The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc

Astrid P.A.M. Eijkelenboom*†, Fusinita M.I. van den Ent‡†, Arnold Vos‡, Jurgen F. Doreleijers*, Karl Hård*§, Thomas D. Tullius*¶, Ronald H.A. Plasterk‡, Robert Kaptein* and Rolf Boelens*

Background: Integrase mediates a crucial step in the life cycle of the human immunodeficiency virus (HIV). The enzyme cleaves the viral DNA ends in a sequence-dependent manner and couples the newly generated hydroxyl groups to phosphates in the target DNA. Three domains have been identified in HIV integrase: an amino-terminal domain, a central catalytic core and a carboxy-terminal DNA-binding domain. The amino-terminal region is the only domain with unknown structure thus far. This domain, which is known to bind zinc, contains a HHCC motif that is conserved in retroviral integrases. Although the exact function of this domain is unknown, it is required for cleavage and integration.

Results: The three-dimensional structure of the amino-terminal domain of HIV-2 integrase has been determined using two-dimensional and three-dimensional nuclear magnetic resonance data. We obtained 20 final structures, calculated using 693 nuclear Overhauser effects, which display a backbone root-mean square deviation versus the average of 0.25 Å for the well defined region. The structure consists of three α helices and a helical turn. The zinc is coordinated with His12 via the $N^{\varepsilon 2}$ atom, with His16 via the $N^{\delta 1}$ atom and with the sulfur atoms of Cys40 and Cys43. The α helices form a three-helix bundle that is stabilized by this zinc-binding unit. The helical arrangement is similar to that found in the DNA-binding domains of the trp repressor, the prd paired domain and Tc3A transposase.

Conclusion: The amino-terminal domain of HIV-2 integrase has a remarkable hybrid structure combining features of a three-helix bundle fold with a zinc-binding HHCC motif. This structure shows no similarity with any of the known zinc-finger structures. The strictly conserved residues of the HHCC motif of retroviral integrases are involved in metal coordination, whereas many other well conserved hydrophobic residues are part of the protein core.

Addresses: *Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. ‡Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

Present addresses: §Astra Structural Chemistry Laboratory, S-43183 Mölndal, Sweden. ¶Department of Chemistry, Boston University, Boston, Massachusetts 02215, USA. (On leave from: Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland 21218, USA.)

†These authors contributed equally to this work.

Correspondence: Robert Kaptein
E-mail: kaptein@nmr.chem.ruu.nl

Received: **8 July 1997**
Revised: **1 August 1997**
Accepted: **1 August 1997**

Published: **5 September 1997**

Current Biology 1997, 7:739–746
http://biomednet.com/elecref/0960982200700739

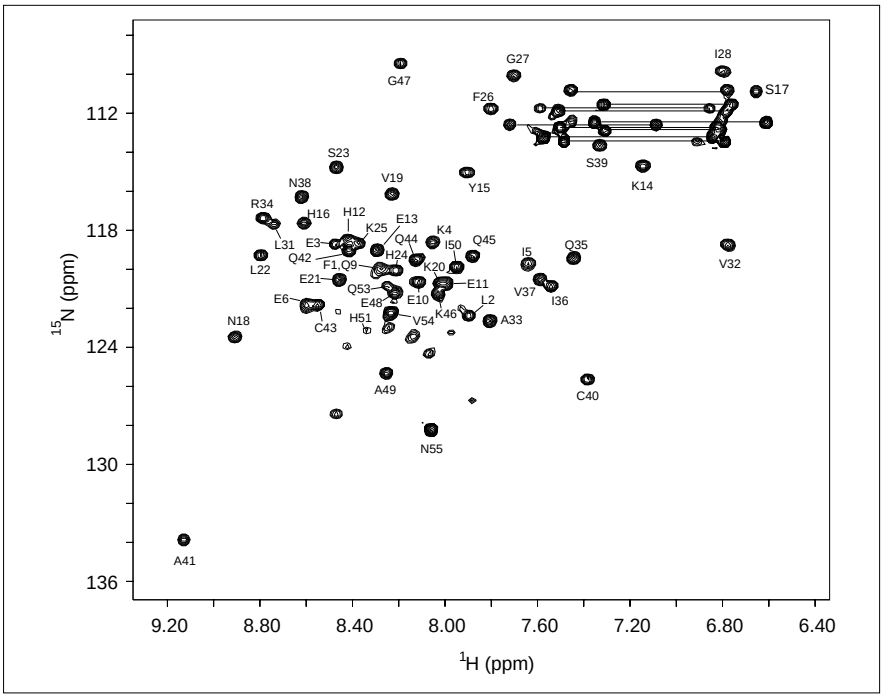
Background

Current successes in chemotherapy for acquired immune deficiency syndrome (AIDS) are partly based on structure–function analysis of the replication enzymes reverse transcriptase and protease. The structure of the third *pol*-encoded replication enzyme, integrase, has been elusive thus far, but the structures of two of its three domains, the catalytic core and the carboxy-terminal DNA-binding domain, have been solved. Here, we report the solution structure of the remaining amino-terminal domain.

Integrase catalyses two distinct reactions: site-specific cleavage of two nucleotides from both 3′ ends of the viral DNA (cleavage reaction), and integration of the recessed viral DNA into the target DNA (integration reaction) [1–4]. Three domains of the protein have been identified by partial proteolysis [5] and deletion analysis [6,7] as well as by *in vitro* complementation between mutant integrase proteins [8,9]. The X-ray structure of a mutant of the catalytic core of HIV-1 integrase has been solved [10], as well as that of the core of integrase from avian sarcoma virus [11]. These structures resemble the structure of other polynucleotidyl transferases such as ribonuclease H, RuvC resolvase and MuA transposase [12–14]. The solution structure of the carboxy-terminal DNA-binding domain as determined by nuclear magnetic resonance (NMR) has been reported to have a Src homology 3 (SH3)-like fold [15,16].

