Solution Structure of the DNA Binding Domain of HIV-1 Integrase†

Patricia J. Lodi,‡ James A. Ernst,‡ John Kuszewski,‡ Alison B. Hickman,§ Alan Engelman,§ Robert Craigie,§ G. Marius Clore,*,‡ and Angela M. Gronenborn*,‡

Laboratory of Chemical Physics and Laboratory of Molecular Biology, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

Received May 10, 1995; Revised Manuscript Received June 5, 1995⊗

ABSTRACT: The solution structure of the DNA binding domain of HIV-1 integrase (residues 220−270) has been determined by multidimensional NMR spectroscopy. The protein is a dimer in solution, and each subunit is composed of a five-stranded β-barrel with a topology very similar to that of the SH3 domain. The dimer is formed by a stacked β-interface comprising strands 2, 3, and 4, with the two triple-stranded antiparallel β-sheets, one from each subunit, oriented antiparallel to each other. One surface of the dimer, bounded by the loop between strands β1 and β2, forms a saddle-shaped groove with dimensions of approximately 24 × 23 × 12 Å in cross section. Lys264, which has been shown from mutational data to be involved in DNA binding, protrudes from this surface, implicating the saddle-shaped groove as the potential DNA binding site.

The integrase protein of the human immunodeficiency virus (HIV) mediates a key step in the life cycle of the virus, namely the integration of a DNA copy of the viral genome into a host chromosome (see Goff, 1992; Vink & Plasterk, 1993; for reviews). HIV-1[1] integrase is composed of three functional domains (Bushman et al., 1993; Engelman et al., 1993; van Gent et al., 1993; Vink et al., 1993): a small N-terminal domain that contains a His_2Cys_2 zinc binding motif, a central catalytic domain whose crystal structure has recently been solved (Dyda et al., 1994), and a C-terminal DNA binding domain. While the catalytic core domain can carry out a simple polynucleotidyl transfer termed disintegration (Bushman et al., 1993), all three domains are required for the 3′ processing and DNA strand transfer activities that accomplish integration of the viral genome (Drelich et al., 1992; Schauer & Billich, 1992; Vink et al., 1993). Consequently, the N- and C-terminal domains provide additional potentially useful targets for rational drug design aimed at inhibiting HIV integration into the host genome.

The function of the N-terminal domain is at present unknown, and in the case of the related Rous Sarcoma virus integrase at least, its integrity appears not to be essential for *in vitro* integration activity as it can be replaced by unrelated fusion peptides comprising polyhistidine sequences (Bushman & Wang, 1994). The C-terminal domain, on the other hand, displays the same DNA binding characteristics and affinity for both viral and nonspecific double stranded DNA as the intact integrase (van Gent et al., 1991; Mumm & Grandgenett, 1991; Schauer & Billich, 1992). The minimal DNA binding domain has recently been shown to comprise residues 220−270 (Puras-Lutzke et al., 1994). In this paper we present the solution structure of this domain ($Int^{220-270}$) by multidimensional heteronuclear NMR spectroscopy.

EXPERIMENTAL PROCEDURES

Sample Preparation. The DNA coding sequence for $Int^{220-270}$ was cloned into the *Escherichia coli* vector PET15b (Novagen) which introduces a histidine tag at the N-terminus. The construct was expressed in the host strain BL21(DE3) grown in minimal medium using $^{15}NH_4Cl$ and/or [$^{13}C_6$]-glucose as the sole nitrogen and carbon sources, respectively. Cells were grown to an A_{550} of about 1 OD unit, induced by the addition of 1 mM isopropyl β-D-thiogalactopyranoside, and grown for a further 7 h. Cells were harvested by centrifugation at 3800*g* for 8 min, resuspended in "1+1 lysis buffer" (1 M urea, 1 M NaCl, 20 mM Tris·HCl, pH 8.0, 5 mM imidazole, 5 mM benzamidine), frozen at −90 °C, thawed at room temperature, passed through a French press, and centrifuged at 1000*g* for 1 h. A freshly charged chelating Sepharose column (Pharmacia) containing bound nickel was prepared and equilibrated with "1+1 binding buffer" (1 M urea, 1 M NaCl, 20 mM Tris·HCl, pH 8.0, 5 mM imidazole). The supernatant was loaded and the column washed first with 20 column volumes of 1+1 binding buffer and then with 20 column volumes of 1+1 binding buffer containing an additional 60 mM imidazole. $Int^{220-270}$ was eluted using 1+1 binding buffer and an imidazole gradient from 65 to 805 mM imidazole. The fractions containing $Int^{220-270}$ (as determined by SDS electrophoresis using the Pharmacia Phast system) were pooled and passed over uncharged chelating Sepharose to remove excess nickel ions, then made up to 10 mM EDTA, and dialyzed against buffer comprising 500 mM NaCl, 20 mM NaH_2PO_4, pH 7.5, and 0.1 mM EDTA. The dialysate was diluted to a final NaCl concentration of 250−325 mM with 20 mM NaH_2PO_4 pH 7.5, and 0.1 mM EDTA. The histidine tag was removed by adding 70 μL of thrombin (1000 units/mL) and incubating the solution at room temperature for 2.5 h. A benzamidine Sepharose column (Pharmacia) was equilibrated with the

† This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C., A.M.G., and R.C.). P.J.L. is a recipient of a Cancer Research Institute-Elisa Heather Halpern postdoctoral fellowship.

* Authors to whom correspondence should be addressed.
‡ Laboratory of Chemical Physics.
§ Laboratory of Molecular Biology.

⊗ Abstract published in *Advance ACS Abstracts,* July 1, 1995.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; NMR, nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser enhancement; $Int^{220-270}$, minimal DNA binding domain of HIV-1 integrase comprising residues 220−270.

