Two Domains of the Epstein-Barr Virus Origin DNA-binding Protein, EBNA1, Orchestrates Sequence-specific DNA Binding*

Jennifer Cruickshank, Kathy Shire, Alan R. Davidson, Aled M. Edwards, and Lori Frappier

From the Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Ontario M5S 1A8 and the Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario MSG 1L6, Canada

The EBNA1 (for Epstein-Barr nuclear antigen 1) protein of Epstein-Barr virus governs the replication and partitioning of the viral genomes during latent infection by binding to specific recognition sites in the viral origin of DNA replication. The crystal structure of the DNA binding portion of the EBNA1 protein revealed that this region comprises two structural motifs: a core domain, which mediates protein dimerization and is structurally homologous to the DNA binding domain of the papillomavirus E2 protein, and a flanking domain, which mediated all the observed sequence-specific contacts. To test the possibility that the EBNA1 core domain plays a role in sequence-specific DNA binding not revealed in the crystal structure, we examined the effects of point mutations in potential hydrogen bond donors located in an α-helix of the EBNA1 core domain whose structural homologue in E2 mediates sequence-specific DNA binding. We show that these mutations severely reduce the affinity of EBNA1 for its recognition site, and that the core domain, when expressed in the absence of the flanking domain, has sequence-specific DNA binding activity. Flanking domain residues were also found to contribute to the DNA binding activity of EBNA1. Thus, both the core and flanking domains of EBNA1 play direct roles in DNA recognition.

Epstein-Barr virus (EBV) infects most people worldwide and is associated with several types of cancer (1). The infection usually takes a latent form in which the viral DNA episomes are replicated once every cell cycle and are maintained at a constant copy number per cell (1, 2). During latent infection, various combinations of as many as nine viral proteins are expressed, but only one viral protein, Epstein-Barr nuclear antigen 1 (EBNA1), is expressed in all cases (3). EBNA1 is the only viral protein required for the replication and stable maintenance of EBV genomes (4). EBNA1 also regulates the expression of itself and other latent gene products by both transactivating and repressing transcription (5–7).

The replication, segregation, and transactivation functions of EBNA1 require the binding of EBNA1 dimers to 18-bp palindromic recognition sites present in multiple copies in the viral “plasmid” origin of replication, oriP (8–10). EBNA1 residues that mediate DNA binding and dimerization have been localized to the C-terminal portion of the protein between amino acids 459 and 607 (11–14). Mutagenesis of this region identified residues 459–487 as important for sequence-specific DNA recognition (13, 14).

The crystal structure of the EBNA1 DNA binding and dimerization region was solved both in the presence and absence of its 18-bp recognition site (15–17). These structures revealed that this region comprised two closely associated domains, referred to as the core domain and the flanking domain (see Fig. 1). The core domain (amino acids 504–604) contains an eight-stranded antiparallel β-barrel, comprised of four strands from each monomer and two α-helices per monomer. The core domain forms the dimerization interface but was not observed to mediate contacts with the DNA bases. The flanking domain (amino acids 461–503) comprises an α-helix and an extended chain, and mediates all of the base contacts observed in the co-crystal structure. The helix is oriented perpendicular to the axis of the DNA and its N-terminal amino acid (Lys477) contacts two bases in the outer portion of the DNA recognition site through the major groove. Five additional base contacts are made by three amino acids (Lys461, Gly463, and Arg469) in a portion of the extended chain that tunnels along the base of the minor groove.

Although the EBNA1 co-crystal structure did not reveal a direct role for the core domain in sequence recognition, the core domain is hypothesized to play an important role in DNA binding for several reasons. First, the core domain is structurally homologous (root mean square deviation 0.908 Å) to the complete DNA binding domain of the E2 protein of bovine papillomavirus; this structural homology suggests a common function and DNA binding mechanism. In E2, each monomer contains a helix, termed the recognition helix, which makes all of the base contacts with the DNA (18). The analogous helix (also termed the recognition helix) in the EBNA1 core domain contains three amino acids (Lys514, Tyr516, and Arg522) that are potential hydrogen bond donors (15) (Fig. 1). Second, the potential hydrogen bond donors in the EBNA1 recognition helix are preferentially conserved in the distantly related EBNA1 protein from herpesvirus papio (19). Third, an EBNA1 DNA binding domain mutant lacking most of the flanking domain binds some EBNA1 recognition sites with high affinity, despite the fact that it lacks the residues that mediate five out of the seven base-specific contacts (12). Finally, weak DNA binding has been observed with an EBNA1 fragment lacking the complete flanking domain (11).
In this paper we investigated the role of the EBNA1 core domain in DNA binding by mutating potentially important hydrogen bond donors in the core domain and by analyzing the core domain in the absence of the flanking domain. Our data reveal that the core domain plays a direct role in DNA recognition.

