Nucleic Acids Research, Vol. 19, No. 18 4901–4908

Purification of the DNA binding domain of herpes simplex virus type 1 immediate-early protein Vmw175 as a homodimer and extensive mutagenesis of its DNA recognition site

Roger D.Everett, Margaret Elliott, Graham Hope and Anne Orr
Medical Research Council Virology Unit, Church Street, Glasgow G11 5JR, UK

Received June 21, 1991; Revised and Accepted August 14, 1991

ABSTRACT

The herpes simplex virus type 1 (HSV-1) Immediate-Early (IE) polypeptide Vmw175 is essential for the activation of transcription from viral early and late promoters. Vmw175 also reduces the activity of its own (IE-3) promoter in transfection assays. Both transactivation and repression mediated by Vmw175 require the integrity of a conserved domain of the polypeptide which has been shown to bind to specific DNA sequences. We have investigated the DNA sequence requirements for Vmw175 binding using a randomly mutated target. The results indicate that the binding site covers a region of 13 nucleotides divided into proximal and distal parts which are consistent with the consensus ATCGTNNNNNYSG. We have also expressed several different constructs encompassing the DNA binding domain of Vmw175 in bacteria, and obtained preparations of greater than 90% purity. The DNA binding domain is a dimer in solution, and binds DNA with a specificity similar to that of the intact protein, although the smallest DNA binding competant protein has a slightly reduced specificity.

INTRODUCTION

Herpes simplex virus type 1 (HSV-1) encodes at least 72 distinct genes in a genome of approximately 150kb (1). These genes can be divided into three differentially regulated groups (Immediate-Early (IE), early and late) based on the time course of their expression and whether they can be transcribed in the absence of viral protein synthesis or viral DNA replication (for reviews see references 2–4). The IE genes are transcribed first, and the function of at least some of the IE polypeptides is required for the normal expression of the later classes of genes. Of the five IE gene products, Vmw175 (expressed from gene IE-3) is perhaps the most important since its inactivation results in failure to transcribe early or late genes and an apparent over-expression of the IE genes (5,6). Thus Vmw175 is directly or indirectly required for the normal regulation of all other HSV-1 genes.

The properties of Vmw175 have been investigated in some detail. It is a phosphorylated nuclear protein of 1298 predicted residues which binds to DNA in crude cell extracts and is associated with chromatin (7–9). Vmw175 has been purified as a homodimer (10,11) and has been shown to bind specifically to selected DNA sequences (12–23). The conclusion from these studies is that many (but not all) of the sites bound by Vmw175 include the motif ATCGTC, that this sequence alone is not sufficient for DNA binding, and that a further sequence further 3′ is also required. An extended binding consensus of RTCGT-CNNYNYSG has recently been described (24). However, the contributions to binding strength of each base within this consensus has not been rigorously examined. One aim of the present study was to conduct a thorough mutagenic analysis of a single binding site in order to determine the relative importance of both conserved and non-conserved bases within and outside the consensus.

The DNA binding domain of Vmw175 has been defined by mutagenic studies (25–28), by its protease resistance (29) and by its expression as a fusion protein in bacteria (23,30). It consists of the entire conserved region 2 of the polypeptide (31), which is found in all the related proteins encoded by members of the sub-familiy alphaherpesviridae (32–34), and also the distal portion of the less conserved region 1. We have previously reported the expression in bacteria of non-fusion proteins comprising the DNA binding domain of Vmw175 (23). In this report we describe the purification to apparent homogeneity of the bacterially expressed polypeptide. The bacterial proteins showed similar DNA binding specificity to that of Vmw175 expressed in virus-infected cells, although there was some relaxation of specificity using the smallest binding-competent protein studied.

Vmw175 expressed in infected cells can be purified in solution as a dimer or higher order oligomers (10). The intact protein also binds DNA as a dimer (17,35). In this study we found by physical analysis that the isolated DNA binding domain is also a dimer, which raises the question whether, like many other DNA binding proteins, dimerisation of Vmw175 is important for DNA binding. The role of DNA binding by Vmw175 in regulating transcription is discussed.

MATERIALS AND METHODS

Plasmids, bacteria, cells and viruses

Plasmid p175 and its insertion mutants pI9, pI10, pI11 and pI12 have been described previously (26). Plasmid p585.4, a T7 expression vector derived from pET-8c (36) with HindIII and BamHI sites present after the NcoI site at the initiation of translation, was kindly provided by Dr. C.M.Preston. All plasmids were constructed and maintained in E. coli strain HB101. For expression experiments, the plasmids were transformed into E. coli strain BL21 (DE3) pLysS (36). Bacteriophage M13mp18 was used as the vector for oligonucleotide directed mutagenesis. The host bacterium JM101 was used for the growth and maintenance of all M13 bacteriophage.

HSV-1 strain 17 syn+ was grown and titrated in BHK C13 cells which were propagated in Glasgow modified Eagles medium (GMEM) supplemented with 10% new born calf serum, 10% tryptose phosphate broth and penicillin (100μ/ml) and streptomycin (100μg/ml). Crude extracts containing Vmw175 were prepared from HSV-1 infected HeLa cells (Flow Laboratories) grown in GMEM with 10% foetal calf serum and penicillin and streptomycin as above.

Construction of T7 expression plasmids and crude extract preparation of induced bacteria

The construction of the expression plasmids pT7I9X, pT7I10X and pT7I11 have been described previously (23). The T7 expression plasmids were transformed into E. coli host BL21 (DE3) pLysS and maintained as glycerol stocks. Bacteria were grown, induced with IPTG, harvested in 0.01 volumes of 10mM Tris-HCl pH 8.0, 1mM EDTA, 1mM PMSF, 0.1mM DTT and crude extracts prepared as described before (23).

