been mapped precisely, it was shown that under normal growth conditions, the switch occurs between the origin and nucleotide 245 (*15*). In our constructions, the adduct is located at nucleotide position 363 and 432 in the leading and lagging strand constructions, respectively. Therefore, replication past the AAF adducts in our constructions is likely to be Pol III–mediated.

Alternatively, mutations may be generated at similar rates in both strands, but a repair mechanism may act preferentially on the leading strand after replication. However, we found that the bias in mutation frequency between the two strands was not affected when the long patch mismatch repair (LPMR) pathway was inactivated (Table 1, *mutS* allele) (*16*). Although this result shows that LPMR does not cause the bias, it does not rule out the possibility that an unknown postreplication correction mechanism contributes to the bias.

No difference in fidelity between leading and lagging strand synthesis was found for spontaneous base substitution errors and deletions in a human HeLa cell extract in vitro: Exonucleolitic proofreading occurred with the same efficiency in both strands (*17*). On the other hand, the lagging strand in *E. coli* is more susceptible than the leading strand to replication-dependent deletions triggered by palindromes (*18*).

REFERENCES AND NOTES

1. R. K. Kaderlik *et al.*, *Proc. Am. Assoc. Cancer Res.* **34**, 152 (March 1993).
2. R. P. P. Fuchs, N. Schwartz, M. P. Daune, *Nature* **294**, 657 (1981).
3. N. Koffel-Schwartz *et al.*, *J. Mol. Biol.* **177**, 33 (1984).
4. I. B. Lambert, R. L. Napolitano, R. P. P. Fuchs, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1310 (1992).
5. D. Burnouf, P. Koehl, R. P. P. Fuchs, *ibid.* **86**, 4147 (1989).
6. G. Maenhaut-Michel, R. Janel-Bintz, R. P. P. Fuchs, *Mol. Gen. Genet.* **235**, 373 (1992).
7. J. Vieira and J. Messing, *Gene* **19**, 259 (1982).
8. J.-I. Tomizawa, Y. Sakakibara, T. Kakefuda, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2260 (1974).
9. N. Koffel-Schwartz and R. P. P. Fuchs, *Mol. Gen. Genet.* **215**, 306 (1989).
10. X. Veaute and R. P. P. Fuchs, unpublished results.
11. A. Garcia, I. B. Lambert, R. P. P. Fuchs, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
12. K. O. Johanson and C. S. McHenry, *J. Biol. Chem.* **259**, 4589 (1984).
13. H. Maki, S. Maki, A. Kornberg, *ibid.* **263**, 6570 (1988).
14. K. T. Marians, *Annu. Rev. Biochem.* **61**, 67 (1992).
15. K. Backman *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **43**, 69 (1978).
16. R. S. Lahue, K. G. Au, P. Modrich, *Science* **245**, 160 (1989).
17. J. D. Roberts, D. C. Thomas, T. A. Kunkel, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3465 (1991).
18. T. Q. Trinh and R. R. Sinden, *Nature* **352**, 544 (1991).
19. The plasmids used to construct the different gapped duplexes were obtained by replacement of the Eco RI–Hind III fragment of the polylinker region of pUC8 by sequences 5′-AATTCAGTCGACTA, 5′-AATTCAGTCATACCCGGGACAGACTA, and 5′-AATTCAGTCATCACCGGCGCCACAGACTA, generating plasmids pUC-H(I), pUC-3G(I), and pUC-N(I), respectively. Similarly, the replacement of the Eco RI–Hind III fragment in pUC8(L) by the same sequences generated plasmids pUC-H(L), pUC-3G(L), and pUC-N(L), respectively.
20. The monomodified oligonucleotides (5′-ATCACCGGCG^{AAF}CCACA and 5′-ATACCCGGG^{AAF}ACATC) were used for the constructions at the Nar I and GGG run context, respectively. The monomodified oligonucleotides were purified by high-performance liquid chromatography and characterized as previously described (*4*, *5*).
21. We thank J. Lindsley for comments on the manuscript, I. B. Lambert and R. L. Napolitano for helpful discussions, and I. Pinet for technical assistance. Supported in part by a grant from the Human Frontier Science Program and from the Association pour la Recherche sur le Cancer.

