PROTEINS: Structure, Function, and Genetics 36:556–564 (1999)

Refined Solution Structure of the C-Terminal DNA-Binding Domain of Human Immunovirus-1 Integrase

Astrid P.A.M. Eijkelenboom,[1] Remco Sprangers,[1] Karl Hård,[1] Ramon A. Puras Lutzke,[2] Ronald H.A. Plasterk,[2] Rolf Boelens,[1] and Robert Kaptein[1]*

[1]*Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
[2]*Division of Molecular Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands*

ABSTRACT **The structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase has been refined using nuclear magnetic resonance spectroscopy. The protein is a dimer in solution and shows a well-defined dimer interface. The folding topology of the monomer consists of a five-stranded β-barrel that resembles that of Src homology 3 domains. Compared with our previously reported structure, the structure is now defined far better. The final 42 structures display a back-bone root mean square deviation versus the average of 0.46 Å. Correlation of the structure with recent mutagenesis studies suggests two possible models for DNA binding. Proteins 1999;36:556–564.**

Key words: nuclear magnetic resonance spectroscopy; protein structure; dimer

INTRODUCTION

Integration of viral c-DNA into the genome of the infected cell is an essential step in the life cycle of the human immunodeficiency virus (HIV; for recent reviews, see Asante-Appiah and Skalka[1] and Puras Lutzke and Plasterk[2]). This process is catalyzed by the viral enzyme integrase (IN), which mediates two reactions. First, two nucleotides are removed from both 3′ ends of the blunt-ended viral DNA next to a conserved CA dinucleotide (3′ processing or cleavage reaction). Subsequently, the newly generated 3′-hydroxyl groups are coupled to phosphates in the target DNA (strand-transfer or integration reaction). This latter reaction is not sequence-specific but depends on structural features of the target DNA. Finally the integration intermediate is repaired, presumably by cellular enzymes.

In vitro, IN catalyzes the cleavage and integration reactions by using oligonucleotides that mimic the viral DNA ends and magnesium or manganese as a divalent metal cofactor.[3–5] IN also can catalyze the reversal of the integration reaction, termed disintegration, in vitro,[6] which is taken as a measure for the catalytic activity of IN.

Three domains have been identified in HIV integrase:[7,8] an N-terminal, zinc-binding HHCC domain; a central catalytic core; and a C-terminal DNA-binding domain. All three domains are required for cleavage and integration,[8,9] but the disintegration reaction can be performed by the catalytic core alone.[7,8]

The isolated N-terminal, zinc-binding domain is folded and dimeric only in the presence of zinc.[10,11] In the full-length protein, zinc induces tetramerization and enhances catalytic activity.[12–14] This observation, as well as results from other studies,[15,16] indicate that the N-terminal domain is involved in protein-protein interactions in the active multimer. Furthermore, this domain also seems to contribute to the recognition of the viral DNA ends.[16,17] The structure of the N-terminal domain (residues 1–55) has been solved by using nuclear magnetic resonance (NMR) spectroscopy, both for HIV-1[10] and HIV-2 IN.[11] It consists of a three-helix bundle, which is stabilized by the zinc-binding HHCC motif. The three-helix bundle fold is similar to that found in several DNA-binding proteins in which the second and the third helices form a helix-turn-helix motif. These proteins use the third helix to bind DNA, whereas this helix is part of the dimer interface of HIV-1 IN.

The catalytic core domain contains a DD(35)E motif that is essential for all catalytic activities.[18–22] Furthermore, this domain is involved in binding of the termini of the viral DNA and also is responsible for target site selection.[23–27] X-ray structures of the catalytic core with and without metal cofactor ions of HIV-1 IN[28–30] and avian sarcoma virus (ASV) IN[31,32] have been solved. These structures resemble that of other polynucleotidyl transferases, such as ribonuclease H, RuvC resolvase, and MuA transposase.[33,34]

The C-terminal domain can bind DNA in a nonspecific manner,[8,35–38] and the minimal region that is required for DNA binding has been mapped to residues 220–270.[38] Recent photo-cross-linking studies using full-length IN showed that a region comprising residues 247–270 contacts bases A(4) and T(5) of the U5 viral DNA end.[24,39] This suggests that the C-terminal domain is involved directly in stabilizing the interaction of IN with the viral DNA ends.

Grant sponsor: Netherlands Organization of for Scientific Research; Grant sponsor: The European Union; Grant number: ERBFMGECT950032.
Karl Hård's present address is Astra Structural Chemistry Laboratory, S-43183 Mölndal, Sweden.
Ramon A. Puras Lutzke's present address is Department of Biochemistry and Molecular Biology, Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands.
*Correspondence to: Robert Kaptein, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. E-mail: kaptein@nmr.chem.uu.nl
Received 12 February 1999; Accepted 11 May 1999

The C-terminal domain also seems to be important for functional multimerization of IN. Although both the isolated catalytic core domain and the C-terminal domain are dimeric,[28,40,41] a construct comprising both domains exists in a dimer-tetramer equilibrium.[42] In full-length IN, mutations within the C-terminal domain that shift the dimer-tetramer equilibrium to the dimer state give rise to strongly reduced catalytic activities.[43] The structure of the DNA-binding domain has been solved by using NMR spectroscopy.[40,41] It is a homodimer, and the folding topology of the monomers closely resembles that of Src homology 3 (SH3) domains. SH3 domains are approximately 60-residue protein modules that mediate protein-protein interactions by binding to proline-rich sequences (for recent reviews, see Dalgarno et al.[44] and Kuriyan and Cowburn[45]). An SH3-like fold is found in various proteins with different biologic functions, despite the lack of sequence homology (for an overview, see the SCOP database[46]). Among these are the BirA repressor[47] and the diphtheria toxin repressor,[48] which have an SH3-like fold at the C-terminus and a three-helix-bundle fold containing the helix-turn-helix motif at the N-terminus, just like IN. Here, we report on the structure refinement of the C-terminal DNA-binding domain (IN-DBD), and we correlate the structure to the results of mutagenesis studies on this domain.[38,43,49]

MATERIALS AND METHODS

IN-DBD Preparation

The cloning, expression, purification, and sample preparation of the used, unlabelled ^{15}N and the $^{15}N/^{13}C$-labelled IN-DBD protein samples (residues 220–270) were described previously.[38,40] In addition, a 10% ^{13}C-labelled sample was prepared by growing cells in minimal medium with 10% [U-^{13}C]-glucose and 90% unlabelled glucose as the sole carbon source. All protein constructs contained an extra Met at the N-terminus and a C-terminal His-tag (6*His). For our NMR studies, the protein was dissolved in 100 mM NaCl and 2 M glycine-d_5, and the pH was adjusted to 4.7.[40]

NMR Spectroscopy and Structure Calculations

NMR spectra were recorded at 298 K on a Bruker AMX-600 and a Varian Unity+ 750-MHz spectrometer equipped with triple-resonance gradient probes. Assignment of proton, carbon, and nitrogen chemical shifts has been described previously.[40] In addition, methyl group H and C resonances of all Leu (3) and all Val (5) residues were assigned stereospecifically using a ^{13}C-heteronuclear single quantum coherence (HSQC) of the 10% ^{13}C-labelled protein, as described by Neri et al.[50] Stereospecific assignment of the V259 methyl groups was possible only based on the carbon chemical shifts as a result of overlapping proton shifts.

Approximate interproton distances were derived from two-dimensional (2D) ^1H nuclear Overhauser effect spectroscopy (NOESY) spectra, recorded with mixing times of 60 msec, 100 msec, and 160 msec. NOE crosspeaks were classified as strong (up to 2.6 Å), medium (up to 3.5 Å), weak (up to 5.0 Å), or very weak (up to 6.5 Å). In cases of overlap, NOE distance restraints were derived from either sensitivity-enhanced, three-dimensional (3D) N-NOESY-HSQC or 3D C-NOESY-HSQC spectra, using NOE distance restraint upper bounds of 3.0 Å, 4.0 Å, and 5.0 Å. All lower-bound distance restraints were effectively the sum of the van der Waals radii (1.8 Å).

Intermonomer NOE crosspeaks were identified from 2D and 3D, N/C double-half-filtered NOESY spectra acquired on a 1:1 mixture of N/C-labelled protein and unlabelled protein. During the refinement procedure, additional intermonomer and comonomer (which have both an intramonomer and an intermonomer contribution) distance restraints were derived from conventional 2D and 3D NOESY spectra based on the calculated structures.

Phi dihedral angle restraints were derived from $^3J_{HNH\alpha}$ coupling constants obtained from an HNHα experiment.[51] Furthermore, protein back-bone hydrogen bond restraints (two per hydrogen bond) were included in the final structure calculations based on the regularly observed secondary structure, using distance restraints up to 2.5 Å for NH-O and from 2.4 Å to 3.5 Å for N-O distances.

Structures were calculated with X-PLOR version 3.851[52] using the dynamically simulated annealing protocol,[53] in which two extra terms were introduced to handle the symmetric dimer.[54] Noncrystallographic symmetry (NCS) restraints were included to keep the monomers superimposable, and a global symmetry potential was added to keep the dimer symmetric.[54] Sum averaging[54] was used to correct for multiplicity and to include distance restraints resulting from ambiguous NOE crosspeaks. Corrections on the upper bounds of restraints containing nonstereospecifically assigned diastereotopic groups were made as described by Fletcher et al.[55] During the final refinement step, the following force constants were used in the X-PLOR target function for minimization: 1,000 kcal mol^{-1} $Å^{-2}$ for bond lengths, 500 kcal mol^{-1} rad^{-2} for angles and impropers, 100 kcal mol^{-1} $Å^{-2}$ for NCS restraints,[41] 200 kcal mol^{-1} rad^{-2} for dihedral angle restraints, 50 kcal mol^{-1} $Å^{-2}$ for NOE distance restraints, 1 kcal mol^{-1} $Å^{-2}$ for symmetry table distance restraints,[54] and 4 kcal mol^{-1} $Å^{-2}$ for the van der Waals repulsion term.

Analysis of the resulting structures was done by using the program PROCHECK-NMR,[56] and secondary structure elements were identified according to the DSSP algorithm.[57] The program MOLMOL[58] was used to produce Figures 1, 3, and 4. The coordinates of the structures and restraint data have been submitted to the to the Brookhaven Protein Data Bank.

RESULTS AND DISCUSSION

Structure Calculations and Description of the Structure

By using 2,156 distance restraints (21 per residue) and 36 dihedral angle restraints per dimer, we calculated 50 structures, of which 42 were selected on the basis of low overall energy. A superposition of the 42 structures is shown in Figure 1a, a summary of the structural statistics

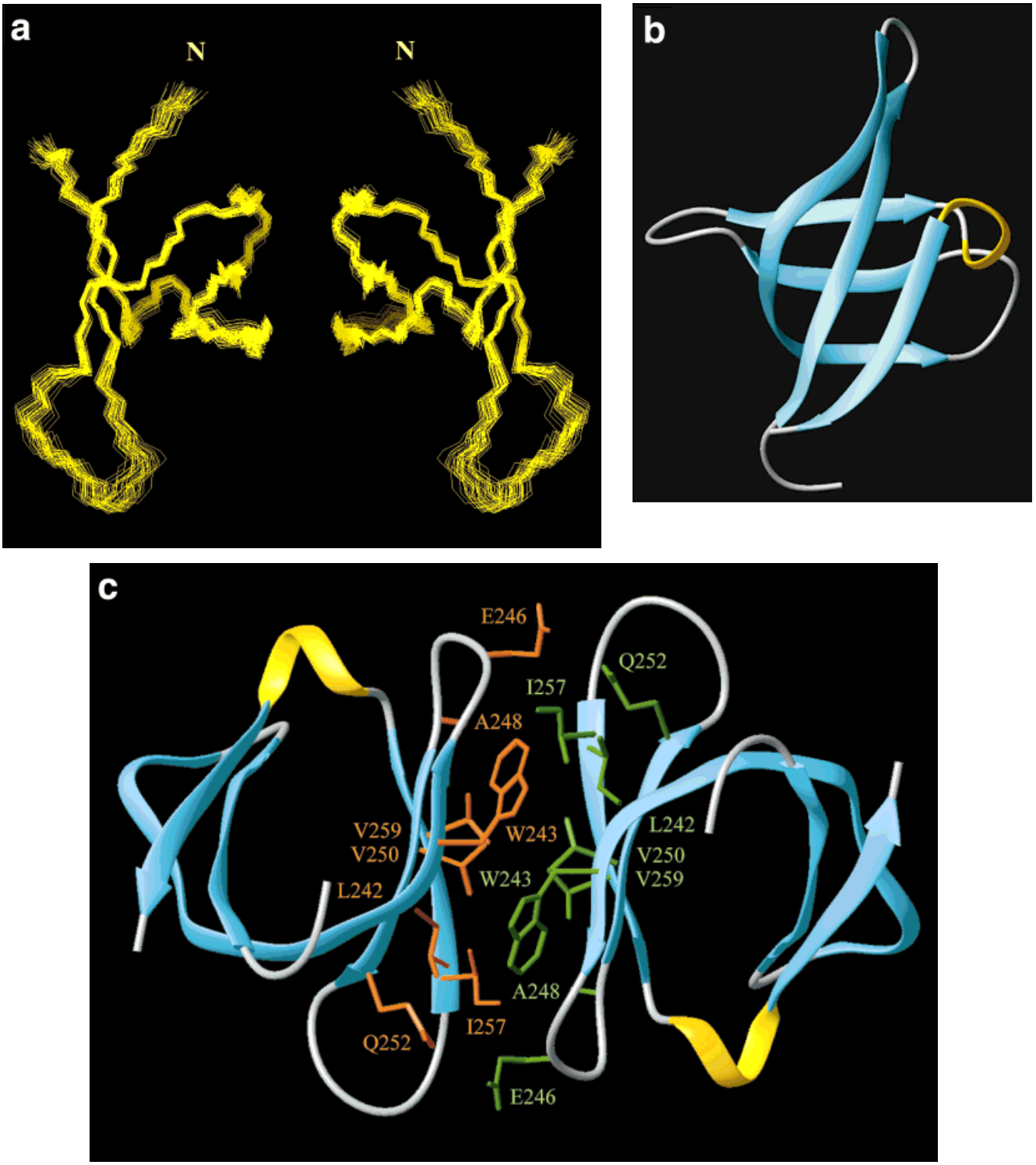

Fig. 1. Structure of the integrase (IN) DNA-binding domain (IN-DBD). **a:** Superposition of back-bone traces of 42 structures. The two-fold axis of the dimer is approximately parallel to the plane of the paper. N designates the N-terminal residue. **b:** Folding topology of the monomer. **c:** Structure of the dimer showing the interfacial residues. The two-fold axis of the dimer is approximately perpendicular to the plane of the paper.

is given in Table I, and an overview of structural parameters is shown in Figure 2. The precision of the structures has improved significantly compared with the previously reported structures.[40] The back-bone root mean square deviation (RMSD) from the average for the dimer has decreased from 0.98 Å to 0.46 Å for residues 220–270 and

TABLE I. Structural Statistics for Human Immunovirus-1 Integrase DNA Binding Domain[†]

Type and number of restraints for the dimer	
Distance restraints	2,156
Intramonomer	1,978
Intraresidual	486
Sequential	474
Medium range	292
Long range	726
Intermonomer	48
Ambiguous[a]	58
Hydrogenbonds	72
Dihedral restraints	36
Maximum experimental violations	
Distance restraints (Å)	0.18
Dihedral restraints (°)	1.0
RMSD from average structure (Å)[b]	
Back bone (N, Cα, C′)	0.46 ± 0.13 (0.27 ± 0.06)
All heavy atoms	0.92 ± 0.12 (0.81 ± 0.07)
Deviation from experimental restraints	
Distance restraints (Å)	0.012 ± 0.001
Dihedral restraints (°)	0.10 ± 0.07
Deviation from idealized covalent geometry	
Bonds (Å)	0.0019 ± 0.0001
Angles (°)	0.38 ± 0.01
Impropers (°)	0.18 ± 0.01
Percentage of residues[c] with ϕ/ψ in	
Most favored region	79.1
Additionally allowed regions	17.3
Generously allowed regions	3.6
Disallowed regions	0.0
Average number of bad contacts per 100 residues	3.4

[†]Given for 42 final structures.
[a]Fifty-two comonomer, 2 intramonomer, and 4 intermonomer. The intramonomer and intermonomer ambiguous distance restraints are a result of the overlapping resonances of 243 HH2 and 243 HE3.
[b]Root mean square deviation (RMSD) for the monomer is given in parentheses.
[c]Excluding glycine and proline residues.

from 0.69 Å to 0.37 Å when comparing residues 222–229 with residues 234–270. This improvement is caused by a considerably greater number of distance restraints (814 additional) and the inclusion of ϕ dihedral angle restraints.

The structure of the monomer consists of a five-stranded, antiparallel β-barrel and a three-residue 3_{10}-helix (see Fig. 1b). The β-strands comprise residues 223–228 (β1), 235–244 (β2), 249–252 (β3), 257–260 (β4), and 265–269 (β5) in at least 85% of the structures. In 50% of the structures, β-1 strands comprise residues 223–229 and β-2 comprises residues 232–244. Strand β2 contains bulges at residues 233–234, 237–238, and 242. Residues 262–264 form the short 3_{10}-helix in 90 % of the structures. The hydrophobic core of the monomer is formed by residues V225, Y227, A239, L241, V249, I251, V260, and A265. Compared with the previous structures, especially residues 229–234 are defined far better, although this is still the least well-defined region of the protein.

The dimer interface is formed by two three-stranded, antiparallel β-sheets that consist of parts of strands β2, β3, and β4. These β-sheets of the individual monomers are oriented approximately antiparallel to each other. The interface is composed of the hydrophobic residues L242, W243, A248, V250, I257, and V259 and the hydrophilic residues E246 and Q252, in which the side chains form a mutual hydrogen bond in 55% of the structures (Fig. 1c). During structure refinement, it became clear that we had to attribute a previously identified intermonomer NOE to an artifact. Consequently, the side chain of residue L241 has changed from a position pointing toward W243 of the other monomer to a position pointing away from the dimer interface. Apart from this, the dimer interface stayed essentially the same.

Comparison With the Structure of IN-DBD by Lodi et al.[41]

The structure of IN-DBD also has been solved by Lodi et al.[41] The protein fragment used by us and by those authors differs for two residues: N232 and L234 in our fragment ($IN\text{-}DBD_{NL}$) are D232 and V234 in the fragment of Lodi et al. ($IN\text{-}DBD_{DV}$). In addition, our construct contained an additional C-terminal, 6-residue His tag, and that of Lodi et al. contained three extra residues (GSH) at the N-terminus. Furthermore, the sample conditions were different for both cases. We used 100 mM NaCl, 2 M glycine, pH 4.7, as NMR sample conditions, whereas Lodi et al. used 100 mM NaCl, 50 mM NH_2PO_4, 0.5 mM EDTA, pH 6.5.

The overall folding topology of the structures is very similar. Both structures consist of a five-stranded β-barrel and a three-residue 3_{10}-helix, in which the β-strands run from residues 223–228, 235–244, 249–252, 257–260, and 265–269. In all structures of Lodi et al.,[41] strands β3 and β4 are extended by one residue to residues 248 and 261, respectively, whereas this extension is found in only 17% of our structures.

The overall quality of the Ramachandran plot is also comparable for both structures, in which 80.1% and 79.1% of the residues excluding glycines and prolines are in the most favorable region for $IN\text{-}DBD_{DV}$ and $IN\text{-}DBD_{NL}$, respectively, as determined by PROCHECK-NMR. In addition, the residues that have well-defined χ_1 dihedral angles in both structures are all in the same rotameric state. However, there are significant local differences that result in a pairwise RMSD of 1.6 Å between the structure that is closest to the average of our structures and the energy-minimized average structure of Lodi et al.[41] for the dimer. Comparing only the monomer structures, the RMSD is 1.0 Å. A superposition of the structures is shown in Figure 3. Significant differences are found mainly in the termini and the regions that connect the β strands. These differences often go together with significant differences in back-bone dihedral angles. The largest differences are observed for residues 221–222 and the less well-defined residues 228–233. Interestingly, for $IN\text{-}DBD_{DV}$, residues 229 and 232 are both an aspartate, whereas, for $IN\text{-}DBD_{NL}$, residue 232 is

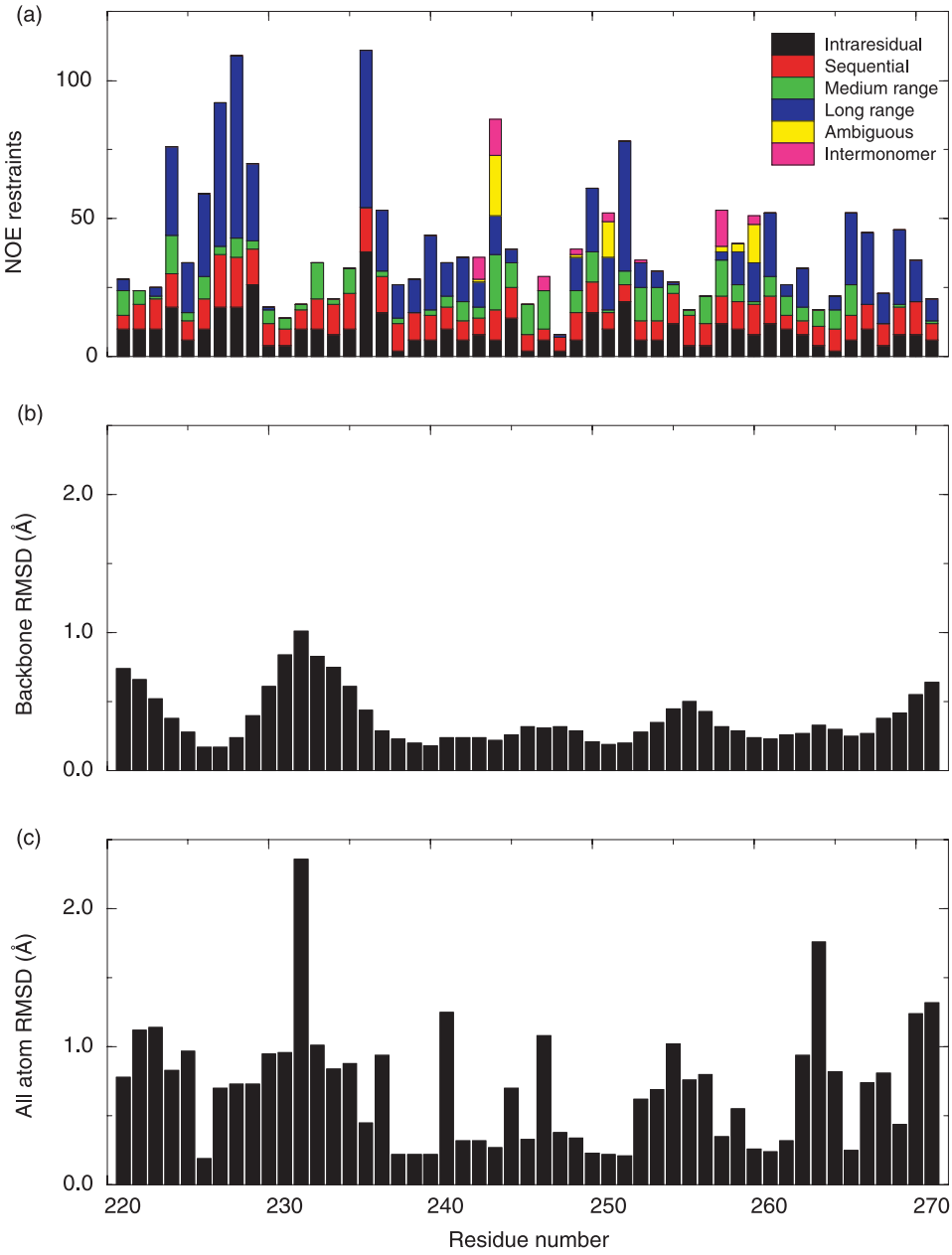

Fig. 2. Structural parameters of IN-DBD. **a:** Number of nuclear Overhauser effect (NOE) distance restraints per residue. The restraints are classified as intraresidual, sequential, medium range, long range, ambiguous, and intermonomer. **b,c:** Root mean square deviation (RMSD) for the backbone N, Cα, C′ atoms and all heavy atoms, respectively.

an asparagine. It is possible that the presence of the two negatively charged residues closely together in space in IN-DBD$_{DV}$ influences the local conformation. The orientation of the two monomers with respect to each other is very similar, although residues 244–248 are pointing more toward the dimer interface for IN-DBD$_{NL}$.

Correlation of the Structure to Mutational Data on IN-DBD

Mutational studies on the C-terminal domain of HIV IN have revealed several residues that are important for DNA-binding as well as for the formation of the active multimer.

In the context of DNA binding by the isolated C-terminal domain of HIV-2 IN, it has been shown that mutant protein K264E has reduced DNA binding affinity.[38] This residue is located at the edge of the saddle-shaped groove of the IN-DBD dimer, and it has been suggested that this groove may be suitable for DNA binding involving residues S230, R231, K258, P261, and K264.[41] Recently, a number of amino acid mutations in IN-DBD of HIV-1 have been studied.[43] For isolated IN-DBD, two mutant proteins were found that showed reduced DNA binding affinity, L234A and R262G. In full-length IN, these mutations resulted in reduced 3′ processing activity compared with wild-type activity.[43] This is consistent with a role for IN-DBD in

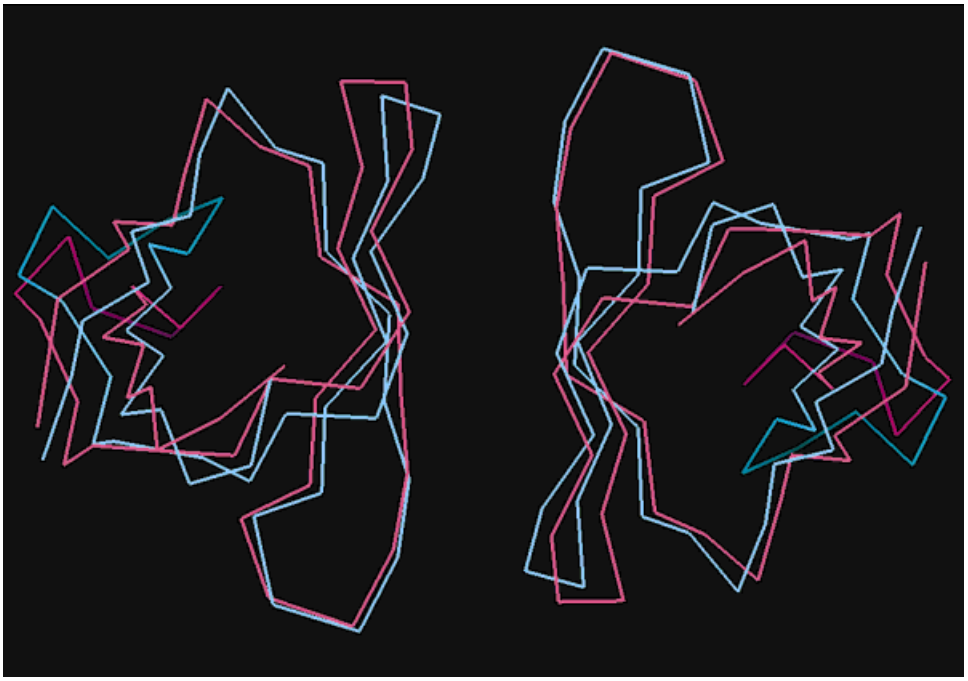

Fig. 3. Superposition of Cα traces of the dimer of IN-DBD residues N232 and L234 (IN-DBD_{NL}; in cyan) and residues D232 and V234 (IN-DBD_{DV}) from Lodi et al.[41] (in magenta). Shown are the structure that is closest to the average and the energy-minimized average, respectively. The structures are superimposed on residues 223–227 and 234–268. The orientation is similar to that shown in Figure 1c.

stabilizing the interaction of IN with the viral DNA. Furthermore, mutation R231A was studied in the context of full-length IN, and this mutant protein showed reduced 3′ processing and DNA strand-transfer activities.[43] Although it is not clear whether this reduction was due to a loss of DNA binding affinity or of intramolecular or intermolecular interactions, a role for R231 in DNA-binding is considered possible. Figure 4a shows that it is possible for residues R231, R262, and K264 to interact with DNA located in the saddle-shaped groove; however, residue L234 seems too far away to contribute to the interactions. However, conformational changes may occur upon DNA binding, and these could bring residue L234 closer to the saddle-shaped groove. Alternatively, DNA also could bind at a surface outside this groove, which is indicated in Figure 4b. This latter model would imply that there are two possible binding sites for DNA in the IN-DBD dimer.

Three mutant, full-length IN proteins were found with strongly reduced 3′ processing and DNA strand-transfer activities: L241A, L242A, and V260E43. Moreover, mutants L241A and V260E also showed disintegration activities comparable to a C-terminal deletion of IN. It has been shown that mutation V260E eliminates intermolecular interactions in IN,[49] and that this mutation of the hydrophobic core residue V260 to a negatively charged glutamate causes misfolding of full-length IN.[43] Mutant proteins L241A and L242A were found to shift the dimer-tetramer equilibrium of wild-type IN to the dimer state, and this effect is especially strong for the former mutant protein.[43] These results support the notion that the C-terminal domain is important for the formation of the active multimer. L242 is part of the dimer interface, pointing toward residue W243, and mutation of this residue could disrupt the dimer interface of IN-DBD. This would imply that the dimerization of IN-DBD is functional for the multimerization of IN. Residue L241 is located just outside the dimer interface. It is packed to the hydrophobic core on one side and is covered by residues I220 and I267 of the termini on the other side. Mutation of L241, therefore, may disrupt the proper folding of IN-DBD, thereby preventing formation of the active multimer.

Two different models have been suggested for tetrameric IN.[10,59] In one model, the IN-DBD dimer acts as a bridge between two catalytic core dimers,[10] consistent with a role of IN-DBD for functional multimerization of IN. In the other model, the IN-DBD dimer is positioned directly on top of the catalytic core dimer, and the catalytic core dimers interact with each other.[59] The authors of that study indicate that the C-terminal residues 271–288, which are not included in the NMR structure, can contact the other core dimer, thereby explaining the importance of the C-terminal domain for tetramerization in this model.

CONCLUSIONS

We have refined the solution structure of the C-terminal DNA-binding domain of HIV-1 IN, which is now defined far better. Mapping of mutagenesis data on the structure locate the residues that are involved in DNA-binding (L234, R262, and K264) as well as residue R231, which is important for 3′ processing and DNA strand-transfer activities. Two models for DNA binding can be suggested. Lodi et al.[41] proposed that DNA could bind to the saddle-shaped groove of the IN-DBD dimer. An alternative model could be that DNA binds to the outer surface of IN-DBD. In that case, there would be two sites available for DNA-binding.

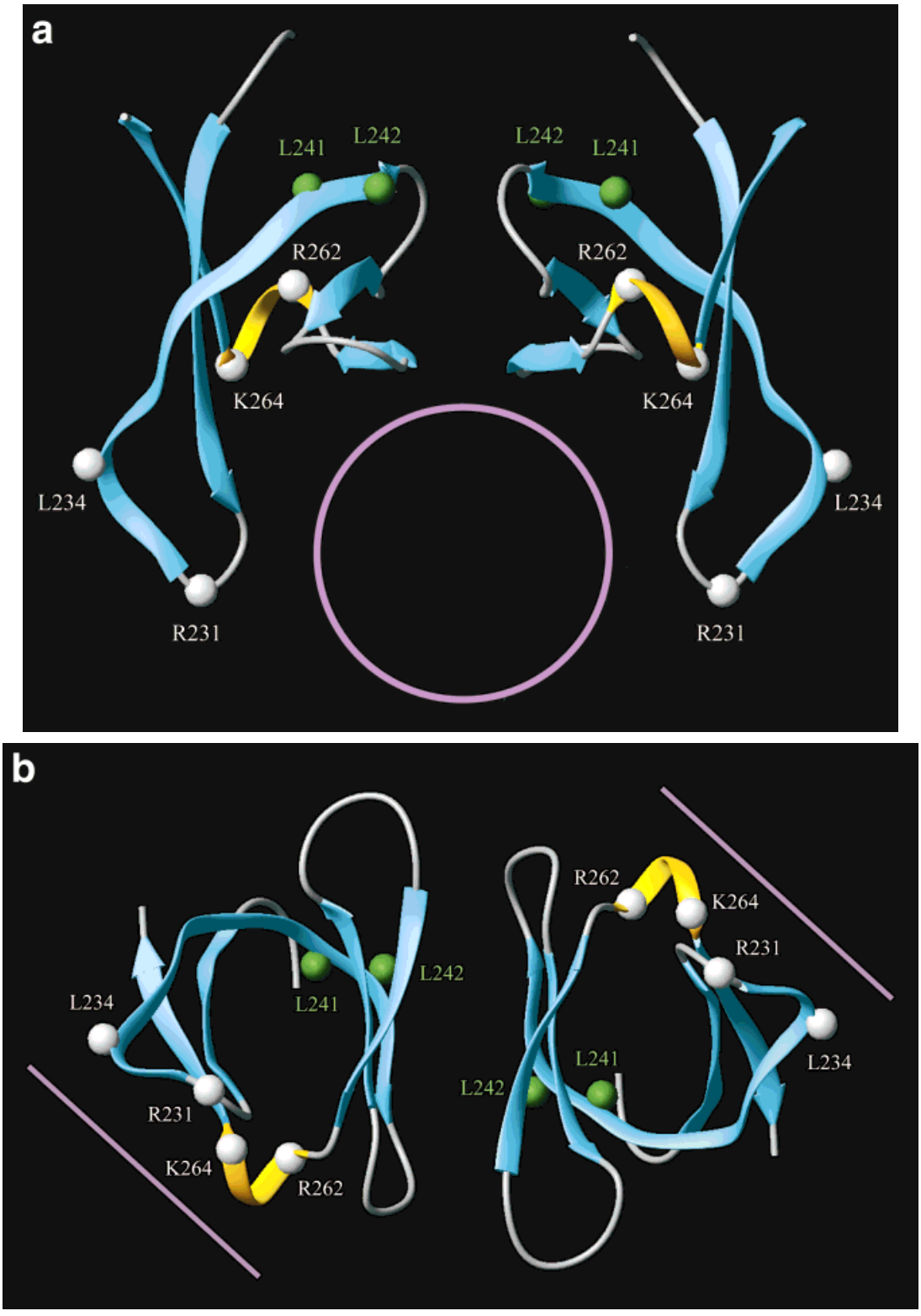

Fig. 4. Functionally important residues of IN-DBD. Mutated residues that are important for DNA binding and multimerization are indicated in white and green, respectively. **a:** Orientation similar to that shown in Figure 1a, showing the groove that potentially binds DNA. **b:** Orientation rotated 180° from that shown in Figure 1c showing the outer side of IN-DBD where residues R231, L234, R263, and K264 cluster. In both cases, the DNA is indicated schematically in purple.

ACKNOWLEDGMENTS

The authors thank Albert George for technical assistance and Alexandre Bonvin for helpful discussions. A.E. was supported by the Netherlands Organization of for Scientific Research (NWO-CW). The 750-MHz spectra were recorded at the SON NMR Large-Scale Facility (Utrecht), which is supported by the Large-Scale Facility Program of the European Union (Contract ERBFMGECT950032).