Purification of protein I10X

Bacteria were grown, induced and harvested as above. After thawing, the bacteria were lysed and treated with DNase I and RNase (final concentrations 20μg/ml and 100μg/ml respectively) on ice for several minutes until the mixture became less viscous. Then NaCl and polymin P were added to final concentrations of 0.5M and 0.2% respectively, and insoluble material was pelleted by centrifugation at 15,000rpm for 30 minutes in the Sorvall SS34 rotor. The clarified supernatant was brought to 25% saturation with ammonium sulphate and precipitated proteins dissolved in 50mM Hepes pH 7.5, 50mM NaCl, 1mM PMSF, 0.1mM DTT, 0.1% CHAPS and then dialysed against 50mM Bis-Tris pH 6.6, 50mM NaCl, 1mM PMSF, 0.1mM DTT, 0.01% CHAPS and stored at −20°C. The volume of extract obtained from a 1L culture was 3ml.

The I10X DNA binding domain peptide of Vmw175 was purified from the initial extract by two steps of Pharmacia FPLC ion exchange chromatography. The first step used a Mono Q column equilibrated in the same buffer as the final crude extract. 15ml of extract was loaded onto an 8ml column and eluted with a 50mM to 1M NaCl gradient in the Bis-Tris buffer. Fractions containing the Vmw175 peptide were identified by SDS-polyacrylamide electrophoresis; the I10X protein eluted mainly in the flow-through. Pooled fractions were dialysed against 50mM Hepes pH 7.5, 50mM NaCl, 1mM PMSF, 0.1mM DTT, 0.01% CHAPS and then loaded onto a 1ml Mono S column equilibrated in the same buffer. A salt gradient was applied, which eluted the I10X protein as a sharp peak at 0.29M NaCl. The purification illustrated in this paper yielded about 0.2mg purified I10X per litre of culture (higher yields have been obtained in more recent experiments). At least 20% of this material was able to bind to DNA in gel retardation experiments (data not shown).

Preparation of probes for gel retardation experiments

Some experiments used end-labelled isolated restriction fragment probes which contained well defined Vmw175 binding sites. The IE-3 probe (19), including the cap-site region of gene IE-3, was isolated as a BamHI-EcoRI fragment (encompassing bases +27 to −108 of the IE-3 promoter) from plasmid pIE3CAT (26) by cutting with BamHI, labelling the end with Klenow and α ^{32}P-dGTP, and then cutting with EcoRI. The labelled fragment was separated on an 8% non-denaturing polyacrylamide gel, located by autoradiography, passively eluted from the gel and ethanol precipitated. For some experiments a 180bp AvaI fragment (spanning nucleotides −129 to +56 of the IE-1 promoter region; the cap site is at position +1 and the A of the ATCGTC consensus is at position −68) which contains another strong Vmw175 binding site (16) was prepared from plasmid pIE1CAT (37).

Comparative gel retardation analysis of binding sites with single point mutations used probes prepared by the prime-cut method (38). Single-stranded M13 templates derived from M13IE3EB (see below) were annealed with the M13 universal primer and then incubated with Klenow polymerase and 5μCi of α ^{32}P-dATP in a mix containing 10mM Tris-HCl pH 8.0, 10mM $MgCl_2$ and dCTP, dGTP and dTTP (each at 0.05mM) and 0.6uM dATP. After incubation at 37°C for 30 min all four dNTPs were added to increase their concentration by a further 0.1mM and incubation continued for a further 10 min. After heating at 70°C for 10 min, suitable restriction enzyme buffer and EcoRI and BamHI were added and the digestion incubated at 37°C for 90 min. Labelled probe fragments were separated on acrylamide gels and eluted as above. Further purification omitted the G50 column step. Radioactivity in the final preparations was measured by Cerenkov counting; probes were dissolved in volumes of water so as to give equal radioactivity concentrations. Since the specific activity of the radioactive dATP was the same for each probe prepared at the same time, this results in equal chemical probe concentrations. In all series of probe preparations, a wild-type control was prepared in an identical manner.

Gel retardation assays

Suitable amounts of bacterial or nuclear extracts were mixed with radiolabelled probes on ice in a buffer containing 10mM Tris-HCl pH 8.0, 1mM EDTA, 100mM NaCl, 1ug polydI/polydC and 0.1% NP40. Incubations included 0.1ng of end-labelled fragment probes or equal quantities of freshly prepared prime-cut probes. After incubation for 20 minutes, the mixtures were applied to a non-denaturing 4% polyacrylamide gel with 0.5×TBE running buffer and run at 200V at 4°C for 2 to 3 hours. DNA-protein complexes and unbound probe were detected by autoradiography of the dried gel. Where appropriate, the extent of binding was determined by densitometric scanning of suitably exposed autoradiographs followed by calculation of the percentage of total probe bound. Some extracts containing the I10X protein were very active in DNA binding reactions; in these cases dilutions of the extract were used so that the proportion of the wild-type probe complexed was not more than 50%.

Oligonucleotide directed mutagenesis

The oligonucleotide 5′ GCCGCGCTCCGTGTGGACGATCG-GGGCGTCCTCGGG 3′ was synthesised using vials of stock precursors which had been contaminated with about 1.1% of each of the other precursors. The product was purified by gel electrophoresis, phosphorylated and used in a mutagenesis reaction using the Amersham M13 mutagenesis kit. The template was M13IE3EB, which has the IE-3 EcoRI-BamHI capsite region (see above) cloned into M13mp18. After mutagenesis, the mixture was transfected into JM101 and plaques picked for screening by DNA sequence analysis. Examination of 400 plaques resulted in the isolation of a cohort of mutants including at least one, and frequently more than one, of the possible single point mutations at each of the nucleotides in the vicinity of the Vmw175 binding site. In addition, a number of useful double mutants were isolated along with a number of single base deletions. In all, about 100 useful mutations were found in the 400 plaques examined.