26 February 1993; 3 June 1993

Signal Sequence Trap: A Cloning Strategy for Secreted Proteins and Type I Membrane Proteins

Kei Tashiro, Hideaki Tada, Ralf Heilker, Michio Shirozu, Toru Nakano, Tasuku Honjo*

A method was developed to clone, without the use of specific functional assays, complementary DNAs (cDNAs) that carry specific amino-terminal signal sequences, such as those encoding intercellular signal-transducing molecules and receptors. The vector used in this system directed the cell surface expression of interleukin-2 receptor fusion proteins when inserts with signal sequences were cloned in-frame with the correct orientation. An expression cDNA library was constructed from a bone marrow stromal cell line, which contained 5′ portion–enriched cDNAs (the average size was 400 base pairs). Two cDNAs that encoded putative cytokine molecules, stromal cell–derived factor–1α (SDF-1α) and SDF-1β, which belong to the intercrine–macrophage inflammatory protein superfamily, were cloned.

Complex interactions between stromal cells and hematopoietic stem (or progenitor) cells are required for hematopoiesis, during which many unknown molecules should be involved in intercellular signal transduction. However, it is difficult and tedious to clone either cDNAs for many different growth factors or adhesion molecules required for transducing specific intercellular signals by establishing a bioassay system specific to each molecule and then measuring the biological activities of unknown molecules. To overcome this problem, we developed a method for trapping signal sequences as a strategy for cloning cDNA for growth factors and type I integral membrane proteins, taking advantage of the specific NH_2-terminal signal sequences that most precursors for secreted factors and transmembrane molecules carry and that are within 400 base pairs (bp) of the 5′ termini of the mRNA (*1*).

The secretory signal sequence leads the NH_2-terminus of the de novo synthesized protein into the endoplasmic reticulum or outside the cell. We constructed the pcDL-SRα-Tac(3′) vector that could direct the cell surface expression of Tac (α chain of the human interleukin-2 receptor) fusion proteins when inserts with signal sequences were cloned in-frame with the correct orientation (*2*). The fusion protein expressed on plasma membranes was easily detected by antibodies to Tac (anti-Tac) (*3*). The pcDL-SRα-Tac(3′) vector had cloning sites between the SRα promoter, a fusion promoter composed of the simian virus 40 early promoter and the R-U5 segment of human T cell leukemia virus type I, and the coding sequence (without a signal sequence) of Tac cDNA (*4*) (Fig. 1A). Using a fragment containing a signal sequence of human granulocyte colony-stimulating factor (hG-CSF) (*5*) as a positive control [hG-CSF(5′)] and a 5′ fragment of the human retinoic acid receptor α (hRAR) (*6*) as a negative control [hRAR(5′)] (*7*), we tested whether the pcDL-SRα-Tac(3′) vector system allowed the transient cell surface expression of the Tac epitope on the transfected cells (Fig. 1B). The Tac antigen was detected by anti-Tac on the surface of COS-7 cells transfected with the hG-CSF(5′)–containing plasmid, which indicates that the fusion protein consisting of hG-CSF(5′) and Tac(3′) was expressed on the cell surface. In contrast, when the hRAR(5′)-containing plasmid was used, the Tac antigen was not detected on the cell surface. When the sensitivity of the detection was tested by serial dilution experiments of the hG-CSF(5′)–containing plasmid with the hRAR(5′)-containing plasmid, one positive clone in a pool of 100 negative clones was easily detected.

We then constructed an expression cDNA library containing 5′ portion–en-

Department of Medical Chemistry, Kyoto University Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto 606, Japan.

*To whom correspondence should be addressed.

riched cDNA (the average size was 400 bp) of a mouse bone marrow stromal cell line (ST-2), which is capable of supporting the growth and differentiation of hematopoietic cells (*8*, *9*). To enrich for 5′ terminal 400-bp cDNAs, we added a deoxycytosine (dC) tail at the 5′ end of the cDNA. This cDNA was primed with a specific primer that contained polydeoxyguanosine (dG) and Eco RI recognition sequences for second strand synthesis, and the product was sonicated, ligated with a Sac I linker, and amplified by polymerase chain reaction (PCR) (*10*, *11*) (Fig. 1A). The amplified fragments were digested with Eco RI and Sac I and ligated to the pcDL-SRα-Tac(3′) vector in the same orientation with the Tac cDNA to obtain an expression library that contained 5′ termini–enriched ST-2 cDNAs of 400 bp.

