Solution structure of the N-terminal zinc binding domain of HIV-1 integrase

Mengli Cai[1], Ronglan Zheng[2], Michael Caffrey[1], Robert Craigie[2], G. Marius Clore[1] and Angela M. Gronenborn[1]

The solution structure of the N-terminal zinc binding domain (residues 1–55; IN^{1-55}) of HIV-1 integrase has been solved by NMR spectroscopy. IN^{1-55} is dimeric, and each monomer comprises four helices with the zinc tetrahedrally coordinated to His 12, His 16, Cys 40 and Cys 43. IN^{1-55} exists in two interconverting conformational states that differ with regard to the coordination of the two histidine side chains to zinc. The different histidine arrangements are associated with large conformational differences in the polypeptide backbone (residues 9–18) around the coordinating histidines. The dimer interface is predominantly hydrophobic and is formed by the packing of the N-terminal end of helix 1, and helices 3 and 4. The monomer fold is remarkably similar to that of a number of helical DNA binding proteins containing a helix-turn-helix (HTH) motif with helices 2 and 3 of IN^{1-55} corresponding to the HTH motif. In contrast to the DNA binding proteins where the second helix of the HTH motif is employed for DNA recognition, IN^{1-55} uses this helix for dimerization.

The human immunodeficiency virus (HIV) integrase is the enzyme responsible for insertion of a DNA copy of the viral genome into host DNA, an essential step in the replication cycle of HIV (see refs 1–4 for reviews). HIV integrase catalyzes the two key chemical steps of the retroviral DNA integration reaction. The first step, termed 3' processing, involves removal of two nucleotides from each 3' end of the reversed transcribed viral DNA. In the second step, termed DNA strand transfer, the resulting recessed 3'-OH ends of the viral DNA are covalently joined to newly created 5' ends in the target DNA[5].

HIV-1 integrase comprises three functional and structural domains. The central core domain, which contains the catalytic site, is flanked by an N-terminal zinc binding domain and a C-terminal DNA binding domain. All three domains are required for 3' processing and DNA strand transfer[6], although the core domain alone can catalyze a phosphoryl transfer reaction termed disintegration[7]. The exact roles, however, of the C- and N-terminal domains are not well understood. Although it has been shown that the C-terminal domain binds DNA non-specifically[8–10], it is not clear whether the substrate comprises viral sequence or target DNA. Three-dimensional structures of the catalytic domain of both HIV-1 integrase[11] and Rous sarcoma virus (RSV) integrase[12] have been determined by X-ray crystallography. The overall fold is similar to that of RNase H, with highly conserved acidic residues located at equivalent positions in the structures. The topological resemblance, the analogous triad of catalytic carboxylate side chains, and the similar transesterification reaction are shared by a superfamily of polynucleotidyl transferases that includes RNase H, the Mu transposase and the Holliday junction resolving enzyme, Ruv C[13,14]. The structure of the C-terminal domain, which has been solved by NMR[15,16] has an SH3 domain fold and reveals a putative DNA binding surface.

The N-terminal domain of integrase contains a His_2Cys_2 motif that is conserved in all retroviral integrases. It binds one equivalent of zinc[17–19] and is necessary for full integrase activity. The isolated domain (residues 1 to 55; IN^{1-55}) is unstructured in the absence of zinc, but folds into a structure with a high α-helical content, as judged by circular dichroism measurements, in the presence of zinc. In the context of full length integrase, zinc enhances both tetramerization and catalytic activity in *in vitro* assays[19], illustrating its functional importance in promoting an active multimeric state of the enzyme.

The precise role of the N-terminal domain in the overall integration reaction is still little understood. It is clearly involved in protein–protein interactions and has also been speculated to interact with DNA substrate. The N-terminal domain functions in partnership with the core domain of another monomer of integrase to catalyze 3' processing and DNA strand transfer. Integrase lacking the N-terminal domain or containing a mutation in the His_2Cys_2 motif is inactive. Similarly, mutations of the conserved acidic residues in the core domain abolish catalytic activity. However, these inactive proteins complement to restore activity *in vitro* when they are both present in the same reaction mixture[20,21]. Therefore, the catalytic domain of one integrase monomer functions in *trans* with the N-terminal domain of another monomer. Similarly, modification of Cys 56 in the core domain by N-ethylmaleimide inactivates integrase, but again complementation occurs with integrase lacking the N-terminal domain, restoring activity[22]. Based on these mixing experiments

Laboratories of Chemical Physics[1] and Molecular Biology[2], Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA

Correspondence should be addressed to G.M.C. or A.M.G. *email: clore@vger.niddk.nih.gov* or *gronenborn@vger.niddk.nih.gov*

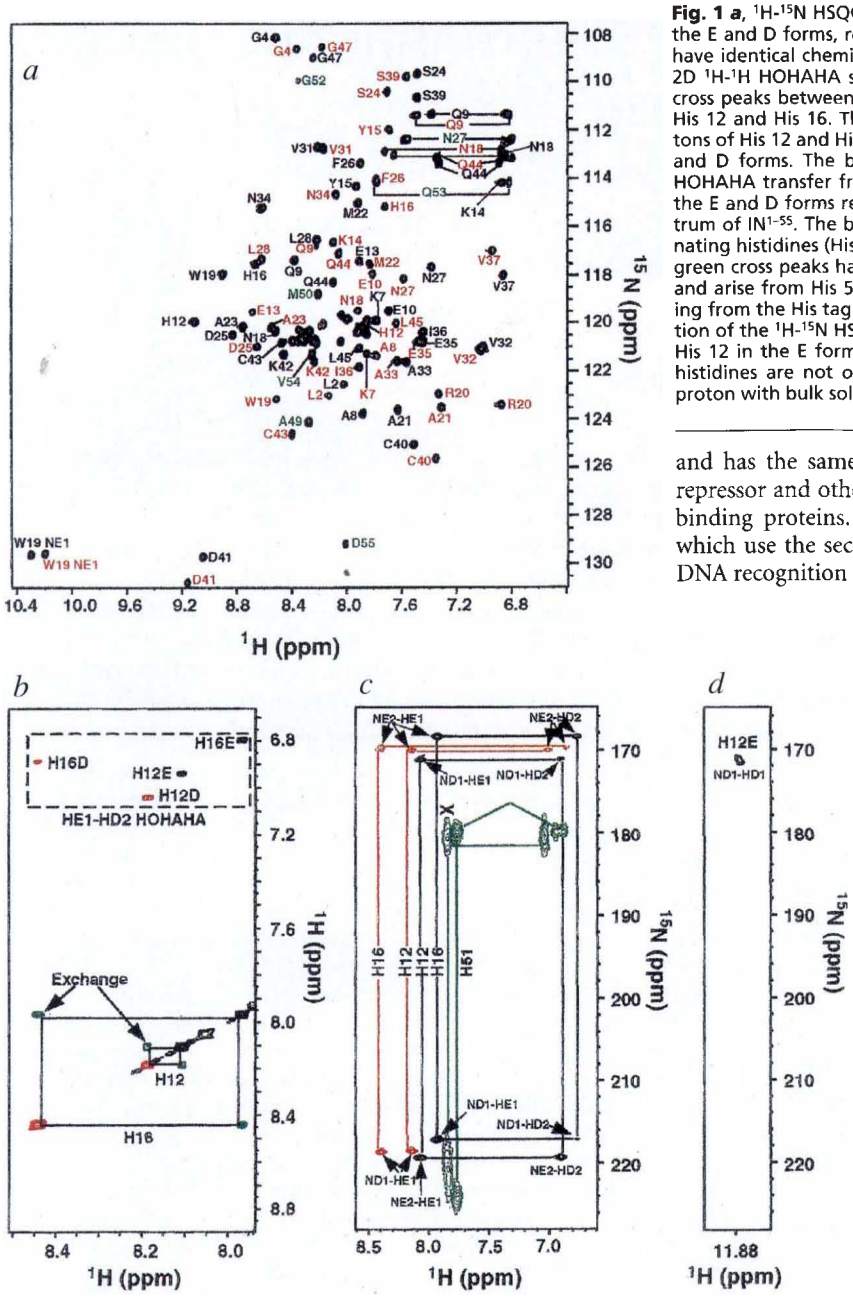


Fig. 1 ***a***, ^{1}H-^{15}N HSQC spectrum of IN$^{1-55}$. Black and red labels refer to the E and D forms, respectively; green labels refer to cross peaks that have identical chemical shifts in the E and D forms. ***b***, Portion of the 2D ^{1}H-^{1}H HOHAHA spectrum of IN$^{1-55}$ illustrating chemical exchange cross peaks between the E and D forms involving the Hε1 protons of His 12 and His 16. The green cross-peaks, which involve the Hε1 protons of His 12 and His 16, arise from chemical exchange between the E and D forms. The black and red diagonal and cross peaks (due to HOHAHA transfer from the Hδ2 to Hε1 protons of histidine) refer to the E and D forms respectively. ***c***, Long range ^{1}H-^{15}N correlation spectrum of IN$^{1-55}$. The black and red cross peaks refer to the zinc coordinating histidines (His 12 and 16) of the E and D forms respectively; the green cross peaks have the same chemical shifts in the E and D forms and arise from His 51 and an additional His at the N terminus resulting from the His tag (see Methods). ***d***, Histidine Nδ1-Hδ1/Nε2-Hε2 portion of the ^{1}H-^{15}N HSQC spectrum showing the Nδ1-Hδ1 cross peak for His 12 in the E form of IN$^{1-55}$; the Nε2-Hε2 cross peaks for the other histidines are not observed due to very rapid exchange of the Hε2 proton with bulk solvent.

and has the same fold as the DNA binding domain of Trp repressor and other structurally related helix-turn-helix DNA binding proteins. In contrast to the DNA binding proteins which use the second helix of the helix-turn-helix motif as a DNA recognition helix, IN$^{1-55}$ uses this helix for dimerization.

and the effects of these mutations on the solubility of the various integrase variants, it has been suggested that active integrase oligomerizes through interactions between the N-terminal and core domains[22]. In addition, competition experiments with monoclonal antibodies[23] suggest that the N-terminal and C-terminal domains are also close; binding of an antibody that is specific to one of the domains can exclude binding of an antibody that is specific to the other domain. The structures of the HIV-1 and RSV integrase core domains also suggest that the N- and C-terminal domains may interact; the ends of the core polypeptide are on the same side of the protein, consistent with such an interaction between the N- and C-terminal domains.