Gel filtration chromatography and glycerol gradient centrifugation

The true molecular weight of purified I10X protein was determined by gel filtration and glycerol gradient analysis. A Pharmacia FPLC Superose 12 column was equilibrated with 50mM Hepes pH 7.0, 0.2M NaCl, 1mM PMSF, 0.1mM DTT, 0.01% CHAPS. Purified I10X protein was loaded with bovine serum albumin and bovine carbonic anhydrase (Sigma, gel filtration standards) as controls. Eluted proteins were detected by the uv monitor, by SDS gel electrophoresis and by gel retardation experiments. Other experiments used only I10X and BSA, and only the two standard proteins. Glycerol gradients contained 10% to 30% glycerol in 50mM Hepes pH 7.0, 0.2M NaCl and were centrifuged in the Sorvall AH650 rotor at 40,000 rpm for 46 hours. Fractions (0.2ml) were collected from the top of the tube and assayed by SDS gel electrophoresis, Western blotting and gel retardation analysis. Gradients were loaded either with I10X and BSA and carbonic anhydrase controls, or with the two standards alone, and were run in parallel. The results from these two approaches were used to derive the Stokes radius and sedimentation coefficient of I10X, and thus the true molecular weight, using equations as described in reference 10.

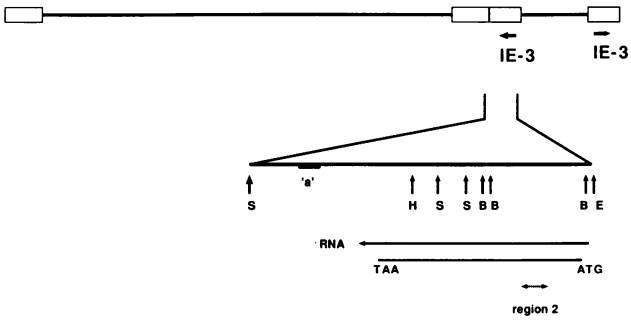

Figure 1. The structure of the HSV-1 genome with the repeat sequences shown as boxes and the positions of the two copies of gene IE-3 marked. Below is an expansion of the IE-3 region with some of the landmark restriction sites; E = EcoRI, B = BamHI, S = SstI, H = HincII, 'a' indicates the 'a' sequence which divides the long and short repeats. The location of the IE-3 transcript is also indicated, with the 1298 amino acid coding region shown below. Conserved region 2 of the polypeptide (codons 315–484) is also marked.

RESULTS

Cloning and expression of the DNA binding domain of Vmw175 in a T7 expression vector, and its subsequent purification

Gene IE-3, which encodes Vmw175, lies in the repeats bounding the short unique segment of the HSV-1 genome (Figure 1). A combination of mutational and functional studies have shown that the DNA binding domain of Vmw175 lies within amino acid residues approximately 276–490 (23, 25–30). This includes the distal part of region 1 and the whole of region 2 of the polypeptide (31).

The cloning of sections of the DNA binding domain of Vmw175 in a T7 expression vector has been presented before (23). As described in the Methods section, plasmids pT7I9X, pT7I10X and pT7I11 contain the coding regions of residues 252–523, 276–523 and 293–523 of Vmw175 respectively. The expressed polypeptides will be referred to as I9X, I10X and I11; all were produced in amounts sufficient to be visible on SDS-polyacrylamide gels stained with Coomassie Blue (Figure 2A).

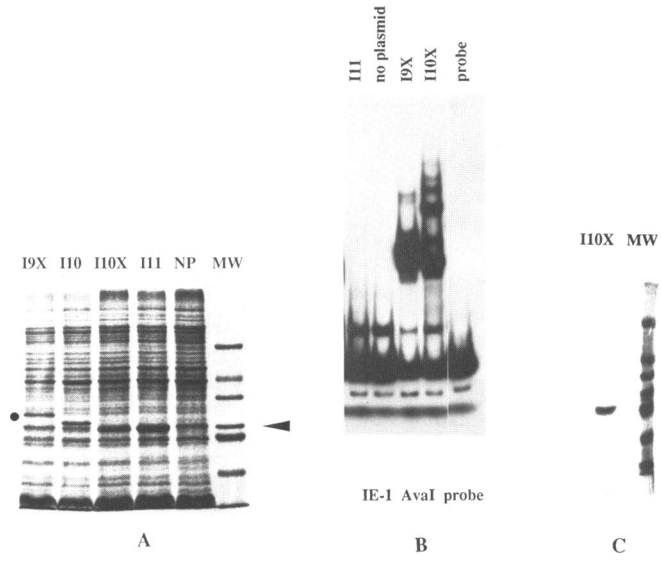

Figure 2. A. Expression of the Vmw175 DNA binding domains in E. coli. Bacteria were grown, expression from the T7 plasmids induced and crude extracts were prepared by ammonium sulphate precipitation as detailed in Methods. The extracts were analysed on a 12.5% SDS-polyacrylamide gel. The amount of each extract loaded corresponded to that produced from about 3ml of induced bacterial culture. I9X, I10, I10X and I11 indicate extracts prepared from bacteria harbouring pT7I9X, pT7I10, pT7I10X and pT7I11 respectively. (pT7I10 is the parent of pT7I10X without the XbaI stop codon linker at the end of the expressed coding region). NP is an extract prepared from induced bacteria which did not carry an expression plasmid. MW indicates the molecular weight standards: BSA, 66kd; egg albumin, 45kd; glyceraldehyde 3 phosphate dehydrogenase, 36kd; carbonic anhydrase, 29kd; trypsinogen, 24kd; trypsin inhibitor, 20.1kd. The I10, I10X and I11 proteins can be seen migrating close to carbonic anhydrase, as indicated by the arrow. The I9X protein is marked by a dot. B. Gel retardation experiments using partially purified non-fusion expression proteins. Suitable volumes of extracts were incubated with the IE-1 AvaI probe and the resultant complexes analysed by gel electrophoresis. The position of the uncomplexed probe is shown by the track on the right; the slower migrating DNA-protein complexes which are specific to the non-fusion expression proteins are obvious. C. Purification of protein I10X. A sample of the final purified product is shown next to molecular weight markers (the additional marker band visible on this gel is lactalbumin, 14.2kd). The marker track was at the edge of the gel, which ran slightly more slowly than tracks in the middle, which explains the apparent slightly higher than expected mobility of the I10X protein.