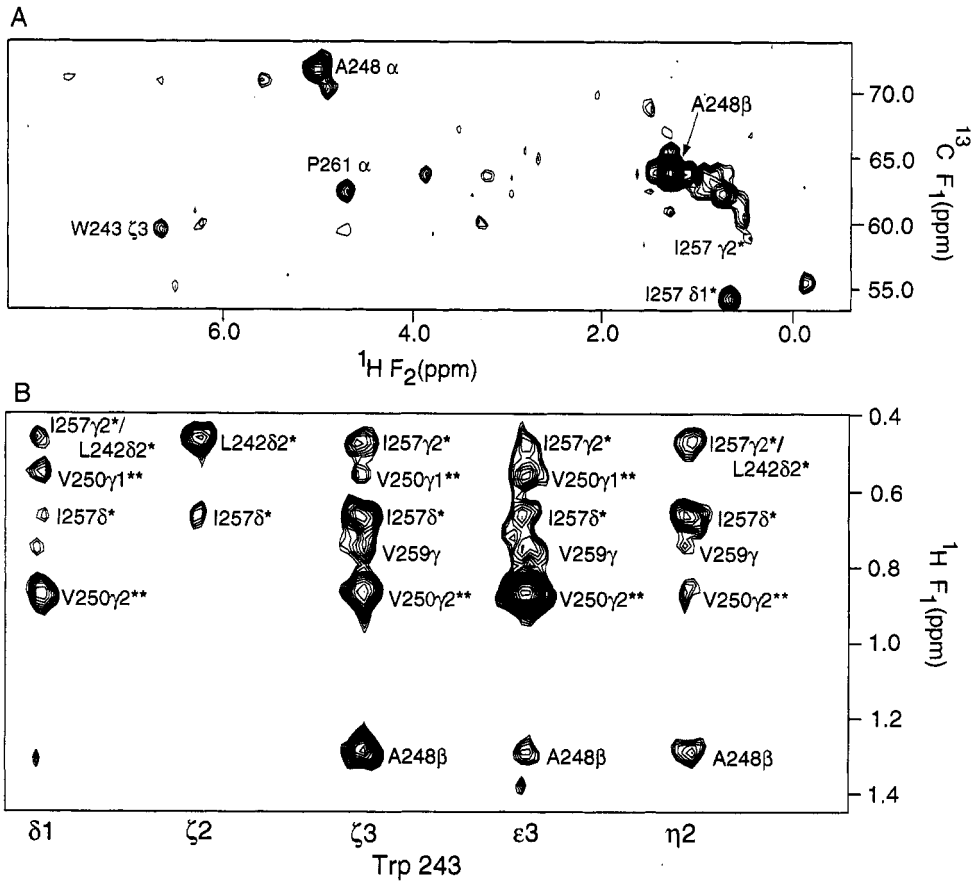

FIGURE 1: Examples of (A) the $^{13}C(F_1)-^1H(F_2)$ plane at the $^{13}C(F_3)-^1H(F_4)$ frequencies of Ala248($C\beta H_3$) and (B) strips taken from the 3D ^{13}C-separated NOE spectrum of $Int^{220-270}$ illustrating intra- and intersubunit NOEs from the methyl protons of Ala248 and the ring protons of Trp243, respectively. Unlabeled peaks in (A) have their maximum intensity in a neighboring plane. Peaks with no asterisk and with one asterisk indicate unambiguous intrasubunit and intersubunit NOEs, respectively. Peaks labeled with two asterisks indicate NOEs with contributions from both intra- and intersubunit interactions. Note that in (A) extensive folding is employed in the ^{13}C dimensions so that the ^{13}C chemical shifts are given by $x \pm n$SW, where n is an integer and SW is the sweep width (20.71 ppm).

dialysis buffer, and the cleaved protein solution was passed over the column at 4 °C to remove thrombin. SDS electrophoresis showed that the cleavage reaction was complete but that a small amount of nonspecific cleavage had occurred, yielding some smaller protein fragments. These fragments and the cleaved histidine tag were removed from the protein solution by precipitating the protein with 4 M ammonium sulfate and subsequent dialysis of the resuspended precipitate with a 3500 molecular weight cutoff Disposo-Dialyser (Spectrum) into 10 mM NaCl, 5 mM NaH_2PO_4, pH 6.5, and 0.05 mM EDTA. The resulting solution containing pure $Int^{220-270}$ with a four-residue leader sequence (GlySerHisMet) at the N-terminus was lyophilized and taken up in either 90% H_2O/10% D_2O or 100% D_2O at a final buffer concentration of 100 mM NaCl, 50 mM NaH_2PO_4, pH 6.5, and 0.5 mM EDTA and a final protein concentration of about 1 mM.

NMR Spectroscopy. All NMR experiments were carried out at 25 °C on a Bruker AMX600 spectrometer equipped with a z-shielded gradient triple resonance probe. The sequential assignment of the ^1H, ^{13}C, and ^{15}N chemical shifts was achieved by means of through-bond heteronuclear correlations along the backbone and side chains using the following 3D experiments: ^{15}N-separated HOHAHA, HNHA, CBCANH, CBCA(CO)NH, HBHA(CO)NH, C(CO)NH, HCCH-COSY, and HCCH-TOCSY. Details of these experiments, as well as the multidimensional nuclear Overhauser enhancement (NOE) experiments, together with the original references are provided in the following reviews: Bax and Grzesiek (1993), Clore and Gronenborn (1991, 1994). $^3J_{HN\alpha}$, $^3J_{CC}$, $^3J_{C\gamma N}$, $^3J_{C\gamma CO}$, and $^3J_{NH\beta}$, coupling constants were obtained by quantitative J correlation spectroscopy (Bax et al., 1994). NOEs were assigned from 3D ^{15}N-separated, 3D ^{13}C-separated, 4D $^{15}N/^{13}C$-separated, and 4D $^{13}C/^{13}C$-separated NOE spectra (mixing times of 130, 120, 120, and 120 ms, respectively). ^{15}N T_1 and T_2 relaxation times were measured using pulse sequences described previously (Kay et al., 1992).