After transformation, 36 individual colonies were replated on a 9-cm agar plate in a matrix format (six rows by six lines) and assigned to one pool. Plasmid DNAs of each pool were used for transfection of COS-7 cells, and fusion proteins expressed on the cell surface were microscopically detected by immunostaining with anti-Tac. After a pool was identified as positive with fluorescent microscopy, 12 smaller pools consisting of six individual clones in each row or line of the matrix were tested to identify a single positive clone (*12*). We screened about 600 clones and obtained 6 immunologically positive cDNA clones. Nucleotide sequencing showed that they all contained 5′ terminal ST-2 cDNA sequences, the length of which varied from 350 to 420 bp. The sequence information was used for analysis of hydrophobicity (*13*) (Fig. 2) and compared with sequences in the GenBank database for homology. Clone G2a, which has an open reading frame (ORF) of 97 residues, and clone M5b, with an ORF of 95 residues, turned out to be identical to the 5′ part of the cDNA encoding the JE cytokine of the intercrine β–macrophage inflammatory protein (MIP) cytokine family (*14*, *15*). Clones C5c (with an ORF of 84 residues), D3b (with an ORF of 103 residues), and H5a (with an ORF of 83 residues) were identical to the 5′ part of the cDNA for the Gla protein, one of the major bone matrix proteins (*16*). The JE and Gla proteins are secreted and bear typical NH_2-terminal signal sequences (Fig. 2, A and B). The other clone, D2d, encoded 68 residues of an unknown sequence that had a hydrophobic profile characteristic of a signal sequence (Fig. 2C) and four cysteine residues positioned identically to those of the intercrine α–MIP cytokine family (*15*), including the stem cell inhibitor (SCI) factor (*17*) (Fig. 3).

We probed an oligo(dT)-primed ST-2 cDNA library with clone D2d, obtained two different full-length cDNA clones, stromal cell–derived factor (SDF)–1α and SDF-1β, and determined their cDNA sequences. The SDF-1α cDNA was 1797 bp long and had an ORF of 89 residues. The SDF-1β cDNA was 3152 bp long and had an ORF of 93 residues. The SDF-1α and SDF-1β cDNAs had identical sequences in the 5′ untranslated (UT) region and the ORF, whereas the ORF in SDF-1β had an additional four residues. No significant identity was found in the 3′ UT region. The ORF of SDF-1α and SDF-1β had no hydrophobic regions other than the 5′ putative signal sequence contained in D2d. No glycosylation sites nor retention signals to endoplasmic reticulum were found (*18*). The amino acid sequences of the deduced mature proteins of SDF-1α and SDF-1β are 32% identical to those of human interleukin-8 and mouse MIP-1α (also known as

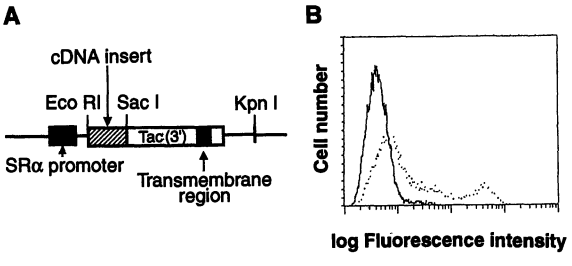

Fig. 1. Strategy of the signal sequence trap method. (**A**) Schematic diagram of the pcDL-SRα-Tac(3′) vector containing a 5′ terminal cDNA fused to the Tac(3′) sequences. The pcDL-SRα-Tac(3′) vector was constructed from pcDL-SRα (*26*) as described (*2*). (**B**) Detection with anti-Tac of a heterologous signal sequence as a fusion protein with the Tac protein. The pcDL-SRα–hG-CSF(5′)–Tac(3′) and pcDL-SRα-hRAR(5′)-Tac(3′) plasmids were used as positive and negative control plasmids, respectively. The flow cytometry profiles are those of COS-7 cells transfected with positive (dotted lines) and negative (solid lines) control plasmids. Positive and negative control fragments in pcDL-SRα-Tac(3′) were used for transfection of COS-7 cells by the DEAE-dextran method (*12*). After 48 to 60 hours, the cells were harvested and immunostained with anti-Tac (*3*). Flow cytometry was performed with a FACScan (Becton Dickinson).