To check that these expressed proteins were biologically active, the extracts were used in gel retardation assays. The probe was an AvaI fragment from the IE-1 promoter which includes a strong Vmw175 binding site. Figure 2B confirms that extracts containing I9X and I10X produced DNA-protein complexes which were neither present in extracts containing I11 nor in extracts from bacteria without an expression plasmid. The absence of complexes formed by protein I11 confirms earlier observations (23,30) that truncation of Vmw175 at residue 293 removes sequences important for DNA binding. Extracts with high binding activity tended to produce additional multiple complex bands of slower mobility which are most likely due to multiple protein binding. It is not known whether this represents protein-protein, or additional protein-DNA interactions. Confirmation of the expression of active forms of the Vmw175 DNA binding domain in this system allowed the expansion of these studies to attempt their purification.

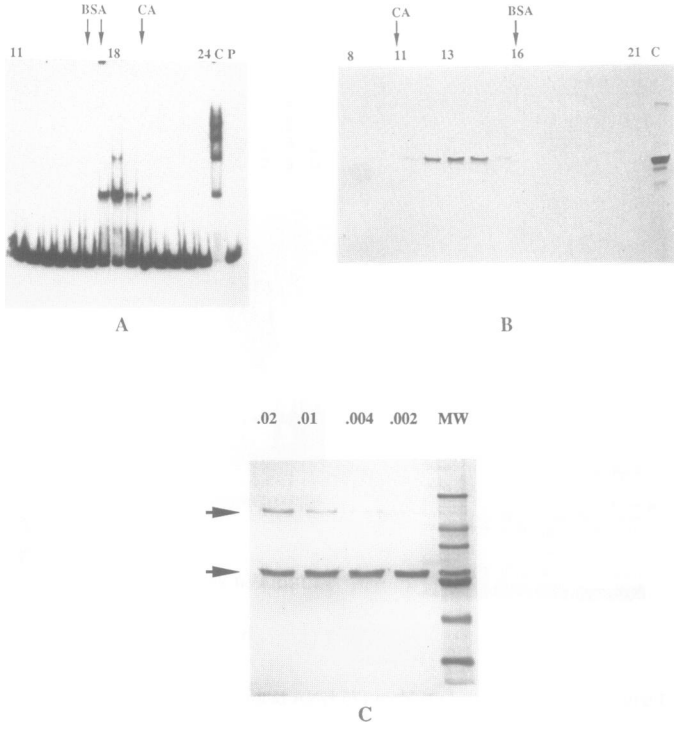

Figure 3. The I10X protein is a dimer in solution. (A) 20μg of BSA and 20μg of carbonic anhydrase (CA) were loaded onto a Superose column with 10μg of purified I10X. Fractions (0.5ml) were collected and appropriate samples (based on the uv scan) were analysed by SDS-gel electrophoresis; the positions of the BSA and CA peaks are indicated. The same samples (fractions 11 to 24) were analysed for active I10X by a gel retardation experiment using the IE-3 probe. The peak of I10X activity is shown in fraction 18. Track C contains a control with approximately 20ng of purified I10X protein as a marker (the slower mobility bands are discussed in the text). Lane P contains free probe. (B) 20μg of BSA and 20μg of CA were loaded onto a 5ml 10−30% glycerol gradient with 10μg I10X. After centrifugation at 40k rpm for 46 hours, 0.2ml fractions were collected and analysed by SDS gel electrophoresis and Western blotting (using a rabbit antiserum generated by immunisation with purified I10X protein). The results with fractions 8 to 21 are shown here; the positions of the peaks of BSA and CA on a parallel stained gel are shown in fractions 16 and 11 respectively. The I10X peak is in fraction 13. The control lane C contains unpurified I10X protein from a crude bacterial extract. C. Glutaraldehyde cross linking of protein I10X. The protein was treated with 0.02, 0.01, 0.004 and 0.002% glutaraldehyde (as indicated) at room temperature for 2 hours and then analysed on an acrylamide gel. The positions of monomer and cross-linked dimer proteins are marked. Molecular weight standards are as in Figure 2.

We have purified to apparent homogeneity protein I10X. The procedures used are described in detail in the Methods section; Figure 2C shows a Coomassie blue stained gel of the final preparation.

The purified Vmw175 DNA binding domain is a dimer in solution

Native, intact Vmw175 is a dimer (10,35). Given that many DNA binding proteins bind to DNA as dimers, it was of interest to determine whether the dimerisation of Vmw175 was an intrinsic property of the DNA binding domain itself.

A Pharmacia FPLC Superose 12 column was equilibrated with a suitable buffer (see Methods) and then used to calculate the Stokes radius of protein I10X, using bovine serum albumin and bovine carbonic anhydrase as controls (Stokes radii 3.55nM and 2.28nM respectively). The elution volumes of the three proteins were determined by uv detection, SDS gel electrophoreis (data not shown) and gel retardation experiments (Figure 3A). The I10X protein eluted as a shoulder of the BSA peak, and its Stokes radius was calculated as 3.0nM. Glycerol gradient centrifugation of the same three proteins was used to determine the sedimentation coefficient of I10X; BSA and carbonic anhydrase are 4.3S and 3.2S respectively. Comparison of sedimentation rates allowed the calculation of the sedimentation coefficient of I10X as 3.7S (Figure 3B). The Stokes radius of I10X was then used to determine its diffusion coefficient, which was then used to calculate the real molecular weight of I10X; this was found to be 48kd, close to the predicted dimeric molecular weight of 52kd.

In confirmation of these experiments, purified I10X protein was treated with increasing concentrations of the cross-linking agent glutaraldehyde, and the products analysed by gel electrophoresis. The results (Figure 3C) showed clear formation of I10X dimers. It is not known why the cross-linking was not 100% efficient. It was not apparently a consequence of the I10X molecules being in an equilibrium of dimers and monomers since gel filtration analysis of cross-linked material showed that the cross-linked dimers and the remaining un-linked protein eluted as a single peak at the position expected of a uniform preparation of dimers (data not shown).