In this paper we present the three-dimensional solution structure of the N-terminal domain of HIV-1 integrase (IN$^{1-55}$). We show that IN$^{1-55}$ is dimeric, exists in two interconverting forms which arise from different modes of zinc coordination to His 12,

Interconverting states of IN$^{1-55}$

The rotational correlation time for IN$^{1-55}$ at 20 °C, derived from ^{15}N T_1 and T_2 relaxation measurements[24], is ~7 ns. Taking into account the temperature dependence of the viscosity of water[26], this corresponds to a protein of ~13,000 M_r in molecular mass, as expected for a dimer of IN$^{1-55}$. A dimeric structure is also consistent with the results from gel filtration and dynamic light scattering experiments (data not shown). The ^{1}H-^{15}N correlation spectrum of IN$^{1-55}$ is shown in Fig. 1*a* and displays clear evidence for heterogeneity with two sets of cross peaks for many resonances. The two species, denoted as the E and D forms, represent interconverting conformational states, as evidenced by the presence of exchange cross peaks in the 2D ^{1}H-^{1}H HOHAHA spectrum (Fig. 1*b*). From the time dependence of the exchange cross peaks and corresponding diagonal peaks observed in a series of 2D ^{1}H-^{1}H NOE spectra recorded with mixing times ranging from 25–300 ms, the overall rate constant for the exchange process at 20 °C is 3.9 ± 0.8 s^{-1} (Fig. 2*a*). At 20 °C the equilibrium constant for the interconversion of the D and E forms of IN$^{1-55}$ (given by K_{DE} = [D]/[E]) is ~0.8, yielding rate constants for the conversion of the D to the E form and of the E to the D form of ~2.2 and 1.7 s^{-1} respectively. From the temperature dependence of the volumes of nine pairs of well-resolved ^{1}H-^{15}N cross peaks for the two forms (Fig. 2*b*), we obtain values of ΔG_{DE}(298 K), ΔH_{DE} and ΔS_{DE} of 0.017 ± 0.050 kcal mol^{-1}, 8 kcal mol^{-1} and 26.7 cal mol^{-1} K^{-1} respectively. Below ~300 K, the predominant species is the E form, while above ~300 K, the D form is the major one.

The structural origin of the two forms of IN$^{1-55}$ can be deduced from the pattern of cross peaks and ^{15}N chemical shifts of the Nε2 and Nδ1 atoms of the histidine residues (His 12 and 16) that coordinate the zinc (Fig. 1*c*). In a long range ^{1}H-^{15}N correlation spectrum[26], cross peaks of approximately

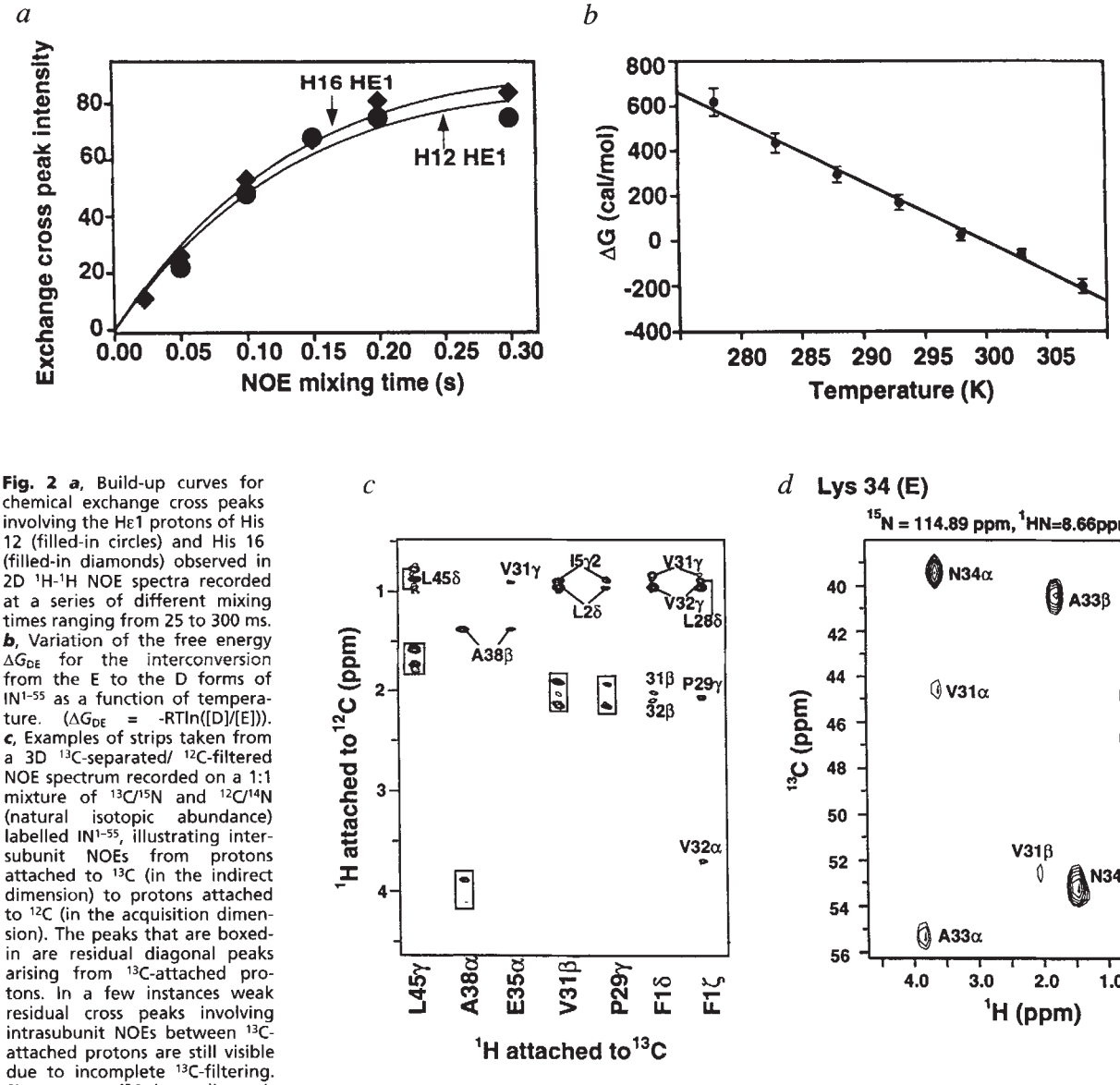


Fig. 2 ***a***, Build-up curves for chemical exchange cross peaks involving the Hε1 protons of His 12 (filled-in circles) and His 16 (filled-in diamonds) observed in 2D ^{1}H-^{1}H NOE spectra recorded at a series of different mixing times ranging from 25 to 300 ms. ***b***, Variation of the free energy ΔG_{DE} for the interconversion from the E to the D forms of IN^{1-55} as a function of temperature. (ΔG_{DE} = -RTln([D]/[E])). ***c***, Examples of strips taken from a 3D ^{13}C-separated/ ^{12}C-filtered NOE spectrum recorded on a 1:1 mixture of $^{13}C/^{15}N$ and $^{12}C/^{14}N$ (natural isotopic abundance) labelled IN^{1-55}, illustrating intersubunit NOEs from protons attached to ^{13}C (in the indirect dimension) to protons attached to ^{12}C (in the acquisition dimension). The peaks that are boxed-in are residual diagonal peaks arising from ^{13}C-attached protons. In a few instances weak residual cross peaks involving intrasubunit NOEs between ^{13}C-attached protons are still visible due to incomplete ^{13}C-filtering. Since no ^{13}C-decoupling is employed in the acquisition dimension, such peaks occur in pairs, split by the one-bond ^{13}C-^{1}H coupling (~125–150 Hz), and are therefore easily recognized. An example of incomplete filtering is provided in the L45γ strip where the two weak bracketed peaks, separated by ~125 Hz and located on either side of the intersubunit L45γ-L45δ NOE cross peak, arise from theNOE between the γ and δ protons of L45 of the ^{13}C-labelled subunits. ***d***, A typical ^{1}H-^{13}C plane taken from a 4D $^{13}C/^{15}N$-separated NOE spectrum (120 ms mixing time) recorded on $^{15}N/^{13}C$-labelled IN^{1-55}. The plane is taken at the ^{15}N and HN shifts of Lys34 (E form).

equal intensity are observed for the Nε2-Hε1, Nε2-Hδ2 and Nδ1-Hε1 two bond correlations ($^{2}J_{NH}$ ~ 6–10 Hz), while either a very weak or absent cross peak is observed for the Nδ1-Hδ2 three-bond correlation ($^{3}J_{NH}$ ~2–3 Hz)[26]. For a neutral histidine, the protonated nitrogen resonates at ~168 p.p.m., while the unprotonated nitrogen resonates at ~250 p.p.m.[26]. From the data shown in Fig. 1*c*, it is evident that His 16 is in the Nε2-H tautomer in both the E and D forms. In contrast, His 12 is in the Nδ1-H tautomer in the E form, but in the Nε2-H tautomer in the D form. Since histidine can only coordinate to zinc through an unprotonated ring nitrogen, we can conclude that the Nδ1 atom of His 16 is coordinated to zinc in both forms of IN^{1-55}, while the Nε2 atom of His 12 is coordinated to zinc in the E form and the Nδ1 atom of His 12 is coordinated to zinc in the D form. Since the Nε2-H tautomer is the more stable of the two tautomers for an unperturbed histidine ring[26], a partially unfolded structure around the coordinating zinc would consequently comprise His 12 with the proton predominantly located at the Nε2 position. Weak coordination to the metal will thus occur through the Nδ1 atom in this locally unfolded state, while the more stable, folded species exhibits coordination through the Nε2 atom with the proton attached to the Nδ1 position of the imidazole ring. In this respect, it is also of interest that the exchangeable Hδ1 proton of His 12 is observed in the E form (Fig. 1*d*), providing strong evidence that it is involved in a hydrogen bond, which would stabilize

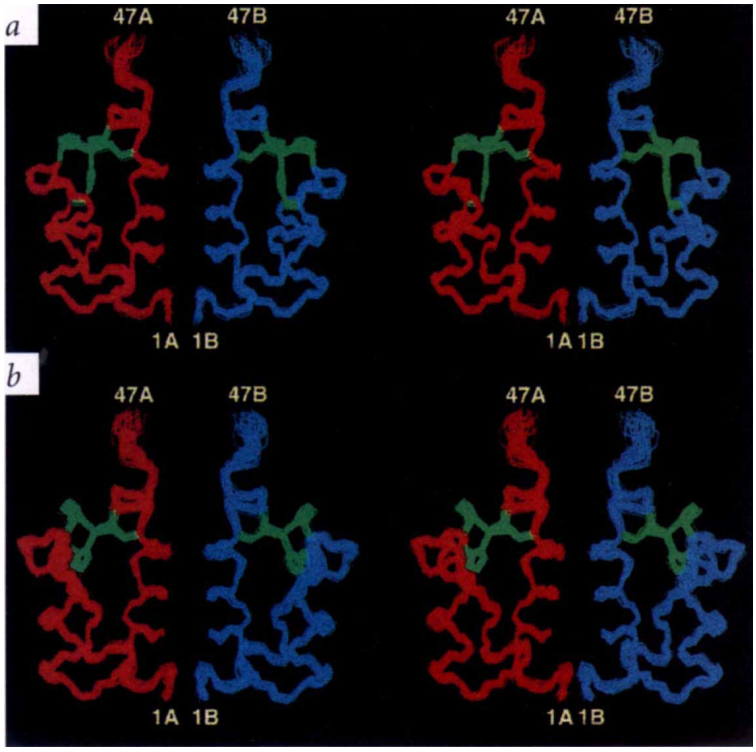


Fig. 3 Streoviews showing superpositions of the backbone atoms, zinc and coordinating histidines and cysteines of the IN^{1-55} dimer of ***a***, the E and ***b***, the D forms with one subunit in red and the other in blue. Residues 1–47 of each subunit are displayed and the zinc and coordinating His and Cys residues are shown in green. The C-terminal region extending from residues 48-55 is not shown since it is not defined by the present data and appears to be disordered.

this particular orientation of the imidazole ring for Nε2 coordination.

Structure determination

The solution structure of the IN^{1-55} dimer was solved by multidimensional heteronuclear NMR spectroscopy[27–30], making use of uniformly labelled ^{15}N and $^{15}N/^{13}C$-labelled protein, as well as heterodimers comprising a 1:1 mixture of uniformly labelled ($^{15}N/^{13}C$) and unlabelled ($^{14}N/^{12}C$ at natural isotopic abundance) subunits to identify intersubunit NOEs. Complete assignments for the two forms of IN^{1-55} were obtained, and structures for the E and D forms were calculated separately by simulated annealing on the basis of 851 and 640 experimental NMR restraints (per monomer) respectively. The quality of the NMR data is illustrated in Fig. 2 which depicts strips from a 3D ^{13}C-separated/^{12}C-filtered NOE experiment (Fig. 2*c*) and a typical plane from a 4D $^{13}C/^{15}N$-separated NOE spectrum (Fig. 2*d*). The 3D ^{13}C-separated/^{12}C-filtered NOE experiment, which is recorded on a heterodimer comprising a 1:1 mixture of labelled ($^{15}N/^{13}C$) and unlabelled ($^{14}N/^{12}C$) IN^{1-55} partitioned in a binomial 1:2:1 distribution of labelled/labelled, labelled/unlabelled, and unlabelled/unlabelled dimers, specifically detects intersubunit NOEs from ^{13}C-attached protons (in the indirect dimension) to ^{12}C-attached protons (in the acquisition dimension). A summary of the structural statistics is given in Table 1, and stereoviews for the two forms are shown in Figs. 3 and 4. No non-sequential NOEs were observed beyond Lys 46 so that the C-terminal region from residues 47–55 could not be defined by the present data and appears to be disordered. The precision of the backbone coordinates for residues 1–46 is ~0.3 Å for the E form and ~0.4 Å for the D form, and the percentage of residues in the most favourable region of the Ramachandran plot is ~94% for the E form and ~89% for the D form.

Description of the monomer

Each monomer is composed of four helices. Helices 2, 3 and 4 are the same in the E and D forms of IN^{1-55} and comprise residues 19–25, 30–39 and 41–45. Helix 1 extends from residue 2–14 in the E form and from 2–8 in the D form. The loop connecting helices 1 and 2 in the E form is replaced by a four residue helical turn (residues 14–17) in the D form. The overall backbone atomic r.m.s. difference (residues 1–46 of both subunits) between the E and D forms is ~2 Å, and the corresponding value for all atoms is ~2.9 Å. These large values are entirely due to the difference in conformation of the polypeptide chain comprising residues 9–18. Excluding the latter 10 residues reduces the atomic r.m.s. difference between the two forms for the dimer to ~0.6 Å for the backbone and ~0.9 Å for all atoms, which is within the errors of the coordinates (Fig. 5*a*,*b*).

The lower region of each subunit is stabilized by a hydrophobic core, while the upper portion is stabilized by the coordination of the zinc to His 12, His 16, Cys 40 and Cys 43 (Fig. 6*a*,*b*). The hydrophobic core is formed by the packing of helices 1, 2 and 3, and comprises Ile 5 and Ala 8 of helix 1, Met 22 of helix 2, Phe 26 and Leu 28 located in the loop between helices 2 and 3, and Val 32 and Ile 36 of helix 3 (Fig. 5*a*).

The zinc atom is tetrahedrally coordinated to His 12, His 16, Cys 40 and Cys 43 in the two forms of IN^{1-55} (Fig. 6*c*,*d*), in agreement with the observation of Co^{2+} d–d transitions in the visible range of the optical spectrum of IN^{1-55} reconstituted with Co^{2+} (ref. 17). The positions of the zinc and the two coordinating cysteine residues (Cys 40 and 43) are identical in the E and D forms of IN^{1-55}, and in both cases the $C\gamma_1$ methyl group of Val 37 is packed against the Cβ methylene of Cys 43 (Figs 5*b*, 6*c*, 6*d*). However, the relative positions of His 12 and 16 in the two forms are reversed (Figs 5*b*, 6*c*, 6*d*). In the E form, His 12 is located in helix 1 (residues 2–14), buried within the protein interior and packed against the Cγm methyl group of Ile 36, its Nε2 atom is coordinated to zinc, and its Hδ1 proton donates a

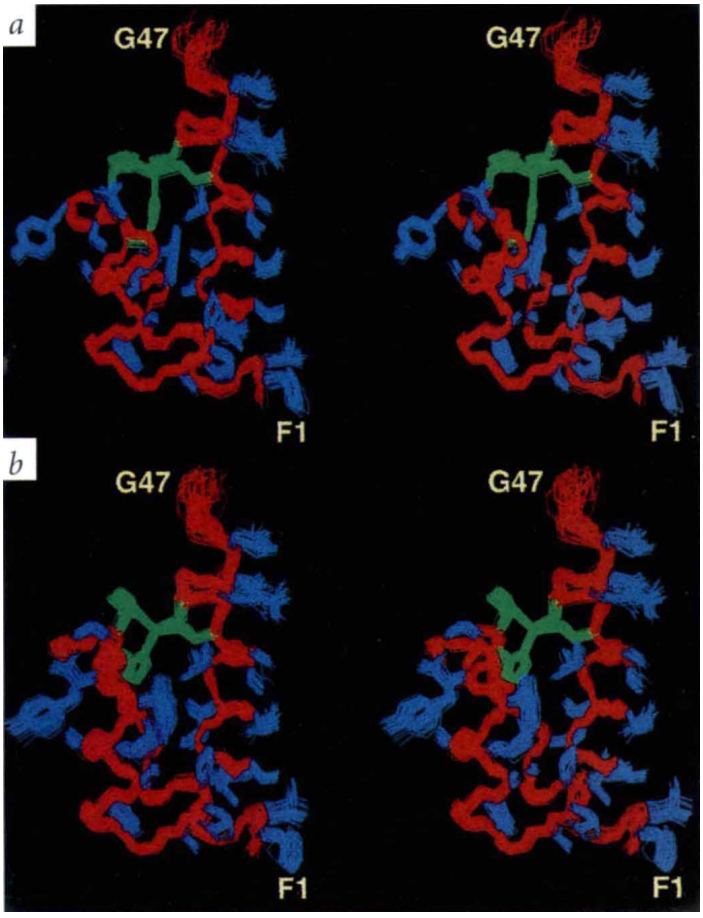


Fig. 4 Stereoviews showing superposition of the IN^{1-55} monomer (residues 1–47) of ***a***, the E and ***b***, the D forms with the backbone in red, ordered side chains in blue, and the zinc and coordinating His and Cys residues in green.

hydrogen bond to the Sδ atom of Met 22 ($r_{Hδ1\text{-}Sδ}$ ~ 2.9 Å, $r_{Nδ1\text{-}Sδ}$ ~ 3.7 Å), in agreement with the observation of the His 12 Nδ1-Hδ1 cross peak in the long range ^{1}H-^{15}N correlation spectrum (Fig. 1*d*). In the D form, His 12 is located in a loop connecting a shortened helix 1 (2–8) to a four residue helical turn (14–17), exposed to solvent and coordinated to the zinc through its Nδ1 atom. In both the E and D forms, His 16 is coordinated to zinc through its Nδ1 atom. In the E form, His 16 is located in the loop connecting helices 1 and 2 and is exposed to solvent. In the D form, His 16 is located in a four residue helical turn (14–17) and is buried within the protein interior, occupying approximately the same position as that occupied by His 12 in the E form. There is no hydrogen bond involving the Sδ of Met 22 in the D form which is ~6 Å away from the Nε2 atom of His 16.

It thus appears that the alternative coordination geometry arrangement observed for the two histidine imidazole rings is correlated with a helix-coil transition for helix 1. In the E form, which predominates at low temperatures, a stable continuous helix is observed for residues 2–14 with metal coordination by His 12 through the less usual Nε2 atom. At higher temperatures where the D form predominates, partial unfolding of helix 1 occurs: a continuous helix is only observed for residues 2–8 and is disrupted by a loop comprising His 12 which coordinates the zinc in its more usual Nε2-H tautomeric form through the Nδ1 atom.

On the external surface of the protein, Trp 19 lies on top of Ala 23 and Ala 33, and is surrounded by Arg 20, Lys 34, Pro 30 and Val 37 (Fig. 5*a*). The position of the tryptophan ring is slightly different in the E and D forms, which accounts in part for chemical shift differences between the two forms. In particular, the observation of distinct chemical shifts for regions relatively remote from the Zn cluster may be due to ring current shifts induced by the tryptophan ring.

The dimer interface

The dimer interface is identical in the E and D forms of IN^{1-55} (Fig. 5*c*). We note that all the intersubunit contacts involve side chains whose ^{13}C and ^{1}H resonances have virtually identical chemical shifts in the two forms. Consequently, only one set of intersubunit NOEs is observed and we cannot resolve homodimers (that is, E/E and D/D) from heterodimers (E/D). However, we note that the time scale of interconversion between the E and D forms is much faster than the rate of subunit exchange so that these two processes occur independently of each other. Since the conformation of the backbone and side chains for the residues that form the interface is identical within coordinate errors for the two forms, these exchange phenomena have no impact on the structure of the dimer interface.

The two subunits are arranged approximately parallel to each other and intermolecular contacts occur between the N-terminal end of helix 1 and helices 3 and 4. The accessible surface area per subunit buried upon dimerization is ~550 $Å^2$, and the calculated free energy of dimerization[31] per subunit is ~ -7.6 kcal mol^{-1}. All the intersubunit interactions between helices 1 and 3 are hydrophobic in nature. In particular Phe 1 of one subunit interacts with Phe 1', Pro 29' and Val 32' of the second subunit, Leu 2 with Pro 29', Pro 30' and Val 31', and Ile 5 with Val 31'. Helix 3 of one subunit is oriented at ~60° to helix 3 of the other subunit, and the packing of the two helices is stabilized mainly by hydrophobic interactions: Val 31 and Val 32 interact with Val 31' and Val 32', and Ala 38 with Glu 35(Cα)' and Ala 38'. In addition, there is a potential salt bridge between Glu 35 and Lys 34'. The orientation of helix 4 of one subunit relative to helix 4 of the second subunit is ~110°, with Leu 45 packed against Leu 45', and a potential hydrogen bond between the side carboxyamide groups of Gln 44 of the two subunits. In this regard, it is worth noting that only a single set of chemical shifts are observed for the side chain of Gln 44 in the E form, and likewise for the D form. This suggests that there is rapid conformational exchange associated with a 180° flip about the Cγ–Cδ bond (χ_3 angle), such that in one conformation the side chain carboxyl of Gln 44 is hydrogen bonded to the side chain amide of Gln 44', while in the other the converse occurs. (Note that the energy barrier for rotation about χ_3 of Gln is small). A similar phenomenon has been observed for an Asn with flipping about the Cβ–Cγ bond in the leucine zipper domain of the c-Jun homodimer[32].

Relationship to other retroviral integrases

A comparison of the sequence of HIV-1 IN^{1-55} with the N-terminal domains of other retroviral integrases[33] indicates that the spacings between the two zinc coordinating histidines,

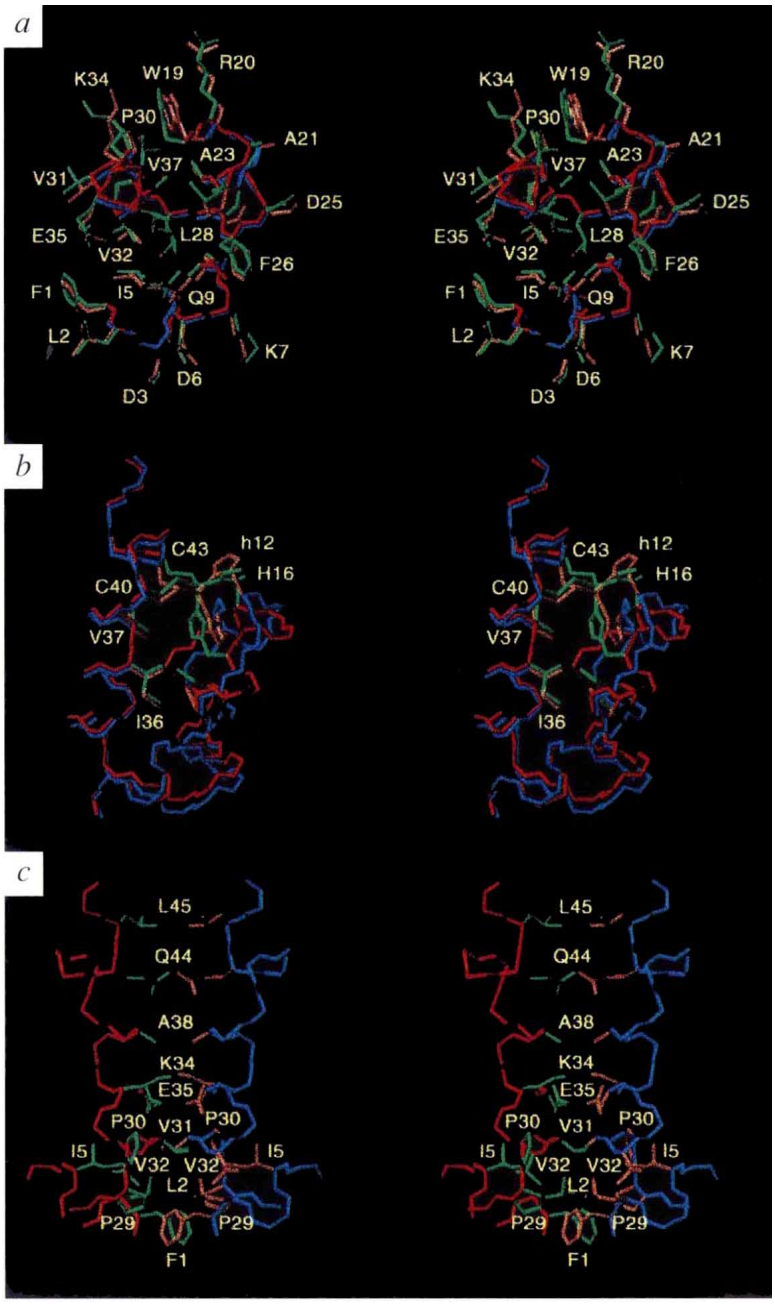


Fig. 5 Stereoviews showing superpositions of the restrained regularized mean structures of the E and D forms of IN^{1-55} for ***a***, the hydrophobic core and ***b***, the region surrounding the zinc. The backbone and sidechains are shown in red and green, respectively, for the E form, and in blue and pink, respectively for the D form. In (*b*) h12 refers to His 12 of the D form, and H16 refers to His 16 of the E form. ***c***, Stereoview of the dimer interface of IN^{1-55}. The E form is displayed and the backbone of one subunit is shown in red and the other in blue; the corresponding sidechains are in green and pink respectively. The structure of the dimer interface is identical in the E and D forms of IN^{1-55}.

Similarity of IN^{1-55} to helix-turn-helix proteins

The monomer fold of IN^{1-55} is remarkably similar to that of a number of helical DNA binding proteins containing a helix-turn-helix (HTH) motif. There is, however, no significant sequence identity between IN^{1-55} and these DNA binding proteins (the percentage sequence identity ranges from 6–12%), and none of these DNA binding proteins bind zinc. The closest structural similarity found was that with helices C, D and E of the Trp repressor[34] (Fig. 7), and the Cα r.m.s. difference for 40 residues is 1.7 Å (with residues 1–15 and 17–41 of IN^{1-55} best-fitted to residues 50–64 and 66–90 of the Trp repressor). Good Cα superpositions were also obtained for the DNA binding domains of the *Drosophila* paired protein[35] (1.9 Å for 39 residues), γδ-resolvase[36] (1.3 Å for 32 residues), the *E. coli* response regulator NarL[37] (2.1 Å for 34 residues), and the engrailed homeodomain[38] (2.3 Å for 32 residues). In each case, helices 2 and 3 of IN^{1-55} correspond to the first and second helices, respectively, of the HTH motif. In contrast to the HTH proteins which use the second helix of the HTH motif to recognize DNA, the equivalent helix in IN^{1-55} is used for dimerization. To our knowledge, IN^{1-55} represents the first example of the use of a HTH motif for protein–protein recognition.

Relationship of IN^{1-55} to the core and C-terminal domains

Structures of all three domains of HIV-1 integrase have now been determined and all are dimeric. In each case the accessible surface area buried at the dimer interface is substantial (~550, 1300 and 360 $Å^2$ per subunit for the N-terminal, catalytic core[11] and C-terminal[15] domains respectively), suggesting that the dimers observed by NMR and crystallography are likely to be functionally relevant. The inter-domain interactions, however, may not necessarily be identical for its different functions at each stage of the viral replication cycle. Initially, integrase constitutes part of the pol polyprotein, which, subsequent to cleavage by HIV protease, is released as an isolated native enzyme. While probably free of DNA early on, at later times after infection it forms a tight multimeric complex with DNA substrate. Based on these observations it seems unlikely that a single unique multimeric form of integrase persists throughout these transitions.

If symmetry is not required, there are clearly many possible ways in which the individual domains of HIV-1 integrase can be arranged in space in the multimeric integration complex. Nevertheless the locations of the N and C termini of the dimeric forms

His-X_3-His, and the two zinc coordinating cysteines, Cys-X_2-Cys, are preserved. The spacing between the second histidine and the first cysteine of the His_2Cys_2 motif is 23 in HIV-1 integrase, but can vary from 22–32 in other retroviral integrases. Any additional residues can be accommodated in the loop connecting helices 2 and 3, or by N-terminal extension of helix 3. The percentage sequence identity with HIV-1 IN^{1-55} varies from 20% for HTLV-I integrase to ~50% for HIV-2 and SIV integrase. In all cases, the hydrophobic residues that form the core of the monomer are either preserved or substituted conservatively. This is also true, with the exception of HTLV-I and HTLV-II integrase, for those residues that constitute the dimer interface. Thus, it is very likely that the monomer fold, and in the majority of cases, the dimer topology of the N-terminal domain are preserved throughout the retroviral integrases.

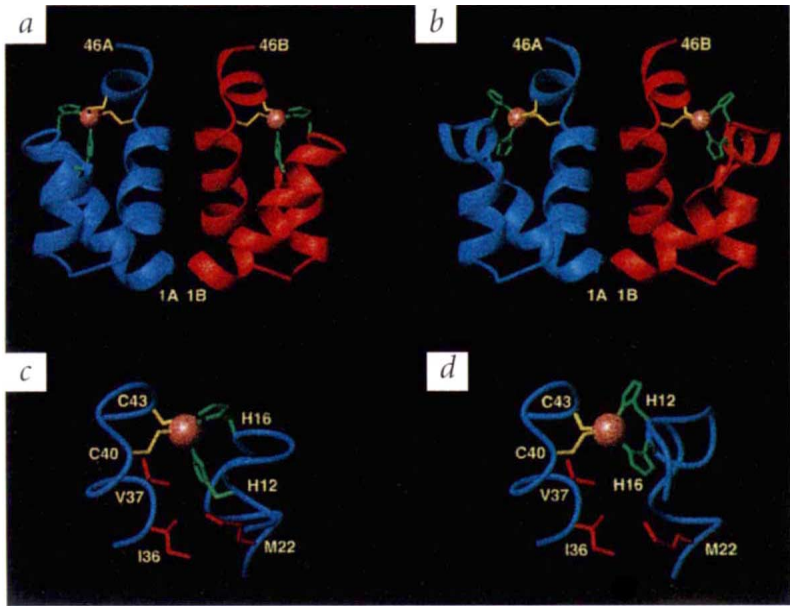


Fig. 6 Ribbon diagrams illustrating ***a***, the E and ***b***, the D forms of the IN^{1-55} dimer (residues 1–46). Detailed views of the region surrounding the zinc in ***c***, the E and ***d***, the D forms of IN^{1-55}. The backbone of one subunit is shown in blue and of the other subunit in red, the zinc atom is displayed as a pink colored ball, selected sidechains are shown in red, and the coordinating cysteines and histidine residues are shown in yellow and green respectively.

of the individual domain structures place certain restrictions on the arrangement of the domains and provide clues as to their positioning. On the other hand, if changes in dimer interfaces or major conformational changes are involved in higher order multimerization and association with DNA substrate, any speculation on our part on the organization of the complex based on the available dimeric structures of the individual domains may prove to be incorrect. With these caveats in mind, what is a likely domain assembly?

The ordered regions in the structures of the N-terminal, catalytic core[11] and C-terminal[15,16] domains extend from Phe 1–Gly 47, Pro 58–Ile 208, and Ile 220–Asp 270 respectively. Thus, the linkers connecting the N- and C-terminal domains to the catalytic core are 10–12 residues long. The C termini of the N-terminal domain dimer, Gly 47 and Gly 47', are separated by ~13 Å and lie on the same face of the molecule. Likewise, the N-termini of the C-terminal domain dimer, Ile 220 and Ile 220', are separated by ~17 Å and located on the same face of the dimer[5]. The C termini of the catalytic core dimer, Ile 208 and Ile 208', are also situated on the same face of the molecule, separated by ~17 Å. However, the N termini of the catalytic core dimer, Pro 58 and Pro 58', are found on opposite faces of the molecule and are separated by ~24 Å. Moreover, the C-terminal helices of the two subunits of the catalytic core dimer are located in between the two N termini. Given this arrangement of ends, it is potentially possible for the C-terminal domain dimer to sit directly on top of the catalytic core dimer[15], but it is not feasible for the N-terminal domain dimer to be connected to the catalytic core dimer, even when the linker length is taken into account. Hence, the N-terminal domain dimer must bridge a pair of core domain dimers, consistent with the observation that the N-terminal domain enhances tetramerization of intact integrase in the presence, but not in the absence, of zinc[19,39]. Moreover, biophysical studies have shown that while both the isolated catalytic core and C-terminal domains are dimeric[15,40,41], a construct comprising both the core and C-terminal domains is in a dimer–tetramer equilibrium[39], strongly suggesting that the C-terminal domain dimer also bridges a pair of catalytic core dimers.

Assuming that the integrase tetramer is symmetric, we were able to construct a model that is consistent with the available structural information (Fig. 8). The four monomer subunits, A, B, C and D are colour coded in yellow, red, green and blue, respectively, and both the N- and C-terminal domains bridge a pair of core domain dimers. The two catalytic core dimers are

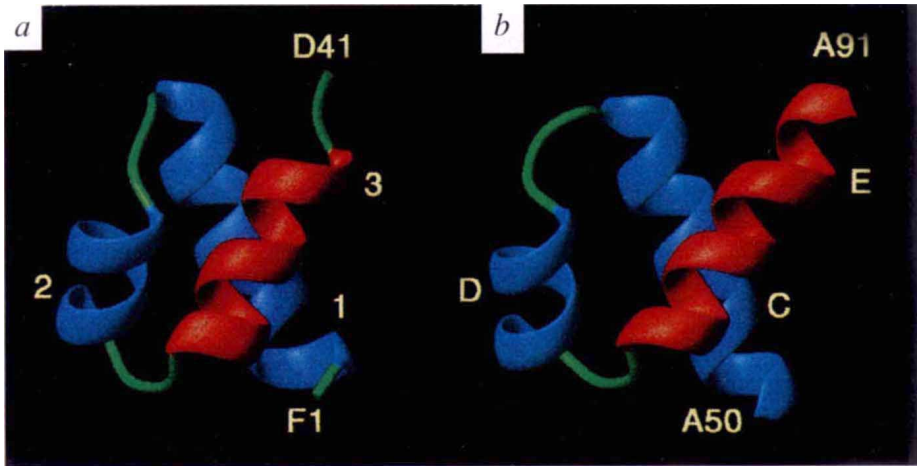
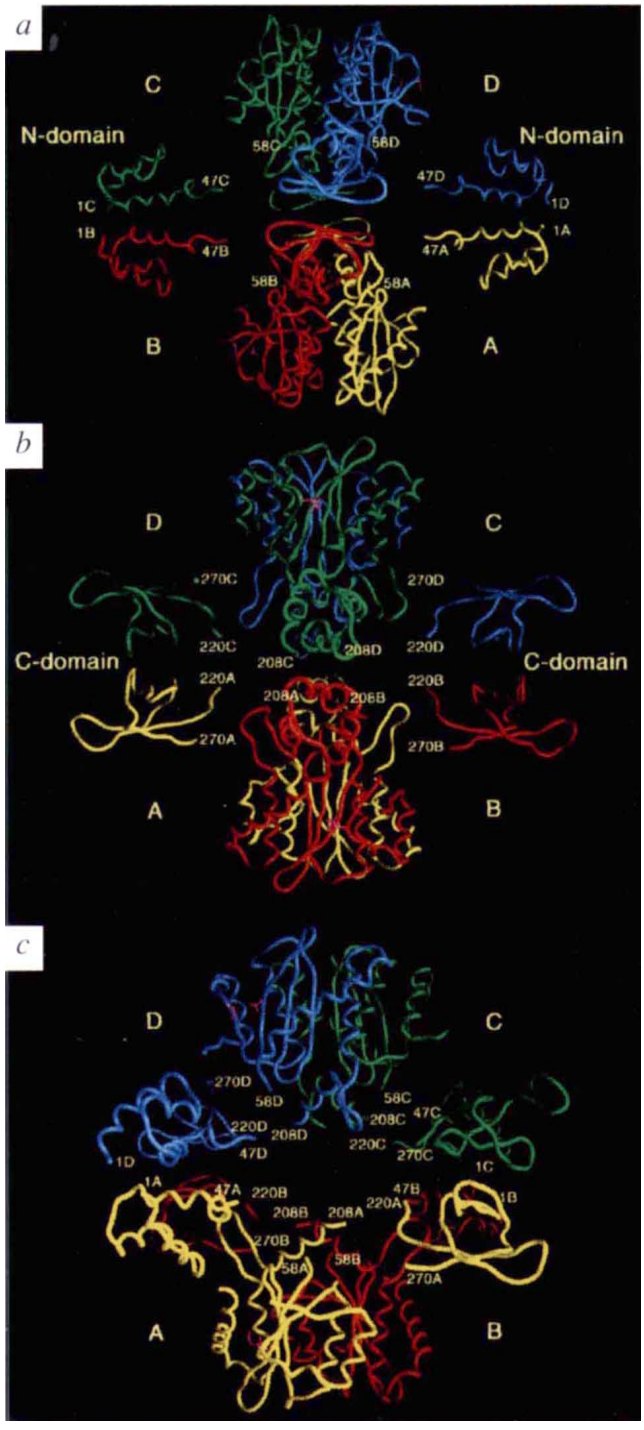


Fig. 7 Comparison of the topology of ***a***, the IN^{1-55} monomer and ***b***, the DNA binding region of Trp repressor. The DNA recognition helix (helix E) of the helix-turn-helix motif of Trp repressor corresponds to the dimerization helix (helix 3) of IN^{1-55}; these two helices are shown in red. The coordinates of the Trp repressor are taken from ref. 38

formed by subunits A (yellow) and B (red), and subunits C (green) and D (blue); the two N-terminal domain dimers bridge subunits A (yellow) and D (blue), and subunits B (red) and C (green), while the two C-terminal domain dimers bridge subunits A (yellow) and C (green), and subunits B (red) and D (blue).

In addition to being compatible with the available biophysical data on the multimeric state of various integrase constructs[19,39,40], the model of the tetramer is also compatible with biochemical data. This includes the observation of trans-complementation between the domains in relation to integration activity[20–22], and the finding that antibodies directed against the N-terminal domain interfere with binding of antibodies directed against the C-terminal domain and *vice versa*[23].

It should be noted, however, that in the model of the tetramer there is no pair of active sites (comprising the Asp 64, Asp 116 and Glu 152 triad) in sufficiently close proximity (~15 Å) to permit five base pair staggered cleavage of the target DNA required for DNA strand transfer[42]. Such close proximity could be achieved by the further assembly of two tetramers into an octamer.



Methods

Expression and purification. IN^{1-55} (ref. 43) was cloned into the *E. coli* vector Pet-15b (Novagen), introducing a His tag at the N terminus. The construct was expressed in the host strain BL21(DE3) grown in minimal medium. Uniform ^{15}N and ^{13}C labelling was obtained using $^{15}NH_4Cl$ and $^{13}C_6$-glucose as the sole nitrogen and carbon sources respectively. Cells were grown at 37 °C, induced with isopropyl β-D-thiogalactopyranoside, and harvested by centrifugation, and resuspended in 'lysis' buffer (20 mM HEPES pH 7.5, 5 mM imidazole, 0.5 M NaCl, 2 mM β-mercaptoethanol and 0.4 mg m^{-1} lysozyme) for 20 min at 0 °C. The lysed cells were homogenized and sonicated, and the cell walls removed by centrifugation at 40,000 g for 1 h. The supernatant was filtered through a 0.2 μm filter, and loaded onto a freshly charged Ni chelating sepharose column containing bound nickel. The column was first washed overnight with 2 l of 'binding buffer' (20 mM HEPES pH 7.5, 0.5 M NaCl, 2 mM β-mercaptoethanol) containing 40 mM imidazole. The protein was eluted in binding buffer using an imidazole gradient from 40–500 mM at a flow rate of 2 ml min^{-1}. Frations containing IN^{1-55} were pooled, and dialyzed overnight at 4 °C against 20 mM HEPES pH 7.5, 10 mM NaCl, 5 mM EDTA and 1 mM dithiothreitol (DTT). The His tag was removed by incubating the purified protein with thrombin (50 units mg^{-1} of protein) at room temperature for 1 h. Thrombin was removed by passing the cleaved protein solution over a benzamidine bead column. The eluted protein, containing the first 55 residues of HIV-1 integrase and an additional 3 residues at the N terminus (GlySerHis) arising from the remains of the His tag, was further purified by HPLC.

Sample preparation. HPLC purified IN^{1-55} (1 mg ml^{-1}) was folded by dialysis against buffer containing 25 mM Tri-HCl pH 7.4, 0.1 mM ZnCl, 0.25 M NaCl, 1 mM DTT and 50 mM glycerol. Oxygen was eliminated by purging the buffer with helium for 1 h and keeping all solutions under helium throughout the dialysis procedure. After two changes of buffer, the Zn reconstituted IN^{1-55} was concentrated to 2–3 mM (in monomer). The mixed dimer containing a 1:1 mixture of $^{14}N/^{12}C$ (natural isotopic abundance) and $^{15}N/^{13}C$ subunits was prepared by mixing equal amounts of $^{14}N/^{12}C$ and $^{15}N/^{13}C$-labelled protein after HPLC purification which was then refolded by dialysis as above. The following samples were prepared: ^{15}N-labelled IN^{1-55} in 95% H_2O/5% D_2O; $^{15}N/^{13}C$-labelled IN^{1-55} in 95% H_2O/5% D_2O and 99.996% D_2O; heterodimer containing a 1:1 mixture of $^{15}N/^{13}C$ and $^{14}N/^{12}C$-labelled IN^{1-55} in 99.996% D_2O.

NMR spectroscopy. NMR experiments were carried out at 20 °C on Bruker DMX500, DMX600 and DMX750 spectrometers equipped with x,y,z-shielded gradient triple resonance probes. Spectra were processed with the NMRPipe package[44], and analyzed using the programs PIPP, CAPP and STAPP[45]. ^1H, ^{13}C and ^{15}N sequential resonance assignment was achieved by means of through-bond het-

Fig. 8 Three views of a model of the integrase tetramer . The four subunits, A, B, C, and D are shown in yellow, red, green and blue respectively. The active site residues (Asp 64 and Asp 116) of each catalytic core are shown in magenta. (Note that the third catalytic residue, Glu 152, is not visible in the electron density map). Views (*b*) and (*c*) are rotated by 90° and 45°, respectively, relative to view (*a*); the rotation is about an axis that is parallel to the dimer interfaces of the two catalytic core dimers. The residues shown for the N-terminal, catalytic and C-terminal domains are 1–47, 58–208 and 220–270 respectively. The coordinates for the X-ray structure of the catalytic core dimer and the NMR structure of the C-terminal domain dimer are taken from refs. 11 and 15 respectively.

Table 1 Structural statistics[1]

	<SA> E form	<SA> D form
Structural Statistics		
R.m.s. deviations from experimental distance restraints (Å)[2]		
All (623/415)	0.033±0.002	0.029±0.003
Intrasubunit		
interresidue sequential ($\|i-j\|=1$) (201/156)	0.034±0.004	0.026±0.005
interresidue short range ($1<\|i-j\|\leq 5$) (224/137)	0.034±0.004	0.027±0.003
interresidue long range ($\|i-j\|>5$) (98/47)	0.019±0.004	0.033±0.005
intraresidue (17/15)[3]	0.021±0.014	0.020±0.013
H-bonds (32/14)	0.021±0.005	0.0004±0.002
Intersubunit (28/28)[4]	0.054±0.011	0.043±0.011
Ambiguous intra- and intersubunit (23/23)[5]	0.042±0.023	0.027±0.015
R.m.s. deviations from experimental		
dihedral restraints (°) (91/92)[2]	0.039±0.048	0.105±0.114
R.m.s. deviations from experimental		
$^3J_{HN\alpha}$ coupling constants (Hz) (40/36)[2]	0.57±0.03	0.64±0.04
R.m.s. deviations from experimental ^{13}C shifts		
$^{13}C\alpha$ (p.p.m.) (50/50)	0.98±0.05	1.23±0.05
$^{13}C\beta$ (p.p.m.) (47/47)	0.88±0.05	0.92±0.07
Deviations from idealized covalent geometry		
bonds (Å) (855/855)	0.003±0.0002	0.003±0.0004
angles (°) (1551/1551)	0.461±0.017	0.369±0.020
impropers (°) (450/450)	0.464±0.029	0.359±0.038
Measures of Structure Quality		
$E_{L\text{-}J}$ (kcal mol^{-1})[6]	–465±10	–457±12
PROCHECK[7]		
% residues in most favorable region		
of Ramachandran plot	93.8±1.5	88.5±1.9
Number of bad contacts/100 residues	2.7±1.6	3.2±2.1
H-bond energy	0.98±0.04	0.79±0.08
Coordinate Precision of the dimer[8]		
backbone (Å)	0.32±0.08	0.41±0.10
all atoms (Å)	0.70±0.07	0.80±0.09

[1]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 (residues 1–46 of both subunits) best fitted to each other; $(\overline{SA})r$ is the restrained minimized mean structure obtained by restrained regularization of the mean structure SA. The number of terms for the various restraints *per monomer* is given in parentheses; the first number refers to the E form, the second to the D form. The final force constants employed for the various terms in the target function used for simulated annealing are as follows: 1,000 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 non-crystallographic 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[64]), 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 1.0 for the conformational database potential. The later is based on the populations of various combinations of torsion angles observed in a database of 70 high-resolution (1.75 Å or better) X-ray structures and biases sampling to conformations that are energetically possible by effectively limiting the choice of dihedral angles to those that are known to be physically realizable[56,57].