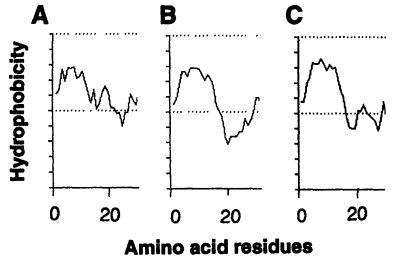

Fig. 2. Hydropathy profiles of clones obtained by the signal sequence trap method. Hydropathy profiles of NH_2-terminal portions of positive clones are shown. (**A**) G2a and M5b. (**B**) C5c, D3b, and H5a. (**C**) D2d. All the Tac-positive clones obtained from the signal sequence–trapped cDNA library in the pcDL-SRα-Tac(3′) vector were sequenced in both directions with the automated fluorescence-based sequencing system (Model 373A, Applied Biosystems). Hydropathy profiles were analyzed with a computer program based on that previously described (*13*).

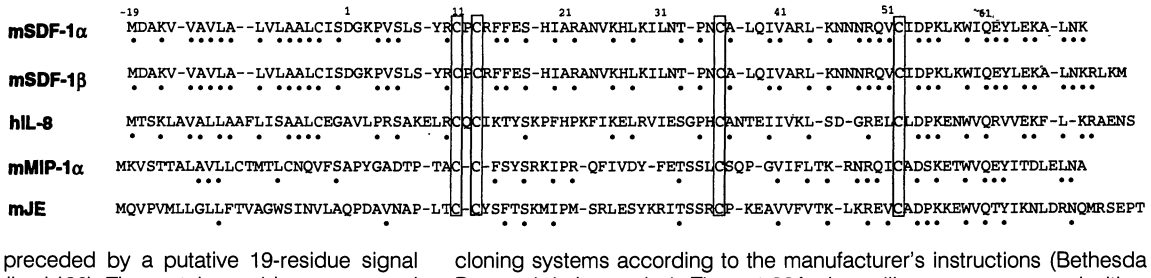

Fig. 3. Alignment of the deduced amino acid sequences of the mature proteins of SDF-1α and SDF-1β (mSDF-1α and mSDF-1β) together with those of murine intercrine-MIP cytokine members (mMIP-1α and mJE) (*27*). Amino acids are numbered starting at Asp[1] preceded by a putative 19-residue signal sequence predicted as described (*28*). The cysteine residues conserved in the CXC or CC family are boxed (*14*, *15*) (X is any amino acid). Amino acids identical to those in SDF-1α or SDF-1β are shown by dots. Dashes indicate gaps introduced to maximize homology. The GenBank accession numbers of the mSDF-1α and mSDF-1β sequences are L12029 and L12030, respectively. From 5 μg of ST-2 mRNA, an oligo(dT)-primed cDNA library was constructed with Super Script cDNA synthesis and cloning systems according to the manufacturer's instructions (Bethesda Research Laboratories). The λgt 22A phage library was screened with a D2d probe. Positive clones with 1.8- and 3.2-kb inserts were named SDF-1α and SDF-1β, respectively. After recloning into a Bluescript II plasmid, deletion mutants of these clones were made in both directions with a deletion kit containing exo III/mung bean nuclease (Stratagene). Deletion mutants of SDF-1α and SDF-1β were sequenced by the automated fluorescence-based sequencing system.