The conclusion from these experiments is that I10X is a dimer in solution, and that the ability of Vmw175 to dimerise is most probably a consequence of a dimerisation surface within the DNA binding domain. Whether dimerisation is essential for DNA binding is not currently known. It is intriguing that the DNA binding site recognised by Vmw175 is bi-partite (see below), although the two parts have little in common. It should also be noted that I10X forms multiple complexes on a probe molecule, especially when present in large amounts (Figure 4A). If the highest mobility complex contains one dimer of I10X, then those further up the gel probably represent multiples of the basic dimer, either containing additional dimers which are recognising other sites within the probe molecule or the result of further protein-protein interactions.

Extensive mutagenesis of the Vmw175 binding site at the cap site of the IE-3 promoter reveals a functional bi-partite structure

Previous analyses of the sequence requirements of Vmw175 binding sites have shown that a proportion of these sites contain the consensus ATCGTC, that this sequence alone is insufficient for binding, that alterations in some but not all of the bases in the consensus disrupt binding and that a sequence further 3′ is

also required (17,20,23,27, 37,39). The directed mutagenesis strategy of these studies suffers from the limitation that the importance of nucleotides outside the conserved sequence could be overlooked, and the effect of alternative substitutions has not generally been investigated. The theoretical recognition of consensus sequences, although very useful, is limited because the theoretical predictions should be confirmed experimentally. In order to expand upon the previous studies, we have used a random mutagenesis method to create a family of Vmw175 binding site mutations with single point mutations at each position in the consensus and its near environment.

The site chosen for study was that at the cap site of the IE-3 promoter (19) which has been shown to have a role in the autoregulation of the IE-3 promoter, at least in transfection assays (20). An oligonucleotide encompassing the region from −18 to +17 was synthesised using nucleotide precursor reservoirs which had been contaminated with low levels of each of the other precursors. The resultant randomly mutated oligonucleotide was then used as a primer for an M13 mutagenesis reaction using the EcoRI-BamHI (−108 to +27) IE-3 fragment as template. The progeny M13 phage were screened by DNA sequence analysis. Examination of 400 plaques produced 100 useful mutants which had substitution or deletion mutations in the region of the consensus. In particular, point mutations were isolated in all bases over a 30 nucleotide stretch starting 11 bases 5′ of the A of the ATCGTC motif; at 20 of these positions alternative

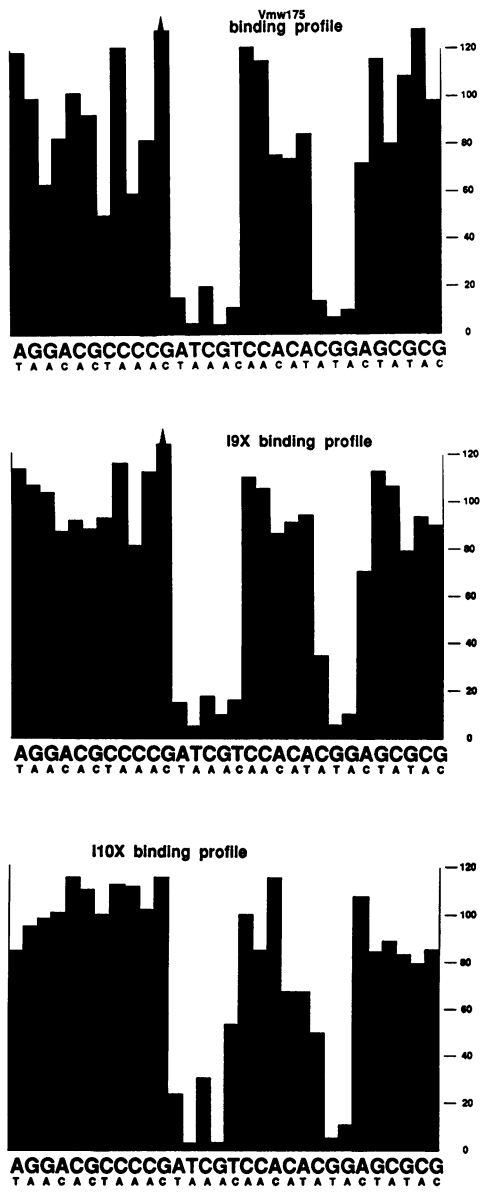

Figure 4. The effect of point mutations in the IE-3 cap site probe on the binding efficiency of Vmw175 and the I9X and I10X proteins. The sequence in the region of the Vmw175 binding site is shown in large type at the bottom of each panel. The vertical bars above each base represent the percentage efficiency of binding (as indicated on the right, compared to the wild type probe), of the relevant protein when that base was mutated to the base shown in smaller type below. In most instances the indicated change was the only one in the entire probe fragment. The family of mutant probes was used to analyse binding by Vmw175, I9X and I10X as indicated. These histograms represent a partial abstraction of the information in Table 1.

TABLE 1.

