Recombining the structures of HIV integrase, RuvC and RNase H

The recently reported crystal structures of two recombination enzymes, the catalytic domain of HIV integrase and *Escherichia coli* RuvC, an endonuclease, are surprisingly similar to that of ribonuclease H suggesting the possibility that they have a common enzymatic mechanism.

Structure 15 February 1995, **3**:131–134

Thirty years after the proposal of the Holliday junction as an intermediate in homologous recombination of DNA, the crystal structures of *Escherichia coli* RuvC [1], an endonuclease that specifically cleaves Holliday junctions, and HIV integrase [2], (which integrates HIV DNA into the chromosome) have been determined. The AIDS epidemic has greatly stimulated HIV-related research, and HIV integrase has become one of the best studied integration enzymes since its discovery 10 years ago. In parallel, RuvC has also become one of the best characterized enzymes involved in homologous recombination. The crystal structures of these two recombination enzymes show that their folding topology and active-site geometry are similar to those of the ribonuclease H (RNase H) enzymes from HIV [3] and *E. coli* [4,5] despite the lack of detectable amino acid sequence homology. HIV integrase and RuvC may therefore employ the general two-metal-ion trans-esterification mechanism proposed for RNase H [3,5,6], the 3′ to 5′ exonuclease of Klenow fragment [7], polymerases [8] and ribozymes [9]. The crystal structure of HIV integrase provides the basis for rational drug design against HIV integrase which, unlike the viral protease or reverse transcriptase presently used in drug design, has no host homolog.

Retroviral integration and integrase

An essential process in the retroviral lifecycle is the integration of the viral double-stranded DNA into host chromosomes [10,11]. This integration process is completed in three discrete steps: firstly, the 3′ ends of the viral DNA are processed to expose the absolutely conserved dinucleotides CA-$_{OH}^{3'}$; secondly, strand-transfer of the processed viral DNA into the target DNA takes place, during which the staggered cleavage of target DNA and ligation of the viral 3′ and target 5′ ends are concerted; and finally, the viral 5′ ends are processed and the gaps between the viral 5′ and target 3′ ends are repaired. The retroviral-encoded integrase is able to catalyze the first two steps of integration *in vitro* in the presence of the divalent cations, Mg^{2+} or Mn^{2+}, and accomplishes both the hydrolysis and strand-transfer of DNA in a trans-esterification reaction using first OH^- and then the 3′-OH of DNA as nucleophiles. The reverse of the second step, called the disintegration reaction, cleaves off an invading viral DNA branch and reseals the target DNA [12] and is used as an assay of integrase activity. The third step of integration is believed to be carried out by host DNA-repair enzymes.

HIV integrase contains 288 amino acids and can be divided into three functional domains [13]. The N-terminal domain (residues 1–49) contains a zinc-binding motif with two histidine and two cysteine residues as putative Zn^{2+} ligands, which is conserved among all retroviral and retrotransposon integrases. The central core domain (residues 50–212) contains three acidic residues that are invariant among integrases and are arranged in the same conserved sequence pattern as that of the catalytic residues in bacterial transposases [14]. Mutation of any one of the three acidic residues renders an integrase inactive. In contrast, the remaining C-terminal residues of the integrase family have quite diverse sequences. The central core domain alone is capable of catalyzing the disintegration reaction but not the 3′ processing and strand-transferring reactions [13], whereas the N- and C-terminal domains are required for the viral integration and are thought to be critical for substrate specificity. Integrase forms dimers, and multimers of dimers, in solution: these are probably the functional units, because it has been shown that inactive integrase molecules that are defective in different domains can complement one another [15,16].

The determination of the crystal structure of HIV integrase core domain [2] is a triumph of the combination of mutagenesis and multiwavelength anomalous diffraction (MAD) techniques [3]. The major barrier to obtaining the crystal structure of this enzyme, which has prevented many laboratories from pursuing it, has been the insolubility of both intact integrase and the catalytic core domain. The problem was overcome by systematically mutating each of the hydrophobic residues in every hydrophobic patch to a lysine, (this was achieved by Craigie and Jenkins, two of the authors of [2]). The substitution of Phe186 by lysine resulted in a much more soluble, monodispersed integrase core domain that readily formed diffraction-quality crystals. Davies' group then solved the catalytic core domain structure in two months using the MAD technique on the selenomethionine-derivatized protein [2].