**EXPERIMENTAL PROCEDURES**

**The EBNA1 Mutants—** EBNA1 mutants with alanine substitutions in arginine 469 (R469A), tyrosine 518 (Y518A), arginine 522 (R522A), or lysine 514, tyrosine 518, and arginine 522 (RH) were constructed by amplifying the EBNA1 gene in two fragments; the 3' end to the mutation site and the mutation site to the 5' end. The polymerase chain reaction products were phosphorylated and ligated together, and EBNA1 sequences between amino acids 452 and 641 were reamplified using oligonucleotides that placed an NdeI site just N-terminal to residue 452 and a BamHI site C-terminal to residue 641. The polymerase chain reaction products were digested with NdeI and BamHI and cloned between the NdeI and BamHI sites of PET15b (Novagen). The resulting constructs expressed EBNA1 residues 452–641 (with point mutations) fused to a hexahistidine tag. All constructs were sequenced to confirm the positions of the point mutations.

The EBNA1 fragment corresponding to amino acids 496–641 was generated by polymerase chain reaction amplification using primers that placed an NdeI site prior to codon 496 and a BamHI site following codon 641. The EBNA496–641 expression construct was generated by digesting the amplified EBNA1 fragment with NdeI and BamHI and ligating it between the NdeI and BamHI sites of PET15b.

**Protein Purification—** EBNA452–641 was overproduced and purified from Escherichia coli as described previously (20). R469A, Y518A, R522A, RH, and EBNA496–641 were expressed in E. coli strain BL21(DE3) pLysS and purified on heparin-agarose and metal chelating acrylamide gel electrophoresis.

**DNA Binding Assays—** DNA binding assays used a single EBNA1 recognition site corresponding to site 1 of the oriP dyad symmetry element. To generate this site, the 20-mer oligonucleotide 5′-CCGGAACGATGCTACCGC-3′ was end-labeled with [γ-32P]ATP and then annealed to its complementary oligonucleotide. For equilibrium binding assays, EBNA1 proteins were incubated with 10 fmol of end-labeled DNA recognition site for 10 min at room temperature in 20-μl reactions containing 10 mM HEPES (pH 7.5), 5 mM MgCl2, 300 mM NaCl (binding buffer). Glycerol was then added to a final concentration of 5%, and the reactions were loaded onto a 12% polyacrylamide gel containing 0.5× TBE. After electrophoresis, bound and unbound DNA was visualized by autoradiography and quantified using a PhosphorImager (Molecular Dynamics) and Imagequant software.

**RESULTS**

**A Role for the EBNA1 Recognition Helix in DNA Binding—** The EBNA1 core domain contains a helix that is structurally homologous to the “recognition helix” in E2. In E2, this helix harbors all of the residues that make base-specific contacts with the DNA. To investigate the role of the EBNA1 recognition helix, we generated three different EBNA1 mutants in which potential hydrogen bond donors of the recognition helix were mutated to alanines (see Fig. 1). In one mutant, called RH, all three potential hydrogen bond donors were disrupted (Lys514, Tyr518, and Arg522); in the second mutant, tyrosine 518 was targeted (Y518A); and the third mutant contained a point mutation in arginine 522 (R522A). In the EBNA-DNA co-crystal structure, the three targeted amino acids are oriented toward the DNA but are too far from the nearest H-bond acceptor in the bases (more than 6 Å) to form H-bonds (16). Arg522 does, however, form a H-bond with one phosphate residue.