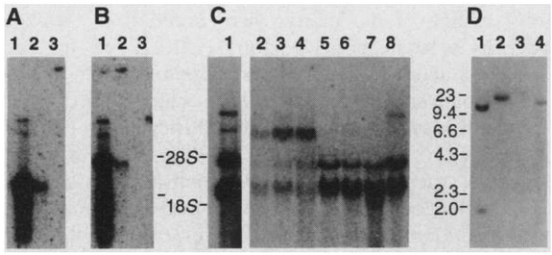

Fig. 4. Characterization of mRNA and gene coding for SDF-1α and SDF-1β. Northern hybridization analysis of SDF-1α mRNA (**A**) and SDF-1β mRNA (**B**). The following cells were used to prepare polyadenylate [poly(A)] RNA: lanes 1, ST-2; lanes 2, mouse fibroblast cell line NIH 3T3 (*24*); and lanes 3, mouse hematopoietic progenitor cell line LyD9 (*9, 19*). In (A), the 3′ UT region of SDF-1α cDNA (nucleotides 1005 to 1613) was used as a probe. In (B), the 3′ UT region of SDF-1β cDNA (nucleotides 2312 to 3068) was used as a probe. (**C**) Northern hybridization analysis of SDF-1α and SDF-1β mRNA. Lane 1, ST-2 poly(A) RNA. The following mouse organs were used to prepare the total RNA used in lanes 2 to 8: lane 2, brain; lane 3, heart; lane 4, lung; lane 5, kidney; lane 6, thymus; lane 7, spleen; and lane 8, liver. D2b cDNA was used as a probe. (**D**) Southern hybridization of SDF-1α and SDF-1β. Lanes 1 and 2, mouse lymphoma cell line BW5147 (*29*); lanes 3 and 4, human pro-B cell line FLEB14 (*30*). Restriction enzymes used were Eco RV (lanes 1 and 3) and Eco RI (lanes 2 and 4). The coding region of SDF-1α cDNA (nucleotides 1 to 315) was used as probe. Poly(A) RNA (2 μg) or total RNA (30 μg) was electrophoresed through a 1.0% agarose gel and blotted onto a nitrocellulose filter (*12*). Probes labeled with ^{32}P were used for hybridization under stringent conditions. Autoradiograms were analyzed with a Bio-image analyzer (BAS 2000, Fuji Film, Tokyo, Japan), and the DNAs (5 μg) were digested with various restriction enzymes, separated by electrophoresis in a 0.8% agarose gel, and transferred to a nitrocellulose filter. The filter was hybridized with the ^{32}P-labeled probe as described above and washed with 0.3 M NaCl and 30 mM sodium citrate containing 0.1% SDS at 50°C. Size markers are indicated to the left in kilobases.

mMIP-1α–SCI) and 29% identical to those of mouse JE (Fig. 3), which indicates that SDF-1α and SDF-1β are members of the intercrine α cytokine family (*15*).

Northern (RNA) blot analysis of ST-2 cell mRNA revealed a 1.9-kb SDF-1α mRNA and a 3.4-kb SDF-1β mRNA that hybridized to the SDF-1α–specific 3′ UT probe and SDF-1β–specific 3′ UT probe, respectively (Fig. 4, A and B). The SDF-1α and SDF-1β mRNAs could be transcripts of two related genes or alternative splicing products of a single gene. Expression of SDF-1α and SDF-1β mRNAs was not observed in hematopoietic progenitor cell lines or lymphocytes, such as LyD9 (*9, 19*), HL-60 (*20*), SSK41 (*21*), and HIL-3 (*22*) (Fig. 4, A and B) (*23*). In the mouse fibroblast cell line NIH 3T3 (*24*) SDF-1α and SDF-1β mRNA expression was detected, although it was much fainter than that in ST-2 cells (Fig. 4, A and B). The SDF-1α and SDF-1β mRNAs were in all organs tested, but this could be due to expression in fibroblast cells located in the various tissues (Fig. 4C). Southern (DNA) blot analysis showed that the SDF-1α and SDF-1β genes are conserved in mouse and human sequences (Fig. 4D). Our preliminary data showed that the amino acid sequences of human and mouse SDF-1α and SDF-1β were more than 95% conserved (*23*), which suggests that SDF-1α and SDF-1β might play some important biological roles.

No intercrine α cytokine members except for SDF-1α and SDF-1β have been reported to be expressed in bone marrow stromal cells. Although many known biological functions of intercrine cytokine family members are related to inflammation, MIPs also regulate hematopoiesis (*15, 25*). MIP-1α–SCI, a member of the intercrine β cytokine family, inhibits hematopoietic stem cells from proceeding into the S phase (*17*). Whether SDF-1α and SDF-1β are involved in the regulation of bone marrow hematopoiesis is not yet known.