RuvC and the Holliday junction

The Holliday junction is a four-way cruciform-like structure formed during homologous recombination between two DNA duplexes by RecA-catalyzed homologous pairing of these duplexes and strand exchange [17,18]. *E. coli* RuvC is a 172 amino acid endonuclease that binds specifically to the Holliday junction and

Fig. 1. (opposite). Stereo ribbon diagrams (left), and topological diagrams (right), of **(a)** *E. coli* HIV RNase H, **(b)** HIV RNase H domain, **(c)** the core domain of HIV integrase and **(d)** *E. coli* RuvC [36]. Secondary structures are defined using program DSSP [37] and the four structures are superimposed based on the least-squares fitting of the two structurally conserved carboxylates and the first four strands in the β-sheet [38]. The topologically conserved β-sheet and helices are shown in dark gray, the disordered loops in HIV RNase H and HIV integrase are represented by the dashed lines; the three conserved carboxylates required for the enzymatic activity are drawn in ball-and-stick representation. In HIV integrase (c), the third carboxylate is located in the disordered loop, and therefore is not shown. The fourth carboxylates in RuvC and RNases H are shown lighter. The two Mn^{2+} ions bound in the active site of HIV RNase H (b) are also shown as light gray spheres. In the topological diagrams of the four enzymes, β-strands are shown as thick arrows, α-helices as rectangular boxes and loops as thin lines. The location of the active-site carboxyl triads are marked by asterisks and the fourth carboxylates marked by plus signs.

resolves it into two separated DNA duplexes [19,20]. The binding of RuvC to Holliday junctions or structurally equivalent cruciform DNAs does not require divalent cations, but RuvC is catalytically active only when the two duplex DNAs are homologous and Mg^{2+} is present [19,20]. The binding of RuvC to the Holliday junction changes its conformation and makes it hypersensitive to hydroxyl-radical cleavage, particularly around the base pairs adjacent to the crossover point [21]. The cleavage of the four-way DNA junction by RuvC then occurs symmetrically on two strands of the same polarity and preferentially at the 3′ side of a thymidine [19–22].

RNase H from HIV and *E. coli*

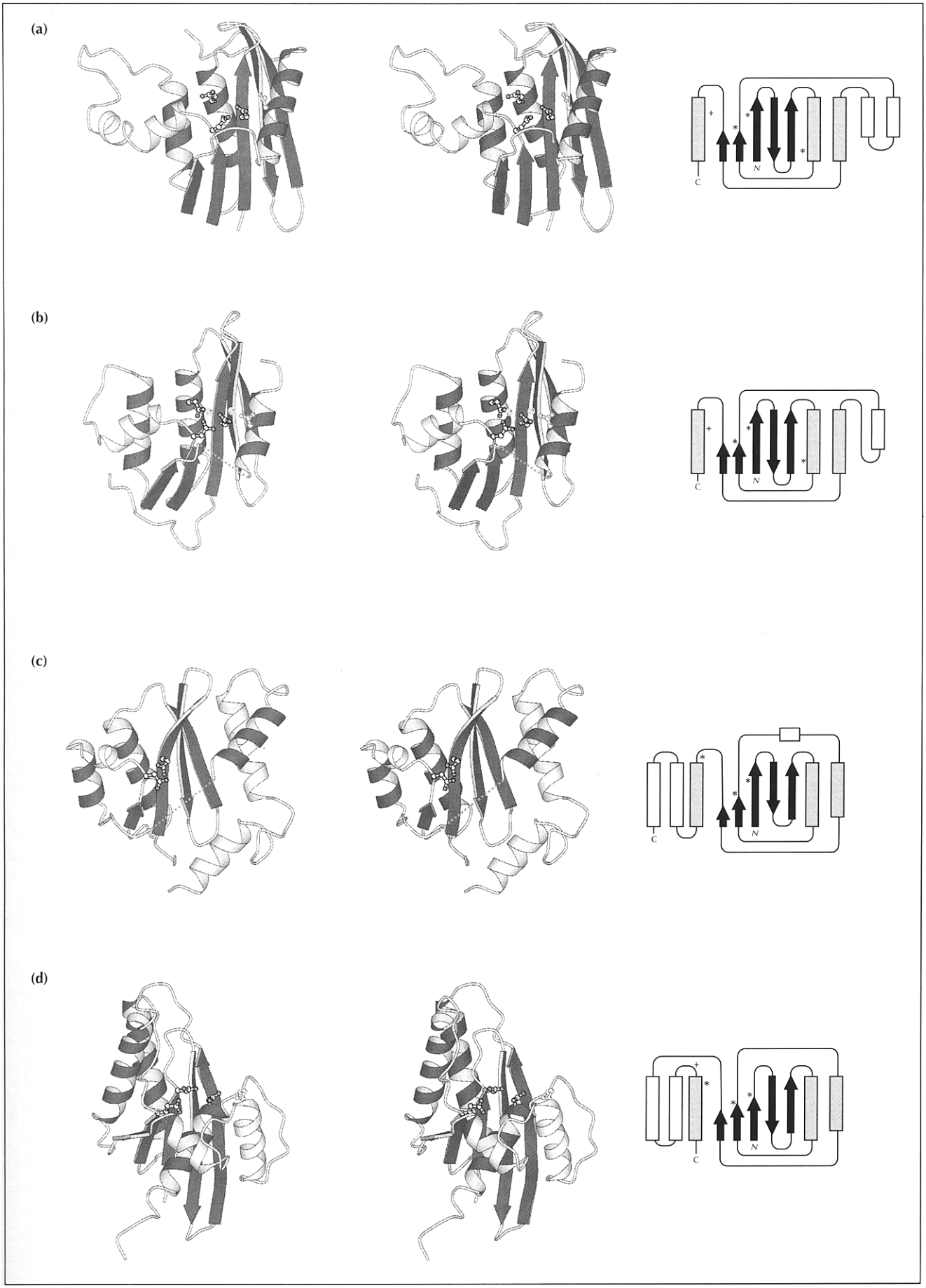
RNase H specifically hydrolyzes RNA strands in RNA–DNA hybrids in the presence of either Mg^{2+} or Mn^{2+}; its reaction mechanism differs from most ribonucleases but resembles deoxyribonucleases and many ribozymes, as it does not use the 2′-OH of ribose as a nucleophile to form a 2′ to 3′ cyclic phosphate intermediate that leads to a 3′ phosphate but rather it produces 3′-OH and 5′ phosphate groups [23,24]. The *E. coli* RNase H comprises a single polypeptide chain of 155 amino acid residues [24], whereas the HIV RNase H is a 130 amino acid domain of the viral reverse transcriptase and is indispensable for the replication of genomic RNA to double-stranded DNA [25]. Crystal structures of both RNases H were determined several years ago [3–5] and shown to be homologous with a high degree of structural similarity despite their relatively low sequence identity [3].

