Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75

Peter Cherepanov^{*†}, Andre L. B. Ambrosio^{*§}, Shaila Rahman^{*}, Tom Ellenberger[‡], and Alan Engelman^{*†¶}

^{*}Department of Cancer Immunology and AIDS, Dana–Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; [†]Department of Pathology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115; [‡]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115; and [§]Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, CEP 13560-970, São Carlos–São Paulo, Brazil

Edited by Robert M. Stroud, University of California, San Francisco, CA, and approved September 28, 2005 (received for review August 10, 2005)

Integrase (IN) is an essential retroviral enzyme, and human transcriptional coactivator p75, which is also referred to as lens epithelium-derived growth factor (LEDGF), is the dominant cellular binding partner of HIV-1 IN. Here, we report the crystal structure of the dimeric catalytic core domain of HIV-1 IN complexed to the IN-binding domain of LEDGF. Previously identified LEDGF hotspot residues anchor the protein to both monomers at the IN dimer interface. The principal structural features of IN that are recognized by the host factor are the backbone conformation of residues 168–171 from one monomer and a hydrophobic patch that is primarily comprised of α -helices 1 and 3 of the second IN monomer. Inspection of diverse retroviral primary and secondary sequence elements helps to explain the apparent lentiviral tropism of the LEDGF-IN interaction. Because the lethal phenotypes of HIV-1 mutant viruses unable to interact with LEDGF indicate that IN function is highly sensitive to perturbations of the structure around the LEDGF-binding site, we propose that small molecule inhibitors of the protein–protein interaction might similarly disrupt HIV-1 replication.

integration | structure | retrovirus | transcription | host factor

Like all retroviruses, HIV-1 must integrate a reverse-transcribed copy of its viral RNA genome into a host cell chromosome to establish a productive infection. Integration is mediated by the viral integrase (IN) protein acting on the DNA attachment sites at the ends of the linear reverse transcript. IN acts within the context of a higher-order preintegration complex (PIC) that is derived from the core of the infecting virion. IN catalyzes two sequential reactions, initially removing 3' terminal GT nucleotides from both ends of HIV-1 cDNA. After nuclear entry, IN inserts the processed 3' termini into opposing strands of chromosomal DNA. Repair of single-strand gaps by host cell enzymes completes the integration process (for a review, see ref. 1). In addition to IN, other viral and numerous cellular proteins are believed to play auxiliary roles in HIV-1 PIC assembly, transport into the nucleus, and targeting integration into transcriptionally active regions of chromatin (2–4).

HIV-1 IN is a 32-kDa protein composed of three structural domains (5, 6). The N-terminal domain (NTD; residues 1–49) is a helical bundle stabilized by the coordination of a single zinc atom (7). The catalytic core domain (CCD; residues 50–212) belongs to a superfamily of DNA/RNA strand transferases/nucleases that in addition to retroviral INs includes divergent bacterial transposases, Holiday junction resolvase RuvC, and RNase H (8, 9). Several independent retroviral IN CCD structures reveal a similar dimer interface, strongly arguing that it is physiologically relevant. The CCD contains the essential active site residues, Asp-64, Asp-116, and Glu-152 in HIV-1, collectively referred to as the DDE triad. The carboxyl-terminal domain (residues 213–288), which is the least conserved among divergent retroviruses, possesses an Src homology 3-like fold (10, 11). Each domain contributes to HIV-1 IN multimerization and is essential for 3' processing and DNA strand transfer activities (6, 12–17). Although the functional multimeric state of IN *in vivo*

is not known, when extracted from human cells, the HIV-1 protein is a stable tetramer (18).

Ectopically expressed HIV-1 IN accumulates in the nuclei of human cells (19, 20), associates with chromatin (21), and forms a tight complex with endogenous transcriptional coactivator p75 (2, 18), which is also referred to as lens epithelium-derived growth factor (LEDGF). Binding to LEDGF appears to account for the characteristic intracellular distribution of IN. RNA interference-mediated depletion of LEDGF or overexpression of a nuclear localization-defective mutant of LEDGF disrupted nuclear/chromosomal localization of HIV-1 IN (22–24). Conversely, the interaction-deficient IN mutant Q168A was excluded from condensed chromosomes (25). Furthermore, interaction with LEDGF protected HIV-1 IN from degradation through the ubiquitin-proteasome pathway (26). So far, LEDGF was demonstrated to bind the INs from HIV-1, HIV-2, and the divergent lentivirus feline immunodeficiency virus (FIV), but not the INs from the gammaretrovirus Moloney murine leukemia virus, alpharetrovirus Rous sarcoma virus (RSV), or deltaretrovirus human T cell leukemia virus (2, 18, 24, 27). Thus, binding to LEDGF seems to be a conserved property of lentiviral INs. Consistent with a potential role in lentiviral replication, LEDGF cofractionated with PICs isolated from HIV-1- or FIV-infected cells (24).

LEDGF binds HIV-1 IN via a small, approximate 80-residue IN-binding domain (IBD) within its C-terminal region (28, 29). The IBD is both necessary and sufficient for the interaction with HIV-1 IN (28). The three-dimensional structure of the IBD was recently determined by NMR spectroscopy, revealing an α -helical domain that is topologically similar to a pair of HEAT repeats (30). To differentiate it from HEAT repeat proteins that tend to be large and contain numerous repeated elements (31), the IBD was classified as a pseudo-HEAT repeat analogous topology (PHAT) domain, similar to the PHAT domain in *Drosophila* Smaug protein (30, 32). Residues Ile-365, Asp-366, and Phe-406, located in the interhelical loops within the HEAT repeats, play critical roles in HIV-1 IN recognition (30).

The CCD of IN possesses the main determinants for interacting with LEDGF, although the NTD increased the affinity of the interaction (22). Mutations within the CCD that disrupted the interaction collectively pointed to residues 165–173 as an approximate location for the interaction site on IN (2, 25, 30). The failure of these IN mutants to bind LEDGF correlated with catastrophic preintegration blocks of the corresponding mutant

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CCD, catalytic core domain; IN, integrase; IBD, IN-binding domain; LEDGF, lens epithelium-derived growth factor; NTD, N-terminal domain; PDB, Protein Data Bank; PIC, preintegration complex; RSV, Rous sarcoma virus.

Data deposition: Atomic coordinates and diffraction data have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 2B4J).

[¶]To whom correspondence should be addressed. E-mail: alan.engelman@dfci.harvard.edu.

© 2005 by The National Academy of Sciences of the USA

Fig. 2. Molecular mechanism of the IN-LEDGF interaction. (A) The overall structure of the CCD-IBD complex. IN chains A and B are colored blue and green, respectively; the IBD subunits are violet. The side chains of the DDE catalytic triad are shown as yellow sticks. (B) Key intermolecular contacts at the CCD-IBD interface. Selected residues are shown as sticks. The water molecule hydrogen-bonded to main-chain carbonyl groups of LEDGF residue Ile-365 and IN residue Thr-125 is shown as a red sphere. Hydrogen bonds discussed in the text and the salt bridge between IN residues Glu-69 and Arg-166 are indicated by dotted lines. (C) The pocket at the CCD dimer interface. LEDGF hotspot residues Ile-365, Asp-366, and Phe-406 are shown as sticks (Upper) or in space-fill mode (Lower). The IN subunits are shown as semitransparent surfaces. Selected IN residues are indicated. (D) Sequence alignment of HIV-1, HIV-2, and feline immunodeficiency virus (FIV) INs. Identical residues are white on red background; residues with conserved properties are bold on yellow background. Residue numbering, secondary structure elements, and the position of the $\alpha 4/5$ connector in HIV-1 IN are shown above the alignment; structural elements are colorized as in A–C. Open circles and filled boxes under the alignment indicate residues that make contacts to the LEDGF IBD through side-chain and main-chain atoms, respectively. The alignment was printed by using ESPRIT-2.2 (56).

dran plot, and the remaining 5% were in additionally allowed regions. Figs. 1 and 2C and Fig. 4, which is published as supporting information on the PNAS web site, were drawn with PYMOL (W. L. DeLano, www.pymol.org). Figs. 2A and B and 3 were prepared with the program MOLMOL (41).

Results and Discussion

Crystallization and Structure Determination of the IN-LEDGF Complex.

The CCD of HIV-1 IN (residues 50–212) is both necessary and sufficient for the interaction of IN with LEDGF (22). Furthermore, the F185K mutation, which greatly improved the solubility of recombinant HIV-1 IN protein (42), did not significantly affect the binding affinity of LEDGF (30). Consistent with these observations, the results of sedimentation equilibrium and gel

filtration experiments revealed that IN^{50–212}(F185K) readily formed a complex with the LEDGF IBD (data not shown). Crystals of the IN^{50–212}(F185K)-LEDGF^{347–442} complex diffracted x-rays to ≈ 2 -Å resolution at a synchrotron light source. The structure of the complex was solved through molecular replacement and refined to 2.02 Å with an *R* factor of 18.0% and an *R*_{free} of 22.6% (Table 1). A fragment of the final electron density map showing selected residues at the CCD-IBD interface is shown in Fig. 1. The color scheme of blue and green for chains A and B of the IN dimer, respectively, and violet for the LEDGF IBD, is used throughout this work.

The asymmetric unit of the crystals contained a dimer of IN CCDs (chains A and B) and a pair of LEDGF IBD molecules (chains C and D) bound at the CCD dimer interface (Fig. 2A).

Gln-168 are expected to be important for the structural integrity of the $\alpha 4/5$ connector. On the other hand, Val-165 and Leu-172 are buried within the IN core structure; mutating either of them to the smaller Ala residue is also expected to compromise local structure.

Typically, formation of protein complexes is not accompanied by major structural changes of the interacting proteins (44). Similarly, binding of the LEDGF IBD did not significantly alter the structure of the CCD (Fig. 3A). Rotation of the Trp-131 side chain was the only significant adjustment that occurred upon complex formation. In previous CCD structures that retained wild-type Trp at position 131, this exposed residue was involved in crystal packing interactions and its orientation varied from crystal to crystal (8, 45). On the other hand, the conformation of the $\alpha 4/5$ connector was preserved among all reported HIV-1 IN structures and in simian immunodeficiency virus IN (46). The LEDGF-binding site of HIV-1 IN therefore appears to adopt a stable conformation closely resembling that in the complex with LEDGF. In contrast, RSV IN, which does not bind to LEDGF (27), has an extended loop (RSV IN residues 176–182) connecting $\alpha 4$ and $\alpha 5$ that is pulled away from the potential interaction site (Fig. 3B). The conformation of this loop is preserved in the known RSV IN structures with accompanying low temperature factors of loop residues (47, 48), arguing that it reflects the native state of RSV IN. Several solvent-exposed aromatic and hydrophobic residues occupy the cleft between $\alpha 4$ and $\alpha 5$ of RSV IN (Fig. 3B). The shape and chemical features of this cleft raise the possibility of an interaction with some alternate binding partner.

Implications for IN Function and Inhibitor Design. LEDGF enhanced the *in vitro* enzymatic activity, solubility, and DNA-binding activity of HIV-1 IN (18, 27, 28). The interaction with LEDGF is essential for the association of ectopically expressed IN with chromatin (22, 24) and the stability of the viral protein in human cells (22, 26). Moreover, mutations within the IN $\alpha 4/5$ connector, which abolished interactions with LEDGF, cause catastrophic viral replication defects (25, 30, 49, 50). Intriguingly, the mutant INs retained enzymatic activity *in vitro* and when tested, efficiently cross-complemented the infectivity defects of IN active-site mutant viruses in cells (25, 49, 50). Although the crystal structure of the IN CCD–LEDGF IBD complex elucidates why these mutant proteins fail to bind LEDGF, we cannot unambiguously explain the replication deficiency of the corresponding mutant viruses. One obvious possibility is that the interaction with LEDGF is essential for HIV-1 replication (25). However, RNA interference-mediated depletion of endogenous LEDGF did not yield a concomitant reduction in HIV-1 replication (24), arguing against, although not ruling out, this hypothesis. Alternatively, the LEDGF-binding site on IN could mediate other essential protein–protein interactions during HIV-1 infection, such as contacts between IN protomers within higher-order complexes during PIC assembly.

Notwithstanding the final outcome of this riddle, the phenotypes of IN mutant viruses carrying changes in the $\alpha 4/5$ connector underscore the fact that the structural organization of the LEDGF-binding site has a profound effect on HIV-1 replication. The LEDGF-binding site pocket accommodates IBD residues Ile-365 and Asp-366 (Fig. 2C). We speculate that a small molecule inhibitor designed to fit into this pocket is likely to induce potent HIV-1 replication defects, similar to those observed for mutant viruses unable to interact with LEDGF. Persuasively, the potential inhibitor target site contains both strong nonpolar and hydrogen-bonding components (Fig. 2B). It might also be possible for an inhibitor with appropriate hydrogen-bonding groups to replace the water molecules buried at the LEDGF–IN interface. Because LEDGF binds to full-length IN even when the enzyme is in complex with its cognate DNA substrate (28, 30), we further speculate that the LEDGF-binding site may be available for drug binding during the early steps of viral replication.

Soaking IN CCD crystals in the presence of tetraphenylarsonium and its hydroxylated derivative resulted in binding of the compounds at the CCD dimer interface at a position that directly overlapped the LEDGF-binding site (51). Furthermore, acetylated analogs of L-chicoric acid inhibited HIV-1 IN activity (52) and modified Lys-173 (53), which is adjacent to the LEDGF-binding site. Because of the relatively low affinity of these compounds for IN, they did not interfere with formation of the IN–LEDGF complex (P.C. and A.E., unpublished observations). However, these known examples of small molecules that bind at or close to the LEDGF-interaction site strongly suggest that this site is suitable for interaction with drug-like compounds. It may be possible to identify molecules that bind to the LEDGF-binding site with high affinity and antagonize the protein interaction by using simple *in vitro* assays (22, 24, 28). Because mutations of the LEDGF-binding site did not drastically compromise the *in vitro* catalytic activities of HIV-1 IN (25, 49, 50), enzymatic assays are less likely to identify this class of drug. Recent successes in identifying highly specific small molecule inhibitors of protein–protein interactions (54, 55) encourage efforts to exploit virus–host protein interactions for the development of novel antiretroviral drugs.

We thank R. Craigie (National Institutes of Health, Bethesda), D. Hazuda (Merck Research Laboratories), and M. Kvaratskhelia (Ohio State University, Columbus) for sharing reagents; H. Aihara, L. Briebe, and T. Biswas (Harvard Medical School) for generous advice and helpful discussions; N. Vandegraaff, G. Maertens, and R. Lu for critical reading of the manuscript; and the staff of the X26C beam line at the National Synchrotron Light Source for assistance in x-ray data collection. This work was supported by National Institutes of Health Grants AI39394 and AI62249 (to A.E.) and GM59902 (to T.E.), and state of São Paulo Research Foundation Grant 03/00231-0 (to A.L.B.A.).

- Craigie, R. (2002) in *Mobile DNA II*, eds. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M. (Am. Soc. Microbiol., Washington, DC), pp. 613–630.
- Turlure, F., Devroe, E., Silver, P. A. & Engelman, A. (2004) *Front. Biosci.* **9**, 3187–3208.
- Engelman, A. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 1275–1276.
- Bushman, F. D. (2003) *Cell* **115**, 135–138.
- Engelman, A. & Craigie, R. (1992) *J. Virol.* **66**, 6361–6369.
- Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 3428–3432.
- Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1997) *Nat. Struct. Biol.* **4**, 567–577.
- Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994) *Science* **266**, 1981–1986.
- Rice, P. A. & Baker, T. A. (2001) *Nat. Struct. Biol.* **8**, 302–307.
- Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1995) *Biochemistry* **34**, 9826–9833.
- Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R. & Hard, K. (1995) *Nat. Struct. Biol.* **2**, 807–810.
- Vink, C., Oude Groeneger, A. M. & Plasterk, R. H. (1993) *Nucleic Acids Res.* **21**, 1419–1425.
- Engelman, A., Bushman, F. D. & Craigie, R. (1993) *EMBO J.* **12**, 3269–3275.
- Hickman, A. B., Palmer, I., Engelman, A., Craigie, R. & Wingfield, P. (1994) *J. Biol. Chem.* **269**, 29279–29287.
- Zheng, R., Jenkins, T. M. & Craigie, R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13659–13664.
- Jenkins, T. M., Engelman, A., Ghirlando, R. & Craigie, R. (1996) *J. Biol. Chem.* **271**, 7712–7718.
- Lee, S. P., Xiao, J., Knutson, J. R., Lewis, M. S. & Han, M. K. (1997) *Biochemistry* **36**, 173–180.
- Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E. & Debyser, Z. (2003) *J. Biol. Chem.* **278**, 372–381.

19. Petit, C., Schwartz, O. & Mammano, F. (1999) *J. Virol.* **73**, 5079–5088.
20. Pluymers, W., Cherepanov, P., Schols, D., De Clercq, E. & Debyser, Z. (1999) *Virology* **258**, 327–332.
21. Cherepanov, P., Pluymers, W., Claeys, A., Proost, P., De Clercq, E. & Debyser, Z. (2000) *FASEB J.* **14**, 1389–1399.
22. Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z. & Engelborghs, Y. (2003) *J. Biol. Chem.* **278**, 33528–33539.
23. Maertens, G., Cherepanov, P., Debyser, Z., Engelborghs, Y. & Engelman, A. (2004) *J. Biol. Chem.* **279**, 33421–33429.
24. Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M. & Poeschla, E. M. (2004) *J. Virol.* **78**, 9524–9537.
25. Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempe, D., Vandekerckhove, L., Moisan, F., Ben-Slama, L., Witvrouw, M., *et al.* (2005) *J. Biol. Chem.* **280**, 25517–25523.
26. Llano, M., Delgado, S., Vanegas, M. & Poeschla, E. M. (2004) *J. Biol. Chem.* **279**, 55570–55577.
27. Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F. & Debyser, Z. (2005) *J. Biol. Chem.* **280**, 17841–17847.
28. Cherepanov, P., Devroe, E., Silver, P. A. & Engelman, A. (2004) *J. Biol. Chem.* **279**, 48883–48892.
29. Vanegas, M., Llano, M., Delgado, S., Thompson, D., Peretz, M. & Poeschla, E. (2005) *J. Cell Sci.* **118**, 1733–1743.
30. Cherepanov, P., Sun, Z. Y., Rahman, S., Maertens, G., Wagner, G. & Engelman, A. (2005) *Nat. Struct. Mol. Biol.* **12**, 526–532.
31. Andrade, M. A., Petosa, C., O'Donoghue, S. I., Muller, C. W. & Bork, P. (2001) *J. Mol. Biol.* **309**, 1–18.
32. Green, J. B., Gardner, C. D., Wharton, R. P. & Aggarwal, A. K. (2003) *Mol. Cell* **11**, 1537–1548.
33. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276**, 307–326.
34. CCP4 (1994) *Acta Crystallogr. D* **50**, 760–763.
35. Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R. & Davies, D. R. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 9150–9154.
36. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54**, 905–921.
37. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) *Acta Crystallogr. D* **53**, 240–255.
38. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003) *Methods Enzymol.* **374**, 300–321.
39. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47**, 110–119.
40. Emsley, P. & Cowtan, K. (2004) *Acta Crystallogr. D* **60**, 2126–2132.
41. Koradi, R., Billeter, M. & Wuthrich, K. (1996) *J. Mol. Graphics* **14**, 51–55, 29–32.
42. Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlando, R., Davies, D. R. & Craigie, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6057–6061.
43. Wang, J. Y., Ling, H., Yang, W. & Craigie, R. (2001) *EMBO J.* **20**, 7333–7343.
44. Lo Conte, L., Chothia, C. & Janin, J. (1999) *J. Mol. Biol.* **285**, 2177–2198.
45. Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C. & Mikol, V. (1998) *J. Mol. Biol.* **282**, 359–368.
46. Chen, Z., Yan, Y., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B., Witmer, M., Felock, P., Wolfe, A., Sardana, V., *et al.* (2000) *J. Mol. Biol.* **296**, 521–533.
47. Lubkowski, J., Dauter, Z., Yang, F., Alexandratos, J., Merkel, G., Skalka, A. M. & Wlodawer, A. (1999) *Biochemistry* **38**, 13512–13522.
48. Yang, Z. N., Mueser, T. C., Bushman, F. D. & Hyde, C. C. (2000) *J. Mol. Biol.* **296**, 535–548.
49. Bouyac-Bertoia, M., Dvorin, J. D., Fouchier, R. A., Jenkins, Y., Meyer, B. E., Wu, L. I., Emerman, M. & Malim, M. H. (2001) *Mol. Cell* **7**, 1025–1035.
50. Lu, R., Limon, A., Devroe, E., Silver, P. A., Cherepanov, P. & Engelman, A. (2004) *J. Virol.* **78**, 12735–12746.
51. Molteni, V., Greenwald, J., Rhodes, D., Hwang, Y., Kwiatkowski, W., Bushman, F. D., Siegel, J. S. & Choe, S. (2001) *Acta Crystallogr. D* **57**, 536–544.
52. Lin, Z., Neamati, N., Zhao, H., Kiryu, Y., Turpin, J. A., Aberham, C., Strebel, K., Kohn, K., Witvrouw, M., Pannecouque, C., *et al.* (1999) *J. Med. Chem.* **42**, 1401–1414.
53. Shkriabai, N., Patil, S. S., Hess, S., Budihas, S. R., Craigie, R., Burke, T. R., Jr., Le Grice, S. F. & Kvaratskhelia, M. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 6894–6899.
54. Vassilev, L. T., Vu, B. T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., *et al.* (2004) *Science* **303**, 844–848.
55. Lepourcelet, M., Chen, Y. N., France, D. S., Wang, H., Crews, P., Petersen, F., Bruseo, C., Wood, A. W. & Shivdasani, R. A. (2004) *Cancer Cell* **5**, 91–102.
56. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999) *Bioinformatics* **15**, 305–308.