We describe here a strategy for cloning, without functional assays, cDNAs bearing signal sequences. In this method, the 5′ portions of cDNAs were enriched and ligated in a proper orientation to reduce the appearance of stop codons and to increase the production of fusion proteins. Our strategy has two additional advantages. (i) The selected fragments are short enough to be sequenced without recloning, which allowed us to select new genes easily by comparison with databases. (ii) Asymmetric adaptor ligation allows amplification of inserts with PCR whenever necessary. We also screened another library containing shorter fractions (80 to 250 bp) of NH_2-terminal cDNA of ST-2 cells. A cDNA clone (247 bp) encoding the NH_2-terminal signal sequence of SDF-1α was obtained together with some other cDNA clones encoding signal sequences. This result ensures the reproducibility of this method, although the library of 400-bp cDNA is preferable because cDNAs with hydrophobic sequences in unnatural ORFs appeared when we used the shorter fractions (80 to 250 bp) of processed cDNA. This signal sequence trapping strategy provides a general approach for the cloning of intercellular signal-transducing molecules involved in complex phenomena, for which analysis by means of biological assays is difficult. This protocol may be particularly helpful for studying biological systems such as minute organs in embryos, from which it is difficult to obtain large amounts of mRNA.

REFERENCES AND NOTES

1. D. D. Sabatini, G. Kreibich, T. Morimoto, M. Adesnik, *J. Cell Biol.* **92**, 1 (1982).
2. The pcDL-SRα-Tac(3′) vector was made by digestion of Tac cDNA (*4*) at the Sac I site that is located immediately downstream of the signal sequence. Signal sequence–deleted Tac cDNA (nucleotides 243 to 1309) (*4*), designated Tac(3′), was ligated downstream of the SRα promoter in the pcDL-SRα vector (*26*). After digestion with Eco RI and Sac I, the Eco RI and Sac I sites between the SRα promoter and Tac(3′) were used for the insertion of cDNAs.
3. T. Uchiyama, S. Broder, T. A. Waldmann, *J. Immunol.* **126**, 1393 (1981).
4. T. Nikaido *et al.*, *Nature* **311**, 631 (1984).
5. S. Nagata *et al.*, *ibid.* **319**, 415 (1986).
6. V. Giguere, E. S. Ong, P. Segui, R. M. Evans, *ibid.* **330**, 624 (1987).
7. A positive control hG-CSF(5′) fragment that encoded residues 1 to 78, including the signal sequence of G-CSF (*5*), was obtained by Eco RI–Sac I digestion of a PCR product of the pSP72–hG-CSF plasmid template with two primers, 5′-GGAGATATCGAGCTCCTCGGGGTGGCACAG and 5′-GATTTAGGTGACACTATAG. A negative control hRAR(5′) fragment encoding residues 1 to 61 of hRAR (*6*), which does not have the signal sequence, was obtained by Eco RI–Sac I digestion of a PCR product of the pGEM-3-hRAR plasmid template with two primers, 5′-GGAGATATCGAGCTCAATGGTGGCTGGGGATG and 5′-GATTTAGGTGACACTATAG.
8. M. Ogawa *et al.*, *EMBO J.* **7**, 1337 (1988); S.-I. Nishikawa, M. Ogawa, S. Nishikawa, T. Kunisada, H. Kodama, *Eur. J. Immunol.* **18**, 1767 (1988); T. Sudo *et al.*, *J. Exp. Med.* **170**, 333 (1989).
9. T. Kinashi *et al.*, *Int. Immunol.* **1**, 11 (1989); K. H. Lee *et al.*, *J. Exp. Med.* **173**, 1257 (1991).
10. R. K. Saiki *et al.*, *Science* **330**, 1350 (1985).
11. We made the signal sequence–enriched cDNA library by converting ST-2 cell mRNA (5 μg) to cDNA with the use of random primers; a dC tail was then added to the cDNA at the 3′ end with terminal deoxynucleotidyl transferase. The second strand was synthesized by priming with 5′-GCGGCCGCGAATTCTGACTAACTGAC-$(dG)_{17}$, which contained an Eco RI linker and oligo(dG). After sonication, DNA fragments around 400 bp were isolated by gel electrophoresis in a 1.8% agarose. After blunting, the Sac I linkers