Comparison of HIV integrase, RuvC and RNase H structures

The crystal structures of HIV integrase and RuvC are very similar to those of *E. coli* and HIV RNase H [1,2]. All four structures share a similar αβ-fold containing a central five-stranded mixed β-sheet surrounded by α-helices on both sides, which are in the same topological order (Fig. 1): three antiparallel β-strands followed by two α-turn-β units that position the two parallel helices on one side of the β-sheet and finish with a third helix on the opposite side of the sheet. However, the lengths of the secondary-structure elements, the twist of the β-sheet and the relative positions of helices, vary among the four structures (Fig. 1a). The RNase H structures also have additional helices and loops inserted between the two α-turn-β units, which form part of the substrate-binding surface inferred from structural studies. Integrase and RuvC, in contrast, have two additional helices on the other side of the β-sheet next to the C-terminal helix. In RuvC, these three C-terminal helices together with the β-sheet form a deep cleft that Morikawa and his colleagues suggest is suitable for DNA binding [1]. Finally, although HIV integrase and RuvC form functional dimers in solution [15,16,19,20], in the crystal structures they dimerize using completely different interfaces [1,2]. These four enzymes have the same folding topology as observed for several other proteins, such as the nucleotide-binding domains of a yeast hexokinase [26], the heat shock protein, hsp70, [27] actin [28] and the connection domain of HIV reverse transcriptase [29,30].

Despite the lack of close resemblance, these four enzymes have remarkably similar active sites that consist of a set of three highly conserved carboxylates that are absolutely required for catalytic activity. In all four structures, two of the three carboxylates are located in identical positions relative to the tertiary structure, one in the first β-strand and the other next to the C terminus of the fourth β-strand (Fig. 1). The location of the third carboxylate varies: in RuvC it is on the fifth helix, while in the integrase it occurs on a disordered loop leading into the third helix, which is actually in a vicinity similar to that in RuvC; in RNase H, however, the third carboxylate is on the first helix. If the four structures are aligned by superposition of their β-sheets and the first two carboxylates, the third carboxylate in RNase H is on the opposite side of the β-sheet from those in HIV integrase and RuvC. However, disregarding the alignment of the protein tertiary structures, Morikawa and his coworkers have reported that all atoms of the carboxyl triad from RuvC can be superimposed onto those of *E. coli* RNase H with a root mean square deviation of 2.1 Å [1]. RuvC and RNase H also have a fourth conserved carboxylate that is required for catalysis in RuvC [1]. It can be replaced by a carboxyl amide or imidazole, which could serve as a metal ligand, (though less optimally than the carboxylate) in *E. coli* RNase H [31].