The amino terminus of integrase contains a putative zinc-finger motif, $HisX_{3-7}HisX_{23-32}CysX_2Cys$ (HHCC), which is conserved in all retroviral integrases and also in other integrases such as those of Ty3 from yeast or copia from *Drosophila* [17–19]. Although the exact role of this domain is not known, it is required for cleavage and integration.

Figure 1



Two-dimensional (^{1}H,^{15}N)-HSQC spectrum of $IN_{1–55}$ in the presence of zinc at pH 6.5. The residues of the folded form are labelled by residue names. The side chain NH_2 signals of glutamine and asparagine residues are indicated by lines only. The unmarked NH cross-peaks belong to the unfolded form.

Zinc binding by the first 55 amino acids of integrase has been demonstrated in a zinc blot and by atomic absorption spectroscopy [6,20,21]. Zinc induces folding in the amino-terminal domain as well as in the full-length protein and binds in an equimolar ratio [20,22]. In the full-length protein, the presence of zinc promotes multimerization and stimulates Mg^{2+}-dependent cleavage [22–24]. Mutations in the HHCC motif reduce zinc binding and impair cleavage and integration [6,25,26]. These studies strongly implicate the HHCC motif in zinc binding.

The primary sequence of the amino-terminal domain does not show homology with any of the zinc-finger topologies characterized thus far [27–30]. Previously, it has been suggested that the structure of this domain might have similarities with the structure of the CCHH zinc-finger domain, which is found in transcription factor IIIA-like peptides [20]. However, the three-dimensional solution structure of the amino-terminal domain of HIV-2 integrase, as reported here, shows no structural similarity with any of the known zinc-finger structures.

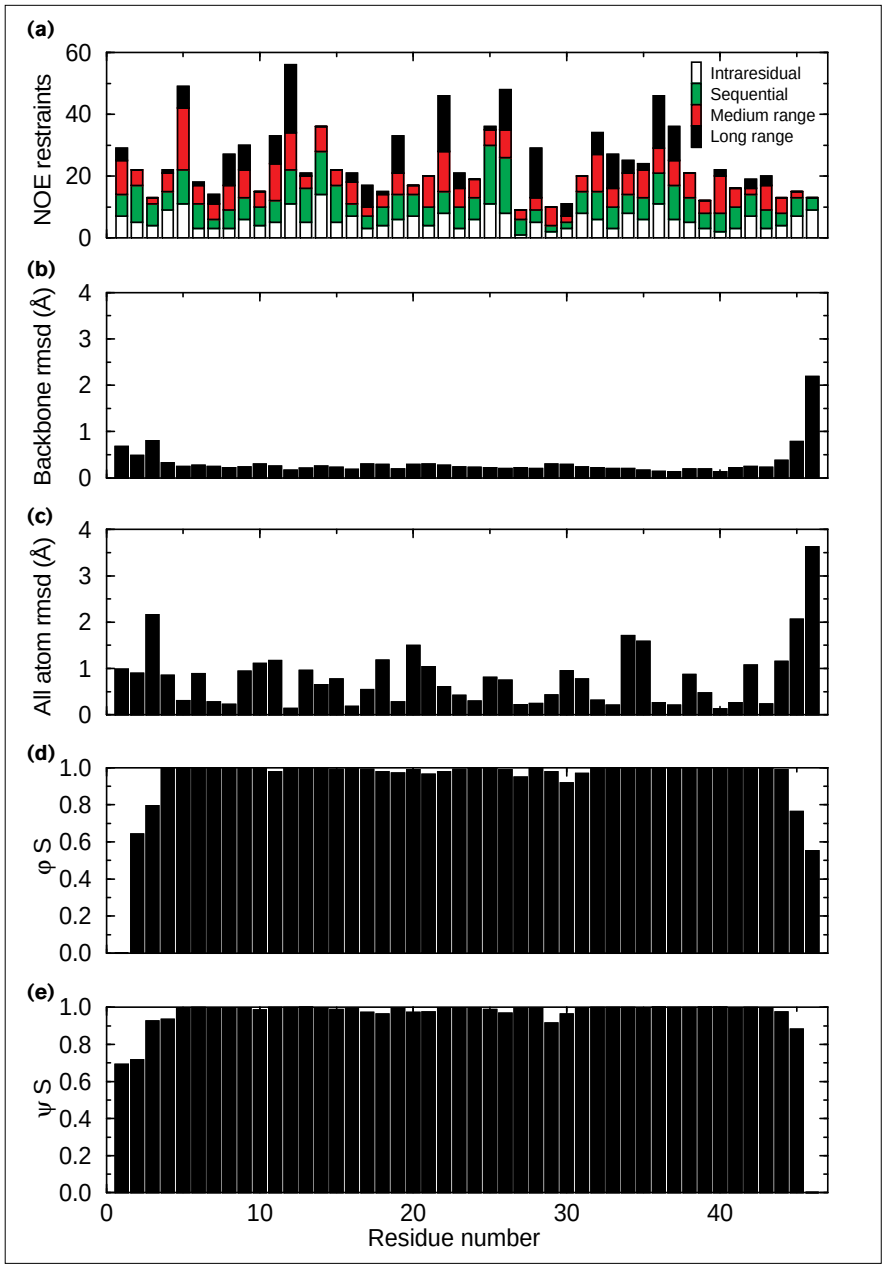
Results and discussion

Zinc induces folding of $IN_{1–55}$

NMR measurements showed that the amino-terminal domain of HIV-2 integrase (hereafter named $IN_{1–55}$) is folded and stable over a pH range 5–8 in the presence of $ZnCl_2$. Removal of Zn^{2+} by EDTA results in unfolding of the protein. Under reducing conditions the protein can be refolded by adding Zn^{2+}. A two-dimensional (^{1}H,^{15}N)-HSQC spectrum of $IN_{1–55}$ in the presence of zinc at pH 6.5, at which the structural studies were performed, is shown in Figure 1. Apart from signals corresponding to amides of the folded protein, some weak peaks from the unfolded form can be observed as well, indicating that the affinity of the $IN_{1–55}$ fragment for Zn^{2+} is relatively low. In ROESY experiments, cross-peaks can be observed between the Zn^{2+}-containing folded and the unfolded form of $IN_{1–55}$, demonstrating that these two forms are in slow exchange (data not shown).