    5′-CCGCGAGCTCGATATCAAGCTTGTAC
    3′-GGCGCTCGAGCTATAGTTCGAACATGGAG

    were ligated. The fragments were then amplified by PCR with two primers, 5′-GAGGTACAAGCTTGATATCGAGCTCGCGG and 5′-GCCGCGAATTCTGACTAACTGAC. Fifteen cycles of PCR reaction (94°C for 30 s, 55°C for 2 min, and 72°C for 2 min) were done with a Thermal Cycler (Perkin-Elmer/Cetus) according to the manufacturer's protocol. Amplified DNA fragments were recovered after electrophoresis in a 1.8% agarose gel. After digestion with Eco RI and Sac I, the fragments were ligated to the Eco RI and Sac I sites of the pcDL-SRα-Tac(3′) vector.
12. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).
13. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105 (1982).
14. B. J. Rollines, E. D. Morrison, C. D. Stiles, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3738 (1988).
15. J. J. Oppenheim, C. O. C. Zacharial, N. Mukaida, K. Matsushima, *Annu. Rev. Immunol.* **9**, 617 (1991); S. D. Wolpe and A. Cerami, *FASEB J.* **3**, 2565 (1989).
16. T. Ikeda, A. Yamaguchi, T. Icho, N. Tsuchida, S. Yoshiki, *J. Bone Miner. Res.* **6**, 1013 (1991).
17. G. J. Graham *et al.*, *Nature* **344**, 442 (1990).

18. H. R. R. Pelham, *Annu. Rev. Cell Biol.* **5**, 1 (1989).
19. R. Palacios, H. Karasuyama, A. Rolink, *EMBO J.* **6**, 3687 (1987); K. Tashiro *et al.*, *Immunol. Lett.* **28**, 147 (1991).
20. S. J. Collins *et al.*, *Nature* **270**, 347 (1977).
21. P. Sideras *et al.*, *Int. Immunol.* **1**, 631 (1989).
22. M. Morita *et al.*, *Blood* **77**, 1766 (1991).
23. M. Shirozu, unpublished data.
24. J. L. Jainchill *et al.*, *J. Virol.* **4**, 549 (1969).
25. H. Broxmeyer *et al.*, *Blood* **76**, 1110 (1990).
26. Y. Takebe *et al.*, *Mol. Cell. Biol.* **8**, 466 (1988).
27. Single-letter abbreviations for the amino acid residues are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
28. G. von Heine, *Nucleic Acids Res.* **14**, 4683 (1986).
29. P. Ralph, *J. Immunol.* **110**, 1470 (1973).
30. S. Katamine *et al.*, *Nature* **309**, 369 (1984).
31. We thank S. Okazaki, M. Saitoh, and T. Satoh for their technical assistance; S. Nagata for the pSP72–hG-CSF plasmid; K. Umesono for the pGEM-3-hRAR plasmid; Y. Takebe for the pcDL-SRα; and T. Uchiyama for the anti-Tac.

2 April 1993; accepted 4 June 1993

Ixr1, a Yeast Protein That Binds to Platinated DNA and Confers Sensitivity to Cisplatin

Steven J. Brown, Patti J. Kellett, Stephen J. Lippard*

Structure-specific recognition proteins (SSRPs) bind to DNA containing intrastrand cross-links formed by the anticancer drug cisplatin. A yeast gene encoding an SSRP, designated *IXR1*, was cloned and sequenced. The Ixr1 protein, a member of the high mobility group–box protein family, bound specifically to DNA modified with cisplatin but not inactive platinum compounds. A yeast strain with an inactivated *IXR1* gene was half as sensitive to cisplatin and accumulated one-third as many platinum-DNA lesions after treatment with cisplatin as the parental strain. These findings suggest that SSRPs play a role in mediating the cytotoxicity of cisplatin.

The antitumor drug *cis*-diamminedichloroplatinum(II) (*cis*-DDP or cisplatin) is widely used in cancer chemotherapy and is especially effective in the treatment of testicular and ovarian cancer (*1*). We have identified SSRPs that bind to the major cisplatin-DNA adducts, 1,2-intrastrand d(GpG) and d(ApG) cross-links, but not to DNA damaged by inactive platinum compounds such as *trans*-diamminedichloroplatinum(II) (*trans*-DDP) (*2*). The SSRPs contain the recently identified high mobility group (HMG)–box motif (*3*). Several HMG-box proteins bend DNA (*4*, *5*), and we have suggested that HMG-box proteins recognize and bind specifically to locally bent and unwound DNA, such as DNA that has been modified by cisplatin (*6*).

Elucidation of a role for SSRPs in the mechanism of action of platinum antitumor drugs could greatly facilitate the design of new cancer therapies. Several models for the modulation of cisplatin cytotoxicity by SSRPs have been proposed (*7*). One model postulates that when SSRPs bind to platinated DNA they shield the DNA lesions from repair. As a result, the platinum-DNA adducts persist, block DNA replication, and ultimately lead to cell death. In this model, removal of the SSRPs would confer cisplatin resistance to the cell but would have no effect on the cytotoxicity of inactive platinum compounds.