REFERENCES

1. Asante-Appiah E, Skalka AM. Molecular mechanisms in retrovirus DNA integration. Antiviral Res 1997;36:139–156.
2. Puras Lutzke RA, Plasterk RHA. HIV integrase: a target for drug discovery. Genes Funct 1998;1:289–307.
3. Bushman FD, Craigie R. Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. Proc Natl Acad Sci USA 1991;88:1339–1343.
4. Sherman PA, Fyfe JA. Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. Proc Natl Acad Sci USA 1990;87:5119–5123.
5. LaFemina RL, Callahan PL, Cordingley MG. Substrate specificity of recombinant human immunodeficiency virus integrase protein. J Virol 1991;65:5624–5630.
6. Chow SA, Vincent KA, Ellison V, Brown O. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. Science 1992;255:723–726.
7. Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. Proc Natl Acad Sci USA 1993;90:3428–3432.
8. Vink C, Oude Groeneger AM, Plasterk RHA. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. Nucleic Acids Res 1993;21:1419–1425.
9. Schauer M, Billich A. The N-terminal domain region of HIV-1 integrase is required for integration activity, but not for DNA-binding. Biochem Biophys Res Commun 1992;186:874–888.
10. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn, AM. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. Nature Struct Biol 1997;4:567–577.
11. Eijkelenboom APAM, van den Ent FMI, Vos A, Doreleijers JF, Hård K, Tullius TD, Plasterk RHA, Kaptein R, Boelens R. The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. Curr Biol 1997;7:739–746.
12. Zheng R, Jenkins TM, Craigi R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. Proc Natl Acad Sci USA 1996;93:13659–13664.
13. Lee SP, Han MK. Zinc stimulates Mg^{2+}-dependent 3′-processing activity of human immunodeficiency virus type 1 integrase in vitro. Biochemistry 1996;35:3837–3844.
14. Lee SP, Xiao J, Knutson JR, Lewis MS, Han MK. Zn^{2+} promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. Biochemistry 1997;36:173–180.
15. Ellison V, Gerton J, Vincent KA, Brown PO. An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. J Biol Chem 1995;270:3320–3326.
16. van den Ent FMI, Vos A, Plasterk RHA. Dissecting the role of the N-terminal domain of HIV integrase by trans complementation analysis. J Virol 1999;73:3176–3183.
17. Vincent KA, Ellison V, Chow SA, Brown PO. Characterization of human immunodeficiency virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino-terminal mutations. J Virol 1993;67:425–437.
18. Drelich M, Wilhelm R, Mous J. Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein in vitro. Virology 1992;188:459–468.
19. Engelman A, Craigie R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. J Virol 1992;66:6361–6369.
20. Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol 1992;12:2331–2338.
21. Leavitt AD, Shiue L, Varmus HE. Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. J Biol Chem 1993;268:2113–2119.
22. van Gent DC, Groeneger AA, Plasterk RHA. Mutational analysis of the integrase protein of human immunodeficiency virus type 2. Proc Natl Acad Sci USA 1992;89:9598–9602.
23. Gerton JL, Ohgi S, Olsen M, DeRisi J, Brown PO. Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions. J Virol 1998;72:5046–5055.
24. Heuer TS, Brown PO. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. Biochemistry 1997;36:10655–10665.
25. Jenkins TM, Esposito D, Engelman A, Craigie R. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. EMBO J 1997;16:6849–6859.
26. Katzman M, Sudol M. Mapping viral DNA specificity to the central region of integrase by using functional human immunodeficiency virus type 1/visna virus chimeric proteins. J Virol 1998;72:1744–1753.
27. Shibagaki Y, Chow SA. Central core domain of retroviral integrase is responsible for target site selection. J Biol Chem 1997;272:8361–8369.
28. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science 1994;266:1981–1986.
29. Maignan S, Guilloteau JP, Zhou-Liu Q, Clement-Mella C, Mikol V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. J Mol Biol 1998;282:359–368.
30. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. Proc Natl Acad Sci USA 1998;95:9150–9154.
31. Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, Katz RA, Skalka AM. High-resolution structure of the catalytic domain of avian sarcoma virus integrase. J Mol Biol 1995;253:333–346.
32. Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, Katz RA, Skalka AM. The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. Structure 1996;4:89–96.
33. Yang W, Steitz TA. Recombining the structures of HIV integrase, RuvC and RNase H. Structure 1995;3:131–134.
34. Rice P, Craigie R, Davies DR. Retroviral integrases and their cousins. Curr Opin Struct Biol 1996:6;76–83.
35. Engelman A, Hickman AB, Craigie R. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. J Virol 1994;68:5911–5917.
36. Khan E, Mack JP, Katz RA, Kulkosky J, Skalka AM. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res 1991;19:851–860.
37. Woerner AM, Marcus-Sekura CJ. Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. Nucleic Acids Res 1993;21:3507–3511.
38. Puras Lutzke RA, Vink C, Plasterk RHA. Characterization of the minimal DNA-binding domain of the HIV integrase protein. Nucleic Acids Res 1994;22:4125–4131.
39. Esposito D, Craigie R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. EMBO J 1998;17:5832–5843.
40. Eijkelenboom APAM, Puras Lutzke RA, Boelens R, Plasterk RHA, Kaptein R, Hård K. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. Nature Struct Biol 1995;2:807–810.
41. Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM. Solution structure of the DNA binding domain of HIV-1 integrase. Biochemistry 1995:34;9826–9833.

42. Jenkins TM, Engelman A, Ghirlando R, Craigie R. A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. J Biol Chem 1996; 271:7712–7718.
43. Puras Lutzke RA, Plasterk RHA. Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. J Virol 1998;72:4841–4848.
44. Dalgarno DC, Botfield MC, Rickles RJ. SH3 domains and drug design: ligands, structure, and biological function. Biopolymers 1997;43:383–400.
45. Kuriyan J, Cowburn D. Modular peptide recognition domains in eukaryotic signaling. Annu Rev Biophys Biomol Struct 1997;26: 259–288.
46. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995:247;536–540.
47. Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW. *Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. Proc Natl Acad Sci USA 1992;89:9257–9261.
48. Qiu X, Pohl E, Holmes RK, Hol WG. High-resolution structure of the diphtheria toxin repressor complexed with cobalt and manganese reveals an SH3-like third domain and suggests a possible role of phosphate as co-corepressor. Biochemistry 1996;35:12292–12302.
49. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science 1994;266:2002–2006.
50. Neri D, Szyperski T, Otting G, Senn H, Wüthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. Biochemistry 1989;28:7510–7516.
51. Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three bond $J(H^NH^{\cdot})$ coupling constants in ^{15}N-enriched proteins. J Am Chem Soc 1993;115:7772–7777.
52. Brünger AT. X-PLOR, version 3.1: a system for X-ray crystallography and NMR. New Haven, CT: Yale University Press; 1992.
53. Nilges M, Clore GM, Gronenborn AM. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. Circumventing problems associated with folding. FEBS Lett 1988;239:129–136.
54. Nilges M. A calculation strategy for the structure determination of symmetric dimers by ^1H NMR. Proteins 1993;17:297–309.
55. Fletcher CM, Jones DN, Diamond R, Neuhaus D. Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. J Biomol NMR 1996;8:292–310.
56. Laskowski RA, Rullman JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the stereochemical quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
57. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
58. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.
59. Heuer TS, Brown PO. Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. Biochemistry 1998;37:6667–6678.