All three mutations were introduced in the context of the EBNA1 fragment 461–607 bound to an 18-bp EBNA1 recognition site was solved previously by Bochkarev et al. (16) (Protein Data Bank entry 1B3T). The flanking domain is shown in red and the core domain in blue; the recognition helices are shown in a darker shade of blue to distinguish them from the rest of the core domain. Side chains of amino acids 514 (Lys), 518 (Tyr), and 522 (Arg) in the recognition helix are also shown.
DNA binding and dimerization region (amino acids 452–641), purified, and titrated onto an EBNA1 recognition site (Fig. 2 and Table I). The DNA binding activity of each mutant, as analyzed by electrophoretic mobility shift assays (EMSA), was compared with the equivalent wild type EBNA1 polypeptide (EBNA\textsubscript{452–641}). All of the mutant proteins showed defects in DNA binding. The triple mutant and R522A had severely diminished EBNA1 DNA binding activity, and the activity of the Y518A mutant was reduced 80-fold compared with the wild type polypeptide. These experiments were performed three times for each mutant with very similar results.

To ensure that the mutations in the recognition helix affected DNA binding specifically rather than disrupted the folding of the protein, we tested the effects of the mutations on protein folding using CD spectroscopy. The CD spectra of the three recognition helix mutants were indistinguishable from the wild type polypeptide (Fig. 3). All four proteins had elliptical minima at 208 and 222 nm, in keeping with their helical content.

We also tested whether the recognition helix mutations were less active because they destabilized the EBNA1 dimer. The active form of EBNA1 is a dimer, and the dimerization interface is highly similar to the E2 protein from papillomavirus (15, 18, 22). In E2, dissociation of the dimer leads to rapid unfolding and inactivation (23, 24). To test the effects of the recognition helix mutations on the stability of the EBNA1 dimer, we monitored the rates of unfolding of the EBNA1 proteins in guanidine using circular dichroism at 222 nm (Table II). EBNA\textsubscript{452–641} was extremely stable, requiring a 33-min incubation in 6.6 M GdnHCl to unfold half of the protein (standard equilibrium stability assays were not possible because the unfolding of EBNA\textsubscript{452–641} was not reversible). The Y518A mutation had no significant effect on the stability of the protein. The triple point mutation had only a small effect on EBNA1 stability, with half of the protein remaining folded after 14 min in 6.6 M guanidine. Our data suggest that all of the EBNA1 mutants remained dimeric and folded during the 10-min DNA binding reaction (in the absence of guanidine). Our results therefore support a direct role for the EBNA1 recognition helix in DNA recognition.

**The Core DNA Binding Domain Is Sufficient for Sequence-specific DNA Binding**—The above data strongly suggest that the core domain recognition helices play a direct role in DNA recognition, but alternative explanations are possible. For example, the recognition helix mutations could have elicited their effects by causing minor structural changes that would not be detected by CD analysis or by disrupting DNA phosphate (as opposed to base) interactions. To directly examine the contribution of the core domain to DNA binding, we expressed and purified an EBNA1 polypeptide containing the core domain but lacking the flanking domain (EBNA\textsubscript{496–641}), and monitored sequence-specific DNA binding by EMSAs. Like the EBNA1 proteins examined above, EBNA\textsubscript{496–641} contained the C-terminal acidic tail in addition to DNA binding domain residues (the EBNA1 core domain produced in the absence of the acidic tail was insoluble).

Purified EBNA\textsubscript{496–641} protein was titrated onto two different DNA fragments, one that contained the EBNA1 binding site and a second 18-bp sequence that contained the DNA recognition site for the E2 protein of bovine papillomavirus. As shown in Fig. 4, EBNA\textsubscript{496–641} bound to the EBNA1 site in a saturable manner but not to the E2 site. Thus, the core domain is capable
The crystal structure of the EBNA1 DNA binding region bound to DNA revealed two domains: the core domain, which mediated dimerization, and the flanking domain, which mediated base contacts with the DNA (16). We have now shown that the core domain also plays a direct role in sequence-specific DNA binding: one that involves the flanking domain and a second that uses the core domain. In this scenario, the flanking domain binding mode would be favored under the conditions of the crystallography, whereas the core domain binding mode would be favored under the conditions of the binding reactions for the EMSAs. This explanation is unlikely because point mutations both in the core domain recognition helices and in the flanking domain extended chain have pronounced effects on DNA binding under the conditions of the EMSAs. Furthermore, the crystal structure of the EBNA1 DNA binding region has been solved on three different EBNA1 recognition sequences (including the site 1 sequence studied here) under varied salt and buffer conditions and, in all cases, the hydrogen bonds observed between EBNA1 residues and the DNA bases were identical (16, 17).2

A second explanation is that both the core and flanking domains mediate DNA contacts under the EMSA and crystallography conditions, but that the recognition helices are dynamically associated with the DNA. In this case, the crystal structure might simply have captured the recognition helices in the dissociated position. Based on the co-crystal structures of EBNA1 and E2, the base contacts made by the EBNA1 recognition helices are expected to occur through the major groove. The possibility that the core and flanking domains contact the DNA at the same time is supported by methylation protection footprints that show protection of a G residue (at position 3′/3′ relative to the axis of symmetry of the palindrome) that is not contacted by flanking domain residues in the crystal structure.