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139.

*To whom correspondence should be addressed.

We have used the yeast *Saccharomyces cerevisiae* to study the role of SSRPs in mediating cisplatin cytotoxicity. We screened a λgt11 yeast expression library with radiolabeled, platinated DNA and isolated one partial clone, λyPt (*8*, *9*). A 3.0-kb Sac I DNA fragment containing the full-length *IXR1* gene was obtained by screening a yeast DNA library (*10*). Sequence analysis revealed a single, large open reading frame (ORF) encoding 591 amino acids. The deduced Ixr1 protein contains several long stretches of Asn and Gln residues (Fig. 1), a feature that is common in transcription factors (*11*, *12*). A database search revealed that Ixr1 is a member of the HMG-box protein family (*3*) and contains two repeated HMG-box motifs within residues 347 to 493. The best alignment of the HMG-box domain of Ixr1 occurs with ABF2 (*ARS*-binding factor 2) from *S. cerevisiae* (*13*), which shares 56 identical residues with Ixr1 over a 151–amino acid region.

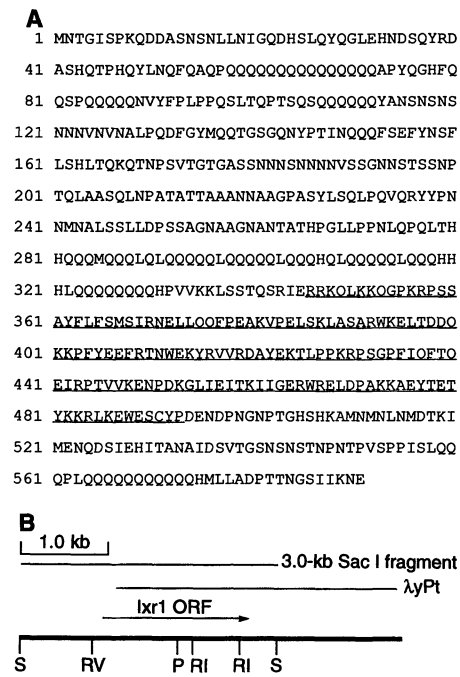

Fig. 1. (**A**) Deduced sequence of the Ixr1 protein. The two consecutive HMG boxes are underlined. Double-stranded plasmid DNA was alkaline denatured for the sequencing reactions. We performed sequencing by the dideoxy chain termination method with T7 DNA polymerase (United States Biochemical). Database homology searches were performed with the BLAST network service provided by the National Center for Biotechnology Information. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. The DNA sequence has been submitted to GenBank (accession number L16900). (**B**) Restriction map of the *IXR1* gene. RI, Eco RI; P, Pst I; RV, Eco RV; and S, Sac I.

Ixr1 protein is a single-copy gene localized on yeast chromosome XI (*14*), and Northern (RNA) blot analysis revealed that it produces a 2.1-kb transcript. We disrupted the *IXR1* gene by replacing the Eco RV–Pst I region of the 3.0-kb Sac I fragment with a copy of the *LEU2* gene. A diploid strain (*15*) with a null allele of *IXR1* was obtained, and tetrad analysis demonstrated that the *IXR1* gene product is not essential for cell growth. The *ixr1* haploid cells were not obviously different in phenotype from *IXR1* cells. For example, they displayed similar growth rates at temperatures ranging from 20° to 40°C, and homozygous *ixr1* diploid cells sporulated normally (*16*, *17*).

To investigate the DNA binding properties of Ixr1, we analyzed λyPt *Escherichia coli* lysogen and whole-cell extracts prepared from the wild-type (*IXR1*) and mutant (*ixr1*) yeast strains (*18*) by a modified Western analysis in which the protein blot is probed with radiolabeled DNA (*8*). The λyPt β-galactosidase fusion protein bound to *cis*-DDP–modified DNA and had an apparent molecular mass of 176 kD (Fig. 2). Extracts from wild-type yeast contained several proteins that bound specifically to *cis*-DDP–modified DNA; these proteins were ~100, 80, 55, and 20 kD in size. The 80-kD protein was absent from the *ixr1* strain and, thus, most likely represents the *IXR1* gene product. The λyPt β-galacto-