Structure Calculations. Approximate interproton distance restraints were derived from the multidimensional NOE spectra, essentially as described previously (Clore & Gronenborn, 1991). NOEs were grouped into four distance ranges, 1.8−2.7 Å (1.8−2.9 Å for NOEs involving NH protons), 1.8−3.3 Å (1.8−3.5 Å for NOEs involving NH protons), 1.8−5.0 Å, and 1.8−6.0 Å, corresponding to strong, medium, weak, and very weak NOEs. Distances involving methyl groups, aromatic ring protons, and nonstereospecifically assigned methylene protons were represented as a $(\Sigma r^{-6})^{-1/6}$ sum (Nilges, 1993). Protein backbone hydrogen-bonding restraints (two per hydrogen bond, r_{NH-O} = 1.5−2.8 Å, r_{N-O} = 2.4−3.5 Å) within areas of regular secondary structure were introduced during the final stages of refinement. A total of 49 ϕ, 10 ψ, 25 χ_1, and 12 χ_2 torsion angle restraints per monomer were derived from the NOE and coupling constant data, and the minimum ranges employed were ±10°, ±50°, ±20°, and ±30°, respectively (Nilges et al., 1990;

Table 1: Structural Statistics[a]

structural statistics	⟨SA⟩	$\overline{(SA)}$r
RMS deviations from exptl distance restraints (Å)[b]		
all (1530)	0.024 ± 0.003	0.022
intrasubunit		
interresidue sequential ($\|i-j\|=1$) (332)	0.021 ± 0.008	0.014
interresidue short range ($1<\|i-j\|\leq 5$) (202)	0.033 ± 0.004	0.030
interresidue long range ($\|i-j\|>5$) (530)	0.014 ± 0.004	0.019
intraresidue (318)	0.017 ± 0.007	0.016
H-bonds (74)	0.035 ± 0.005	0.039
intersubunit (44)	0.043 ± 0.005	0.033
ambiguous (30)	0.047 ± 0.007	0.042
RMS deviations from expt		
dihedral restraints (deg) (192)[b]	0.215 ± 0.068	0.266
$^3J_{HN\alpha}$ coupling constants (Hz) (78)[b]	0.39 ± 0.04	0.52
RMS deviations from exptl ^{13}C shifts		
$^{13}C\alpha$ (ppm) (100)	0.74 ± 0.04	0.75
$^{13}C\beta$ (ppm) (94)	0.83 ± 0.04	0.81
RMS deviations for exptl ^{1}H shifts		
all (ppm) (392)	0.22 ± 0.004	0.22
CαH (ppm) (102)	0.25 ± 0.007	0.24
methyl (ppm) (52)	0.15 ± 0.008	0.16
others (ppm) (238)	0.21 ± 0.005	0.23
deviations from idealized covalent geometry		
bonds (Å) (1788)	0.004 ± 0.0001	0.005
angles (deg) (3246)	0.616 ± 0.012	0.648
impropers (deg) (938)	0.411 ± 0.059	0.393
$E_{L\text{-}J}$ (kcal·mol^{-1})[c]	−433 ± 10	−413

[a] The notation of the NMR structures is as follows: ⟨SA⟩ are the final 40 simulated annealing structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual SA structures (complete dimer) best fitted to each other; $\overline{(SA)}$r is the restrained minimized mean structure obtained by restrained regularization of the mean structure $\overline{SA}$. The number of terms for the various restraints is given in parentheses and applies to the complete dimer. The final force constants employed for the various terms in the target function used for simulated annealing are as follows: 500 kcal·mol^{-1}·Å^{-2} for bond lengths, 500 kcal·mol^{-1}·rad^{-2} for angles and improper torsions (which serve to maintain planarity and chirality), 100 kcal·mol^{-1}·Å^{-2} for the noncrystallographic symmetry, 4 kcal·mol^{-1}·Å^{-4} for the quartic van der Waals repulsion term (with the hard sphere effective van der Waals radii set to 0.8 times their value used in the CHARMM PARAM19/20 parameters), 30 kcal·mol^{-1}·Å^{-2} for the experimental distance restraints (interproton distances and hydrogen bonds), 200 kcal·mol^{-1}·rad^{-2} for the torsion angle restraints, 1 kcal·mol^{-1}·Hz^{-2} for the coupling constant restraints, 0.5 kcal·mol^{-1}·ppm^{-2} for the carbon chemical shift restraints, and 7.5 kcal·mol^{-1}·ppm^{-2} for the proton chemical shift restraints. [b] None of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5°, or $^3J_{HN\alpha}$ coupling constant violations greater than 2 Hz. [c] $E_{L\text{-}J}$ is the Lennard-Jones van der Waals energy calculated with the CHARMM PARAM19/20 protein parameters (Brooks et al., 1983) and is *not* included in the target function for simulated annealing or restrained minimization.

Clore & Gronenborn, 1994). Stereospecific assignments were obtained for 17 out of the 30 non-proline β-methylene groups, for the methyl groups of four of the six Val residues, and for the methyl groups of both Leu residues. The structures were calculated using a modified version of the hybrid distance geometry−dynamical simulated annealing protocol (Nilges et al., 1988) using the program X-PLOR-31 (Brünger, 1993), adapted to incorporate pseudopotentials for $^3J_{HN\alpha}$ coupling constant (Garrett et al., 1994), secondary $^{13}C\alpha$ and $^{13}C\beta$ chemical shift (Kuszewski et al., 1995a), and ^{1}H chemical shift (Kuszewski et al., 1995b) restraints. The target function that is minimized during simulated annealing and restrained regularization comprises only quadratic harmonic potential terms for covalent geometry, noncrystallographic symmetry, $^3J_{HN\alpha}$ coupling constant, secondary $^{13}C\alpha$ and $^{13}C\beta$ chemical shift and ^{1}H chemical shift restraints, square-well quadratic potentials for the experimental distance and torsion angle restraints, and a quartic van der Waals repulsion term for the nonbonded contacts. There were *no* hydrogen-bonding, electrostatic, or 6-12 Lennard-Jones empirical potential energy terms in the target function. As in previous structure determinations from this laboratory (Kraulis et al., 1989; Forman-Kay et al., 1991; Clore & Gronenborn, 1991) an iterative refinement strategy was employed, incorporating more experimental restraints at each successive stage as the quality of the structures improves. This includes taking care that predicted NOEs corresponding to short interproton distances are present in the spectra. The final aim is to ensure that all structurally useful NOEs are extracted from the data (Clore & Gronenborn, 1991).

The coordinates of the 40 final simulated annealing structures, together with the coordinates of the restrained regularized mean structure, $\overline{(SA)}$r, and the complete list of experimental NMR restraints and ^{1}H, ^{15}N, ^{13}C assignments have been deposited in the Brookhaven Protein Data Bank (1IHV, 1IHW, and R1IHVMR, respectively).