In gel-filtration experiments, $IN_{1–55}$ elutes at the position of a dimer at a protein concentration of approximately 0.1 mM under similar conditions as used in the NMR experiments. In an attempt to further characterize the aggregation state, two-dimensional ^{13}C and ^{15}N/^{13}C double-half-filtered nuclear Overhauser effect (NOE) experiments were performed on a mixture of unlabelled and ^{15}N/^{13}C-labelled $IN_{1–55}$. Since the sample was obtained by adding Zn^{2+} to a mixture of unfolded unlabelled and labelled $IN_{1–55}$, heterodimeric molecules should be formed in which one monomer is unlabelled and the other contains the ^{15}N/^{13}C isotope labels. In the half-filtered two-dimensional NOE spectra, however, no clear intersubunit NOEs could be observed above the artefact level. Various possible explanations can be offered for this observation. One explanation would be that the dimer forms weakly, if at all, under NMR conditions; this,

Figure 2



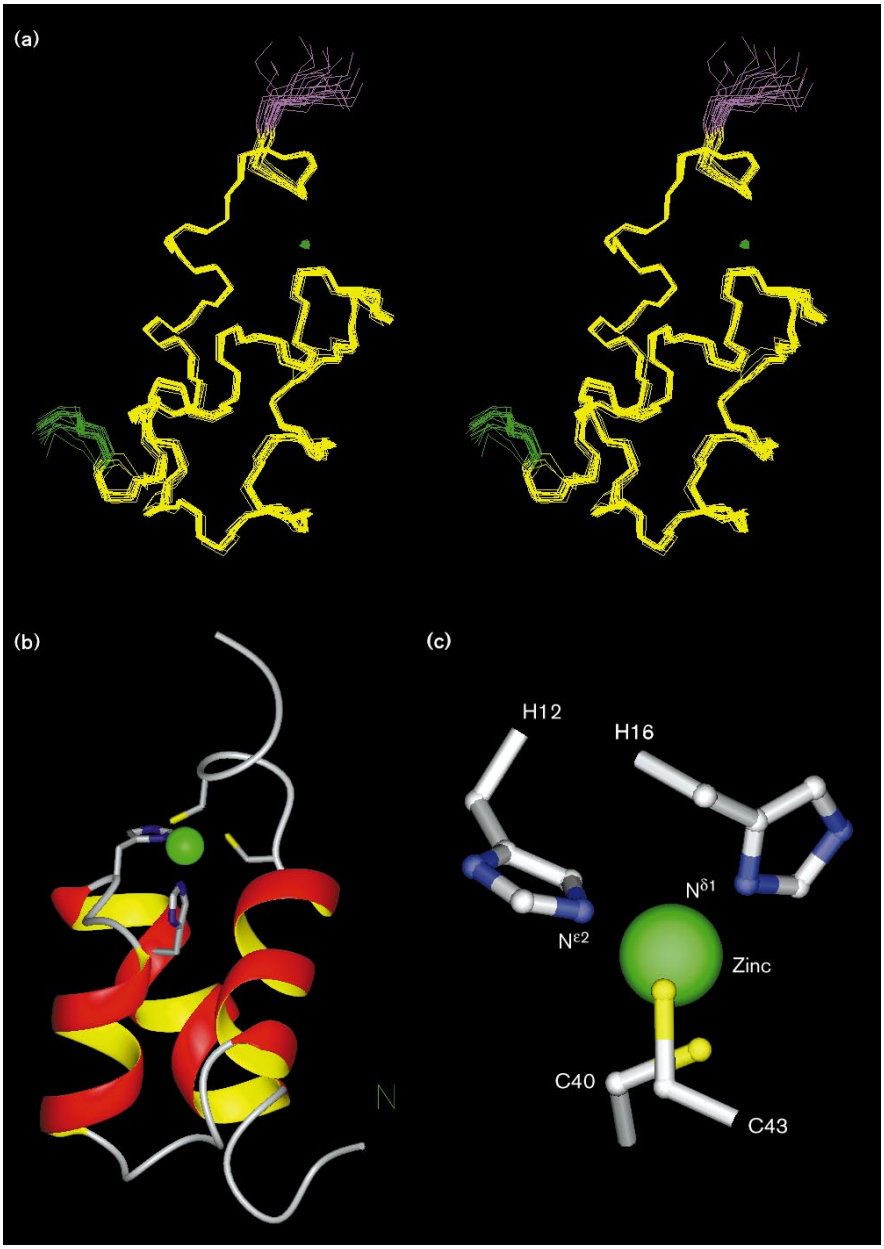
Overview of structural parameters of residues 1–46 of IN_{1-55}. **(a)** Number of distance restraints per residue. The restraints are classified as long range, medium range, sequential and intraresidual. **(b, c)** Root mean square deviation (r.m.s.d.) for the backbone C^{α}, N, C′ atoms and all heavy atoms versus the residue number, respectively. The structures are superimposed on the backbone atoms of residues 3–44. **(d)** The angular order parameter for the backbone angle ϕ. **(e)** The angular order parameter for the backbone angle ψ.

however, contradicts the gel-filtration experiments, which suggest that dimers are already present at a concentration of 0.1 mM. A more likely explanation is that there are significant local motions at the dimer interface, which could selectively broaden resonances originating from the interface beyond detection. Therefore, we have focused in this study on the monomeric fold of the protein.