[2]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. The torsion angles restraints comprise 45 ϕ, 1 ψ, 31 χ_1 and 14 χ_2 angles per monomer for the E form, and 45 ϕ, 5 ψ, 30 χ_1 and 12 χ_2 for the D form.

[3]Only structurally useful intraresidue NOEs are included in the intraresidue interproton distance restraints. Thus, intraresidue NOEs between protons separated by two bonds or between non-stereospecifically assigned protons separated by three-bonds are not incorporated in the restraints.

[4]Intersubunit NOEs from protons attached to ^{13}C (in the indirect dimension) to protons attached to ^{12}C (in the acquisition dimension) were obtained from a 3D ^{13}C-separated/^{12}C-filtered NOE spectrum recorded on a sample containing a 1:1 mixture of $^{15}N/^{13}C$ and $^{14}N/^{12}C$ (natural isotopic abundance) labelled IN^{1-55}.

[5] NOEs where a distinction between intra- and intersubunit effects could not be distinguished were treated as $(\Sigma r^{-6})^{-1/6}$ sums[50].

[6]$E_{L\text{-}J}$ is the Lennard-Jones van der Waals energy calculated with the CHARMM PARAM19/20 protein parameters[64] and is *not* included in the target function for simulated annealing or restrained minimization. In calculating $E_{L\text{-}J}$, the van der Waals radius for the zinc atom was set equal to its ionic radius of 0.75 Å. However, since the zinc is totally buried by the coordinating ligands, the exact value of the van der Waals radius used for the zinc has no impact on the value of $E_{L\text{-}J}$.

[7]The program PROCHECK[65] was used to assess the overall quality of the structures. Less than 10 bad contacts per 100 residues and a H-bond energy of 0.6–1.0 are expected for a good quality structure. The dihedral angle G-factors (which should be greater than -0.5 for a good quality structure) for the ϕ/ψ, χ_1/χ_2, χ_1 and χ_3/χ_4 distributions are 0.57 ± 0.05, 0.53 ± 0.08, 0.46 ± 0.10 and 0.12 ± 0.18 respectively, for the E form, and 0.54 ± 0.05, 0.47 ± 0.10, 0.42 ± 0.15 and 0.10 ± 0.13 respectively, for the D form. The PROCHECK statistics apply to the ordered region of IN^{1-55} comprising residues 1-46 of the two subunits.

[8]The precision of the atomic coordinates is defined as the average rms difference between the 40 final simulated annealing structures and the mean coordinates, $\overline{SA}$. The values given relate to residues 1–46 of the two subunits together. The backbone atoms comprise the N, Cα, C and O atoms. As a non-crystallographic symmetry restraint on all atoms was employed during refinement, the all atom atomic RMS difference between the two subunits is very small (<0.08 Å for the individual simulated annealing structures and 0.002 Å for the restrained regularized mean structures).

eronuclear correlations along the backbone and side chains[27–29]. $^3J_{HN\alpha}$, $^3J_{NH\beta}$, $^3J_{C'C\gamma}$(aromatic, methyl and methylene), $^3J_{NC\gamma}$(aromatic, methyl and methylene) couplings were obtained by quantitative *J* correlation spectroscopy[30,46–48]. Interproton distance restraints were derived from the following spectra: 3D ^{15}N-separated (120 ms mixing time), ^{13}C-separated (50 and 120 ms mixing times), and ^{13}C-separated/^{12}C-filtered (150 ms mixing time) NOE spectra, 3D ^{15}N-separated ROE (40 ms mixing time) spectrum, and 4D ^{13}C/^{15}N-separated (120 and 150 ms mixing times) and ^{13}C/^{13}C-separated (150 ms mixing time) NOE spectra. ^{15}N T_1 and T_2 relaxation times were measured as described[49]. Long range nitrogen-proton correlations involving the histidine rings were observed in a ^{1}H-^{15}N HSQC spectrum recorded with a 22 ms dephasing delay during which time the ^{1}H and ^{15}N signals become antiphase[26].

Structure calculations. Approximate interproton distance restraints were derived from the multidimensional NOE spectra, essentially as described[27]. 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. 0.5 Å was added to the upper bounds for distances involving methyl groups to account for the higher apparent intensity of the methyl resonances. Distances involving methyl groups, aromatic ring protons, non-stereospecifically assigned methylene protons, and groups where a distinction between intermolecular and intramolecular effects could not be distinguished were represented as a $(\Sigma r^{-6})^{-1/6}$ sum[50]. Protein backbone hydrogen bonding restraints (two per hydrogen bond, $r_{NH\text{-}O}$ = 1.5–2.8 Å, $r_{N\text{-}O}$ = 2.4–3.5 Å) within areas of regular secondary structure were introduced during the final stages of refinement using standard NMR criteria based on backbone NOEs and $^3J_{HN\alpha}$ coupling constants, supplemented by secondary ^{13}C shifts[21]. ϕ, ψ, χ_1 and χ_2 torsion angle restraints were derived from the NOE/ROE and homo- and heteronuclear three-bond coupling constant data, and the minimum ranges employed were ±15°, ±40°, ±20° and ±30° respectively[51]. The structures were calculated by simulated annealing[52] using the program XPLOR-3.1[53], adapted to incorporate pseudo-potentials for $^3J_{HN\alpha}$ coupling constant[54] and secondary ^{13}Cα/^{13}Cβ chemical shift[55] restraints, and a conformational database potential[56,57]. Two different protocols were employed with identical end results: a modified version of the hybrid distance geometry-simulated annealing protocol[52]; and a torsion angle dynamics protocol starting from random initial coordinates[58], followed by conventional simulated annealing in cartesian coordinate space[59]. The target function that is minimized during simulated annealing and restrained regularization comprises quadratic harmonic potential terms for covalent geometry, non-crystallographic symmetry, and $^3J_{HN\alpha}$ coupling constant and secondary ^{13}Cα and ^{13}Cβ chemical shift restraints, square-well quadratic potentials for the experimental distance and torsion angle restraints, a quartic van der Waals repulsion term for the non-bonded contacts, and a conformational database potential. There were *no* hydrogen-bonding, electrostatic or 6-12 Lennard-Jones empirical potential energy terms in the target function. In the final stages of refinement terms for maintaining the tetrahedral coordination geometry of the zinc were added to the covalent geometry restraints[60].

Figures were generated using the programs MOLMOL[61] and GRASP[62]. Optimized best-fit Cα superpositions of IN$^{1-55}$ with helix-turn-helix proteins was carried out with the program O[63].

The coordinates of the final 40 simulated annealing structures for the D and E forms (accession codes 1WJB and 1WJD), together with the coordinates of the corresponding restrained regularized mean structures (accession codes 1WJA and 1WJC), and the complete list of experimental NMR restraints and ^{1}H, ^{15}N, ^{13}C assignments (accession codes R1WJBMR and R1WJDMR) have been deposited in the Brookhaven Protein Data Bank.

Acknowledgements

We thank Dan Garrett and Frank Delaglio for software support; Dan Garrett for help with the figures; Rolf Tschudin for technical support; Ying Huang for help with protein purification; Fred Dyda and David Davies for the coordinates of the catalytic core dimer of HIV-1 integrase; and Ad Bax, David Covell, Stephan Grzesiek, Jin-Shan Hu, and John Kuszewski for useful discussions. 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.).

Received 9 April 1997; accepted 2 June 1997.

1. Whitcomb, J.M. & Hughes, S.H. Retroviral reverse transcription and integration: progress and problems. *Annu. Rev. Cell Biol.* **8**, 275–306 (1992).
2. Goff, S. P. Genetics of retroviral integration. *Annu. Rev. Genet.* **26**, 527–544 (1992).
3. Vink, C. & Plasterk, R.H.A. The human immunodeficiency virus integrase protein. *Trends. Genet.* **9**, 433–438 (1993).
4. Katz, R.A. & Skalka, A.M. The retroviral enzymes. *Annu. Rev. Biochem.* **63**, 133–173 (1994).
5. Engelman, A., Mizuuchi, K. & Craigie, R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67**, 1211–1221 (1991).
6. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P.T. & Craigie, R. Domains of integrase protein of human immunodeficiency virus type-1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl. Acad. Sci. USA* **90**, 3428–3432 (1993).
7. Chow, S.A., Vincent, K.A., Ellison, V. & Brown, P.O. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* **255**, 723–726 (1992).
8. . Vink, C., Oude Groeneger, A.M. & Plasterk, R.H.A. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res.* **21**, 1419–1425 (1993).
9. Engelman, A., Hickman, A.B. & 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.* **68**, 5911–5917 (1994).
10. Woerner, A.M. & Marcus-Sekura, C.J. Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. *Nucleic Acids Res.* **21**, 3507–3511 (1993).
11. Dyda, F., Hickman, A.B., Jenkins T.M., Engelman, A., Craigie, R. & Davies, D.R. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986 (1994).
12. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. & Skalka, A.M. High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* **253**, 333–346 (1995).
13. Yang, W. & Steitz, T.A. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* **3**, 131–134 (1995).
14. Rice, P., Craigie, R. & Davies, D.R. Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* **6**, 76–83 (1996).
15. Lodi, P.J., Ernst J.A., Kuszewski, J., Hickman, A.B., Engelman, A., Craigie, R., Clore, G.M. & Gronenborn, A.M. Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* **34**, 9826–9833 (1995).
16. Eijkelenboom, A.P., Lutzke, R.A., Boelens, R., Plasterk, R.H.A., Kaptein, R. & Hard, K. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nature Struct. Biol.* **2**, 807–810 (1995).
17. Burke, C.J., Sanyal, G., Bruner, M.W., Ryan, J.A., LaFemina, R.L., Robbins, H.L., Zeft, A.S., Middaugh, C.R. & Cordingley, M.G. Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *J. Biol. Chem.* **267**, 9639–9644 (1992).
18. Haugan, I.R., Nilsen, B.M., Worland, S., Olsen, I. & Helland, D.E. Characterization of the DNA-binding activity of HIV-1 integrase using a filter binding assay. *Biochem. Biophys. Res. Commun.* **217**, 802–810 (1995).
19. Zheng, R., Jenkins, T.M. & Craigie, R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13659–13664 (1996)
20. Engelman, A., Bushman, F.D. & Craigie, R. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.* **12**, 3269–3275 (1993).
21. van Gent, D.C., Vink, C., Groeneger, A.A. & Plasterk, R.H.A. Complementation between HIV integrase proteins mutated in different domains. *EMBO J.* **12**, 3261–3267 (1993).
22. Ellison, V., Gerton, J., Vincent, K.A. & Brown, P.O. An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. *J. Biol. Chem.* **270**, 3320–3326 (1995).
23. Bizub-Bender, D., Kulkosky, J. & Skalka, A.M. Monoclonal antibodies against HIV type 1 integrase: clues to molecular structure. *AIDS Research and Human Retroviruses* **10**, 1105–1115 (1994).
24. Clore, G.M., Driscoll, P.C., Wingfield, P.T. & Gronenborn, A.M. Analysis of backbone dynamics of interleukin-1β using two-dimensional inverse detected heteronuclear ^{15}N-^{1}H NMR spectroscopy. *Biochemistry* **29**, 7387–7401 (1990).
25. Lide, D.R. *Handbook of Chemistry and Physics*, p. 6-10, CRC Press, Boca Raton (1993).
26. Pelton, J.G., Torchia, D.A., Meadow, M.D. & Roseman, S. Tautomeric states of the active-site histidines of phophorylated and unphophorylated III^{Glc}, a signal transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci.* **2**, 543–558 (1993).
27. Clore, G.M. & Gronenborn, A.M. Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390–1399 (1991).
28. Bax, A. & Grzesiek, S. Methodological advances in protein NMR *Acct. Chem. Res.* **26**, 131–138 (1993).
29. Gronenborn, A.M. & Clore, G.M. Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy. *CRC Crit Rev. Biochem. Mol. Biol.* **30**, 351–385 (1995).
30. Bax, A. et al. Measurement of homo- and hetero-nuclear J couplings from quantitative J correlation. *Meth. Enzymol.* **239**, 79–106 (1994).
31. Eisenberg, D. & McLaghlan, .D. Solvation energy in protein folding and binding. *Nature* **319**, 199–203 (1986).
32. Junius, F.K., MacKay, J.P., Bubb, W.A., Jensen, S.A., Weiss, M.A.S. & King, G.F. Nuclear magnetic resonance characterization of the Jun Leucine zipper domain: unusual properties of coiled-coil interfacial polar residues. *Biochemistry* **34**, 6164–6174.
33. Khan, E., Mack, J.P.G., Katz, R.A., Kulkosky, J. & Skalka, A.M. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucl. Acids Res.* **19**, 851–860 (1991).
34. Otwinowski Z. *et al.* Crystal structure of the trp repressor/operator complex at atomic resoution. *Nature* **335**, 3321–329 (1988).
35. Xu, W., Rould, M.A., Jun, S., Desplan, C. & Pabo, C.O. Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell* **80**, 639–650 (1995).
36. Yang, W. & Steitz, T.A. Crystal structure of the site-specific recombinase γδ resolvase complexed with a 34 bp cleavage site. *Cell* **82**, 193–207 (1995).
37. Baikalov, I. *et al.* Structure of the *Escherichia coli* response regulator NarL. *Biochemistry* **35**, 11053–11061 (1996).
38. Kissinger, C., Liu, B., Martin-Blanco, E., Kornberg, T. & Pabo, C.O. Crystal structure of the engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579–590 (1990).
39. Jenkins, T.M., 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.* **271**, 7712–7718 (1996).
40. Jenkins, T.M., Hickman, A.B., Dyda, F., Ghirlando, R., Davies, D.R. & Craigie, R. Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc. Natl. Acad. Sci. USA* **92**, 6057–6061.
41. Hickman, A.B., Dyda, F. & Craigie, R. Heterogeneity in recombinant HIV-1 integrase corrected by site-directed mutagenesis: the identification and elimination of a protease cleavage site. *Prot. Engng.* **10**, *in the press.*
42. Vink, C. *et al.* Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA. *J. Virol.* **64**, 5626–5627 (1990).
43. Adachi, A. *et al.* Production of acquired immunodeficiency syndrome-associated retrovirus in human and non-human cells transfected with an infectious molecular clone. *J. Virol.* **59**, 284–291 (1986).
44. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomolec. NMR* **6**, 277–293 (1995).
45. Garrett, D.S., Powers, R., Gronenborn, A.M. & Clore, G.M. A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214–220 (1991).
46. Hu, J.-S., Grzesiek, S. & Bax, A. Two-dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from thee-bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ couplings. *J. Am. Chem. Soc.* **119**, 1803–1804 (1997).
47. Hu, J.-S. & Bax, A. Determination of ϕ and χ_1 angles in proteins from ^{13}C-^{13}C three-bond *J* couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond. *J. Am. Chem. Soc. in the press.*
48. Hu, J.-S. & Bax, A. χ_1 angle information from a simple two-dimensional NMR experiment which identifies trans $^{3}J_{NC\gamma}$ couplings in isotopically enriched proteins. *J. Biomol. NMR in press* (1997).
49. Tjandra, N., Kuboniwa, H., Ren, H. & Bax, A. Rotational dynamics of calcium-free calmodulin studied by ^{15}N-NMR relaxation measurements. *Eur. J. Biochem.* **230**, 1014–1024 (1995).
50. Nilges, M. A calculational strategy for the structure determination of symmetric dimers by ^{1}H NMR. *Proteins Struct. Funct. Genet.* **17**, 297–309 (1993).
51. Nilges, M., Clore, G.M. & Gronenborn, A.M. ^{1}H-NMR stereospecific assignments by conformational database searches. *Biopolymers* **29**, 813–822 (1990).
52. Nilges, M., Clore, G.M. & Gronenborn, A.M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* **229**, 317–324 (1988).
53. Brünger, A.T. X-PLOR Version 3.1: A system for X-ray crystallography and NMR. (Yale University Press, New Haven, Connecticut; 1993).
54. Garrett, D.S. *et al.* The impact of direct refinement against three-bond HN-CαH coupling constants on protein structure determination by NMR. *J. Magn. Reson. Series B* **104**, 99–103 (1994).
55. Kuszewski, J., Qin, J., Gronenborn, A.M. & Clore, G.M. The impact of direct refinement against $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Series B* **106**, 92–96 (1995).
56. Kuszewski, J., Gronenborn, A.M. & Clore, G.M. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Prot. Sci.* **5**, 1067–1080 (1996).
57. Kuszewski, J., Gronenborn, A.M. & Clore, G.M. Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J. Magn. Reson.* **125**, 171–177 (1997).
58. Stein, E.G., Rice, L.M. & Brünger, A.T. Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J. Magn. Reson.* **124**, 154–164 (1997).
59. Nilges, M., Gronenborn, A.M., Brünger, A.T. & Clore, G.M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Prot. Engng.* **2**, 27–38 (1988).
60. Omichinski, J., Clore, G.M., Appella, E., Sakaguchi, K. & Gronenborn, A.M. High resolution three-dimensional solution structure of a single zinc finger from a human enhancer binding protein in solution. *Biochemistry* **29**, 9324–9334. (1990).
61. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 52–55 (1996).
62. Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
63. Jones, T.A. & Kjeldgaard. *O - The Manual*. Version 5.8.1, University of Uppsala, Sweden (1992).
64. Brooks, B.R. *et al.* CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1993).
65. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).