RESULTS AND DISCUSSION

Structure Determination. Int$^{220-270}$ is a dimer in solution. Analytical ultracentrifugation indicates that the dimer dissociation constant is significantly less than 1 μM (under the same salt and pH conditions used in the NMR experiments), as no evidence of monomer could be detected (P. T. Wingfield, personal communication). In addition, the rotational correlation time at 25 °C determined from an analysis of ^{15}N T_1 and T_2 relaxation times (Kay et al., 1989; Clore et al., 1990) is ~8 ns, consistent with the molecular mass of a dimer (M_r ~ 13 kDa). The NOE data indicated unambiguously that the two monomers are arranged in an antiparallel manner at the dimer interface. There were 22 NOEs which corresponded to distances greater than 8.5 Å in the monomer, of which 13 were greater than 11 Å. These NOEs therefore had to be attributed unambiguously to intersubunit interac-

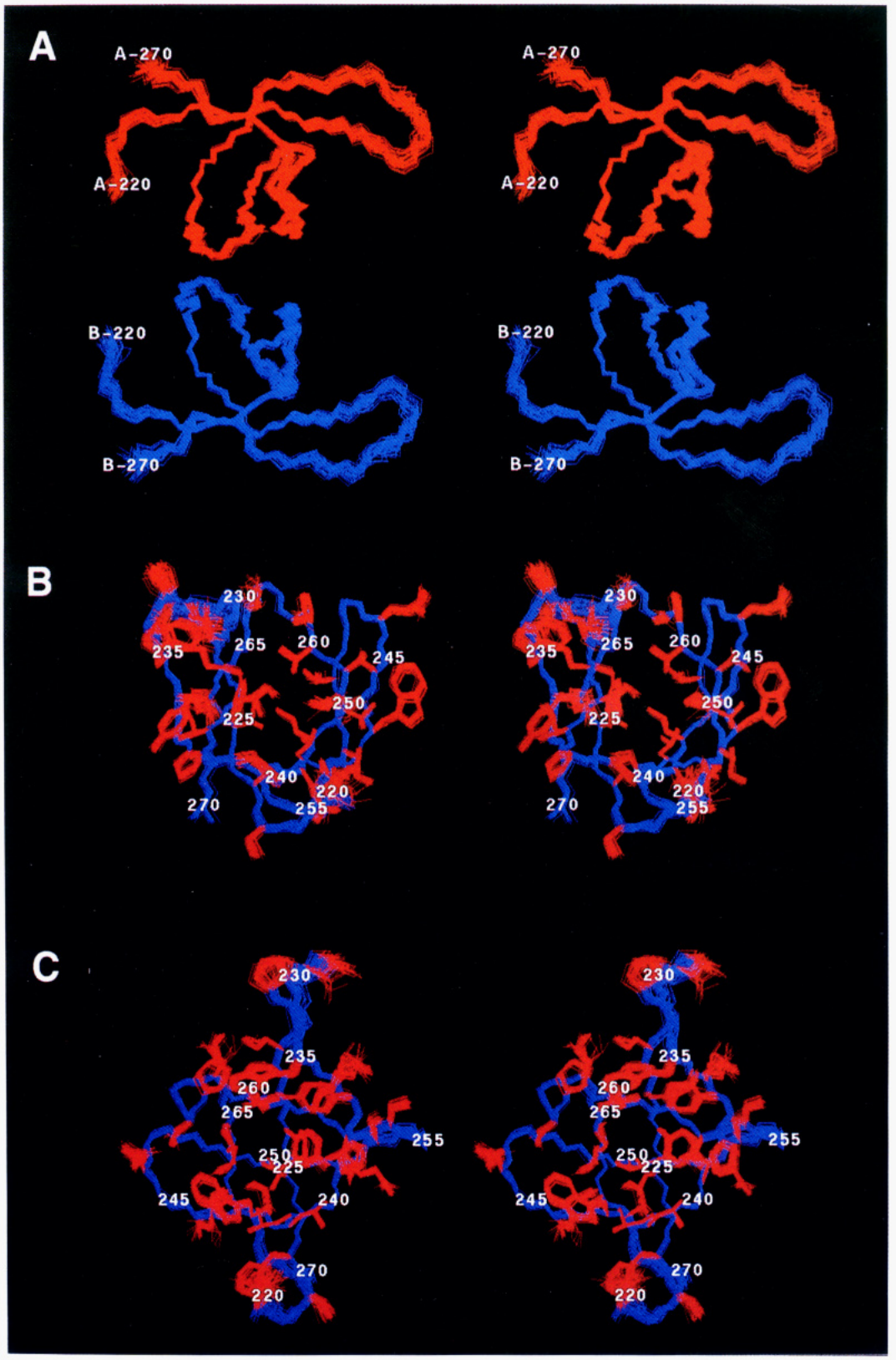

FIGURE 2: Superpositions of 40 simulated annealing structures of $Int^{220-270}$. (A) Backbone atoms of the dimer with one subunit in red and the other in blue. (B and C) Two views of the monomer with the backbone in blue and ordered side chains in red. The figure was generated with the program AVS-XPLOR (Brünger & DeLano, 1993).

tions. These comprised NOEs from the side chain of Ile257 to the ring protons of Trp243, to Gly245(CαH), and to the side chains of Glu246, Ala248, and Val259, and from the $C\delta 2H_3$ methyl group of Leu242 to the ring protons of Trp243. Some of these NOEs are illustrated in the spectra shown in Figure 1. In addition, there were 15 NOEs that could potentially arise from both intra- and intersubunit interactions. These were treated using a $(\sum r^{-6})^{-1/6}$ sum (Nilges, 1993). As Trp243 is located at the dimer interface and is in close proximity to the methyl groups of Leu242, Val250, and Ile257, refinement against ^1H chemical shifts (Kuszewski et al., 1995b) proved quite useful in helping to