IN_{1-55} structure determination

Sequence-specific resonance assignments were obtained using two-dimensional and three-dimensional NMR (see Materials and methods). NOE distance restraints were obtained from two-dimensional NOE and three-dimensional NOESY-(^{1}H,^{15}N)-HSQC spectra recorded on 600 and 750 MHz spectrometers. Various simulated annealing calculations were performed using these restraints. Initial calculations without any restraints for zinc coordination showed that His12 coordinates zinc through the $N^{\varepsilon 2}$ atom, whereas His16 coordinates zinc through $N^{\delta 1}$. This finding is confirmed by the cross-peak pattern observed in a (^{1}H,^{15}N)-HSMQC spectrum (see Supplementary mate-rial), which shows that His12 is protonated at the $N^{\delta 1}$ atom, whereas His16 is protonated at the $N^{\varepsilon 2}$ atom [31]. The region corresponding to residues 47–55 is largely disordered due to a

Figure 3

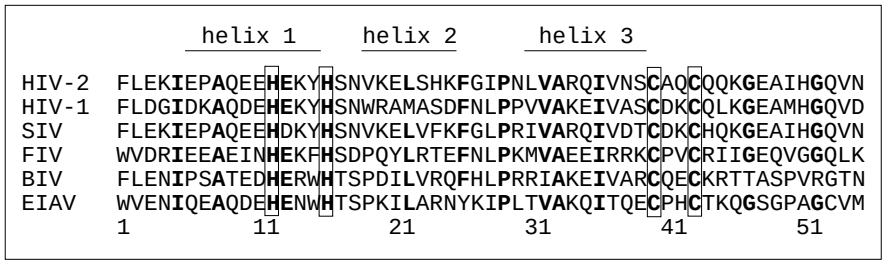


Solution structure of the amino-terminal domain of HIV-2 integrase. **(a)** Stereoview of a superposition of backbone traces (residues 1–46) of 20 calculated structures. The well defined region is coloured yellow, the zinc ions and the N-termini (residues 1–2) are coloured green and the C-termini (residues 45–46) are coloured purple. **(b)** Schematic view of the structure. The structure closest to the average is presented. **(c)** Coordination topology of the zinc centre.

low amount of NOEs, and therefore this part of the protein was not included in the final structure calculations. These final calculations were performed with 693 NOE-derived distance restraints (approximately 15 restraints per residue) and were supplemented by a set of restraints to define the proper tetrahedral coordination geometry around the zinc ion. Figure 2a shows the distribution of NOEs over residues 1–46 of the $IN_{1–55}$ sequence. Of 48 structures calculated, 20 structures were selected on the basis of low number of NOE violations and low overall energies. Figure 3a shows a superposition of the 20 selected structures, and a summary of structural statistics is given in Table 1 and Figure 2. Table 1 shows that very few NOE violations larger than 0.35 Å are present in the structures, indicating that the implicit assumption that most NOEs would correspond to a single monomer or subunit was a valid one. The region in the structure corresponding to residues 3–44 is well defined and displays a root-mean square deviation (r.m.s.d.) from the average structure for the backbone atoms of 0.25 Å and for all heavy atoms of 0.87 Å (see Figure 2b and 2c). Angular order parameters for the backbone angles ϕ and ψ are shown in Figure 2d and 2e, respectively. Figure 2 shows that significant disorder in the backbone conformation is limited to the amino and carboxyl termini. This disorder

Figure 4

Amino acid sequences of integrase from human immunodeficiency virus type 2 (HIV-2), type 1 (HIV-1), simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), bovine immunodeficiency virus (BIV) and equine infectious anaemia immunodeficiency virus (EIAV). Residues that are identical in at least five sequences are shown in bold. The position of the three α helices of IN_{1-55} is indicated above the sequences.



can be accounted for by the low number of long-range NOEs. A PROCHECK analysis of the structures [32] showed that more than 77% of all residues (excluding glycines and prolines) are in the most favoured region of the Ramachandran plot and less than 0.5% of all residues are in the disallowed region. The coordinates have been deposited at the Brookhaven protein databank, pdb accession code 1AUB.

Structure of the HHCC domain

Figure 3b shows a ribbon diagram of the folding topology of IN_{1-55}. The structure consists of three α helices, which range from residues 6–15, 19–25 and 31–39. In addition, medium-range NOEs indicate the presence of a helical turn for residues 41–44. The three-helix bundle is stabilized by a zinc-binding unit that is located near the carboxyl termini of the first and the third helix. The zinc-binding unit is formed by His12, which coordinates zinc through the $N^{\varepsilon 2}$ atom, by His16, which coordinates zinc via $N^{\delta 1}$, and by Cys40 and Cys43, which coordinate zinc through their sulfur atoms. Figure 3c shows the metal coordination of the HHCC motif. The $N^{\varepsilon 2}$-coordinated His12 is located in the last turn of the first helix, whereas the $N^{\delta 1}$-coordinated His16 is located just outside this helix (see Figure 3b). This coordination is consistent with the observation that helical histidine residues generally coordinate a metal through $N^{\varepsilon 2}$, whereas non-helical histidines can coordinate metals via $N^{\delta 1}$ or $N^{\varepsilon 2}$ [33]. We note that this coordination is different from previous zinc-finger structures. In the classical CCHH zinc fingers, the two helical histidines coordinate zinc through $N^{\varepsilon 2}$, while $N^{\delta 1}$ coordination was found for the non-helical histidine in the CCHC zinc-binding motif of the LIM domain [34], as well as in the CHCC zinc-binding motif of the RING finger domain [35,36]. The residues of the HHCC motif are strongly conserved in all retroviral integrases and the previous conclusion that these residues would be involved in zinc coordination is strongly supported by our analysis of the solution structure.

Table 1

Structural statistics for HIV-2 IN_{1-55}*.