A. Bochkarev, A. M. Edwards, and L. Frappier, unpublished results.
but is predicted to be contacted by Lys14 in the recognition helix (16, 25).

The third possibility is that EBNA1 binds DNA by a two-step mechanism in which both the core and flanking domains interact sequentially with the EBNA1 recognition site. In this mechanism, the core domain recognition helices would make the first sequence-specific DNA contacts, positioning the protein such that the flanking domain extended chain can be loaded into the minor groove. After the flanking domain is assembled on the DNA, the position of the recognition helices would shift such that direct contacts with the bases are no longer stable. Thus, in the assembled EBNA1-DNA complex that was crystallized, no base contacts with the core domain would be observed. During dissociation of EBNA1 from the DNA, the recognition helices might again make transient DNA contacts, facilitating the release of the flanking domain from the DNA.

A key feature of models two and three is that the recognition helices and the flanking domain extended chain are flexible. Evidence for the former comes from NMR studies on the E2 DNA binding domain, which revealed that the E2 recognition helices are mobile (26). The flexibility of the EBNA1 flanking domain extended chain are revealed in the crystal structure of EBNA470–607 in the absence of DNA (15). In this structure, the extended chains in the two monomers were disordered. Thus, DNA binding models two or three are likely, and further kinetic analyses will be necessary to differentiate between the two.

Since EBNA1 dimerization is a requirement for DNA binding, we have addressed the possibility that the recognition helix mutations affect DNA binding indirectly by destabilizing the dimerization of the protein. In these experiments, we used protein unfolding as a measure of disruption of the dimerization interface. Several lines of evidence indicate that unfolding will occur if the dimerization of EBNA1 is disrupted. First, the EBNA1 core domain that mediates dimerization is almost identical in structure to the DNA binding and dimerization domain of the papillomavirus E2 protein. Careful studies on the folding of E2 have clearly shown that dissociation of the dimer occurs when the domain unfolds (23, 24). Second, the recognition helix point mutations are distant from the residues in the three β-sheets known to mediate dimerization, and therefore any effects on dimerization would be expected to involve destabilization of the domain (15). Third, EBNA1 is only found as a dimer under native conditions; folded monomers of EBNA1 have never been observed (10). Finally, the dimeric interface forms the hydrophobic core of the domain and, thus, folding of an EBNA1 monomer is energetically very unlikely. We believe, therefore, that the defect associated with the EBNA1 recognition helix mutations is in DNA binding and not in dimerization. One caveat to our EBNA1 dimerization analysis was that the CD experiments were only conducted at one protein concentration (10 μM); therefore, we have not ruled out the possibility that the EBNA1 dimers dissociate at lower protein concentrations. However, the DNA binding analyses of the RH and R522A mutants were conducted at protein concentrations similar to those used in the CD analysis (up to 26 μM for RH and up to 6 μM for R522A); therefore, the DNA binding defects exhibited by these mutants are unlikely to be due to disruption of dimerization.

DNA binding mechanisms involving two DNA interaction modes or steps are unusual but not completely novel. For example, the Myb oncoprotein and T7 and E. coli RNA polymerases have some dynamic characteristics. Tryptophan fluorescence studies indicate that the initial rapid interaction of Myb with DNA is followed by a slower step, which involves a structural reorganization of the protein (27). Similarly, kinetic studies with T7 RNA polymerase indicate that this protein forms an initial weak complex with promoter DNA, then undergoes a conformational change resulting in tighter DNA binding (28). The multisubunit E. coli RNA polymerase has at least two biochemically defined nucleic acid binding domains that interact transiently during the course of the polymerization reaction and stabilize the ternary elongation complex (29–31); the dynamic nature of these interactions underlies the regulation of RNA synthesis. Perhaps the dynamics of the interaction of the EBNA1 domains with DNA also plays a regulatory role in the initiation of DNA replication.