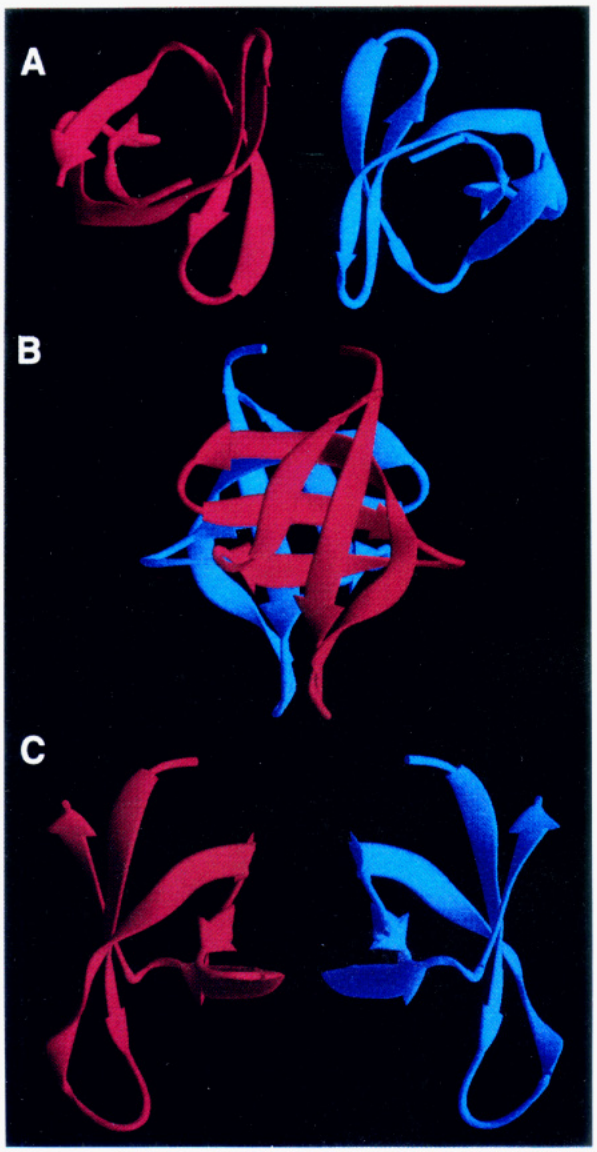

FIGURE 3: Ribbon diagrams illustrating three views of the $Int^{220-270}$ dimer. The figure was generated with the program RIBBONS (Carson, 1987).

verify the orientation of the two subunits with respect to each other. Consider, for example, the Leu242($C\delta1H_3$), Val250-($C\gamma1H_3$), and Ile257($C\gamma2H_3$) methyl groups and the β-methine protons of Val250 and Ile257 located at the dimer interface. Their observed ^1H chemical shifts are 0.02, 0.57, 0.48, 1.92, and 1.59 ppm, respectively. The deviations between the calculated and observed ^1H shifts for these five resonances are all greater than 0.3 ppm for the monomer (+0.59, +0.34, +0.33, +0.77, and +0.46 ppm, respectively). In the dimer, however, the deviations between calculated and observed ^1H shifts for these methyl protons are significantly improved (+0.06, −0.17, −0.23, +0.39, and −0.20 ppm, respectively), principally as a result of ring current shifts arising from Trp243 of the two subunits.

The structure calculations, using simulated annealing, were based on a total of 2386 experimental NMR restraints for the complete dimer (i.e., 1193 per monomer). A summary of the structural statistics is provided in Table 1, and best fit superpositions of the final 40 simulated annealing structures are shown in Figure 2. The precision of the dimer coordinates (that is, the atomic RMS distribution of the 40 simulated annealing structures about their mean coordinate positions) is 0.45 ± 0.12 Å for the backbone (N, Cα, C, O) atoms, 0.96 ± 0.10 Å for all atoms, and 0.48 ± 0.12 Å for all ordered side chains. The corresponding values for the monomer coordinates are 0.29 ± 0.06, 0.87 ± 0.07, and 0.46 ± 0.11 Å, respectively. As a noncrystallographic symmetry restraint on all atoms was employed during the refinement, the all atom atomic RMS difference between the two subunits is very small (<0.08 Å for the individual simulated annealing structures and 0.0006 Å for the restrained regularized mean structure). Eighty-two percent of the residues lie within the most favored region of the Ramachandran ϕ,ψ plot, 16% in the additionally allowed regions, and 2% in the generously allowed regions (Laskowski et al., 1993).

Description of the Structure. The monomer is composed of five β-strands (residues 222−229, 232−245, 248−253, 256−262, and 266−270) which are arranged in an antiparallel (a) manner and form a five-stranded β-barrel (Figures 2, 3, and 5). The overall topology is $\beta5(a)\beta1(a)\beta2(a)\beta3-(a)\beta4(a)$. Strand $\beta2$ has a bulge at residues 242 and 243. The core of the β-barrel is formed by Val225, Tyr227, Ala239, Val249, Ile251, Val260, and Ala265. The β-barrel is closed off at the front (in the view shown in Figure 2B) by Ile220, Phe223, Leu241, Lys244, and Ile267 and at the rear by the aliphatic portions of the side chains of Arg231 and Lys258. There is one potential salt bridge between the side chains of Lys236 and Asp253.

The dimer is formed by a stacked β-interface involving β-strands 2 (C-terminal half), 3, and 4 (Figures 2A, 3, and 4), with the two three-stranded antiparallel β-sheets, one from each subunit, arranged in an antiparallel orientation. This orientation is defined unambiguously by the NOEs from the side chain of Ile257 to the side chains of Trp243 and Ala248 (Figure 1). The interface is stabilized by hydrophobic contacts between Trp243 of subunit A and Leu242, Val250, and Ile257 of subunit B, between Ala248 of subunit A and Ile257 of subunit B, between Val250 of subunit A and Val250 and Ile257 of subunit B, and between Val259 of subunit A and Ile257 and Val259 of subunit B (Figure 4). In addition, there is a hydrogen bond between the carboxylate of Glu246(A) and the carboxamide of Gln252(B). The accessible surface area per subunit buried upon dimerization is about 360 $Å^2$, which is consistent with the linear relationship $ASA_i = 0.06\ M_r$, where ASA_i is the interface accessible surface area and M_r the monomer molecular weight (Jones & Thornton, 1985). The calculated solvation free energy of dimerization (Eisenberg & McLaghlan, 1986) is approximately −8 kcal·mol^{-1}, consistent with a tight dimer stabilized predominantly by hydrophobic interactions. As a result of dimer formation, a large saddle-shaped groove (about 24 × 23 × 12 Å in cross section), bounded by the loop connecting strands $\beta1$ and $\beta2$ (residues 228−235), is formed comprising several positively charged residues (Figures 2A and 3A).

