T Antigen Origin-Binding Domain of Simian Virus 40: Determinants of Specific DNA Binding†

Elizabeth M. Bradshaw,‡ David G. Sanford,‡ Xuelian Luo,§ James L. Sudmeier,‡ Zachary A. Gurard-Levin,‡ Peter A. Bullock,‡ and William W. Bachovchin*,‡

Department of Biochemistry, Tufts University School of Medicine, Boston, Massachusetts 02111, and Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75235-9041

Received October 27, 2003; Revised Manuscript Received February 19, 2004

ABSTRACT: To better understand origin recognition and initiation of DNA replication, we have examined by NMR complexes formed between the origin-binding domain of SV40 T antigen (T-ag-obd), the initiator protein of the SV40 virus, and cognate and noncognate DNA oligomers. The results reveal two structural effects associated with "origin-specific" binding that are absent in nonspecific DNA binding. The first is the formation of a hydrogen bond (H-bond) involving His 203, a residue that genetic studies have previously identified as crucial to both specific and nonspecific DNA binding in full-length T antigen. In free T-ag-obd, the side chain of His 203 has a pK_a value of ∼5, titrating to the $N^{\epsilon 1}H$ tautomer at neutral pH (Sudmeier, J. L., et al. (1996) *J. Magn. Reson., Ser. B 113*, 236−247). In complexes with origin DNA, His 203 $N^{\delta 1}$ becomes protonated and remains nontitrating as the imidazolium cation at all pH values from 4 to 8. The H-bonded $N^{\delta 1}H$ resonates at 15.9 ppm, an unusually large N−H proton chemical shift, of a magnitude previously observed only in the catalytic triad of serine proteases at low pH. The formation of this H-bond requires the middle G/C base pair of the recognition pentanucleotide, GAGGC. The second structural effect is a selective distortion of the A/T base pair characterized by a large (0.6 ppm) upfield chemical-shift change of its Watson−Crick proton, while nearby H-bonded protons remain relatively unaffected. The results indicate that T antigen, like many other DNA-binding proteins, may employ "catalytic" or "transition-state-like" interactions in binding its cognate DNA (Jen-Jacobson, L. (1997) *Biopolymers 44*, 153−180), which may be the solution to the well-known paradox between the relatively modest DNA-binding specificity exhibited by initiator proteins and the high specificity of initiation.

Initiation of DNA replication is a complex process that begins with the binding of an "initiator" protein or protein complex to an origin of DNA replication (*1*, *2*). Biochemical and genetic studies have provided considerable insight into this process (*3*), but at the molecular level, it is still not well-understood. This is especially true with respect to higher eukaryotic organisms, where DNA origins have proved difficult to identify and where limited structural information is not yet available on the origin-recognition complex (ORC), the eukaryotic cellular initiator (*3*). In contrast, origins of replication have been identified for several viral systems, and structural information is available for the initiator proteins employed by simian virus 40 (SV40)[1], bovine papillomavirus (BPV), and adeno-associated virus (AAV) (*4*−*6*), including a structure of the papillomavirus initiator protein bound to an origin sequence (*7*). Thus, at present, viral systems offer better prospects for advancing the understanding of molecular mechanisms underlying origin recognition and initiation of replication.

SV40 T antigen (T-ag), a 708 amino acid protein with 2 phosphorylation regions, is the initiator protein of the SV40 virus (*8*−*10*). It performs at least four distinct functions during initiation of DNA replication: (1) origin recognition, (2) DNA melting, (3) recruitment of other proteins needed for replication, and (4) unwinding, serving as the helicase that unwinds the DNA in front of the replication fork. Immediately after recognition, T-ag assembles into hexamers and double hexamers. These higher order complexes are responsible for functions 2−4 above (*8*−*10*). The SV40 core origin is a 64-base-pair segment (*11*) consisting of a central region containing four GAGGC sequences arranged as inverted pairs, flanked by an adenine-thymine (AT)-rich domain and an early palindrome (EP). Origin recognition involves T-ag binding specifically to the GAGGC sequence.

† This work was supported by grants from the National Institutes of Health (5RO1GM53618 and 5RO1GM055397) awarded to W.W.B. and P.A.B., respectively.

* To whom correspondence should be addressed. E-mail: william.bachovchin@tufts.edu, Phone: 617-636-6881, Fax: 617-636-2409.

‡ Tufts University School of Medicine.

§ University of Texas Southwestern Medical Center.

[1] Abbreviations: 1D, one dimensional; 2D, two dimensional; AAV, adeno-associated virus; BPV, bovine papillomavirus; D_2O, deuterium oxide; DNA, deoxyribonucleic acid; E1-DBD, E1 DNA-binding domain; H-bond, hydrogen bond; H148Q, H192Q, H201Q, and H203Q, T-ag-obd with a His mutated to Gln; HSMQC, heteronuclear single- and multiple-quantum coherence; HSQC heteronuclear single-quantum coherence; IPTG, isopropyl-β-D-thiogalactopyranoside; K_d, dissociation constant; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; pD, negative logarithm of the deuterated hydronium ion concentration; RMSD, root-mean-square deviation; SV40, Simian virus 40; T_2, transverse relaxation; T-ag, T antigen; T-ag-obd, T antigen origin-binding domain; T155S, T-ag-obd with Thr 155 mutated to Ser.

10.1021/bi030228+ CCC: $27.50
Published on Web 05/11/2004

Our overall goal is to understand the molecular mechanisms underlying origin recognition and initiation of DNA replication. Toward this goal, we have previously expressed the T-ag origin-binding domain (T-ag residues 131−260) of SV40 T-ag and solved its solution structure using NMR spectroscopy (*4*). This domain, termed "T-ag-obd", interacts sequence-specifically with the SV40 origin, with an affinity comparable to full-length T-ag (*12, 13*).

The surface of the T-ag-obd that interacts with DNA was mapped by monitoring the changes in the NMR spectrum upon binding of oligonucleotides containing the specific recognition pentanucleotide GAGGC sequence or a nonspecific AGAAT sequence. The NMR results confirmed previous mutagenesis data that both sequence-specific and nonspecific binding is largely mediated by two elements in T-ag-obd, termed "A1" and "B2" (*14*). These elements exhibit the largest chemical-shift displacements on binding DNA and are adjacent in the structure. A1 forms a loop, and B2 is part of a helix and a loop. The fact that the same surface of T-ag-obd is involved in both specific origin recognition and nonspecific DNA binding is not surprising, owing to the substantial chemical and structural similarities shared by all DNA sequences. In fact, it is difficult to imagine how a site-specific DNA-binding protein could be devoid of nonspecific DNA-binding activity, which nearly all sequence-specific DNA-binding proteins exhibit to some degree (*15*). However, this raises an important question: How do sequence-specific DNA-binding proteins manage to function in the presence of such immense excesses of competing, nonspecific DNA binding sites? The most obvious explanation is that sequence-specific binding proteins bind their cognate sequences with much higher affinities than their noncognate sequences. T-ag-obd does bind more tightly to a GAGGC- than to an AGAAT-containing oligonucleotide (*4, 13*). Titolo et al. reported K_d values of 31 nM and 0.1 μM, respectively, using fluorescence spectroscopy. This difference is insufficient to account for the high specificity of initiation in vivo. This inconsistency seems to apply to all initiator proteins, a fact which has been pointed out and discussed in detail by Stenland (*16*).