The topological resemblance among the structures of HIV integrase, RuvC and RNase H, particularly their analogous catalytic carboxylates, their shared requirement for divalent cations, and the similar trans-esterification function, raises the possibility that they employ a common 'carboxylate-metal ion' catalytic mechanism, similar to that proposed for RNase H, although neither of the two new crystal structures have been studied in the presence of divalent cations [1,2]. A 'carboxylate-chelated two-metal-ion' catalytic mechanism, originally proposed for the 3′ to 5′ exonuclease of Klenow fragment [7,32] and later extended to the polymerases [8], ribozymes [7,9] and RNase H [3,5], would be feasible for the hydrolysis as well as the polynucleotidyl transfer on both RNAs and DNAs. Two divalent metal ions, spaced 3.8–4.0 Å apart, have been observed in the active site of


(a)
(b)
(c)
(d)
+
*
*
*
N
C

the isolated HIV RNase H domain [3] and in the intact HIV reverse transcriptase (TA Steitz *et al.*, unpublished data). Although only one Mg^{2+} ion has been detected thus far in *E. coli* RNase H [33], the second metal ion might be missing because of its weak binding in the absence of substrates. The two metal ions chelated by carboxylates, as shown in the 3′ to 5′ exonuclease [7], would play a role in activating hydroxyl nucleophiles and stabilizing the penta-covalent phosphate intermediates of the nucleotidyl transfer reactions.

In summary, RNase H, RuvC and HIV integrase represent a structural and catalytic motif for a class of nucleases and polynucleotidyl transferases that use a general folding scheme and a similar catalytic center to accomplish a trans-esterification reaction on diverse substrates by what may turn out to be a common enzymatic mechanism. An additional example is provided by the crystal structure of the catalytic domain of phage Mu transposase, a member of a large integration enzyme family, which also contains an RNase H-like fold and a conserved catalytic triad (P Rice and K Mizuuchi, personal communication).

Prospects for anti-HIV drug design

Anti-HIV drug design has thus far targeted the viral protease and reverse transcriptase because of the advanced knowledge of their three-dimensional structures. Progress in designing highly potent drugs specifically against the viral protease is promising [34,35]. Until now, screening for inhibitors of HIV integrase has been carried out without the benefit of structural knowledge. In the light of the new atomic structure of the catalytic domain of HIV integrase and the suggestion of a possible catalytic mechanism, drugs specifically targeting the HIV integrase active site can be contemplated. The challenge with this enzyme, as with the other two HIV enzymes, will be to find an inhibitor that is not rendered inactive by mutation of the enzyme.

Acknowledgements: We thank Dr Davies' and Dr Morikawa's groups for providing us with the coordinates of HIV integrase and *E. coli* RuvC. This work was supported in part by NIH grant GM22778.