BASE	17+					I9X					I10X				
	A	C	G	T	D	A	C	G	T	D	A	C	G	T	D
A	X	/	/	118[a]	/	X	/	/	113[a]	/	X	/	/	84[a]	/
G	97	/	X	/	74	106	/	X	/	100	94	/	X	/	78
G	62	/	X	50	/	103	/	X	54	/	98	/	X	85	/
A	X	82	/	50[a]	/	X	87	/	83[a]	/	X	100	/	80[a]	/
C	101	X	/	84	/	92	X	/	87	/	115	X	/	111	/
G	58	92	X	83	NT	86	87	X	100	NT	108	110	X	98	NT
C	/	X	/	50	/	/	X	/	92	/	/	X	/	99	/
C	122	X	/	/	/	116	X	/	/	/	112	X	/	/	/
C	59	X	/	91	/	81	X	/	138	/	111	X	/	96	/
C	81	X	156	/	/	112	X	119	/	/	102	X	91	/	/
G	71	185	X	125	155	94	356	X	202	200	84	115	X	105	155
A	X	/	/	15[a]	14	X	/	/	14[a]	8	X	/	/	23[a]	14
T	5	/	/	X	1	5	/	/	X	5	3	/	/	X	4
C	20	X	19	/	/	17	X	21	/	/	30	X	47	/	/
G	4	7[a]	X	10[a]	2	10	14[a]	X	10[a]	4	3	11[a]	X	20[a]	5
T	29[a]	11	10[a]	X	5	20[a]	16	16[a]	X	32	35[a]	53	41[a]	X	57
C	121	X	94	105	/	110	X	115	99	/	99	X	83	59	/
C	115	X	97	91	12	105	X	69	145	110	84	X	80	81	85
A	X	75[a]	/	74	/	X	86[a]	/	101	/	X	115[a]	/	119	/
C	73	X	/	/	/	91	X	/	/	/	66	X	/	/	/
A	X	/	/	84	3	X	/	/	94	31	X	/	/	66	35
C	13	X	/	81	/	36	X	/	116	/	49	X	/	92	/
G	/	34[a]	X	7	9	/	18[a]	X	6	12	/	55[a]	X	5	5
G	10	/	X	15	/	10	/	X	22	/	11	/	X	13	/
A	X	72	/	/	/	X	70	/	/	/	X	108	/	/	/
G	50	/	X	118[a]	121	84	/	X	113[a]	134	94	/	X	84[a]	125
C	80	X	85	79	/	107	X	90	77	/	89	X	112	102	/
G	/	/	X	109	/	/	/	X	79	/	/	/	X	83	/
C	129	X	/	/	/	93	X	/	/	/	79	X	/	/	/
G	50	98[a]	X	/	/	86	89[a]	X	/	/	133	85[a]	X	/	/

The wild type sequence of the IE-3 capsite region is shown vertically on the left. The binding efficiency of intact Vmw175 (17^+) and the I9X and I10X proteins to the mutant probes is shown as a percentage of wild type. A slash indicates that the mutation was not isolated, NT = isolated but not tested, X = the wild type base. The numbers given for bases encompassing the core region GATCGTCCACACGG are mostly the averages of at least two separate determinations. The nature of the mutation is shown across the top row (replacement of the usual base with A, C, G or T as indicated); D represents a single base deletion. All mutations are single base changes except where indicated by a superscript a; in these cases the probe contains two point mutations. The results with double mutations have been included only where any reduction can be attributed unambiguously to the position where the result is shown in the Table. This was possible in those instances where a single change at the other position (not indicated on the Table), in an independent mutant, produced no reduction in binding. Ambiguous double mutants have not been included in this Table.

substitutions were also found, while in 8 cases all possible base substitutions were isolated. In addition, 10 single base deletions in the region of the consensus were discovered. These mutations allowed a rigorous examination of the DNA sequence requirements for binding by Vmw175 at this site.

Radiolabelled probes were prepared from all the mutant M13 templates by the prime-cut method (38). A reference wild type probe was included in each batch of probes synthesised as a control; because of the procedure used the concentrations and specific activities of all probes in a batch could be made identical, thus allowing a true comparison of relative binding affinities. The probes were used in gel retardation experiments using HSV-1 infected nuclear extracts; the results of several hundred individual incubation reactions are summarised in Figure 4 and Table 1 and the accompanying footnotes. Considering binding by intact Vmw175 in crude nuclear extracts, the mutations can be classified as follows:

(a) Positions at which all mutations which have little or no effect on binding efficiency. While the results of the gel retardation experiments were in general reproducible, the intrinsic errors involved in scanning the gels and other experimental manipulations means that, for the purposes of this analysis, this group of positions is taken to include all those which bind at the level of 50% or better. Mutations falling into this class include all those in bases more than one base 5′ of the A of ATCGTC, the last C of ATCGTC and the next 4 nucleotides, and all positions 3′ of the distal part of the consensus (CGG). The lack of importance of the last C of the ATCGTC of the IE-3 site is consistent with an identical observation with gD site II (23).

(b) Positions at which some but not all base substitutions are at least partially tolerated. Alteration of the second T of ATCGTC to an A resulted in a reduction to 30% normal binding, while changes to C or G reduced binding to less than 10%. Similarly, alteration of the C of the distal CGG part of the consensus to a T had virtually no effect, while a change to an A gave a drastic reduction. Thus a pyrimidine is sufficient at this position. Alteration of the first G of CGG to a C was also partially tolerated, in keeping with the existence of a C at this position in the gD site II (23).

(c) Positions at which all available alterations reduced binding. This could could be sub-divided into those in which mutations reduced binding to below 10% of normal (the first T and the G of ATCGTC) and those which appeared important but less crucial (the A and first C of ATCGTC, and the second G of CGG). It should be noted that not one of the single point mutations resulted in complete elimination of Vmw175 binding; extended exposure of the autoradiograms always revealed a low level of complex formation (data not shown).

(d) Positions at which base substitutions increase the efficiency of binding. The only candidate for this effect is the G immediately 5′ of the A of ATCGTC; alteration of this to a pyrimidine gave binding efficiencies of between 150 and 200%. This observation is consistent with the fact that deletion of this G (thus effectively replacing it with a C) also result in binding at about 150% of wild type.

(e) The effect of single base deletions. Single base deletions at several locations in the mutated region were isolated (Table 1). Those 5′ or 3′ of the consensus region had little effect, except that removing the G 5′ to the A of ATCGTC which, as noted above, resulted in an increase in binding efficiency. Deletion of the first T of ATCGTC gave the most severe reduction in binding efficiency of all mutations isolated, with a specific complex barely visible even after prolonged exposure of the gel. Other deletions in the proximal and distal parts of the consensus also reduced binding by at least 90%. These severe effects can be explained by an additional requirement for a precise spacing between the proximal and distal portions, since deletion of two separate single bases in the 5 nucleotide spacer (in which no point mutations had an effect) resulted in low binding efficiencies.

Therefore this analysis clearly shows that IE-3 Vmw175 binding site consists of a proximal and a distal portion, that (at least for the intact protein) the distance between the two parts is important, and that there are some relatively crucial bases within these motifs. However, there is sufficient flexibility in the binding sequence such that sequences apparently quite divergent from the consensus could bind Vmw175. In summary, the IE-3 Vmw175 binding site can be written as ATcGtnnnnnYgG where lower case letters depict bases where at least some alterations retain more than 15% binding. This is very similar to the previously published consensus sequences, but it has been arrived at after extensive experimental testing.

The bacterially expressed Vmw175 DNA binding domain proteins exhibit some relaxation of binding specificty

Binding of the I9X and I10X proteins to the family of mutant probes was also analysed. Overall, the results obtained were similar to those with intact Vmw175 (Figure 4) but with some notable differences. Firstly, the strict requirement of Vmw175 for precise spacing between the two parts of the consensus was not observed with either I9X or I10X (Table 1). Secondly, replacement of the C of CGG with an A had only a moderate effect on binding of both I9X and I10X. Thirdly, replacement of the second T of ATCGTC by any base had only moderate effects on binding by I10X (I9X gave results intermediate between I10X and whole Vmw175 with these mutations). Thus, consistent with earlier results (23,30), the binding specificities of the bacterially expressed proteins are very similar to those of the whole protein, but there is some relaxation of specificity. The lack of requirement for precise spacing of the two parts of the consensus suggests that the isolated DNA binding domain may be more flexible than the intact protein; this suggests that other parts of Vmw175 might, in principle, be able to influence the DNA binding activity of the intact protein (40).

DISCUSSION

This paper describes the expression and purification of the DNA binding domain of the major transcriptional regulator of HSV-1 and the characterisation of its binding site. It is of interest to discuss the results in terms of the nature of the DNA binding domain itself and of its target sequence, how the results with the isolated sub-fragment of the protein relate to those with the whole protein, and how DNA binding in general relates to the function of Vmw175.

Although the DNA binding domain of Vmw175 has been expressed as a fusion protein in bacteria in previous studies (30), the non-fusion proteins described in this paper have some distinct advantages. Firstly, they can be purified in milligram amounts, which opens the possibilities of future structural studies.

Secondly, that they can be expressed as stable proteins suggests that they represent a separable functional domain of the protein. It is probably relevant that incorporation into the expression vector of sequences progressing further towards the C-terminus of Vmw175 results in the synthesis of unstable polypeptides which are readily degraded (during subsequent purification) to peptides of very similar size to I9X and I10X (results not shown). That I9X and I10X are relatively stable is consistent with the release by proteolysis of the DNA binding domain of Vmw175 (29) in a fragment slightly smaller than I10X (results not shown).

The study of the DNA sequence requirements for Vmw175 binding described here is the most thorough examination of this subject to date. The results show, perhaps not surprisingly, the importance of most of the bases identified in previous consensus sequences derived from known binding sites. However, this study convincingly demonstrates the requirement for both proximal and distal parts of the consensus and that, although the region protected from DNAase I digestion extends several bases 5′ and 3′ from the consensus (13), bases outside the conserved region have little or no sequence-specific contribution to Vmw175 binding. This analysis is also consistent with the results of methylation interference experiments which identified important bases within both proximal and distal parts of the consensus (19). We have also shown that the bases within the consensus are not equal in their contribution to binding affinity, and that no single base is absolutely required. The implication is that, within a binding affinity covering approximately an order of magnitude, that the binding specificity of Vmw175 is quite relaxed. This is consistent with the conclusions of a recent study (24) which suggested, on the basis of a statistical comparison of known Vmw175 binding sites, that sites could occur at frequent locations throughout the HSV-1 genome. It would have been possible to have continued to isolate mutations until all possible single base changes had been found. However, given the flexibility in binding, more information might be obtained by using methods which would detect all possible sequences (including those not obviously related to the IE-3 cap site) that are actually bound by Vmw175.

The slight relaxation of binding specificity observed with the non-fusion proteins compared to the intact protein, including the loss of a requirement for exact spacing between the proximal and distal parts of the consensus, suggests some interesting possibilities. Indeed, the relaxation of binding specificity is even more pronounced with the proteinase K resistant domain of Vmw175. This exhibits a similar binding specificity to the whole protein when wild type and mutant oligonucleotide probes are used, but when such mutants are incorporated into a larger probe fragment, binding after PK treatment becomes less specific (data not shown). If Vmw175 were normally to undergo proteolysis in the infected cell, it seems likely that the DNA binding domain might remain as a relatively stable entity which could then bind to a variety of sites throughout the genome. This could have a role in the replicative cycle of the virus.

Lastly, it is pertinent to discuss the role of DNA binding in Vmw175 function. While much early work stressed the correlation of mutations which affected DNA binding with defects in transactivation and repression by Vmw175 (25,28) some recent results suggest that the situation is not quite so simple. In particular, the ability of Vmw175 to bind to the 'non-consensus' sites upstream of the tk promoter is dispensible for tk promoter activation during infection (22), and deletion from the viral genome of three binding sites in the vicinity of the gD promoter has no affect on the levels of gD expression during infection (manuscript in preparation). It is possible to reconcile these results by suggesting that Vmw175 acts as a trans-acting enhancer, such that binding to any one of the potentially large number of binding sites in the coding or regulatory regions of a gene would enable its function. Thus lack of binding at any one site would not be detrimental. However, a virus has been isolated which grows normally in tissue culture yet contains two mutations in DNA binding domain of Vmw175 which greatly reduce its ability to bind to DNA (41). While this result seems to suggest strongly that DNA binding by Vmw175 is not obligately connected to its function, it does not explain why the DNA binding domain of Vmw175 is so highly conserved in all related proteins (31–34) and why so many of the mutations which affect DNA binding also disrupt function. Clarification of this point, and definition of the functions of the DNA binding domain of Vmw175, requires a thorough mutagenic study of the domain itself. This work is in progress.

ACKNOWLEDGEMENTS

The authors are members of the Medical Research Council Virology Unit. We are grateful for the encouragement of Professor J.H.Subak-Sharpe, for the synthesis of oligonucleotides by Dr. J.McLaughlan, and for advice on the use of the FPLC from Dr. J.Connor. Dr Howard Marsden and Miss Gillian McVey kindly helped with the generation of a polyclonal rabbit antiserum specific for the I10X protein. Dr Ben Luisi provided much helpful criticism of the manuscript.

REFERENCES

1. McGeoch, D.J., Dalrymple, M.A., Davison, A.J., Dolan, A., Frame, M.C., McNab, D., Perry, L.J., Scott, J.E. and Taylor, P. (1988). J. Gen. Virol. 69, 1531–1574.
2. Everett, R.D. (1987). Anticancer Research 7, 589–604.
3. Wagner, E.K. (1985). In: The Herpesviruses. Ed. Roizman, B. Plenum Press, New York/London. 3, 45–104.
4. Roizman. B. and Sears, A.E. (1990). In: Virology, 2nd edition. Eds Fields, B.N., Knipe, D.M. et al., Raven Press, New York.
5. Preston, C.M. (1979). J. Virol. 29, 275–284.
6. DeLuca, N.A., McCarthy, A.M. and Schaffer, P.A. (1985). J. Virol. 56, 558–570.
7. Pereira, L., Wolff, M.H., Fenwick, M. and Roizman, B. (1977). Virology, 77, 733–749.
8. Hay, R.T. and Hay, J. (1980). Virology, 104, 230–234.
9. Wilcox, K.W., Kohn, A., Sklyanskay, E. and Roizman, B. (1980). J. Virol. 33, 167–182.
10. Metzler, D.W. and Wilcox, K.W. (1985). J. Virol. 55, 329–337.
11. Kattar-Cooley, P. and Wilcox, K.W. (1989). J. Virol. 63, 696–704.
12. Faber, S.W. and Wilcox, K.W. (1986). Nucl. Acids Res. 14, 6067–6083.
13. Faber, S.W. and Wilcox, K.W. (1988). Nucl. Acids Res. 16, 555–570.
14. Beard, P., Faber, S., Wilcox, K.W. and Pizer, L.I. (1986). Proc. Nat. Acad. Sci., U.S.A. 83, 4016–4020.
15. Kristie, T.M. and Roizman, B. (1986). Proc. Nat. Acad. Sci., U.S.A. 83, 3218–3222.
16. Kristie, T.M. and Roizman, B. (1986). Proc. Nat. Acad. Sci., U.S.A. 83, 4700–4704.
17. Michael, N. and Roizman, B. (1989). Proc. Nat. Acad. Sci., U.S.A. 86, 9808–9812.
18. Michael, N., Spector, D., Mavromara-Nazos, P., Kristie, T.M. and Roizman, B. (1988). Science, 239, 1531–1534.
19. Muller, M.T. (1987). J. Virol. 61, 858–865.
20. Roberts, M.S., Boundy, A., O'Hare, P., Pizzorno, M.C., Ciufo, D.M. and Hayward, G.S. (1988). J. Virol. 62, 4307–4320.
21. Tedder, D.G., Everett, R.D., Wilcox, K.W., Beard, P. and Pizer, L.I. (1989). J. Virol. 63, 2510–2520.
22. Imbalzano, A.N., Shepard, A.A. and DeLuca, N.A. (1990). J. Virol. 64, 2620–2631.

23. Pizer, L.I., Everett, R.D., Tedder, D.G., Elliott, M. and Litman, B. (1991). Nucl. Acids Res. 19, 477–483.
24. DiDonato, J.A., Spitzner, J.R. and Muller, M.T. (1991). J. Mol. Biol. 219, 451–470.
25. Paterson, T. and Everett, R.D. (1988). Nucl. Acids Res. 16, 11005–11025.
26. Paterson, T. and Everett, R.D. (1988). Virology, 166, 186–196.
27. DeLuca, N.A. and Schaffer, P. (1988). J. Virol. 62, 732–743.
28. Shepard, A.A., Imbalzano, A.N. and DeLuca, N.A. (1989). J. Virol. 63, 3714–3728.
29. Everett, R.D., Paterson, T. and Elliott, M. (1990). Nucl. Acids Res. 18, 4579–4585.
30. Wu, C.L. and Wilcox, K.W. (1990). Nucl. Acids Res. 18, 531–538.
31. McGeoch, D.J., Dolan, A., Donald, S. and Brauer, D.H.K. (1986). Nucl. Acids Res. 14, 1727–1764.
32. Cheung, A.K. (1989). Nucl. Acids Res. 17, 4637–4646.
33. Grundy, F.J., Baumann, R.P. and O'Callaghan, D.J. (1989). Virology, 172, 223–236.
34. Vlcek, C., Paces, V. and Schwyzer, M. (1989). Virus Genes 2, 335–346.
35. Shepard, A.A. and DeLuca, N.A. (1989). J. Virol. 63, 1203–1211.
36. Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990). Methods in Enzymology 185, 60–89.
37. Everett, R.D. and Orr, A. (1991). Virology 180, 509–517.
38. Biggin, M., Farrell, P.J. and Barrell, B.G. (1984). EMBO J. 3, 1083–1090.
39. Resnick, J., Boyd, B.A. and Haffey, M.L. (1989). J. Virol. 63, 2494–2503.
40. Paterson, T., Preston, V.G. and Everett, R.D. (1990). J. Gen. Virol. 71, 851–861.
41. Sheperd, A.A. and DeLuca, N.A. (1991). J. Virol. 65, 787–795.