Solving the structure of the complex between T-ag-obd and origin DNA by NMR has been a specific objective of our efforts to understand origin recognition and initiation. Unfortunately, for reasons not yet fully understood, the T_2 values of the α-carbon atoms in the complex are too short to permit NMR experiments needed to assign side-chain resonances, which has thus far prevented the solution of the structure. We have, however, been able to use NMR to identify and characterize interactions that distinguish the specific DNA binding of T-ag-obd from nonspecific DNA binding. One such difference is that specific complexes exhibit a ^{1}H NMR signal at 15.9 ppm (*4*). Such downfield-shifted proton NMR signals are quite rare in proteins and normally indicative of "catalytic" hydrogen bonds (H-bonds), such as that of the Asp-His dyad in the active sites of serine proteases (*17*). A second difference is that T-ag-obd selectively displaces the H-bonded proton resonance in the A/T base pair of the recognition pentanucleotide upfield by 0.6 ppm, suggesting a selective weakening of the A/T base-pair interaction. This paper reports and characterizes these interactions and discusses their role in origin recognition and how they might solve the paradox between apparently low DNA-binding specificity of T-ag-obd and other initiator proteins and the high specificity of initiation.

MATERIALS AND METHODS

Preparation of T-ag-obd. The cloning, expression, and purification of T-ag-obd (residues 131−260 of T-ag) have been described previously (*12*). Briefly, the DNA coding this fragment of T-ag was cloned into the pGEX-1λT vector and transformed into *Escherichia coli* BL21 cells. The cells were grown to mid-log phase and induced with IPTG. Harvested cells were lysed and the glutathione-*S*-transferase/T-ag-obd fusion protein was separated from the lysate by adsorption onto glutathione-agarose affinity beads. The T-ag-obd was removed from the glutathione-*S*-transferase by digestion with thrombin. The affinity-purified material was further purified by gel-filtration chromatography on a S-100 column (Pharmacia) and concentrated to between 0.35 and 2.2 mM by ultrafiltration (Amicon Centriprep three filters). Protein concentration was determined by the UV absorbance at 280 nm using an extinction coefficient of 9140 M^{-1} cm^{-1}, calculated from the amino acid sequence.

^{15}N Labeling of T-ag-obd. For uniform enrichment of T-ag-obd with ^{15}N, the cells were grown in minimal M9 media that contained $^{15}NH_4Cl$ as the sole nitrogen source. For specific labeling of the $N^{\delta 1}$ of the histidines, the pGEX-1λT−T-ag-obd plasmid DNA was transformed into *E. coli* SG298, which is an auxotroph for histidine, arginine, and leucine. These cells were grown in a rich, defined media that contained 120 mg/L of singly labeled $^{15}N^{\delta 1}$-histidine (L-histidine-ring-π-^{15}N, Cambridge Isotope Laboratories). Also, unlabeled arginine and leucine were added. The specifically labeled T-ag-obd was purified from these cells as described above.

Site-Directed Mutagenesis. Site-directed mutants H148Q, H192Q, H201Q, and H203Q were generated from pGEX-1λT−T-ag-obd using the QuikChange site-directed mutagenesis kit (Stratagene). A total of 5−50 ng of pGEX-1λT−T-ag-obd plasmid DNA were amplified with 125 ng of each primer and 2.5 units of Pfu DNA polymerase (Stratagene). The primers for H201Q were forward 5′-CTGTTTTTTCTTACTCCACAAAGGCATAGAGTGTCTGCT-3′; reverse, 5′-AGCAGACACTCTATGCCTTTGTGGAGTAAGAAAAAACAG-3′. The primers for H203Q were forward, 5′-CTTACTCCACACAGGCAAAGAGTGTCTGCTATTAATAAC-3′; reverse, 5′-AATAGCAGACACTCTTTGCCTGTGTGGAGTAAG-3′. The primers for H148Q were forward, 5′-CTAAGTTTTTTGAGTCAAGCTGTGTTTAGTAATAG-3′; reverse, 5′-CTATTACTAAACACAGCTTGACTCAAAAAACTTAG-3′. The primers for H192Q were forward, 5′-CATAACAGTTATAATCAAAACATACTGTTTTTTCTTACTCC-3′; reverse, 5′-GGAGTAAGAAAAAACAGTATGTTTTGATTATAACAGTTATG-3′. The reactions were cycled 2 times from 95 °C for 30 s, to 37 °C for 1 min, and then to 68 °C for 16 min and then cycled 12 times from 95 °C for 30 s, to 55 °C for 1 min, and then to 68 °C for 11 min. The PCR products were digested with 20 units of *Dpn* I (New England Biolabs) for 2 h at 37 °C. The digested PCR products were transformed into XL1-blue electrocompetent cells. The plasmid DNA was isolated using a plasmid mini kit (Qiagen) and sequenced to confirm the His to Gln mutations. This plasmid DNA was

then transformed into *E. coli* BL21 cells, and the protein was purified as described for the wild-type T-ag-obd. The T155S mutant was made previously (*18*).

Oligonucleotides. Deoxyoligonucleotides used in T-ag-obd binding studies were purchased from Oligos Etc. The two complementary oligonucleotides were d(GCAGAGGCCGA) and d(TCGGCCTCTGC), herein referred to as "GAGGC". Also, the duplex of d(GCAAGAATCGA) and d(TCGATTCTTGC) was prepared and is referred to as "AGAAT". The DNA for "CAGGC" is d(GCACAGGCCGA) and d(TCGGCCTGTGC); "GTGGC" is d(GCAGTGGCCGA) and d(TCGGCCACTGC); "GACGC" is d(GCAGACGCCGA) and d(TCGGCGTCTGC); "GAGCC" is d(GCAGAGCCCGA) and d(TCGGGCTCTGC); and "GAGGG" is d(GCAGAGGGCGA) and d(TCGCCCTCTGC). The two complementary oligonucleotides were mixed in a 1:1 stoichiometry based on the UV absorbance at 260 nm in H_2O. The duplex oligomers to be added to the T-ag-obd sample were then dried on a Speed-Vac evaporator (Savant) and redissolved in the NMR sample.

NMR Experimental Details. All NMR samples contained 100 mM potassium chloride, 10 mM potassium phosphate (pH 7.2), 1 mM dithiothreitol, and 7 mM magnesium chloride. Except for the pH titrations, all samples were adjusted to pH 5.5 with dilute hydrochloric acid. The 1D ^{1}H NMR spectra were collected either at 600 MHz on a Bruker DRX 600, at 500 MHz on a Bruker AMX 500 spectrometer, or at 300 MHz on a DPX 300 spectrometer. All 2D spectra were recorded on the AMX 500 spectrometer. Spectra were recorded at 30 °C, and ^{1}H chemical shifts were referenced versus HOD at 4.707 ppm. ^{15}N chemical shifts are reported relative to liquid NH_3.

To observe the low-field protons, which exchange rapidly with H_2O, solvent excitation was avoided using the "hard-11" pulse sequence, and the dynamic range problem was overcome by oversampling (*19*, *20*). Some spectra were recorded with selective ^{15}N decoupling by continuous wave (CW) irradiation of the ^{15}N spectrum at low power while observing the low-field proton spectrum. The CW power level was adjusted to achieve effective decoupling of no more than ±2.0 ppm. The offset frequency of the decoupler was then incremented by 2 ppm through the region of the histidine side-chain ^{15}N resonances to determine the chemical shift of the nitrogen atom attached to the low-field proton.

The 2D ^{1}H/^{15}N correlation between the amide protons and the covalently bound nitrogen atoms was achieved by heteronuclear single-quantum coherence (HSQC) (*21*) spectroscopy with a spin-locking pulse to suppress the water signal (*22*). The 2D correlation between ^{15}N and carbon-bound ^{1}H was done with heteronuclear single- and multiple-quantum coherence (HSMQC) experiments (*23*, *24*). A fifth order polynomial baseline correction ("ABS2") was performed in the ^{1}H dimension. These spectra were processed in the ^{1}H dimension with a Lorentzian−Gaussian window function using constants LB = −10 Hz and GB = 0.08, and the ^{15}N dimension is processed with a sin^2 window function using constant SSB = 2.

RESULTS

Assignment of the Low-Field Proton. The low-field ^{1}H resonance at 15.9 ppm exhibited by the T-ag-obd, GAGGC

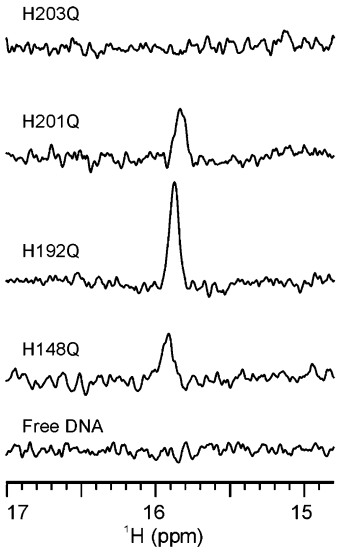

FIGURE 1: Low-field "Hard-11" 500-MHz proton spectra of ^{15}N T-ag-obd mutants 0.4 mM H148Q, 0.6 mM H192Q, 0.6 mM H201Q, and 0.8 mM H203Q with a molar excess of GAGGC at pH 5.5 and 30 °C. The mutations of His 148, 192, and 201 to Gln do not eliminate the low-field resonance.

DNA complex shows ^{15}N spin coupling with a $^1J_{H-N}$ of 97 Hz when the protein is uniformly ^{15}N-labeled. This rules out the possibility that the signal arises from the DNA. Selective ^{15}N-decoupling experiments determined the ^{15}N chemical shift to be 182.2 ppm (data not shown) and independent of pH over the range 5.2−7.6. HSQC spectra independently confirmed these findings. The ^{15}N chemical shift suggests that the proton is a histidine N−H because only imidazole ring nitrogen atoms normally resonate in this region of the spectrum. The resonance remains a doublet in the spectra of T-ag-obd specifically ^{15}N-labeled only at $N^{\delta 1}$ of the histidines, confirming that the nitrogen causing the splitting arises from a histidine and further identifying it as a histidine $N^{\delta 1}$ as opposed to a $N^{\epsilon 2}$ nitrogen. Because neither the low-field proton nor the nitrogen to which it is attached undergoes chemical shift changes with pH, the histidine is one that does not titrate in the complex.

Histidine Mutants. To help identify which of the six histidines gives rise to the low-field ^{1}H signal, each of the histidines was individually mutated to glutamine, but only four clones, H148Q, H192Q, H201Q, and H203Q, yielded sufficient protein for a ^{1}H NMR sample. The first three of these, H148Q, H192Q, and H201Q, exhibited the 15.9-ppm signal on addition of GAGGC (Figure 1), eliminating these histidines as the source of the low-field signal. The H203Q mutant does not exhibit the low-field signal on addition of GAGGC (Figure 1). However, this mutant appears to be devoid of all DNA binding activity as evidenced by the absence of chemical-shift changes and broadening of both protein and DNA signals associated with binding of wild-type T-ag-obd to GAGGC. Therefore, the absence of the low-field signal in the His 203 mutant does not prove that His 203 is the origin of the proton signal. Interestingly, HSQC spectra of the H203Q mutant is quite similar to wild-type T-ag-obd, indicating that H203Q folds correctly and has an overall structure very similar to that of wild-type T-ag-obd (Figure 2). The defect in DNA binding is therefore a specific, local effect of the H203Q mutation and not due to gross structural alterations. In support of this conclusion, H203Q-

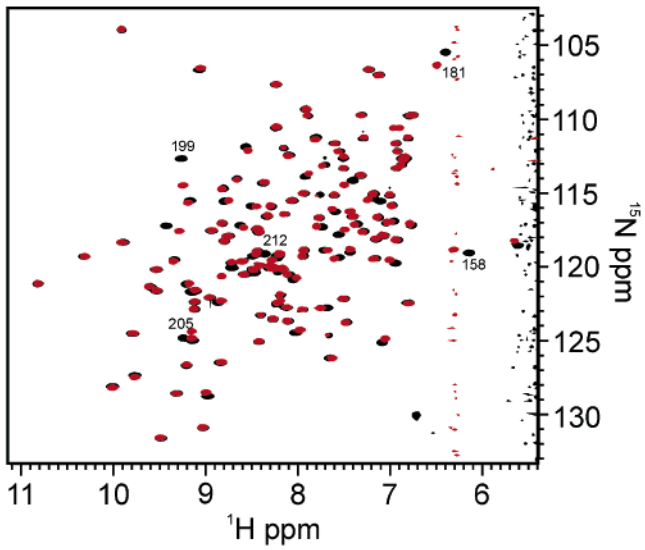

FIGURE 2: Overlaid ^{1}H/^{15}N HSQC 300-MHz spectra of 0.8 mM ^{15}N H203Q (red) and 1.3 mM ^{15}N wild-type T-ag-obd (black) at pH 5.5 and 30 °C. While the majority of peaks do not shift in the mutant, a small number do in the wild type, mostly peaks that are close to His 203 in the structure.

induced chemical-shift displacements are largely confined to residues that are spatially close to His 203, such as Val 205, Thr 199, and Cys 158.

^{15}N pH Titrations of the Histidines. Histidines are chemically active at physiological pH; either of the two imidazole ring nitrogen atoms can donate or accept protons, engage in H-bonding as either a H-bond donor or acceptor, or act as a nucleophile in attacking and binding electrophilic centers such as metals. The ^{15}N chemical shifts of the imidazole ring nitrogen atoms are remarkably informative about these interactions (*25*). Previously, Sudmeier et al. reported the resonance assignments for all of the histidine side-chain atoms (carbon, nitrogen, and proton) in free T-ag-obd at 35 °C in D_2O at a fixed pH (now known to be 5.8 when corrected for several effects, including the deuterium isotope effect upon glass-electrode response, not 5.5 as reported) (*26*).

In free T-ag-obd, none of the histidine ring nitrogen atoms have a chemical shift of 182.2, which indicates that DNA binding must alter the NMR parameters of at least one of the six histidines. Figure 3A shows the titration curves of His 187, His 203, and His 255 in free T-ag-obd. One of these three histidines must be the source of the low-field proton in the complex, because the mutagenesis experiments have ruled out the other three. Note that His 187 does not titrate and exists exclusively as the $N^{\delta 1}$-tautomeric form (*27*). Binding of T-ag-obd to GAGGC severely broadens all resonances, including the histidine ^{15}N signals. In fact, the signals from His 203 are broadened beyond detection at most pH values, in either HSQC or direct ^{15}N-observed spectra. However, pH curves could be completed for His 187 and His 255 in the complex (Figure 3B). The results show these are essentially the same as the corresponding curves in free T-ag-obd, indicating that neither His 187 nor His 255 can be the source of the low-field peak. Therefore, His 203 must be the source of the low-field proton signal.

Formation of the His 203 H-Bond Requires the Middle G/C Base Pair of the Recognition Pentanucleotide. To

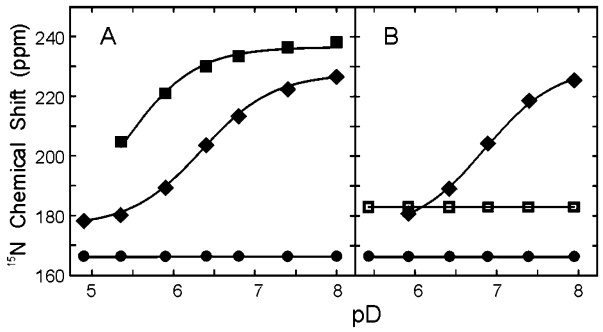

FIGURE 3: pH titrations of histidine $N^{\delta 1}$ resonances of 2 mM ^{15}N,^{13}C T-ag-obd and of T-ag-obd bound to GAGGC at 30 °C as determined by 500-MHz HSMQC experiments correlating carbon-bound protons to nitrogens. (A) $N^{\delta 1}$ free and (B) $N^{\delta 1}$ bound. Histidine symbols are as follows: (●) 187, (■) 203, and (◆) 255. In B the chemical shifts of His 203 are added from the selective decoupling experiment, and the symbol is (□).

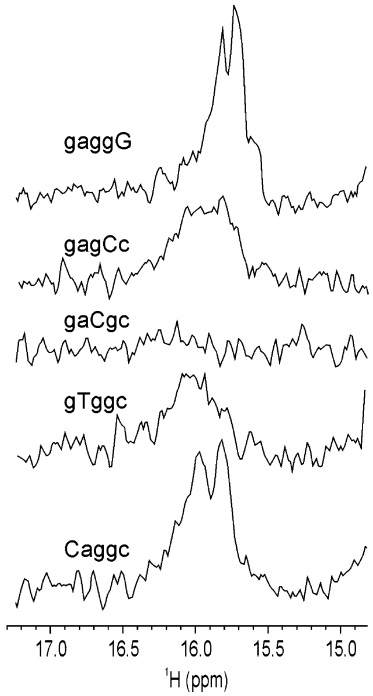

FIGURE 4: Low-field "Hard-11" 500-MHz proton spectra of 0.5 mM ^{15}N T-ag-obd complexed to 2.0 mM DNA at pH 5.5 and 30 °C, with transversion changes one base pair at a time of the recognition pentanucleotide, GAGGC. The 15.9-ppm peak is observed in all of the DNA except GACGC.

identify the DNA requirements for the appearance of the low-field signal, one base pair at a time of the recognition pentanucleotide was changed. Amide chemical-shift changes indicated that none of the mutations prevent the binding of T-ag-obd to DNA. Mutating the first and last base pairs of GAGGC does not affect the 15.9 ppm peak (Figure 4). Mutating the second and fourth base pairs broadens the peak (Figure 4). Mutating the third base pair, G/C to C/G (Figure 4) or to A/T (data not shown), prevents the appearance of the low-field signal.

DNA Structural Distortions. Binding of T-ag-obd to the GAGGC-containing oligonucleotide causes the proton resonance of the A/T base pair in the recognition pentanucleotide to move upfield 0.6 ppm (Figure 5). The H-bonded ^{1}H resonances of the adjacent base pairs exhibit little or no effect. The large and selective displacement in the position

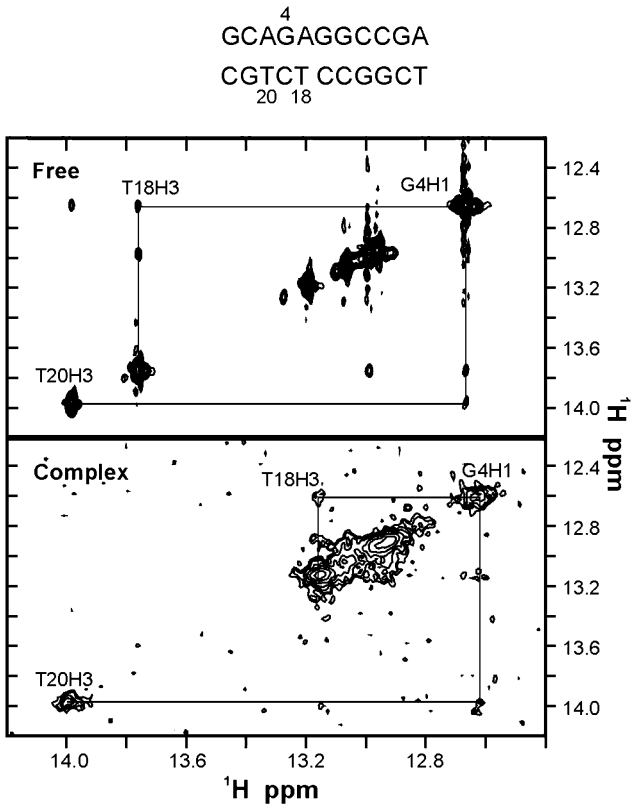

FIGURE 5: Low-field NOESY experiments (500 mHz) show the connectivities between the base pairs. In the free DNA, T18H3 is at 13.75 ppm and upon T-ag-obd binding shifts upfield to 13.15 ppm. The free DNA concentration was 2.5 mM. The complexed sample contained 2.2 mM T-ag-obd and 1.33 mM DNA, and the experiments were examined at pH 5.5 and 30 °C.

of the A/T H-bonded proton indicates T-ag-obd causes a selective structural distortion centered at the A/T base pair. Such a displacement of a base-pair H-bonded resonance does not occur when T-ag-obd binds to the nonspecific AGAAT-containing oligonucleotide (*4*).