Relationship to Catalytic Domain. The catalytic domain is also dimeric with the two subunits oriented parallel to each other (Dyda et al., 1994). The Cα−Cα separation between the C-terminal residue (Ile208) of the catalytic domain (located at the end of a helix) of the two subunits is 21 Å. Likewise, the separation between the N-terminal residue (Ile220) of the two subunits of $Int^{220-270}$ is about 17 Å, and the two N-terminal strands are located on the same face of the dimer (Figure 3). Thus, the catalytic dimer can potentially sit directly on top of the $Int^{220-270}$ dimer in the view

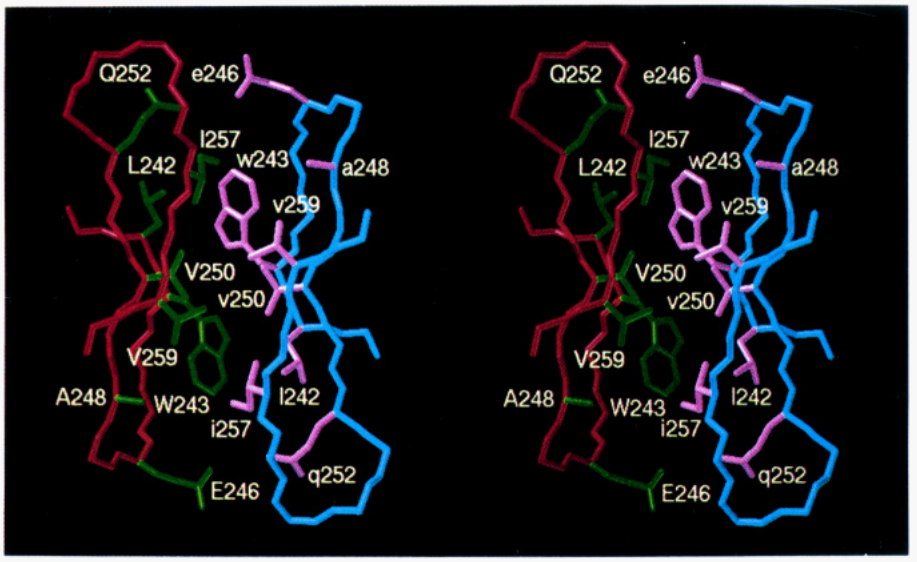

FIGURE 4: Stereoview of the dimer interface of $Int^{220-270}$. The backbone of one subunit is shown in red and the other in blue; the corresponding colors for the side chains are green and magenta. The figure was generated with the program VISP (de Castro & Edelstein, 1992).

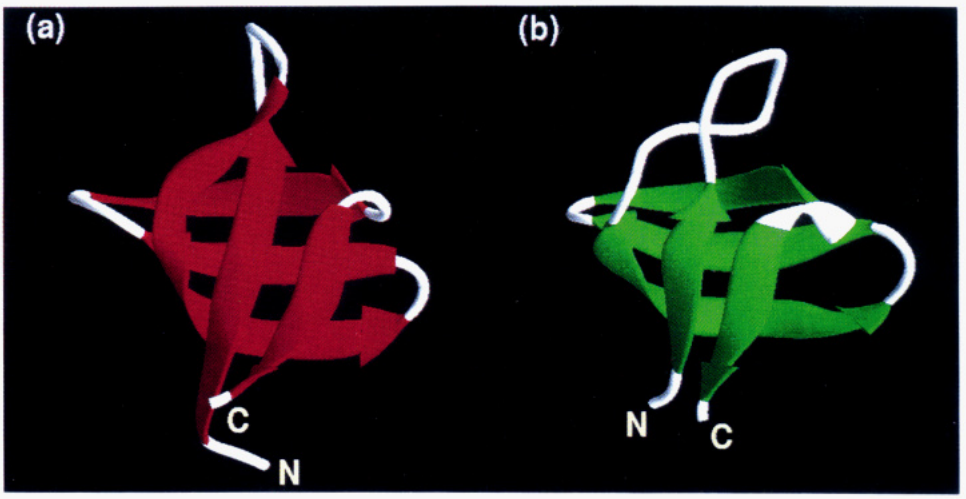

FIGURE 5: Comparison of the topology of the $Int^{220-270}$ monomer (a) and the SH3 domain of spectrin (b). The β-strands in (a) and (b) are shown in red and green, respectively. The coordinates of the SH3 domain of spectrin are from Musacchio et al. (1992). The figure was generated with the program RIBBONS (Carson, 1987).

shown in Figure 3C. It will also be noted that the C-terminal residue of $Int^{220-270}$ is relatively close to the N-terminal one (~11 Å) and that it is positioned such that the remaining C-terminus of the integrase (residues 271−288) can potentially interact with either the catalytic or N-terminal domains.

Relationship to DNA Binding Domains of Other Retroviral Integrases. A comparison of the sequence of $Int^{220-270}$ with the DNA binding domains of other retroviral integrases indicates that the sequence identity ranges from 66% for HIV-2 to 16% for Moloney murine leukemia virus (Puras-Lutzke et al., 1994). Further analysis indicates that both the core residues and the residues involved in the dimer interface are either preserved or substituted conservatively. We therefore suggest that, in all likelihood, the tertiary and possibly the quaternary structures observed for $Int^{220-270}$ are preserved in the other retroviral integrases.

Similarity to SH3 Domains. The topology of the $Int^{220-270}$ monomer is very reminiscent of that of SH3 domains (Kuriyan & Cowburn, 1993) which are found in numerous proteins involved in signal transduction, despite the absence of any significant sequence identity. A comparison of the fold of $Int^{220-270}$ and the SH3 domain of spectrin is shown in Figure 5. The best fit C^{α} RMS difference between $Int^{220-270}$ and the SH3 domains varies from 1.4 Å for 41 residues in the case of spectrin (Musacchio et al., 1992) to 1.9 Å for 39 atoms in the case of the SH3 domain of phospholipase γ (Kohda et al., 1994). Interestingly, the long loop between strands β1 and β2 which closes off the back of the β-barrel in the case of the SH3 domains is much shorter in $Int^{220-270}$. It is also worth noting that a comparison of the structure-based sequence alignment of $Int^{220-270}$ with the SH3 domains reveals that the hydrophobic nature of the core residues is preserved, lending credence to the hypothesis that the protein fold is in part determined by a binary hydrophobic/hydrophilic code (Kamtekar et al., 1993). This is further supported by the finding of SH3-like folds in a wide range of proteins, despite the absence of any significant sequence identity with either SH3 domains or $Int^{220-270}$. These include the carboxy-terminal domains of papain (Kamphuis et al., 1984) and the BirA repressor (Wilson et al., 1992), type II plasmid-encoded R67 dihydrofolate reductase (Matthews et al., 1986), the PsaE polypeptide from photosystem I of cyanobacteria (Falzone et al., 1994), and the small DNA binding protein Sso7d from the archaebac-