Number of distance restraints	
All	693
Intraresidue	264
Sequential	172
Medium range	154
Long range	103
R.m.s.d. from average structure (Å)†	
Backbone (N, C^{α}, C′)	0.25 ± 0.08
All heavy atoms	0.87 ± 0.11
R.m.s.d. from experimental distance restraints (Å)	
All	0.048 ± 0.003
Deviations from idealized covalent geometry	
Bonds (Å)	0.0042 ± 0.0003
Angles (°)	0.61 ± 0.03
Impropers (°)	0.67 ± 0.02
Average number of experimental distance restraints violated by more than 0.35 Å	0.60
Maximum distance restraint violation (Å)	0.50
Percentage of residues with ϕ/ψ‡ in:	
Most favoured regions	77.8%
Additionally allowed regions	16.1%
Generously allowed regions	5.7%
Disallowed regions	0.4%

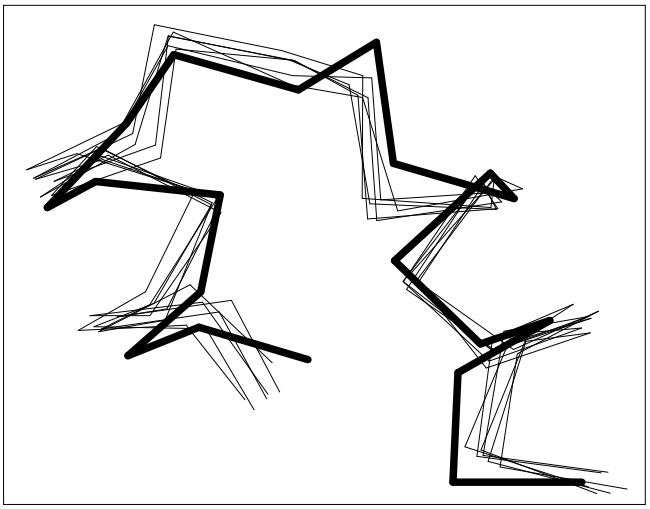
*Structural statistics are presented for the 20 final simulated annealing structures for residues 1–46. †Residues 3–44 (r.m.s.d., root mean square deviation). ‡Residues excluding glycines and prolines.

Apart from the residues of the HHCC motif, various other residues are well conserved in integrases of immunodeficiency viruses (Figure 4). Of these residues, Ile5, Ala8, Leu22, Ile28, Ala33 and Ile36 form part of the hydrophobic core, which explains their importance for structural integrity (note that Ile28 is conservatively substituted by leucine in most sequences). Other well conserved hydrophobic residues are Pro29, which is located in the turn between helix 2 and 3, and Phe26 and Val32. In the structure, Phe26 and Val32 are partly exposed at the surface and, therefore, it is difficult to conclude whether their role is structural or whether they are involved in interactions with other molecules.

Structural similarity to the DNA-binding three-helix bundles

Possible structural similarity of IN_{1-55} to other proteins was investigated with the Dali Web server [37]. The overall

Figure 5

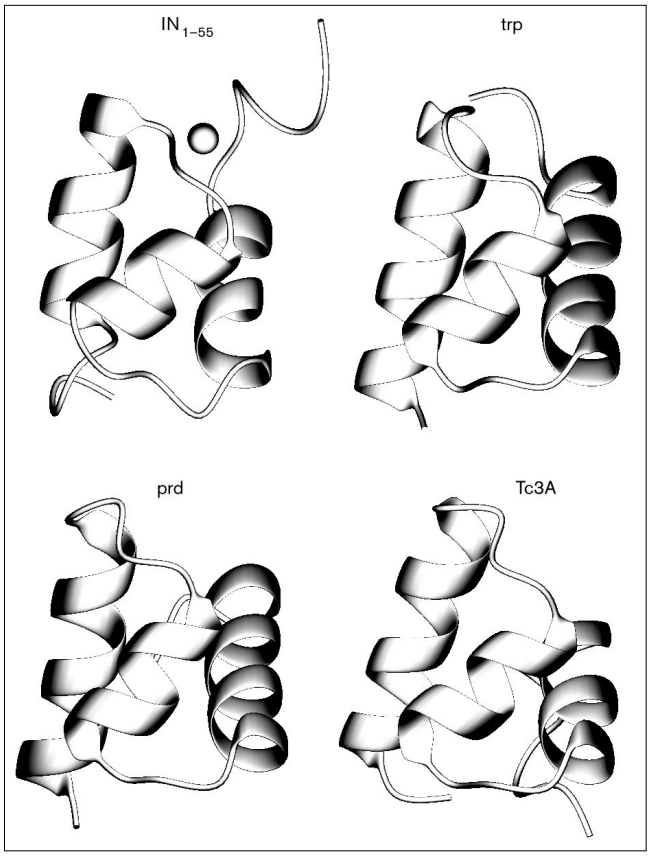


The folding topology of IN_{1-55} and the equivalent regions of three DNA-binding three-helix bundles. Shown are: IN_{1-55}; TRP, trp repressor mutant V581 (pdb entry 1jhg); PRD, amino-terminal subdomain of the prd paired domain (pdb entry 1 pdn); Tc3A, transposase Tc3A.

folding topology turns out to be similar to that of several DNA-binding three-helix bundle folds, as classified by the SCOP database [38]. The backbone atoms of IN_{1-55} superimpose with an r.m.s.d. of 2.1 Å on the equivalent region of the trp repressor (41 residues, [39]) and with 2.2 Å on both the prd paired domain (40 residues, [40]) and the biotin repressor BirA (38 residues, [41]). Interestingly, BirA contains an SH3-like fold at the carboxyl terminus and a three-helix bundle in the amino-terminal part, similarly to integrase. Another remarkable observation is that the amino-terminal DNA-binding domain of the transposase Tc3A, a protein with functional similarity to integrase, has the same type of fold as IN_{1-55} [42]. The structurally equivalent regions of IN_{1-55}, the trp repressor, the prd paired domain and transposase Tc3A are shown in Figure 5. Within the DNA-binding three-helix bundle, the second and third helices form a helix-turn-helix (HTH) motif in which the third, so-called recognition, helix binds DNA. The second and third helices of IN_{1-55} are similar to the HTH motif found in both prokaryotic and eukaryotic DNA-binding proteins. These helices of IN_{1-55} form an angle of 124° with each other, the buried residues at position 4 and 15 of the HTH motif correspond to the core residues Leu22 and Ala33, respectively, and the commonly occurring glycine at position 9 is also a glycine in IN_{1-55} (see Figure 4). The HTH motif of IN_{1-55} superimposes on the equivalent C^{α} atoms of several HTH motifs (Figure 6) with an r.m.s.d. ranging from 1.2 Å for the prd paired domain [40] to 1.4 Å for the trp repressor [39].