The structural distortion also occurs if the A/T base pair is reversed to a T/A base pair (Figure 6A). This complex also exhibits the low-field proton signal at 15.9 ppm (Figure 4). However, replacing the A/T base pair with a C/G pair eliminates both the upfield displacement of a H-bonded proton (which in this case is a C/G base pair) and the low-field proton signal. Interestingly, replacement of the A/T with a G/C base pair does not eliminate the low-field proton signal (Figure 6B). However, owing to overlap in the NMR spectra, we were unable to determine if the G/C base-pair proton resonance is displaced upfield.

The middle G/C base pair, previously demonstrated to be critical for forming the His 203 H-bond, is also critical for the T-ag-induced distortion of the A/T base pair. Even a transversion mutation at this position, i.e., G/C to C/G, prevents both the appearance of the low-field proton signal (Figure 4) and the displacement of the A/T base-pair proton resonance (data not shown).

T155S Binds GAGGC Specifically. Mutating Thr 155 to Ser in full-length T-ag has been reported to result in the loss of specific DNA-binding ability (*14*). The T155S mutant of T-ag-obd, however, on binding GAGGC, exhibits both the low-field proton resonance at 15.9 ppm (data not shown) and the large, selective upfield displacement of the A/T base-pair proton resonance (Figure 7) associated with specific DNA binding of wild-type T-ag-obd. One notable difference between wild-type T-ag-obd and the T155S mutant complexes is the solvent-exchange rate of the thymidine proton resonance. The exchange rate is slow/intermediate on the NMR time scale in the wild-type complex but fast in the T155S mutant complex (Figure 7). This demonstrates that, although T155S binds GAGGC specifically, it does so with reduced affinity. Using line-width and line-shape analysis, we estimate that T-ag-obd has a K_d value of ~100 nM for GAGGC and 30 uM for AGAAT. These values are somewhat higher than those reported by Titolo et al. (*13*) but most likely reflect differences in sample conditions, i.e., pH, ionic

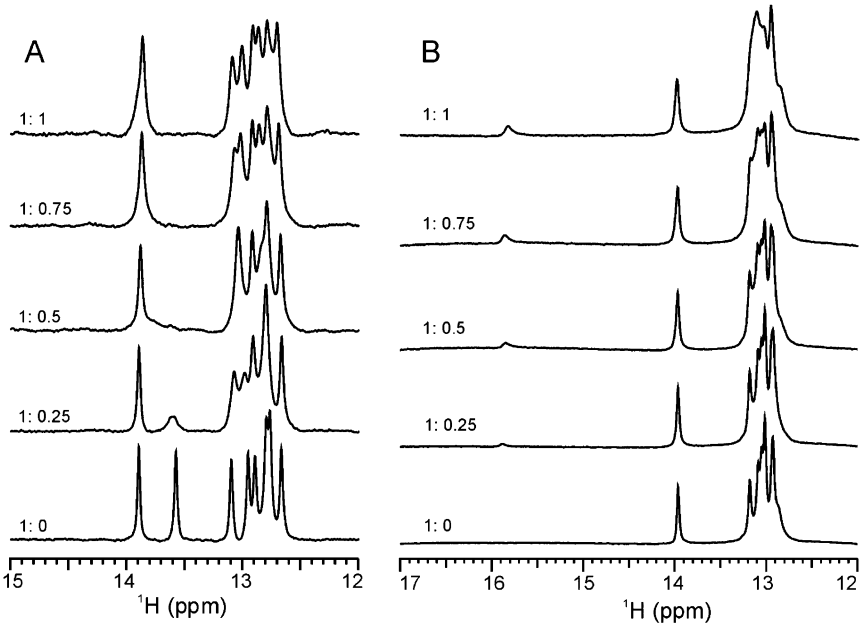

FIGURE 6: (A) 1D low-field "Hard-11" 500-MHz proton spectra of 0.35 mM GTGGC with unlabeled T-ag-obd titrated in at pH 5.5 and 30 °C. The T/A base pair shifts upfield approximately 0.6 ppm, upon binding. (B) 1D low-field "Hard-11" 500-MHz proton spectra of 1.0 mM GGGGC with unlabeled T-ag-obd titrated in at pH 5.5 and 30 °C. The low-field His 203 $N^{\delta 1}H$ resonance is observed; however, whether the second G/C base pair is shifted upfield cannot be determined.

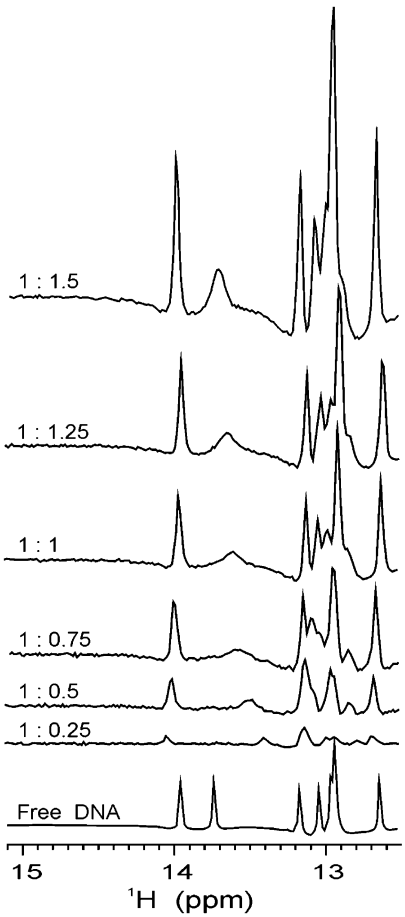

FIGURE 7: 1D "Hard-11" low-field 600-MHz proton spectra of 0.8 mM ^{15}N T155S, with GAGGC added incrementally, at pH 5.5 and 30 °C. The protein−DNA ratio is given, and a spectrum of free GAGGC is shown.

strength, temperature, and concentration. The T155S mutant binds GAGGC about ∼10−100-fold more weakly, with a K_d value of ∼1−10 μM.

Model of T-ag-obd Interacting with GAGGC. Both T-ag and E1 are viral helicases that form double hexamers at their respective origins to initiate replication (*8*, *16*, *28*). Both origins contain repeating sequences inverted to each other. The crystal structure of the E1 DNA-binding domain (E1-DBD) in a complex with its origin recognition sequence has been solved, showing the interactions between the protein and DNA (*7*). T-ag-obd and E1-DBD have high structural homology, even though they have low sequence homology (*5*). On the basis of the structure of E1-DBD bound to its recognition sequence, ATTGTT, we modeled T-ag-obd bound to its recognition sequence, GAGGC. We aligned 92 atoms from the backbone and obtained an RMSD of 1.87 Å. In this model, the His 203 $N^{\delta 1}H$ is positioned to interact with the phosphate group of the fourth base pair (G/C) on the C strand, a position and role assumed by Lys 241 in E1 (*7*).