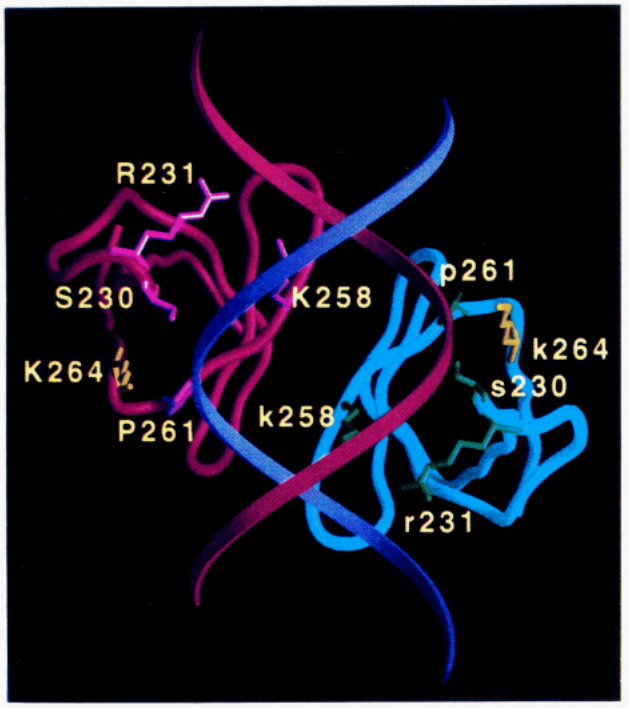

FIGURE 6: Ribbon model of the $Int^{220-270}$ dimer complexed to DNA illustrating that the saddle-shaped groove of the dimer has appropriate dimensions to fit into the major groove of DNA and that the position of Lys264 (shown in yellow), a key residue in the interaction with DNA as defined by mutagenesis, can readily interact with the phosphate backbone. The backbones of the two DNA strands are depicted in red and purple; the backbone of one subunit of $Int^{220-270}$ is shown in red and the other in blue, with the corresponding side chains (except for Lys264) in purple and green, respectively. The structure was generated with the program GRASP (Nicholls, 1993).

terium *Sulfolobus solfatarius* (Baumann et al., 1994). This suggests that the SH3 topology may represent a fold that is ideally suited to graft many different binding properties on, ranging from sites for peptides, to small ligands such as NADP and pteridines, to DNA.

Relationship of Quaternary Structure of $Int^{220-270}$ to Other Dimeric SH3-like Domains. Most of the structures determined for SH3 domains to date confirm the notion that these domains constitute independent folding units of a monomeric nature (Kuriyan & Cowburn, 1993). A recent report, however, of the structure of the mammalian growth factor receptor-binding protein Grb2, which comprises two SH3 domains separated by an SH2 domain, showed that the N- and C-terminal SH3 domains interact with each other, forming a contiguous surface (Maignan et al., 1995). Just like the dimer interface of $Int^{220-270}$, the interaction surface between the two SH3 domains of Grb2 is formed by the stacking of two triple-stranded antiparallel β-sheets (one from each subunit) comprising strands 2, 3, and 4. In contrast to $Int^{220-270}$, where the two β-sheets are stacked antiparallel to each other, the orientation of the two β-sheets in Grb2 is approximately orthogonal. The other example of a dimeric SH3-like domain is that of the homodimeric R67 plasmid-encoded dihydrofolate reductase (Matthews et al., 1986). The arrangement of the two stacked sheets in this dimer seems to be similar to that found in $Int^{220-270}$.

Correlation between Mutational Data and the Putative DNA Binding Surface on $Int^{220-270}$. Recent studies have explored the DNA binding properties of $Int^{220-270}$. The triple and double mutants of [Tyr226,Phe227,Arg228] → [Ala,-Gly,Glu] and of [Trp235,Lys236] → [Lys,Asp] have no effect on DNA binding of HIV-2 integrase (Puras-Lutzke et al., 1994), indicating that the outer surface of the $Int^{220-270}$ dimer formed by strands β1 and β2 is probably not involved in DNA binding. The double mutant [Trp243, Lys244] → [Val,Glu] is insoluble (Puras-Lutzke et al., 1994). It is unlikely that the mutation of Lys244 to Glu in this mutant would have any effect on solubility as Lys244 is a solvent-exposed residue. The replacement of Trp243 by the much smaller Val, however, would be expected to destabilize and probably disrupt the specific dimer interface, thereby exposing a large hydrophobic patch on the surface of the molecule which could potentially undergo nonspecific aggregation both with itself and with a hydrophobic patch formed by Tyr226, Trp235, Pro238, and Ile268 in strands β1, β2, and β5. The Val260 → Glu mutation affects multimeric interactions in intact integrase (Kalpana et al., 1994). This is consistent with the finding that Val260 is located within the hydrophobic core of the $Int^{220-270}$ monomer so that the introduction of a polar negatively charged Glu would be expected to severely destabilize the DNA binding domain. Finally, the mutation of Lys264 → Glu significantly reduced DNA binding (Puras-Lutzke et al., 1994). This residue is located at the edge of the saddle-shaped groove of the dimer. Interestingly, this surface constitutes part of the peptide binding site in the monomeric SH3 domains (Kuriyan & Cowburn, 1993), and the residue at the equivalent position to Lys264 in the SH3 domains, which is invariably a Tyr, is always involved in peptide binding. An area at the dimer interface of the homodimeric R67 dihydrofolate reductase has also been proposed to form the binding surface for cofactors and substrates (Matthews et al., 1986). On this basis, we suggest that Lys258, Pro261, and Lys264, whose equivalents in the SH3 domains are involved in peptide binding, as well as Ser230 and Arg231 located at the tip of the turn connecting strands 1 and 2, may be involved in binding DNA in the case of $Int^{220-270}$.

To determine whether the dimensions of the saddle-shaped groove of the $Int^{220-270}$ dimer are appropriate for DNA binding, we made a model of the complex which is shown in Figure 6. It can be seen that in this model the $Int^{220-270}$ dimer can readily bind to the major groove of the DNA, with the side chains of Ser230, Pro261, Lys258, and Lys264 interacting with the sugar−phosphate backbone, and the side chain of Arg231 interacting with the bases. It is interesting to note that this DNA binding surface is different from that established for the monomeric Sso7d protein (Baumann et al., 1985) where a large portion of the interaction surface corresponds to the dimer interface of $Int^{220-270}$. Hence, this surface would be inaccessible for DNA binding in the case of $Int^{220-270}$ in its dimeric state.

ACKNOWLEDGMENT

We thank Paul Wingfield for carrying out analytical ultracentrifugation on $Int^{220-270}$, Dan Garrett and Frank Delaglio for software support, R. Tschudin for technical support, E. de Castro and S. Edelstein for the program VISP, and Amanda Alteieri, Ad Bax, Andrew Byrd, Robert Clubb, Stefan Grzesiek, Kiyoshi Mizuuchi, Lewis Pannell, and Andy Wang for numerous stimulating discussions.

REFERENCES

Baumann, H., Knapp, S., Lundbäch, T., Ladenstein, R., & Härd, T. (1994) *Nature Struct. Biol. 1*, 808−819.

Baumann, H., Knapp, S., Karshikoff, A., Ladenstein, R., & Härd, T. (1995) *J. Mol. Biol. 247*, 840−846.

Bax A., & Grzesiek, S. (1993) *Acc. Chem. Res. 26*, 131−138.

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R., & Zhu, G. (1994) *Methods Enzymol. 239*, 79−105.

Brooks, B. R., Bruccoleri, R. E. Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.

Brünger, A. T. (1993) *X-PLOR Version 3.1 Manual*, Yale University, New Haven, CT.

Brünger, A. T., & DeLano, W. (1993) *AVS-XPLOR User's Manual*, Yale University, New Haven, CT.

Bushman, F. D., & Wang, B. (1994) *J. Virol. 68*, 2215−2223.

Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. T., & Craigie, R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3428−3432.

Carson, M. (1987) *J. Mol. Graphics 5*, 103−106.

Clore, G. M., & Gronenborn, A. M. (1991) *Science 252*, 1390−1399.

Clore, G. M., & Gronenborn, A. M. (1994) *Protein Sci. 3*, 372−390.

Clore, G. M., Driscoll, P. C., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 7387−7401.

de Castro, E., & Edelstein, S. (1992) *VISP 1.0 User's Guide*, University of Geneva, Geneva, Switzerland.

Drelich, M., Wilhem, R., & Mous, J. (1992) *Virology 188*, 459−468.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., & Davies, D. R. (1994) *Science 266*, 1981−1986.

Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199−203.

Engelman, A., Bushman, F., & Craigie, R. (1993) *EMBO J. 12*, 3269−3275.

Falzone, C. J., Kao, Y.-H., Zhao, J., Bryant, D. A., & Lecomte, J. T. J. (1994) *Biochemistry 33*, 6052−6062.

Forman-Kay, J. D., Clore, G. M., Wingfield, P. T., & Gronenborn, A. M. (1991) *Biochemistry 30*, 2685−2698.

Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P. J., Vuister, G. W., Gronenborn, A. M., & Clore, G. M. (1994) *J. Magn. Reson., Ser. B 104*, 99−103.

Goff, S. P. (1992) *Annu. Rev. Genet. 26*, 527−544.

Jones, S., & Thornton, J. M. (1995) *Prog. Biophys. Mol. Biol. 63*, 31−65.

Kalpana, G. V., Marmon, S., Wang, W., Crabtree, G. R., & Goff, S. P. (1994) *Science 266*, 2002−2006.

Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A., & Drenth, J. (1984) *J. Mol. Biol. 179*, 233−256.

Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., & Hecht, M. H. (1993) *Science 262*, 1680−1685.

Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., & Torchia, D. A. (1992) *J. Magn. Reson. 97*, 359−375.

Kohda, D., Hatanaka, H., Odaka, M., Mandiyan, V., Ullrich, A., Schlessinger, J., & Inagaki, F. (1994) *Cell 72*, 953−963.

Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., & Gronenborn, A. M. (1989) *Biochemistry 28*, 7241−7257.

Kuriyan, J., & Cowburn, D. (1993) *Curr. Opin. Struct. Biol. 3*, 828−837.

Kuszewski, J., Qin, J., Gronenborn, A. M., & Clore, G. M. (1995a) *J. Magn. Reson., Ser. B 106*, 92−96.

Kuszewski, J., Gronenborn, A. M., & Clore, G. M. (1995b) *J. Magn. Reson., Ser. B 107*, 293−297.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.

Maignan, S., Guilloteau, J.-P., Fromage, N., Arnoux, B., Becquart, J., & Ducruix, A. (1995) *Science 268*, 291−293.

Matthews, D. A., Smith, S. L., Baccanari, D. P., Burchall, J. J., Oatley, S. J., & Kraut, J. (1986) *Biochemistry 25*, 4194−4204.

Mumm, S. R., & Grandgenett, D. P. (1991) *J. Virol. 65*, 1160−1167.

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., & Saraste, M. (1992) *Nature 359*, 851−855.

Nicholls, A. J., Sharp, K., & Honig, B. (1991) *Proteins 11*, 281−296.

Nilges, M. (1993) *Proteins: Struct., Funct., Genet. 17*, 297−309.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1990) *Biopolymers 29*, 813−822.

Puras-Lutzke, R. A., Vink, C., & Plasterk, R. H. A. (1994) *Nucleic Acids Res. 22*, 4125−4131.

Schauer, M., & Billich, A. (1992) *Biochem. Biophys. Res. Commun. 185*, 874−888.

van Gent, D. C., Elgersma, Y., Bolk, M. W. J., Vink, C., & Plasterk, R. H. A. (1991) *Nucleic Acids Res. 19*, 3821−3827.

van Gent, D. C., Vink, C., Oude Groeneger, A. A. M., & Plasterk, R. H. A. (1993) *EMBO J. 12*, 3261−3267.

Vink, C., & Plasterk, R. H. A. (1993) *Trends Genet. 9*, 433−437.

Vink, C., Oude Groeneger, A. A. M., & Plasterk, R. H. A., (1993) *Nucleic Acids Res. 21*, 1419−1425.

Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J., & Matthews, B. W. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9257−9261.

BI951055U