It has been suggested that the amino terminus of integrase is involved in the recognition of viral DNA [26,43].

Figure 6



Superposition of C^{α} traces of the HTH motif of IN_{1-55} (residue 19–38) on the HTH-motifs of six DNA-binding three-helix bundles. The bold C^{α} trace is from IN_{1-55}, the others (with the pdb entry given in parentheses) are from: MATa1 homeodomain (1yrn), trp repressor mutant V58I (1jhg), prd paired domain (1pdn), CAP (1ber), Oct-1 POU homeodomain (1oct), and biotin repressor BirA (1bia).

However, at present it is not clear whether the three-helix bundle of IN_{1-55} binds DNA, or whether it interacts with other domains of integrase [44,45], thereby affecting viral DNA recognition.

Conclusions

The amino-terminal domain of HIV-2 integrase has a remarkable hybrid structure combining features of a three-helix bundle fold with a zinc-binding HHCC motif. This three-helix bundle shows considerable structural similarity to several DNA-binding proteins containing an HTH motif. In integrase, the zinc is required for stabilization of this three-helix bundle. This structure presents a new fold for a zinc-binding motif. The strictly conserved residues of the HHCC motif of retroviral integrases are involved in coordinating the metal ion. Many other well conserved hydrophobic residues are part of the protein core, whereas the conserved Pro29 is in the turn between helix 2 and 3.

While this paper was being reviewed, the structure of the amino-terminal region of HIV-1 integrase was published [46]. In contrast to the HIV-2 amino terminus, the amino terminus of HIV-1 (residues 1–55) exists in two interconverting folded states. This difference is clearly visible in the corresponding (^{1}H,^{15}N)-HSQC spectra; in the case of HIV-1, many residues display a doubling of the cross-peaks. One of the forms shows the same folding topology as HIV-2 $IN_{1–55}$. In the other form, the histidines are coordinated differently to the zinc and large conformational differences occur for residues 9–18. Another important difference is that the structure of HIV-1 $IN_{1–55}$ was solved as a dimer. Several of the hydrophobic residues that are part of the interface in HIV-1 $IN_{1–55}$ are not hydrophobic in the HIV-2 protein (Pro30 → N30; A38 → N38; L45 → Q45, see Figure 4) and a potential salt bridge in the interface of HIV-1 $IN_{1–55}$ can not be formed in HIV-2 $IN_{1–55}$ (E35 → Q35). These differences are likely to affect the nature of the dimer interface. A consequence might be an enhanced mobility of the interfacial residues for HIV-2 IN_{1-55}, which may explain our difficulty in detecting intersubunit NOEs by NMR for HIV-2 $IN_{1–55}$.

With the elucidation of the structure of the amino-terminal domain, structural information is now available for all three domains of HIV integrase. Nevertheless, in order to define the topology of the whole protein, it still remains important to determine the structure of intact integrase.

Materials and methods

$IN_{1–55}$ preparation

The coding region of the 55 amino-terminal amino acids of HIV-2 IN was amplified by polymerase chain reaction (PCR) from pRP279 [47] and fused in frame to the 3′ end of the *GST* gene in vector pGEX-2T. The resulting construct (pRP1021) was expressed in *Escherichia coli* BL21(DE3). The cells were sonicated in buffer A, containing 0.5 M NaCl, 25 mM Hepes, pH 7.6, 0.1 mM EDTA and 3 mM β-mercaptoethanol. After centrifugation at 10,000 × *g* for 30 min, the supernatant was diluted 2.5 times with buffer A without NaCl and bound to glutathione Sepharose 4B (Pharmacia). The column was washed with buffer A containing 200 mM NaCl (buffer B) with 0.1% Tween, and with buffer B without Tween. The protein was eluted with 20 mM glutathione in buffer B. Top fractions were pooled and cleaved by thrombin (approximately 2.5 NIH units/mg $IN_{1–55}$) as described by the manufacturer (Sigma). $IN_{1–55}$ was further purified by gel filtration on a HiLoad Superdex 75 column (Pharmacia) and concentrated using Centriprep (Amicon). The protein was dialysed to 3 M guanidine–HCl, 20 mM Tris, pH 8.3, 3 mM β-mercaptoethanol, followed by a reduction step in 20% (v/v) β-mercaptoethanol at 55°C for 1.5 h as described by Burke *et al.* [20]. Subsequently, the protein was purified over a C-18 reversed phase HPLC column and eluted with a gradient of 5–70% (v/v) acetonitrile in 0.05% (v/v) trifluoroacetic acid. The top fractions were lyophilized and kept under N_2 gas to prevent oxidation. In addition to residues 1–55, the protein fragment contains the sequence Gly–Ser–Met at the amino terminus. Uniformly ^{15}N-labelled or ^{15}N/^{13}C-labelled protein was prepared by growing cells in minimal medium with $^{15}NH_4Cl$ and [U-^{13}C]-glucose as sole nitrogen and carbon sources, respectively. For the NMR experiments, the lyophilized protein was dissolved in 50 mM deuterated Tris, 150 mM NaCl and 2 mM β-mercaptoethanol with 1.1 equivalent of $ZnCl_2$, either in 99.99% D_2O or 95% H_2O/5% D_2O, and the pH was adjusted to 6.5. Protein concentrations were approximately 3.5 mM. Three types of NMR samples were used: uniformly ^{15}N-labelled $IN_{1–55}$ in 95% H_2O/5% D_2O; uniformly ^{15}N-labelled $IN_{1–55}$ in 99.99% D_2O; and a 1:1 mixture of unlabelled and uniformly ^{15}N/^{13}C-labelled $IN_{1–55}$ in 95% H_2O/5% D_2O. The last sample was prepared by mixing the purified unfolded labelled and un-labelled proteins after passage through the C-18 reversed phase HPLC column. The lyophilized protein was then folded by dissolving in the zinc-containing buffer as described above.

NMR spectroscopy and structure calculations

NMR spectra were recorded at 300 K on Bruker AMXT-600 and Varian Unity+ 750 MHz spectrometers equipped with triple-resonance gradient probes. Sequential assignments were obtained from sensitivity-enhanced three-dimensional TOCSY-(^{1}H,^{15}N)-HSQC and NOESY-(^{1}H,^{15}N)-HSQC, homonuclear two-dimensional NOE and TOCSY spectra in H_2O. Aromatic side chains were assigned using two-dimensional NOE and TOCSY spectra in D_2O. The protonation states of the histidines were determined from a (^{1}H,^{15}N)-HSMQC spectrum recorded as described by Zuiderweg [48]. The ^{13}C and ^{15}N/^{13}C doubly half-filtered experiments on the 1:1 mixture of unlabelled and uniformly ^{15}N/^{13}C-labelled $IN_{1–55}$ were recorded as described by Folkers *et al.* [49] and Slijper *et al.* [50]. NOE distance restraints were obtained from two-dimensional NOE and three-dimensional NOESY-(^{1}H,^{15}N)-HSQC spectra with mixing times of 50 and 100 msec. In the 50 msec NOE spectra, strong, medium and weak peaks are visible and the corresponding upper distance bounds were set to 2.9 Å, 3.6 Å and 4.5 Å. Additional very weak peaks observed at 100 msec mixing time only were converted into an upper distance bound of 6 Å. The lower distance bounds were set to 1.8 Å. In the final calculations, no hydrogen bond restraints were included. Structure calculations were performed with X-PLOR, version 3.1 [51]. To correct for multiple atom selections we used the sum-averaging option as implemented in X-PLOR. Distance restraints containing diastereotopic groups were corrected as described by Fletcher *et al.* [52]. The structures were calculated using a dynamical simulated annealing protocol starting from randomized coordinates [53]. The final structure calculations were carried out in the presence of zinc. The zinc ion was modelled in the protein structure file with a set of restraints to provide tetrahedral coordination of the zinc atom and to keep the zinc atom within the plane of the histidine rings. Bond lengths were restrained to 2.0 Å for Zn-His$N^{\varepsilon 2}$ and Zn-His$N^{\delta 1}$ and 2.3 Å for Zn-CysS. Bond angles were restrained to 126° for Zn-His$N^{\varepsilon 2}$-His$C^{\varepsilon 1}$/$C^{\delta 2}$ and Zn-His$N^{\delta 1}$-$C^{\varepsilon 1}$/C^{γ} and to 100° for Zn-CysS^{γ}-CysC^{β}. The secondary structure elements were identified with the Kabsch–Sander algorithm [54] as implemented in PROCHECK [32]. The well defined region for the overlays was selected on the basis of an average ϕ and ψ angular order parameter larger than 0.85. Figures 3, 5 and 6 were prepared with the program MOLMOL [55].

Supplementary material

The (^{1}H,^{15}N)-HSMQC spectrum from which the tautomeric states of the zinc-coordinating histidines were obtained is published with this paper on the internet.

Acknowledgements

We thank Albert George and Richard van der Valk for technical assistance and Gertie van Pouderoyen and Titia Sixma for making the Tc3A structure available before publication. We also thank Ramon Puras Lutzke for critical reading of the manuscript. This work was funded in part by a grant from the Netherlands AIDS foundation. A.E. was supported by the Netherlands Foundation for Chemical Research (SON) with financial support from the Netherlands Organization for Scientific Research (NWO). 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.

References

1. Goff SP: **Genetics of retroviral integration.** *Annu Rev Genet* 1992, **26:**527-544.
2. Whitcomb JM, Hughes SH: **Retroviral reverse transcription and integration: progress and problems.** *Annu Rev Cell Biol* 1992, **8:**275-306.
3. Vink C, Plasterk RH: **The human immunodeficiency virus integrase protein.** *Trends Genet* 1993, **9:**433-438.
4. Katz RA, Skalka AM: **The retroviral enzymes.** *Annu Rev Biochem* 1994, **63:**113-173.
5. 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.

