltine, W. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4164−4168.
11. Lee, M. S., and Craigie, R. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 1528−1533.
12. Chen, H., and Engelman, A. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 15270−15274.
13. Furukawa, K. (1999) *J. Cell Sci. 112*, 2485−2492.
14. Wilson, K. L. (2000) *Trends Cell Biol. 10*, 125−129.
15. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) *EMBO J. 14*, 4108−4120.
16. Hollis, T., Ichikawa, Y., and Ellenberger, T. (2000) *EMBO J. 19*, 758−766.
17. Doherty, A. J., Serpell, L. C., and Ponting, C. P. (1996) *Nucleic Acids Res. 24*, 2488−2497.

18. Pelletier, H., Sawaya, M. R., Wolfle, W., Wilson, S. H., and Kraut, J. (1996) *Biochemistry 35*, 12742−12761.
19. Aravind, L., Walker, D. R., and Koonin, E. V. (1999) *Nucleic Acids Res. 27*, 1223−1242.
20. Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L., and Ellenberger, T. E. (1996) *Cell 86*, 321−329.
21. Yamagata, Y., Kato, M., Odawara, K., Tokuno, Y., Nakashima, Y., Matsushima, N., Yasumura, K., Tomita, K., Ihara, K., Fujii, Y., Nakabeppu, Y., Sekiguchi, M., and Fujii, S. (1996) *Cell 86*, 311−319.
22. Guan, Y., Manuel, R. C., Arvai, A. S., Parikh, S. S., Mol, C. D., Miller, J. H., Lloyd, S., and Tainer, J. A. (1998) *Nat. Struct. Biol. 5*, 1058−1064.
23. Hargreaves, D., Rice, D. W., Sedelnikova, S. E., Artymiuk, P. J., Lloyd, R. G., and Rafferty, J. B. (1998) *Nat. Struct. Biol. 5*, 441−446.
24. Roe, S. M., Barlow, T., Brown, T., Oram, M., Keeley, A., Tsaneva, I. R., and Pearl, L. H. (1998) *Mol. Cell 2*, 361−372.
25. Nishino, T., Ariyoshi, M., Iwasaki, H., Shinagawa, H., and Morikawa, K. (1998) *Structure 6*, 11−21.
26. Rafferty, J. B., Sedelnikova, S. E., Hargreaves, D., Artymiuk, P. J., Baker, P. J., Sharples, G. J., Mahdi, A. A., Lloyd, R. G., and Rice, D. W. (1996) *Science 274*, 415−421.
27. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science 264*, 1891−1903.
28. Maciejewski, M. W., Liu, D., Prasad, R., Wilson, S. H., and Mullen, G. P. (2000) *J. Mol. Biol. 296*, 229−253.
29. Cai, M., Huang, Y., Zheng, R., Wei, S. Q., Ghirlando, R., Lee, M. S., Craigie, R., Gronenborn, A. M., and Clore, G. M. (1998) *Nat. Struct. Biol. 5*, 903−909.
30. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., Eds.) pp 307−326, Academic Press, Orlando, FL.
31. Terwilliger, T. C., and Berendzen, J. (1997) *Acta Crystallogr. D53*, 571−579.
32. Furey, W., and Swaminathan, S. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., Eds.) pp 590−620, Academic Press, Orlando, FL.
33. Jones, T. A., Cowan, S., Zou, J.-Y., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
34. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
36. Carson, M. (1991) *J. Appl. Crystallogr. 24*, 958−961.
37. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins 11*, 281−296.
38. Tjandra, N., and Bax, A. (1997) *Science 278*, 1111−1114.
39. Kuszewski, J., Gronenborn, A. M., and Clore, G. M. (1999) *J. Am. Chem. Soc. 121*, 2337−2338.
40. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
41. Shekhtman, A., McNaughton, L., Cunningham, R. P., and Baxter, S. M. (1999) *Struct. Fold Des. 7*, 919−930.
42. Pelletier, H., and Sawaya, M. R. (1996) *Biochemistry 35*, 12778−12787.
43. Rafferty, J. B., Ingleston, S. M., Hargreaves, D., Artymiuk, P. J., Sharples, G. J., Lloyd, R. G., and Rice, D. W. (1998) *J. Mol. Biol. 278*, 105−116.
44. Dodd, I. B., and Egan, J. B. (1990) *Nucleic Acids Res. 18*, 5019−5026.
45. Harrison, S. C. (1991) *Nature 353*, 715−719.

BI000572W