DISCUSSION

His 203 Forms a Sequence-Specific H-Bond to the DNA. In the absence of paramagnetic ions, very few protons in proteins or nucleic acids have chemical shifts as large as 15.9 ppm. The only ones reported thus far are the Asp-His H-bonded protons of serine protease catalytic triads at low pH values (∼4) (*29*). No chemical shift > 13 ppm has previously been reported for any protein−DNA complex. The unusual downfield chemical shift indicates that the $N^{\delta 1}H$ group of His 203 forms a H-bond in the complex of T-ag-obd with GAGGC DNA. The $^{15}N^{\delta 1}$ chemical shift of 182.2 ppm also supports this conclusion (*25*) and further indicates that the imidazole ring is present as the positively charged imidazolium ion. The fact that the His 203 H-bond forms in complexes with GAGGC DNA but not with AGAAT DNA identifies it as an important determinant of origin-specific binding (*4*).

What is the H-bond acceptor? There are no acceptor functional groups in the vicinity of His 203 on the protein, even allowing for small conformational changes on binding DNA, indicating that the interaction must therefore be an intermolecular one, with the acceptor on the DNA. Model building based on the E1-DBD complex suggests that the acceptor is the phosphate of the fourth base pair (G/C). A H-bond interaction between a positively charged imidazolium ion of His 203 and a negatively charged phosphate group would explain the very large chemical shift of the H-bonded proton, because the largest ^{1}H chemical shifts are produced by H-bonds between oppositely charged donors and acceptors (*30*).

However, how can a H-bond to a phosphate group be an important determinant of origin-specific binding? A priori such contacts might seem capable of contributing only to nonspecific DNA binding. Jen-Jacobson has pointed out that nearly all cognate protein−DNA complexes utilize protein−phosphate group contacts in specific recognition (*15*). Often in such cases the protein−phosphate interactions appear to act as "clamps", which in concert with other base-recognition contacts, serve to distort the DNA. To function as a clamp requires that the protein−phosphate interaction tightly constrains the phosphate; therefore, the protein functional group must be relatively immobile. Backbone N−H groups or very short polar side chains would normally be expected to be employed for this purpose, not a histidine side chain with its much greater potential flexibility. However, His 203 in T-ag-obd is very tightly tucked into the protein, with only the $N^{\delta 1}$ nitrogen protruding from the surface. The imidazole ring is partially stacked over and rigidly held in place by the aromatic ring of Phe 159, which is itself entirely buried within the protein, unexposed to the solvent. With such assistance from the surrounding structure, His 203 could well act as such a clamp. Jen-Jacobson has proposed that such interactions can be thought of as "catalytic", rather than as purely binding interactions and that many DNA-binding proteins having an associated function employ such "transition-state-like" interactions to achieve functional specificity (*15*).

His 203 is essential for nonspecific, as well as specific, DNA binding. Mutating this residue to Gln eliminates all DNA binding in T-ag-obd similar to mutating this residue to Asn in full-length T-ag (*31*). However, because the His-203-donated H-bond is formed only in specific complexes, the loss in nonspecific DNA binding cannot come from the loss of this H-bond. It must arise instead from a mutation-induced localized conformational change, a not unexpected result in view of the tight packing around His 203. Selective displacements of backbone N−H resonances in the vicinity of His 203 in HSQC spectra confirm that the H203Q

mutation does not induce global effects on the structure but does cause localized conformational changes around His 203. If such localized conformational effects eliminate nonspecific DNA binding, they are also likely to result in the loss of binding affinity in specific DNA complexes. This further suggests that the His 203 H-bond may be a "catalytic" interaction as discussed above.

T-ag-obd Distorts GAGGC DNA. The 0.6-ppm upfield displacement of the thymidine N-3 proton of the A/T base pair of the recognition pentanucleotide is a significant chemical-shift change and indicates that the ^{1}H involved in this Watson−Crick H-bond experiences a substantial change in its magnetic environment when T-ag-obd binds GAGGC DNA. The fact that T-ag-obd can induce such a large effect without significantly perturbing the neighboring base pairs is remarkable. Such a displacement is consistent with a weakening of the H-bond, perhaps associated with T-ag-mediated strand separation and unwinding, which are known components of initiation (*8*, *28*). Alternatively, the displacement could reflect altered ring currents centered at this position resulting from a distortion in the helical structure of the DNA. BPV E1-DBD induces such a distortion in the helical structure. In this case, all base-pairing H-bonds are maintained, while the helical diameter changes and the major groove becomes deeper (*7*). Unfortunately, the effect on ^{1}H NMR signals in the base-pair H-bonds of E1-DBD bound DNA has not been reported. In our studies, regardless of whether the A/T chemical-shift displacement reflects a selective weakening of the H-bond or a distortion of the helical structure, T-ag-obd appears to catalyze a "functional" structural change, because this occurs only in "specific" complexes.

The fact that the T-ag-obd induces a structural change at the second position of the recognition pentanucleotide regardless of whether this is an A/T or a T/A base pair indicates that no direct specific contacts are made to these bases. This contrasts quite sharply with the situation at the third base pair, where substitution of the G/C with a C/G base pair eliminates both the His 203 H-bond and the T-ag-obd-induced distortion at the A/T site. This suggests that specific contacts are made directly with the G/C base pair and that these contacts are asymmetric or strand-specific.

Replacing the A/T base pair with a C/G base pair eliminates both effects associated with specific DNA binding, i.e., the appearance of the low-field proton signal and the upfield displacement of, in this case, the C/G base pair. Interestingly, however, replacement of the A/T with a G/C does not produce the same result. With G/C, a low-field proton signal appears, indicating the formation of the His 203 H-bond. This suggests that we might also expect to observe a displacement of the G/C H-bonded proton. Unfortunately, this could not be determined owing to the severe overlap in the ^{1}H NMR spectra of the complex. If no such upfield displacement occurs, it would represent the only complex examined in which one of the structural effects associated with specific binding occurs without the other.

$KMnO_4$ oxidation/polymerase pausing studies did not detect the A/T structural alteration on the binding of T-ag (*32*, *33*) or T-ag-obd (*34*) to SV40 origin DNA, a result that may appear to conflict with our NMR results. However, the $KMnO_4$ experiment requires that the distortion lead to the exposure of the reactive 5,6 double bond of thymine to $KMnO_4$ (*32*). Not all structural distortions produce this result. The protein, for example, could block accessibility, or the distortion could be too small to expose the double bond. Such $KMnO_4$ experiments have been used to conclude that eukaryotic initiators (e.g., ORC) do not distort their DNA substrates (*35*). In view of these discrepancies, eukaryotic systems should be re-examined by NMR.

T155S Mutant Binds GAGGC Specifically but Dysfunctionally. In full-length T-ag, mutating Thr 155 to Ser, a mere deletion of a methyl group, eliminates the ability of SV40 to replicate (*36*). Thr 155 lies in the A1 DNA-binding loop of T-ag-obd. In the E1-based model of T-ag-obd complexed to GAGGC, Thr 155 extends into the major groove, clearly in a position to influence specific binding. T155S T-ag has 67% of the nonspecific but only 1% of the specific DNA-binding activity of wild-type T-ag (*31*, *36*), which makes it a so-called class 1 mutant, defined as having lost specific but not nonspecific DNA-binding ability.