6. 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.
7. Vink C, Oude Groeneger AM, Plasterk RH: **Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type 1 integrase protein.** *Nucleic Acids Res* 1993, **21:**1419-1425.
8. van Gent DC, Vink C, Oude Groeneger AM, Plasterk RH: **Complementation between HIV integrase proteins mutated in different domains.** *EMBO J* 1993, **12:**3261-3267.
9. Engelman A, Bushman FD, Craigie R: **Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex.** *EMBO J* 1993, **12:**3269-3275.
10. 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.
11. 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.
12. Yang W, Steitz TA: **Recombining the structures of HIV integrase, RuvC and RNase H.** *Structure* 1995, **3:**131-134.
13. Rice P, Craigie R, Davies DR: **Retroviral integrases and their cousins.** *Curr Opin Struct Biol* 1996, **6:**76-83.
14. Mizuuchi K: **Polynucleotidyl transfer reactions in site-specific DNA recombination.** *Genes to Cells* 1997, **2:**1-12.
15. Eijkelenboom AP, Puras Lutzke RA, Boelens R, Plasterk RH, Kaptein R, Hård K: **The DNA-binding domain of HIV-1 integrase has an SH3-like fold.** *Nature Struct Biol* 1995, **2:**807-810.
16. Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, *et al.:* **Solution structure of the DNA binding domain of HIV-1 integrase.** *Biochemistry* 1995, **34:**9826-9833.
17. Johnson MS, McClure MA, Feng DF, Gray J, Doolittle RF: **Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes.** *Proc Natl Acad Sci USA* 1986, **83:**7648-7652.
18. Doolittle RF, Feng DF, Johnson MS, McClure MA: **Origins and evolutionary relationships of retroviruses.** *Q Rev Biol* 1989, **64:**1-30.
19. 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.
20. Burke CJ, Sanyal G, Bruner MW, Ryan JA, LaFemina RL, Robbins HL *et al.*: **Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase.** *J Biol Chem* 1992, **267:**9639-9644.
21. McEuen AR, Edwards B, Koepke KA, Ball AE, Jennings BA, Wolstenholme AJ, *et al.*: **Zinc binding by retroviral integrase.** *Biochem Biophys Res Commun* 1992, **189:**813-818.
22. Zheng R, Jenkins TM, Craigie R: **Zinc folds the amino-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity.** *Proc Natl Acad Sci USA* 1996, **93:**13659-13664.
23. 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.
24. 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.
25. van Gent DC, Oude Groeneger AM, Plasterk RH: **Mutational analysis of the integrase protein of human immunodeficiency virus type 2.** *Proc Natl Acad Sci USA* 1992, **89**:9598-9602.
26. 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.
27. Klug A, Rhodes D: **Zinc fingers: a novel protein fold for nucleic acid recognition.** *Trends Biochem Sci* 1987, **12:**464-469.
28. Kaptein R: **Zinc fingers.** *Curr Opin Struct Biol* 1991, **1:**63-70.
29. Schwabe JW, Klug A: **Zinc mining for protein domains.** *Nature Struct Biol* 1994, **1:**345-349.
30. Berg JM, Shi Y: **The galvanization of biology: a growing appreciation for the roles of zinc.** *Science* 1996, **271:**1081-1085.
31. Pelton JG, Torchia DA, Meadow ND, Roseman S: **Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques.** *Protein Sci* 1993, **2:**543-558.
32. 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.
33. Chakrabarti P: **Geometry of interaction of metal ions with histidine residues in protein structures.** *Protein Eng* 1990, **4:**57-63.
34. Perez-Alvarado GC, Miles C, Michelsen JW, Louis HA, Winge DR, Beckerle MC, Summers MF: **Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP.** *Nature Struct Biol* 1994, **1:**388-398.
35. Barlow PN, Luisi B, Milner A, Elliott M, Everett R: **Structure of the C_3HC_4 domain by ^{1}H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger.** *J Mol Biol* 1994, **237:**201-211.
36. Borden KL, Lally JM, Martin SR, O'Reilly NJ, Etkin LD, Freemont PS: **Novel topology of a zinc-binding domain from a protein involved in regulating early Xenopus development.** *EMBO J* 1995, **14:**5947-5956.
37. Holm L, Sander C: **Protein structure comparison by alignment of distance matrices.** *J Mol Biol* 1993, **233:**123-138.
38. 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.
39. Lawson CL: **An atomic view of the l-tryptophan binding site of trp repressor.** *Nature Struct Biol* 1996, **3:**986-987.
40. Xu W, Rould MA, Jun S, Desplan C, Pabo CO: **Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations.** *Cell* 1995, **80:**639-650.
41. 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.
42. van Pouderoyen G, Kettting RF, Perrakis A, Plasterk RH, Sixma TK: **Crystal structure of the specific DNA binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA.** *EMBO J 1997,* in press.
43. Mazumder AN, Neamati JO, Ojwang S, Sunder S, Rando RF, Pommier Y: **Inhibition of human immonodeficiency virus type 1 integrase by guanosine quartet structures.** *Biochemistry* 1996, **35:**13762-13771.
44. Ellison V, Gerton J, Vincent KA, Brown, PO: **An essential interaction between distinct domains of of HIV-1 integrase mediates assembly of the active multimer.** *J Biol Chem* 1995, **270:**3320-3326.
45. Bizub-Bender D, Kulkosky J, Skalka AM: **Monoclonal antibodies against HIV type 1 integrase: clues to molecular structure.** *AIDS Res Hum Retrovirus* 1994, **10:**1105-1115.
46. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM: **Solution structure of the amino-terminal zinc binding domain of HIV-1 integrase.** *Nature Struct Biol* 1997, **4:**567-577.
47. van Gent DC, Elgersma Y, Bolk MW, Vink C, Plasterk RH: **DNA binding properties of the integrase proteins of human immunodeficiency viruses types 1 and 2.** *Nucleic Acids Res* 1991, **19:**3821-3827.
48. Zuiderweg ERP: **A proton-detected heteronuclear chemical-shift correlation experiment with improved resolution and sensitivity.** *J Magn Reson* 1990, **86:**346-357.
49. Folkers PJ, Folmer RH, Konings RN, Hilbers CW: **Overcoming the ambiguity problem encountered in the analysis of nuclear Overhauser magnetic resonance spectra of symmetric dimer proteins.** *J Am Chem Soc* 1993, **115:**3798-3799.
50. Slijper M, Kaptein R, Boelens R: **Simultaneous ^{13}C and ^{15}N isotope editing of biomolecular complexes. Application to a mutant lac repressor headpiece DNA complex.** *J Magn Res B* 1996, **111:**199-203.
51. Brünger AT: *X-PLOR, version 3.1: a system for X-ray crystallography and NMR.* Yale University Press: New Haven, CT, USA; 1992.
52. 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.
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. Kabsch W, Sander C: **Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.** *Biopolymers* 1983, **22:**2577-2637.
55. Koradi R, Billeter M, Wüthrich K: **MOLMOL: a program for display and analysis of macromolecular structures.** *J Mol Graph* 1996, **14:**51-55.

Because *Current Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cub – for further information, see the explanation on the contents page.