Acknowledgments—We thank Angela Fleming for the construction and purification of the R469A EBNA1 mutant, and Valerie Booth for generating Fig. 1. We also thank Karen Maxwell and members of the Davidson laboratory for helpful discussions on protein folding.

REFERENCES

1. Kieff, E. (1986) in Fields Virology (Fields, B. N., Knipe, D. M., and Howley, P. M., eds) 3rd ed., pp. 3343–3396, Lippincott-Raven Publishers, Philadelphia.
2. Yates, J. L. (1996) in DNA Replication in Eukaryotic Cells (DePamphilis, M. L., ed), pp. 751–773, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3. Klein, G. (1989) Cell 58, 5–8
4. Yates, J. L., Warren, N., and Sugden, B. (1985) Nature 313, 812–815
5. Reisman, D., and Sugden, B. (1986) Mol. Cell. Biol. 6, 3838–3846
6. Sample, J., Hensen, E. B. D., and Sample, C. (1992) J. Virol. 66, 4654–4661
7. Gahn, T., and Sugden, B. (1995) J. Virol. 69, 2633–2636
8. Rawlins, D. R., Milman, G., Hayward, S. D., and Hayward, G. S. (1985) Cell 42, 359–368
9. Yates, J. L., and Camiolo, S. M. (1988) Cancer Cells 6, 197–205
10. Frappier, L., and O’Donnell, M. (1991) J. Biol. Chem. 266, 7819–7826
11. Ambinder, R. F., Mullen, M., Chang, Y., Hayward, G. S., and Hayward, S. D. (1991) J. Virol. 65, 1466–1478
12. Summers, H., Barwell, J. A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1996) J. Virol. 70, 1225–1231
13. Chen, M.-R., Middeldorp, J. M., and Hayward, S. D. (1993) J. Virol. 67, 4875–4885
14. Chen, M.-R., Zong, J., and Hayward, S. D. (1994) Virology 205, 486–495
15. Bochkarev, A., Barwell, J., Pfuetzner, R., Purye, W., Edwards, A., and Frappier, L. (1995) Cell 83, 39–46
16. Bochkarev, A., Barwell, J., Pfuetzner, R., Bochkareva, E., Frappier, L., and Edwards, A. M. (1996) Cell 84, 791–800
17. Bochkarev, A., Bochkareva, E., Frappier, L., and Edwards, A. M. (1998) J. Mol. Biol. 284, 1273–1278
18. Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (1992) Nature 359, 505–512
19. Yates, J. L., Camiolo, S. M., Ali, S., and Ying, A. (1996) Virology 222, 1–13
20. Barwell, J. A., Bochkarev, A., Pfuetzner, R. A., Tong, H., Yang, D. S. C., Frappier, L., and Edwards, A. M. (1995) J. Biol. Chem. 270, 26536–26559
21. Summers, H., Fleming, A., and Frappier, L. (1997) J. Biol. Chem. 272, 26434–26440
22. Hegde, R. S., and Androphy, E. J. (1998) J. Mol. Biol. 284, 1479–1499
23. Mak, Y., Bycroft, M., and de Prat-Gay, G. (1996) Nat. Struct. Biol. 3, 711–717
24. Mak, Y., de Prat-Gay, G., Butler, P. J., and Bycroft, M. (1996) Protein Sci. 5, 310–319
25. Frappier, L., and O’Donnell, M. (1992) J. Virol. 66, 1786–1790
26. Liang, H., Petro, A. M., Meadrows, R. P., Yoon, H. S., Egan, D. A. W., Walter, K., Holzman, T. F., Robbins, T., and Fesik, S. W. (1996) Biochemistry 35, 2095–2103
27. Zargarian, L., Le Tilly, V., Jamin, N., Chaffotte, A., Gabrielsen, O. S., Toma, F., and Alper, B. (1999) Biochemistry 38, 1921–1929
28. Jia, Y., Kumar, A., and Patel, S. S. (1996) J. Biol. Chem. 271, 30451–30458
29. Chamberlin, M. J. (1992) Harvey Lect. 88, 1–21
30. Johnson, T. L., and Chamberlin, M. J. (1994) Cell 77, 217–224
31. Nudler, E., Avetissova, E., Markovtsov, V., and Goldfarb, A. (1996) Science 273, 211–217