We have shown here that the T155S mutant of T-ag-obd appears to retain specific DNA-binding ability, a result which seems at odds with the full-length T-ag studies. This is demonstrated by the fact that it exhibits both of the structural effects associated with specific complexes, i.e., the His 203 H-bond and the displacement of the A/T H-bonded proton. At a minimum, the observation of these two features shows that essentially all of the T155S is bound to DNA in a single complex type, rather than distributed among different types of complexes characteristic of nonspecific binding.

Is there a discrepancy between genetics studies on full-length T-ag and the current NMR studies on T-ag-obd with respect to specific DNA binding? The NMR results show that although T155S binds GAGGC DNA specifically, it does so with a lower affinity than wild-type T-ag-obd. The evidence for this is that, while wild-type T-ag-obd and the GAGGC DNA are in slow to intermediate exchange on the NMR time scale, T155S is in fast exchange. This means that T155S binds with a 10−100-fold lower affinity than wild-type T-ag-obd. Additionally, band-shift assays demonstrate that T155S binds specific DNA with a lower affinity than T-ag-obd (*18*). Thus, there is no real discrepancy between full-length T-ag and T-ag-obd regarding the effect of the T155S mutation on DNA binding. This mutation appears to decrease the binding affinity of both by similar amounts. However, why is the T155S totally defective with respect to replication in the context of full-length T-ag? One possibility is that formation of a specific complex, while necessary for replication to proceed, might not, in itself, be sufficient. The specific complex might need to be maintained for a period of time to permit subsequent events, i.e., recruitment of other T-ag monomers to form hexamers and double hexamers and of other proteins necessary for initiation and replication. The T155S mutant, though able to bind specifically, might not maintain the complex for a sufficient amount of time to permit these next steps. Such lifetime effects have been demonstrated in other biological systems, notably signal transduction through the T cell receptor. Here, ligands that bind the receptor and induce the correct structural changes nevertheless fail to transmit a signal if the lifetime of the complex is insufficient for recruitment of intracellular participants (*37*−*39*). Alternatively, though the T155S mutant binds specifically, forms the His 203 H-bond, and distorts the A/T base pair, its complex with the origin DNA may

nevertheless be defective and thus nonproductive, in ways still unknown.

Specificity of Origin Recognition and Initiation of Replication. What is the explanation for this apparent discrepancy between the DNA-binding specificity and the specificity of initiation? A common feature of initiator proteins is that they function in association with other cellular proteins. Thus, a potential explanation is that these accessory proteins enable additional protein−protein and protein−DNA contacts that provide higher binding affinity and specificity for origin DNA. The E2 protein appears to provide such help to E1, the papillomavirus initiator protein, increasing its binding specificity for origin DNA ~1000-fold (*40*). Curiously, E2 appears to act not by supplying additional "specific" contacts but by suppressing the nonspecific DNA-binding activity of E1 (*40*). It has been proposed that the suppression of E2 of the nonspecific binding of E1 restores or "unmasks" the DNA-binding specificity intrinsic to the DNA-binding domain of E1 and that other initiators such as SV40 T-ag may employ a similar mechanism. Nevertheless, it is not clear that this 1000-fold increase in specificity is sufficient to account entirely for the specificity of initiation. The full-length E1 protein binds cognate origin DNA only 10−100-fold better than nonspecific DNA (*41*). Improving this by 10^3 via E2 would still yield a specificity, i.e., K_s/K_{ns}, of only 10^4 to 10^5. The vast majority of E1 and E2 would still be bound to nonspecific DNA sites in a mammalian cell with 6×10^9 bp of DNA. It is also not clear if specificity can be generated for the T-ag system by a similar mechanism, because the DNA-binding specificity of T-ag and the T-ag-obd do not differ much and are about the same as that of E1-DBD, i.e., $K_s/K_{ns} \sim 10-100$ (*12*, *13*).

The proposal by Jen-Jacobson that some DNA-binding proteins may achieve specificity through the use of "catalytic" or "transition-state-like" interactions is intriguing (*15*). This idea applies especially to DNA-binding proteins that have a catalytic function such as restriction endonucleases. Initiator proteins, with functions that include DNA melting, unwinding, and the recruitment of other proteins, would fall into this category. The His 203 H-bond and the distortion of the A/T base pair, seen here in specific but not in nonspecific DNA complexes, could well be examples of such catalytic interactions.

How does this mechanism generate specificity for initiation? When an initiator protein binds to its cognate origin DNA, the "catalytic" interactions may be used to distort the DNA or change the protein structure in some way that permits the next steps in the initiation process to proceed. Such interactions would not contribute to the binding affinity. Thus, the binding affinity of an initiator for the cognate DNA site might appear to be a little different than for a nonspecific DNA site. The difference, however, is that a specific complex would be productive, ready, and able to proceed to the next step, while the nonspecific complex would be "nonproductive", unable to proceed along the pathway to initiation, but able to slide along the DNA in search of a productive site.

REFERENCES

1. Jacob, F., Brenner, S., and Cuzin, F. (1963) On the regulation of DNA replication in bacteria, *Cold Spring Harbor Symp. Quant. Biol. 28*, 329−348.
2. Stillman, B. (1994) Initiation of chromosomal DNA replication in eukaryotes. Lessons from λ, *J. Biol. Chem. 269*, 7047−7050.
3. Bell, S. P., and Dutta, A. (2002) DNA replication in eukaryotic cells, *Annu. Rev. Biochem. 71*, 333−374.
4. Luo, X., Sanford, D. G., Bullock, P. A., and Bachovchin, W. W. (1996) Solution structure of the origin DNA-binding domain of SV40 T-antigen, *Nat. Struct. Biol. 3*, 1034−1039.
5. Enemark, E. J., Chen, G., Vaughn, D. E., Stenlund, A., and Joshua-Tor, L. (2000) Crystal structure of the DNA binding domain of the replication initiation protein E1 from papillomavirus, *Mol. Cell 6*, 149−158.
6. Hickman, A., Ronning, D., Kotin, R., and Dyda, F. (2002) Structural Unity among Viral Origin Binding Proteins. Crystal Structure of the Nuclease Domain of Adeno-Associated Virus Rep, *Mol. Cell 10*, 327.
7. Enemark, E. J., Stenlund, A., and Joshua-Tor, L. (2002) Crystal structures of two intermediates in the assembly of the papillomavirus replication initiation complex, *EMBO J. 21*, 1487−1496.
8. Bullock, P. A. (1997) The initiation of simian virus 40 DNA replication in vitro, *Crit. Rev. Biochem. Mol. Biol. 32*, 503−568.
9. Simmons, D. T. (2000) SV40 large T antigen functions in DNA replication and transformation, *Adv. Virus Res. 55*, 75−134.
10. Fanning, E. (1992) Simian virus 40 large T antigen: the puzzle, the pieces, and the emerging picture, *J. Virol. 66*, 1289−1293.
11. Deb, S., DeLucia, A. L., Baur, C. P., Koff, A., and Tegtmeyer, P. (1986) Domain structure of the simian virus 40 core origin of replication, *Mol. Cell. Biol. 6*, 1663−1670.
12. Joo, W. S., Luo, X., Denis, D., Kim, H. Y., Rainey, G. J., Jones, C., Sreekumar, K. R., and Bullock, P. A. (1997) Purification of the simian virus 40 (SV40) T antigen DNA-binding domain and characterization of its interactions with the SV40 origin, *J. Virol. 71*, 3972−3985.
13. Titolo, S., Welchner, E., White, P. W., and Archambault, J. (2003) Characterization of the DNA-binding properties of the origin-binding domain of simian virus 40 large T antigen by fluorescence anisotropy, *J. Virol. 77*, 5512−5518.
14. Simmons, D. T., Loeber, G., and Tegtmeyer, P. (1990) Four major sequence elements of simian virus 40 large T antigen coordinate its specific and nonspecific DNA binding, *J. Virol. 64*, 1973−1983.
15. Jen-Jacobson, L. (1997) Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state, *Biopolymers 44*, 153−180.
16. Stenlund, A. (2003) Initiation of DNA replication: Lessons from viral initiator proteins, *Nat. Rev. Mol. Cell. Biol. 4*, 777−785.
17. Robillard, G., and Shulman, R. G. (1972) High-resolution nuclear magnetic resonance study of the histidine−aspartate hydrogen bond in chymotrypsin and chymotrypsinogen, *J. Mol. Biol. 71*, 507−511.
18. Jones, C. C. (1996) Purification and Partial Characterization of a Mutant within the Simian Virus 40 Large T-antigen DNA Binding Domain, in *Department of Biochemistry, Tufts Medical School* pp 84, Tufts University, Boston, MA.
19. Clore, G. M., Kimber, B. J., and Gronenborn, A. M. (1983) The 1−1 hard pulse; a novel, simple, and effective method of water resonance suppressiion in FT-1H NMR, *J. Magn. Reson. 54*, 170−173.
20. Delsuc, M. A., and Lallemand, J. Y. (1986) Improvement of Dynamic-Range in NMR by Oversampling, *J. Magn. Reson. 69*, 504−507.
21. Bodenhausen, G., and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectrocopy, *Chem. Phys. Lett. 69*, 185−189.
22. Messerle, B., Wider, G., Otting, G., Weber, C., and Wuthrich, K. (1989) Solvent suppression using a spin lock in 2D and 3D NMR spectroscopy with aqueous solutions, *J. Magn. Reson. 85*, 608−613.
23. Xia, B., Cheng, H., Skjeldal, L., Coghlan, V. M., Vickery, L. E., and Markley, J. L. (1995) Multinuclear magnetic resonance and mutagenesis studies of the histidine residues of human mitochondrial ferredoxin, *Biochemistry 34*, 180−187.
24. Zuiderweg, E. R. P. (1990) A proton-detected heteronuclear chemical-shift correlation experiment with improved resolution and sensitivity, *J. Magn. Reson. 86*, 246−357.
25. Bachovchin, W. W., and Roberts, J. D. (1978) Nitrogen-15 nuclear magnetic resonance spectroscopy. The state of histidine in the catalytic triad of α-lytic protease. Implications for the charge-relay mechanism of peptide-bond cleavage by serine proteases, *J. Am. Chem. Soc. 100*, 8041−8047.

26. Sudmeier, J. L., Ash, E. L., Gunther, U. L., Luo, X., Bullock, P. A., and Bachovchin, W. W. (1996) HCN, a triple-resonance NMR technique for selective observation of histidine and tryptophan side chains in $^{13}C/^{15}N$-labeled proteins, *J. Magn. Reson., Ser. B 113*, 236−247.
27. Sudmeier, J. L., Bradshaw, E. M., Haddad, K. E., Day, R. M., Thalhauser, C. J., Bullock, P. A., and Bachovchin, W. W. (2003) Identification of Histidine Tautomers in Proteins by 2D $^{1}H/^{13}C^{\delta 2}$ One-Bond Correlated NMR, *J. Am. Chem. Soc. 125*, 8430−8431.
28. Borowiec, J. A., Dean, F. B., Bullock, P. A., and Hurwitz, J. (1990) Binding and unwinding—How T antigen engages the SV40 origin of DNA replication, *Cell 60*, 181−184.
29. Seavey, B. R., Farr, E. A., Westler, W. M., and Markley, J. L. (1991) A Relational Database for Sequence-Specific Protein NMR Data, *J. Biomol. NMR 1*, 217−236.
30. Metzler, D. E., Metzler, C. M., Mollova, E. T., Scott, R. D., Tanase, S., Kogo, K., Higaki, T., and Morino, Y. (1994) NMR studies of 1H resonances in the 10−18-ppm range for cytosolic aspartate aminotransferase, *J. Biol. Chem. 269*, 28017−28026.
31. Simmons, D. T., Upson, R., Wun-Kim, K., and Young, W. (1993) Biochemical analysis of mutants with changes in the origin-binding domain of simian virus 40 tumor antigen, *J. Virol. 67*, 4227−4236.
32. Borowiec, J. A., and Hurwitz, J. (1988) Localized melting and structural changes in the SV40 origin of replication induced by T-antigen, *EMBO J. 7*, 3149−3158.
33. Parsons, R., Anderson, M. E., and Tegtmeyer, P. (1990) Three domains in the simian virus 40 core origin orchestrate the binding, melting, and DNA helicase activities of T antigen, *J. Virol. 64*, 509−518.
34. Kim, H. Y., Barbaro, B. A., Joo, W. S., Prack, A. E., Sreekumar, K. R., and Bullock, P. A. (1999) Sequence requirements for the assembly of simian virus 40 T antigen and the T-antigen origin binding domain on the viral core origin of replication, *J. Virol. 73*, 7543−7555.
35. Dutta, A., and Bell, S. P. (1997) Initiation of DNA replication in eukaryotic cells, *Annu. Rev. Cell Dev. Biol. 13*, 293−332.
36. Simmons, D. T., Wun-Kim, K., and Young, W. (1990) Identification of simian virus 40 T-antigen residues important for specific and nonspecific binding to DNA and for helicase activity, *J. Virol. 64*, 4858−4865.
37. Rabinowitz, J. D., Beeson, C., Wulfing, C., Tate, K., Allen, P. M., Davis, M. M., and McConnell, H. M. (1996) Altered T cell receptor ligands trigger a subset of early T cell signals, *Immunity 5*, 125−135.
38. Evavold, B. D., Sloan-Lancaster, J., and Allen, P. M. (1993) Tickling the TCR: Selective T-cell functions stimulated by altered peptide ligands, *Immunol. Today 14*, 602−609.
39. Germain, R. N., and Stefanova, I. (1999) The dynamics of T cell receptor signaling: complex orchestration and the key roles of tempo and cooperation, *Annu. Rev. Immunol. 17*, 467−522.
40. Stenlund, A. (2003) E1 initiator DNA binding specificity is unmasked by selective inhibition of nonspecific DNA binding, *EMBO J. 22*, 954−963.
41. Sedman, J., and Stenlund, A. (1995) Cooperative interaction between the initiator E1 and the transcriptional activator E2 is required for replicator specific DNA replication of bovine papillomavirus in vivo and in vitro, *EMBO J. 14*, 6218−6228.

BI030228+