References

1. Ariyoshi, M., Vassylyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H. & Morikawa, K. (1994). Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*. *Cell* **78**, 1063–1072.
2. Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. & Davies, D.R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986.
3. Davies, J.F., Hostomska, Z., Hostomsky, Z., Jordon, S.R. & Matthews, D.A. (1991). Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science* **252**, 88–95.
4. Katayanagi, K., *et al.*, & Morikawa, K. (1990). Three-dimensional structure of ribonuclease H from *E. coli*. *Nature* **347**, 306–309.
5. Yang, W., Hendrickson, W.A., Crouch, R.J. & Satow, Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* **249**, 1398–1405.
6. Katayanagi, K., *et al.*, & Morikawa, K. (1992). Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.* **223**, 1029–1052.
7. Beese, L.S. & Steitz, T.A. (1991). Structural basis for the 3′–5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25–33.
8. Steitz, T.A. (1993). DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3**, 31–38.
9. Steitz, T.A. & Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* **90**, 6498–6502.
10. Goff, S.P. (1992). Genetics of retroviral integration. *Annu. Rev. Genet.* **26**, 527–544.
11. Whitcomb, J.M. & Hughes, S.H. (1992). Retroviral reverse transcription and integration: progress and problems. *Annu. Rev. Cell Biol.* **8**, 275–306.
12. Chow, S.A., Vincent, K.A., Ellison, V. & Brown, P.O. (1992). Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* **255**, 723–726.
13. Bushman, F.D., Engleman, A., Palmer, I., Wingfield, P. & Craigie, R. (1993). Domains of the integrase protein of human immunodeficiency virus type I responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl. Acad. Sci. USA* **90**, 3428–3432.
14. Mizuuchi, K (1992). Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.* **61**, 1011–1051.
15. Engleman, A., Bushman, F.D. & Craigie, R. (1993). Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.* **12**, 3269–3275.
16. Van Gent, D.C., Vink, C., Oude Groeneger, A.A.M. & Plasterk, R.H.A. (1993). Complementation between HIV integrase proteins mutated in different domains. *EMBO J.* **12**, 3261–3267.
17. West, S.C. (1992). Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem.* **61**, 603–640.
18. West, S.C. (1994). The processing of recombination intermediates: mechanistic insights from studies of bacterial proteins. *Cell* **76**, 9–15.
19. Dunderdale, H.J., Benson, F.E., Parsons, C.A., Sharples, G.J., Lloyd, R.G. & West, S.C. (1991). Formation and resolution of recombination intermediates by *E. coli* RecA and RuvC proteins. *Nature* **354**, 506–510.
20. Iwasaki, H., Takahagi, M., Shiba, T., Nakata, A. & Shinagawa, H. (1991). *Escherichia coli* RuvC protein is an endonuclease that resolves the Holliday structure. *EMBO J.* **10**, 4381–4389.
21. Bennett, R.J., Dunderdale, H.J. & West, S.C. (1993). Resolution of Holliday junctions by RuvC resolvase: cleavage specificity and DNA distortion. *Cell* **74**, 1021–1031.
22. Shah, R., Bennett, R.J. & West, S.C. (1994). Genetic recombination in *E. coli*: RuvC protein cleaves Holliday junctions at resolution hotspots *in vitro*. *Cell* **79**, 853–864.
23. Crouch, R.J. & Dirksen, M.-L. (1982). Ribonucleases H. In *Nucleases*. (Linn, S.M. & Roberts, R.J., eds), pp. 211–241, Cold Spring Harbor, NY.
24. Wintersberger, U. (1990). Ribonucleases H of retroviral and cellular origin. *Pharmacol. Ther.* **48**, 259–280.
25. Goff, S.P. (1990). Retroviral reverse transcriptase: synthesis, structure, and function. *AIDS* **3**, 817–831.
26. Fletterick, R.J., Bates, D.J. & Steitz, T.A. (1975). The structure of a yeast hexokinase monomer and its complex with substrates at 2.7 Å resolution. *Proc. Natl. Acad. Sci. USA* **72**, 38–42.
27. Flaherty, K.M., DeLuca-Flaherty, C. & McKay, D.B. (1990). Three-dimensional structure of the ATPase domain of a 70K heat-shock cognate protein. *Nature* **346**, 623–628.
28. Kabsch, W., Mannherz, H.G., Suck, D., Pai, E.F. & Holmes, K.C. (1990). Atomic structure of the actin:DNase I complex. *Nature* **347**, 37–44.
29. Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. & Steitz, T.A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–1790.
30. Wang, J., *et al.*, & Steitz, T.A. (1994). Structural basis of asymmetry in the human immunodeficiency virus type I reverse transcriptase heterodimer. *Proc. Natl. Acad. Sci. USA* **91**, 7242–7246.
31. Haruki, M., *et al.*, & Kanaya, S. (1994). Investigating the role of conserved residue Asp134 in *Escherichia coli* ribonuclease HI by site-directed random mutagenesis. *Eur. J. Biochem.* **220**, 623–631.
32. Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R. & Steitz, T.A. (1988). Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl. Acad. Sci. USA* **85**, 8924–8928.
33. Katayanagi, K., Okumura, M. & Morikawa, K. (1993). Crystal structure of *Escherichia coli* RNase HI in complex with Mg^{2+} at 2.8 Å resolution: proof for a single Mg^{2+}-binding site. *Proteins* **17**, 337–346.
34. Erickson, J. & Kempf, D. (1994). Structure-based design of symmetric inhibitors of HIV protease. *Arch. Virol.* **9**, 19–29.
35. Lam, P.Y.S., *et al.*, & Erickson-Vitanen, S. (1994). Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **263**, 380–384.
36. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
37. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
38. Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**, 922–923.

Wei Yang, Department of Molecular Biophysics and Biochemistry, and Thomas A Steitz, Departments of Molecular Biophysics and Biochemistry, and Chemistry, Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA.