This fact suggests that a large reduction of most of the relevant gene products is required to produce an abnormal phenotype. In contrast, most quantitative variation exhibits some degree of additive inheritance (*21*). Allelic variation at the loci responsible for the dosage effects described here would be additive. The polygenic and additive nature of both phenomena suggests a relation between the two. Thus, it is likely that a substantial fraction of quantitative variation is the result of polymorphism at the loci responsible for the transacting effects.

The overall prevalence of dosage effects on different classes of genes also has implications for the molecular basis of aneuploid syndromes. Monosomics are typically less vigorous than trisomics, which are in turn less vigorous than euploids (*1–3*). This observation correlates with the generalized lowest levels of gene expression found in the respective chromosomal configurations in our study. In other words, the greatest reductions were found in monosomics, and lesser reductions were in trisomics, as compared with the diploids. The finding of a greater set of modulations in the aneuploid series, as opposed to the ploidy series (*5*), suggested that the classically defined imbalance associated with aneuploidy (*1–3*) is a reflection of the underlying regulatory system. We suggest that the reduction of vigor in monosomics and trisomics results from the limiting effects of many gene products encoded on the respective chromosome as well as those from throughout the genome.

REFERENCES AND NOTES

1. A. F. Blakeslee, *Am. Naturalist* **55**, 254 (1921).
2. C. B. Bridges, *ibid.* **59**, 127 (1925).
3. J. T. Patterson, M. S. Brown, W. S. Stone, *Univ. Texas Publ.* **4032** (1940), p. 167.
4. J. A. Birchler, *Genetics* **92**, 1211 (1979).
5. _____ and K. J. Newton, *ibid.* **99**, 247 (1981).
6. W. L. Gerlach *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2981 (1982).
7. D. Schwartz, *ibid.* **56**, 1431 (1966).
8. E. S. Dennis, M. M. Sachs, W. L. Gerlach, E. J. Finnegan, W. J. Peacock, *Nucleic Acids Res.* **13**, 727 (1985).
9. F. C. Belanger and A. L. Kriz, *Plant Physiol.* **91**, 636 (1989).
10. E. Sheldon, R. Ferl, N. Fedoroff, L. C. Hannah, *Mol. Gen. Genet.* **190**, 421 (1983).
11. D. R. McCarty, J. R. Shaw, L. C. Hannah, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9099 (1986).
12. D. E. Geraghty, J. Messing, I. Rubenstein, *EMBO J.* **1**, 1329 (1982).
13. J. A. Birchler and M. R. Alfenito, *J. Heredity* **84** (no. 2), 135 (1993).
14. H. Roman, *Genetics* **32**, 391 (1947).
15. In the male parent, the A segment of the B-A chromosome is marked by a dominant allele of an anthocyanin gene, and the normal A chromosome from the female tester line carries a recessive allele. The two unequivalent sperm produced from the nondisjunction of the B-A chromosome could fertilize the egg cell or the two polar nuclei. Fertilization of the egg cell with the zero-dose gamete and of the polar nuclei with the two-dose gamete gave rise to embryos of one dose and endosperm of four doses of the corresponding A chromosomal segment. This class was identified as kernels with colorless embryos and colored endosperm. The opposite fertilization event produced three-dose embryos and two-dose endosperm, which could be identified as colored embryos and colorless endosperm. Fertilization by equivalent gametes (each carrying one dose of the B-A chromosome) formed by disjunction of the B-A chromosome resulted in normal euploid kernels with two-dose embryos and three-dose endosperm, which could be recognized as those with both pigmented embryos and endosperm.
16. K. C. Cone, F. A. Burr, B. Burr, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 963 (1986).
17. Northern analysis was performed as described [J. A. Birchler and J. C. Hiebert, *Genetics* **122**, 129 (1989)]. For each specific gene, triplicate loadings of each dosage series were analyzed in each blot. The autoradiographic images were quantitated by laser densitometry. We normalized the specific mRNA quantity by calculating the ratio of mRNA to rRNA of the densitometric readings. The normalized values (mRNA:rRNA ratio) of the three replicates were used for statistical analyses.
18. J. Messing, J. Carlson, G. Hagen, I. Rubenstein, A. Oleson, *DNA* **3** (no. 1), 31 (1984).
19. R. H. Devlin, D. G. Holm, T. A. Grigliatti, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1200 (1982).
20. J. A. Birchler, J. C. Hiebert, K. Paigen, *Genetics* **124**, 677 (1990).
21. S. D. Tanksley, *Annu. Rev. Genet.* **27**, 205 (1993).
22. E. H. Coe and M. G. Neuffer, in *Corn and Corn Improvement*, G. F. Sprague, Ed. (American Society of Agronomy, Madison, WI, 1985), pp. 111–223.
23. J. B. Beckett, in *The Maize Handbook*, M. Freeling and V. Walbot, Eds. (Springer-Verlag, New York, 1994), pp. 336–341.

24. We thank L. C. Hannah, A. L. Kriz, R. Phillips, M. M. Sachs, and R. J. Schmidt for providing clones. Supported by a grant from the Department of Energy.

19 April 1994; accepted 3 October 1994

Binding and Stimulation of HIV-1 Integrase by a Human Homolog of Yeast Transcription Factor SNF5

Ganjam V. Kalpana, Shana Marmon, Weidong Wang, Gerald R. Crabtree, Stephen P. Goff*

Upon entry into a host cell, retroviruses direct the reverse transcription of the viral RNA genome and the establishment of an integrated proviral DNA. The retroviral integrase protein (IN) is responsible for the insertion of the viral DNA into host chromosomal targets. The two-hybrid system was used to identify a human gene product that binds tightly to the human immunodeficiency virus–type 1 (HIV-1) integrase in vitro and stimulates its DNA-joining activity. The sequence of the gene suggests that the protein is a human homolog of yeast SNF5, a transcriptional activator required for high-level expression of many genes. The gene, termed *INI1* (for integrase interactor 1), may encode a nuclear factor that promotes integration and targets incoming viral DNA to active genes.

The retroviral integrase enzyme catalyzes two specific reactions: (i) cleavage of the 3′-termini of the viral DNA to produce recessed 3′-OH ends, and (ii) joining of the two newly generated 3′-termini to the 5′-phosphates on each strand of the target sequence in a concerted strand-transfer reaction (*1*). Although recombinant integrase preparations can carry out all the steps known to be required for processing and joining of the viral DNA, there are indications of the involvement of additional factors in integration. For example, the isolated proteins show only very low specific activities for cutting and joining of DNA in vitro (*2*). Furthermore, for some viruses, joining reactions carried out with oligonucleotide substrates result in the transfer of only one 3′-OH to the target DNA, yielding a Y structure, rather than the concerted transfer of two 3′-OH termini to the target. It has been proposed that the viral DNA is restrained from integration into itself (autointegration) by a proposed inhibitor that can be removed by high concentrations of salt (*3*). Finally, integration in vivo is thought to be biased toward targets near transcriptionally active genes or open chromatin (*4*). Although the mechanism for such bias is not clear, transcription factors that bind DNA in a site-specific manner are thought to mediate target site selection by related retrotransposable elements (*5*).

We used the two-hybrid system (*6*) to identify host proteins that bind to the HIV-1 IN. Approximately 10^6 complementary DNAs (cDNAs) of the HL60 macrophage-monocytic cell line were expressed as GAL4AC (activation domain) fusions and tested for coactivation of a reporter gene together with a GAL4DB (DNA binding) IN fusion (*7*). Three iso-

G. V. Kalpana, S. Marmon, S. P. Goff, Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, College of Physicians and Surgeons, 630 West 168th Street, New York, NY 10032, USA.
W. Wang and G. R. Crabtree, Howard Hughes Medical Institute, Department of Developmental Biology, Beckman Center B207, Stanford University, Stanford, CA 94305, USA.

*To whom correspondence should be addressed.

lates of a single cDNA, termed *INI1* (for integrase interactor 1), were identified. To demonstrate the specificity of the IN-Ini1 interaction, we tested the GAL4AC-Ini1 fusion for activation in several settings (Table 1). The GAL4AC-Ini1 protein alone, or with the control GAL4DB protein, did not activate *lacZ* expression, suggesting that the Ini1 protein did not interact directly with the GAL4 DNA-binding domain. GAL4AC-Ini1 was also able to coactivate appropriate reporter genes in concert with a LexADB-IN fusion. These results indicate that activation by Ini1 fusions was not dependent on the particular operator and binding domain used to tether the IN protein to DNA. GAL4AC-Ini1 did not coactivate transcription with unrelated GAL4DB or LexADB fusion proteins.

Various IN mutants were examined to localize the region required for Ini1 binding and to compare the region with one previously shown to be required for IN-IN dimerization (*8*). Mutants lacking either the NH_2-terminal domain of IN, which contains a putative zinc finger region, or the COOH-terminal domain, retained their ability to bind to Ini1 (Fig. 1). Two larger COOH-terminal deletions, and a COOH-terminal point mutation that eliminated IN-IN interactions, did not affect the IN-Ini1 interaction. Two mutants of IN with point mutations in the NH_2-terminal zinc finger region (which did not affect IN-IN interaction), however, were both defective for the IN-Ini1 interaction. These results suggest that the central domain of IN is required for both interactions, but that the two interactions are genetically separable. The IN-IN interaction is affected by point mutations in the COOH-terminus, whereas the IN-Ini1 interaction is affected by mutations at the NH_2-terminus of IN.

To demonstrate that Ini1 interacts directly with IN, we carried out binding studies between recombinant proteins in vitro (9). The *INI1* cDNA isolated from the two-hybrid screen was expressed as a glutathione-S-transferase (GST) fusion, and glutathione-agarose beads (G-beads) containing the protein were used as an affinity matrix for binding IN. The bound protein was visualized with polyclonal antibodies after elution (Fig. 2). The results showed that the recombinant IN bound efficiently to the Ini1 beads and not to the control GST beads or beads alone. The majority of the IN could be bound to the GST-Ini1 (compare lane 1 to lanes 7, 10, or 13 of the lower panel in Fig. 2). Binding was observed over a wide range of salt concentrations (between 0.2 and 1 M NaCl), and with buffers containing 0.5% NP-40 or 0.1% Triton X-100, but not

Table 1. Specificity of IN-Ini1 interaction in yeast. Two plasmids, one encoding a GAL4DB or LexADB fusion and the other encoding a GAL4AC fusion, were cotransformed into either GGY1::171 (for testing the GAL4DB fusions) or CTY10-5d (for testing LexADB fusions). The transformants were scored for β-galactosidase (β-Gal) activity. Fusion protein GAL4DB-IN is encoded by plasmid pMAI, GAL4AC-IN by pGADI, GAL4AC-INI by pD2.1, LexADB by pSH2-1, LexADB-IN by pSHIN, and LexADB-lamin by pLexAlamin (*8*).

Fusion proteins	Promoter	Operator	β-Gal activity
GAL4DB-IN + GAL4AC-IN	*GAL1*	UAS_G	+ + + +
GAL4DB-IN + GAL4AC-INI	*GAL1*	UAS_G	+ + + +
GAL4DB-MG + GAL4AC-INI	*GAL1*	UAS_G	−
GAL4DB + GAL4AC-INI	*GAL1*	UAS_G	−
GAL4AC-INI	*GAL1*	UAS_G	−
LexADB-IN + GAL4AC-IN	*GAL1*	LexA	+ + + +
LexADB-IN + GAL4AC-INI	*GAL1*	LexA	+ + + +
LexADB-lamin + GAL4AC-INI	*GAL1*	LexA	−
LexADB + GAL4AC-INI	*GAL1*	LexA	−
GAL4AC-INI	*GAL1*	LexA	−

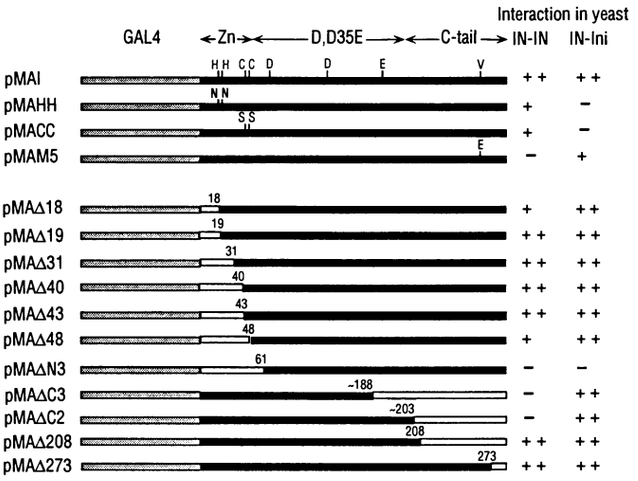

Fig. 1. Interaction of IN mutants with Ini1. Filled bars in the diagram indicate the regions retained in various GAL4DB-IN mutants. Yeast strain GGY1::171 was cotransformed with plasmids encoding each GAL4DB fusion and GAL4AC-Ini1 and assayed for the production of β-galactosidase. The deleted portion of IN in the fusion protein is indicated by the open bar, and the residue at the junction of the deletion is indicated. The substitution mutations are indicated by the residues on top of the relevant bar (*22*). IN-IN, homomeric interactions of mutant INs (*8*); IN-Ini, interaction of mutant IN with Ini1; (++) dark blue, (+) blue, (−) white colony phenotype in the X-gal assay. All proteins indicated were stably expressed in yeast as assessed by protein immunoblot analysis.

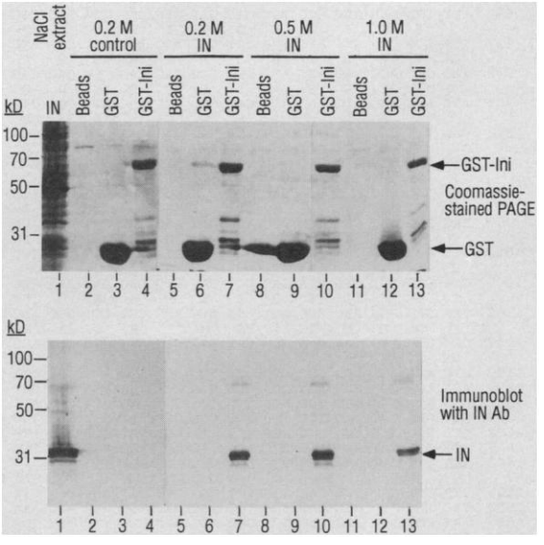

Fig. 2. Interaction of IN with GST-Ini1 in vitro. GST-Ini1 fusion protein was affinity purified on glutathione-agarose beads (G-beads) and was used as an affinity matrix for binding of IN present in bacterial lysates. The bound proteins were separated by SDS-PAGE and subjected to protein immunoblot analysis with polyclonal antibodies specific for HIV-1 IN. (**Top**) Coomassie-stained gel of the bound proteins. (**Bottom**) Protein immunoblot analysis of a duplicate gel with antibodies to IN. IN, lysate of bacterial cultures expressing IN; control, control bacterial lysate not expressing IN; Beads, glutathione beads alone; GST, GST bound to glutathione beads; GST-Ini1, GST-Ini1 bound to glutathione beads. Various concentrations of NaCl used in the binding assays are indicated above the lanes. Lanes 1, total proteins in lysate equivalent to that added to each binding reaction.

0.1% SDS. The interaction of a full-length fusion protein (GST-Ini1F) with IN was also tested; binding activity of the full-length Ini1 was similar to that of the original fragment.

RNAs from HeLa cells, a human B cell tumor line (CB33), and a human T cell line (Hut78) all contained a single major mRNA species ~2.0 kb in size that was homologous to an Ini1 probe (Fig. 3). The HeLa and CB33 lines contained in addition a minor species migrating at ~4.0 kb. Peripheral blood lymphocytes, colon, small intestine, ovary, testis, prostate, thymus, and spleen also expressed substantial amounts of the 2.0-kb mRNA. Long exposures of the autoradiograms revealed small amounts of a species migrating at 1.25 kb that was present in the spleen, and similarly small amounts of a species migrating at about 4 kb in the thymus, prostate, and testes. These results suggest that the *INI1* gene is very widely expressed.

The cDNA inserts in the three GAL4AC plasmids recovered were examined by restriction mapping and partial sequence analysis, and all were found to consist of the identical 1.0-kb fragment, presumably from sibling clones in the original phage library. Overlapping cDNAs spanning 1.85 kb, nearly the size of the full-length mRNA detected by Northern (RNA) blots, were identified by hybridization of phage libraries (*10*). The sequence revealed a single, long open reading frame, curiously beginning with a tandem array of four consecutive ATG codons (Fig. 4A). The first ATG of the array lies in a good match to the consensus sequence for

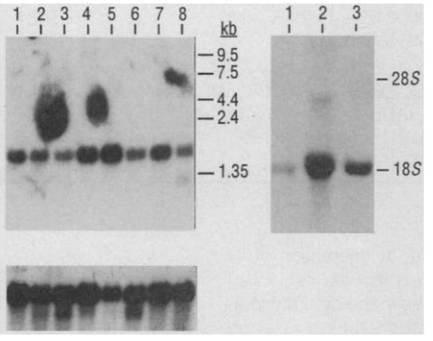

Fig. 3. Analysis of mRNA in human tissues and cell lines. (**Top left**) Northern blot probed with *INI1* cDNA insert. Each lane contains about 2 μg of poly(A)-selected mRNA. Lane 1, peripheral blood lymphocytes; lane 2, colon; lane 3, small intestine; lane 4, ovary; lane 5, testis; lane 6, prostate; lane 7, thymus; and lane 8, spleen. (**Bottom left**) The same blot after stripping and reprobing with a human actin cDNA probe. (**Top right**) Northern analysis of total RNAs from human cell lines hybridized with the *INI1* cDNA probe. Lane 1, HeLa; lane 2, CB33; lane 3, Hut78. The amount of RNA loaded in each lane is not equivalent.

A

```
-69  GCC CCG GCC CCG CCC CAG CCC TCC TGA TCC CTC GCA GCC CGG CTC
                                    ***
-24  CGG CCG CCC GCC TCT GCC GCC GCA ATG ATG ATG ATG GCG CTG AGC
                                     M   M   M   M   A   L   S    7
 22  AAG ACC TTC GGG CAG AAG CCC GTG AAG TTC CAG CTG GAG GAC GAC
      K   T   F   G   Q   K   P   V   K   F   Q   L   E   D   D   22
 67  GGC GAG TTC TAC ATG ATC GGC TCC GAG GTG GGA AAC TAC CTC CGT
      G   E   F   Y   M   I   G   S   E   V   G   N   Y   L   R   37
112  ATG TTC CGA GGT TCT CTG TAC AAG AGA TAC CCC TCA CTC TGG AGG
      M   F   R   G   S   L   Y   K   R   Y   P   S   L   W   R   52
157  CGA CTA GCC ACT GTG GAA GAG AGG AAG AAA ATA GTT GCA TCG TCA
      R   L   A   T   V   E   E   R   K   K   I   V   A   S   S   67
202  CAT GGT AAA AAA ACA AAA CCT AAC ACT AAG GAT CAC GGA TAC ACG
      H   G   K   K   T   K   P   N   T   K   D   H   G   Y   T   82
247  ACT CTA GCC ACC AGT GTG ACC CTG TTA AAA GCC TCG GAA GTG GAA
      T   L   A   T   S   V   T   L   L   K   A   S   E   V   E   97
292  GAG ATT CTG GAT GGC AAC GAT GAG AAG TAC AAG GCT GTG TCC ATC
      E   I   L   D   G   N   D   E   K   Y   K   A   V   S   I  112
337  AGC ACA GAG CCC CCC ACC TAC CTC AGG GAA CAG AAG GCC AAG AGG
      S   T   E   P   P   T   Y   L   R   E   Q   K   A   K   R  127
382  AAC AGC CAG TGG GTA CCC ACC CTG TCC AAC AGC TCC CAC CAC TTA
      N   S   Q   W   V   P   T   L   S   N   S   S   H   H   L  142
427  GAT GCC GTG CCA TGC TCC ACA ACC ATC AAC AGG AAC CGC ATG GGC
      D   A   V   P   C   S   T   T   I   N   R   N   R   M   G  157
472  CGA GAC AAG AAG AGA ACC TTC CCC CTT TGC TTT GAT GAC CAT GAC
      R   D   K   K   R   T   F   P   L   C   F   D   D   H   D  172
517  CCA GCT GTG ATC CAT GAG AAC GCA TCT CAG CCC GAG GTG CTG GTC
      P   A   V   I   H   E   N   A   S   Q   P   E   V   L   V  187
562  CCC ATC CGG CTG GAC ATG GAG ATC GAT GGG CAG AAG CTG CGA GAC
      P   I   R   L   D   M   E   I   D   G   Q   K   L   R   D  202
607  GCC TTC ACC TGG AAC ATG AAT GAG AAG TTG ATG ACG CCT GAG ATG
      A   F   T   W   N   M   N   E   K   L   M   T   P   E   M  217
652  TTT TCA GAA ATC CTC TGT GAC GAT CTG GAT TTG AAC CCG CTG ACG
      F   S   E   I   L   C   D   D   L   D   L   N   P   L   T  232
697  TTT GTG CCA GCC ATC GCC TCT GCC ATC AGA CAG CAG ATC GAG TCC
      F   V   P   A   I   A   S   A   I   R   Q   Q   I   E   S  247
742  TAC CCC ACG GAC AGC ATC CTG GAG GAC CAG TCA GAC CAG CGC GTC
      Y   P   T   D   S   I   L   E   D   Q   S   D   Q   R   V  262
787  ATC ATC AAG CTG AAC ATC CAT GTG GGA AAC ATT TCC CTG GTG GAC
      I   I   K   L   N   I   H   V   G   N   I   S   L   V   D  277
832  CAG TTT GAG TGG GAC ATG TCA GAG AAG GAG AAC TCA CCA GAG AAG
      Q   F   E   W   D   M   S   E   K   E   N   S   P   E   K  292
877  TTT GCC CTG AAG CTG TGC TCG GAG CTG GGG TTG GGC GGG GAG TTT
      F   A   L   K   L   C   S   E   L   G   L   G   G   E   F  307
922  GTC ACC ACC ATC GCA TAC AGC ATC CGG GGA CAG CTG AGC TGG CAT
      V   T   T   I   A   Y   S   I   R   G   Q   L   S   W   H  322
967  CAG AAG ACC TAC GCC TTC AGC GAG AAC CCT CTG CCC ACA GTG GAG
      Q   K   T   Y   A   F   S   E   N   P   L   P   T   V   E  337
1012 ATT GCC ATC CGG AAC ACG GGC GAT GCG GAC CAG TGG TGC CCA CTG
      I   A   I   R   N   T   G   D   A   D   Q   W   C   P   L  352
1057 CTG GAG ACT CTG ACA GAC GCT GAG ATG GAG AAG AAG ATC CGC GAC
      L   E   T   L   T   D   A   E   M   E   K   K   I   R   D  367
1102 CAG GAC AGG AAC ACG AGG CGG ATG AGG CGT CTT GCC AAC ACG GCC
      Q   D   R   N   T   R   R   M   R   R   L   A   N   T   A  382
1147 CCG GCC TGG TAA CCA GCC CAT CAG CAC ACG GCT CCC ACG GAG CAT
      P   A   W  ***                                           385
1192 CTC AGA AGA TTG GGC CGC CTC TCC TCC ATC TTC TGG CAA GGA CAG
1237 AGG CGA GGG GAC AGC CCA GCG CCA TCC TGA GGA TCG GGT GGG GGT
1282 GGA GTG GGG GCT TCC AGG TGG CCC TTC CCG GTA CAC ATT CCA TTT
1327 GTT GAG CCC CAG TCC TGC CCC CCA CCC CAC CCT CCC TAC CCC TCC
1372 CCA GTC TCT GGG GTC AGG AAG AAA CCT TAT TTT AGG TTG TGT TTT
1417 GTT TTG TAT AGG AGC CCC AGG CAG GGC TAG TAA CAG TTT TTA AAT
1462 AAA AGG CAA CAG GTC ATG TTC AAAAAAAAAAA
                                AAT TTC TTA AAT CTA GTG TCT TTA
1507 TTT CTT CTG TTA CAA TAG TGT TGC TTG TGT AAG CAG GTT AGA GTG
1552 CAC AGT GTC CCC AAT TGT TCC TGG CAC TGC AAA ACC AAA TTA AAC
1597 AAT CCC ACA AAG AAT TCT GAC ATC AAT GTG TTT TCC TCA GTC AGG
1642 TCT ATT TCA AGA TTC TAG AAG TTC CTT TTG TAA AAC TTG CCT TTA
1687 AAA CTC TTC CTC CTA ATG CCA TCA GAT CTC TTA ACA TTG GCT CAC
1732 TGT GGG ATC TTT CCT CTT AGG TTG AAT TTC TAC GTG AAT ATC AAA
1777 GTG CCT TTT TC 1787
```

B

```
Ini1  180 ..ASQPEVLVPIRLDMEIDGQK..LRDAFTWNMNEKLMTPEMFSEILCDDLD
SNF5  451 ..NETSEQLVPIRLEFDQDRDRFFLRDTLLWNKNDKLIKIEDFVDDMLRDYR
          <----------Region I---------->
      228  LNPLT...FVPAIASAIRQQIESYPTDSILE......DQSDQRVIIKLNI
      501  FEDATREQHIDTICQSIQEQIQEFQGNPYIELNQDRLGGDDLRIRIKLDI
                                                <----------
      269  HVGNISLVDQFEWDMSEKENSPEKFALKLCSELGLGGEFVTTIAYSIRGQ
      551  VVGQNQLIDQFEWDISNSDNCPEEFAESMCQELELPGEFVTAIAHSIREQ
           ----------Region II----------
      319  LSWHQKTYAFSEN.................PLPTVEIAIRNTGDADQWCP
      601  VHMYHKSLALLGYNFDGSAIEDDDIRSRMLPTITLDDVYRPAAESKIFTP

      352  LLETLTDAEMEKKIRDQDRNTRRMRRLANTAP.. 383
      651  NLQISAAELERLDKDKDRDTRRKRRQGRSNR.. 682
           <----------Region III---------->
```

C

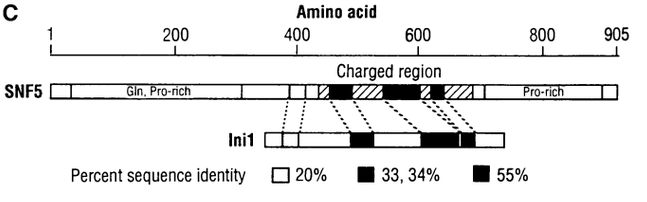

Fig. 4. Sequence analysis of *INI1* cDNA. (**A**) Complete sequence deduced from the overlapping *INI1* cDNA clones (GenEmbl Bank accession number U04847). The A nucleotide of the first methionine codon was considered nt 1. Amino acid residues are numbered on the right side of the diagram and nucleotides on the left (*22*). Potential poly(A) addition signal AATAAA is underlined, and the start and stop codon appear in boldface. The poly(A) stretch in clone pINI.gt is indicated by the stretch of A's in the middle of the 3′ noncoding region. Stop codons are indicated by asterisks. The heptad repeat of leucine-valine residues are in boldface. The potential N-linked glycosylation sites are circled. The original cDNA fragment isolated in the two-hybrid screen spanned nucleotide sequence 322 to 1291. (**B**) A central portion of the Ini1 amino acid sequence is shown aligned with that of the yeast SNF5 sequence. Residues that are identical between the two sequences are indicated by shading. The three regions that show a high degree of sequence similarity between the two proteins (33 to 50% identity) are indicated by the bars underneath. (**C**) Schematic alignment of Ini1 with SNF5. The regions of highest similarity are shaded, and the percent identity given below. The glutamine and proline-rich regions of SNF5 are indicated.

translational initiation (*11*). One clone contained a stretch of polyadenylate [poly(A)] residues at the 3′ junction adjacent to the vector, and a consensus polyadenylation signal, AATAAA, at −25 base pairs (bp) relative to the poly(A), suggesting that most of the *INI1* mRNAs are processed by cleavage and polyadenylation at this position. One cDNA clone (pINI.21), however, extended beyond this region without poly(A) sequences, suggesting that some mRNAs are of extended length and arise through use of alternative poly(A) addition sites further downstream. One clone (pINI.9) lacked a short stretch of 27 bp [nucleotide (nt) 206 to 232] near the 5′ end of the coding region, and might have arisen from an alternatively spliced mRNA lacking an internal exon. Binding of IN by the Ini1 fragment encoded by the original cDNA, both in yeast and in vitro, indicates that the first 107 amino acids of the complete protein are not required for binding.

The long open reading frame predicts the formation of a protein of 44,131 daltons containing 385 amino acids. The sequence revealed the presence of a heptad repeat of three leucine residues near the NH_2-terminus of the encoded protein; these residues could potentially form a leucine zipper structure. The predicted sequence includes no NH_2-terminal secretion signals, no transmembrane segments, and no markedly acidic or basic regions. There are three potential sites for addition of N-linked sugars. The predicted isoelectric point of the protein is 6.15.

Comparison of the predicted sequence of Ini1 with the known sequences in the GenEmbl (Genbank–European Molecular Biology Laboratories) database revealed a single significant match (Fig. 4B), the SNF5 protein of *Saccharomyces cerevisiae* (*12, 13*), a transcription factor thought to act in a complex with several other proteins including SNF2-SWI2, SNF6, SWI1, and SWI3 (the SNF/SWI complex) to activate target gene expression (*14*). The alignment of Ini1 with the SNF5 sequence displayed three regions of close similarity, with 33 to 55% sequence identity and 41 to 71% conserved residues (Fig. 4C). All three regions lay in the central portion of the SNF5 sequence; the flanking proline- and glutamine-rich segments of the yeast protein were not retained in the human gene. On the basis of the pronounced similarity between the yeast and human genes in the core coding region, we propose that *INI1* is a human homolog of the yeast gene encoding SNF5. *INI1* joins two other mammalian genes, *brg1* and *brm*, that have recently been cloned as homologs of another protein in the complex, SNF2 (*15*).

The yeast SNF5 protein is capable of activating a reporter gene when artificially tethered to DNA by fusion to the LexA DNA-binding domain (*13*). For determination of whether Ini1 could also act as a transcriptional activator in this setting, a construct encoding a fusion of GAL4DB-Ini1 was generated and expressed in an indicator strain containing a *Gal1-lacZ* reporter. The transformants expressed high concentrations of β-galactosidase as judged by staining with X-gal, whereas control transformants expressing only the GAL4DB or GAL4AC-Ini1 protein did not.

To determine whether Ini1 protein could affect IN function, we assayed the enzymatic activities of IN in the presence of increasing concentrations of GST-Ini1F (full-length Ini1) fusion protein (*16*). To assay IN DNA-joining activity, we incubated recombinant IN protein with ^{32}P-labeled double-stranded DNA oligonucleotides corresponding to the U5 terminus of the HIV-1 viral DNA as substrate, and with unlabeled plasmid DNA as target. Samples of the reaction mixture were removed at various times and analyzed by agarose gel electrophoresis and autoradiography; radioactivity migrating with the relaxed plasmid DNA represented integration of the labeled oligonucleotide into the target. The addition of increasing levels of GST-Ini1F resulted in a strong and dose-dependent stimulation of joining activity (Fig. 5A). Control experiments with GST showed no such stimulation. In some experiments, addition of very high levels of GST-IniF resulted in inhibition.

To determine whether the native form of Ini1 as present in mammalian cells behaved similarly, we prepared nuclear extracts and partially purified the SNF/SWI complex (Ini1 extract). The presence of Ini1 was monitored by protein immunoblot analysis with polyclonal antisera raised against GST-Ini1. Ini1 cofractionated with Brg1 and the complex through five different conventional columns, and was also retained on Brg1 immunoaffinity columns. The addition of increasing amounts of this preparation to the joining reactions resulted in potent stimulation of IN activity (Fig. 5B). Partial depletion of Ini1 by passage through a Brg1 affinity column resulted in the removal of most of the stimulatory activity (Fig. 5B). The amount of stimulation by the extract was constant over a wide range of target DNA concentrations (Fig. 5C). Maximal stimulation of IN activity (10- to 20-fold) occurred when the IN-Ini1 molar ratio in the reaction mixture was roughly 5:1. Higher concentrations of Ini1 resulted in no fur-

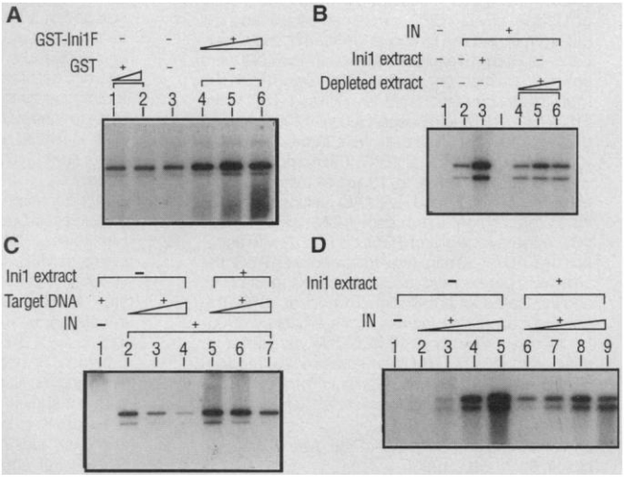

Fig. 5. Stimulation of IN joining activity by GST-Ini1 and mammalian Ini1 extract. All joining reactions in (A) to (D) were carried out as described (*15*), and each contained 15 ng of IN per reaction. (**A**) IN joining reactions carried out with or without the addition of GST and GST-Ini1F. Lanes 4 to 6, 50, 150, and 450 ng of GST-Ini1F, respectively; lanes 1 and 2, 150 and 450 ng of GST. (**B**) Effect of mammalian Ini1 extracts on the IN joining reactions. For isolation of the SNF/SWI complex, a rat liver nuclear extract was prepared as described (*23*) and fractionated on a phosphocellulose P11 column (Whatman). The 0.5 M salt fraction from this column was diluted and loaded onto a DEAE-52 column (Whatman). A 0.3 M KCl eluate from this column was further fractionated on an S-300 column (Pharmacia), and the excluded volume containing Brg1 and Ini1 was collected. Brg1 and Ini1 cofractionated throughout the purification as determined by protein immunoblot analysis with antibodies to Brg1 and Ini1. Depleted extracts were prepared by passing the Ini1 fraction through a Brg1 affinity column; and the flow-through was collected. Lane 3, Ini1 extract (1 μl) containing ~1.5 ng of Ini1 protein as assessed by immunoblot analysis with antibody to Ini1. Lanes 4 to 6, 1, 2, and 4 μl of depleted extract, respectively. Total protein concentration in the depleted extract was approximately half that before depletion. (**C**) Effect of increasing concentration of target DNA on the stimulation of joining activities. The target DNA concentrations used were 10 ng (lanes 1, 2, and 5), 30 ng (lanes 3 and 6), and 90 ng (lanes 4 and 7) per 30-μl reaction. Lanes 5 to 7, Ini1 extract (2 μl) containing ~3 ng of Ini1. (**D**) Effect of increasing concentration of IN on the activity of nuclear extract. The concentrations of IN used were 5 ng (lanes 2 and 6), 15 ng (lanes 3 and 7), 45 ng (lanes 4 and 8), and 145 ng (lanes 5 and 9) per 30-μl reaction. Lanes 6 to 9, 2 μl of Ini1 extract containing ~3 ng of Ini1.

ther stimulation but rather slight inhibition. Maximal stimulation by native Ini1 required lower concentrations than with the recombinant IN. Stimulation by Ini1 was strongly dependent on the Ini1:IN ratio, with strongest stimulation at low IN concentrations, no stimulation at intermediate concentrations, and inhibition at high IN concentrations (Fig. 5D).

The results suggest that a previously unidentified host protein, Ini1, can bind the HIV-1 IN protein and stimulate its DNA-joining activity. The protein shows unexpected sequence similarity to the SNF5 protein of yeast (*13*), which is required for the high-level transcription of many genes, and for the proper functioning of several gene-specific activators (*14*). Genetic and biochemical experiments suggest that SNF5 is part of a very large complex of proteins able to promote transcription both in vitro and in vivo (*14*, *17*). The complex may help reorganize chromatin structure. Mutations in *snf2* and *snf5* are suppressed by mutations affecting histones H2A, H2B, and H3, as well as a nonhistone DNA-binding protein similar to HMG1, and direct biochemical analysis suggests that the complex can alter nuclease sensitivity of chromatin (*18*). The complex has recently been shown to alter chromatin structure and to promote binding of sequence-specific DNA-binding proteins (*19*). We have determined that Ini1 is retained on an affinity column containing antibodies to BRG1 (*20*), suggesting that it is in a complex with BRG1. All these data—the sequence similarity, the ability of Ini1 to activate a reporter gene when tethered to DNA, and its presence in the mammalian SNF/SWI complex—strongly suggest that Ini1 is a functional homolog of the yeast *SNF5* gene.

We suggest that the affinity of Ini1 for the HIV-1 IN might account for the propensity of retroviral DNAs to insert into active genes and their associated open chromatin (*4*). Upon binding to Ini1, the preintegration complex could be stimulated to insert the viral DNA into nearby sites. Finally, we note that Ini1 may provide a target for antiviral therapy: Virus replication might be blocked by drugs that inhibit the IN-Ini1 interaction or by dominant negative alleles of INI1 that bind inappropriately to IN and block its activity.

REFERENCES AND NOTES

1. R. Weiss, N. Teich, H. Varmus, J. Coffin, *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1984); S. P. Goff, *Annu. Rev. Genet.* **26**, 527 (1992); R. Craigie, *Trends Genet.* **8**, 187 (1992).
2. R. Craigie, T. Fujiwara, F. Bushman, *Cell* **62**, 829 (1990); F. D. Bushman and R. Craigie, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1339 (1991); R. A. Katz, G. Merkel, J. Kulkovsky, J. Leis, A. M. Skalka, *Cell* **63**, 87 (1990); F. D. Bushman, T. Fujiwara, R. Craigie, *Science* **249**, 1555 (1990).
3. M. S. Lee and R. Craigie, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9823 (1994).
4. S. Vijaya, D. L. Steffen, H. L. Robinson, *J. Virol.* **60**, 683 (1986); H. Rohdewohld, H. Weiher, W. Reik, R. Jaenisch, M. Breindl, *ibid.* **61**, 336 (1987); U. Scherdin, K. Rhodes, M. Breindl, *ibid.* **64**, 907 (1990); C.-C. Shih, J. P. Stoye, J. M. Coffin, *Cell* **53**, 531 (1988); E. S. Withers-Ward *et al.*, *Genes Dev.* **8**, 1473 (1994).
5. S. B. Sandmeyer, L. J. Hansen, D. L. Chalker, *Annu. Rev. Genet.* **24**, 491 (1990).
6. S. Fields and O.-K. Song, *Nature* **340**, 245 (1989).
7. Construction of pGAL4DB-IN target plasmid was as described (*8*). For construction of the HL60 cell cDNA library, the cDNA inserts were excised from a λZap HL60 cDNA library (Stratagene). A plate lysate of the phage library was prepared, and DNA was isolated from the total phage by standard methods. About 100 μg of DNA was digested with Not I and Xho I, and inserts 0.2 to 3.0 kb in size were isolated by electroelution and ligated to the pGADNot vector (*21*) digested with Not I plus Sal I and phosphatase-treated. DH5α cells were transformed with the ligation products, and the transformants from six individual batches of 100,000 to 500,000 colonies each were pooled separately in LB-Amp (KGLI, pool I to pool VI). This unamplified library in pGADNot vector was pipetted into small vials and frozen until further use. The ratio of non-recombinants to recombinants in the library was determined by comparing the number of transformants obtained with self-ligated vector to that obtained with vector ligated to insert, and by examining plasmids from several individual colonies to determine the presence of insert. Both of these tests indicated that there were >95% recombinants in the library. The plasmid library DNA was isolated from 1-liter cultures of each pool by Plasmid isolation columns (Qiagen, Chatsworth, CA). This DNA was used for transformation into yeast strain GGY1::171. Overnight cultures of GGY1::171 were diluted 1:50 or 1:100 in YPAD [YEPD supplemented with adenine (30 μg/ml)] and incubated at 30°C until the absorbance at 600 nm reached 0.25 to 0.4. The cells were pelleted, washed once with 1/10th volume of 100 mM lithium acetate–10 mM tris-HCl (pH 7.5), 1 mm EDTA (LiAc-TE) and resuspended in 1/200th volume of the same buffer. The cells were further incubated with shaking for 1 hour at 30°C. The competent cells were incubated with 1 to 10 μg of plasmid DNAs encoding GAL4DB and GAL4AC fusions, 20 μg of sonicated salmon sperm carrier DNA (Sigma), and 40% polyethylene glycol (PEG) in LiAc-TE with agitation at 30°C for 30 min. After treatment with PEG, the cells were pelleted and resuspended in 1 ml of YPAD and incubated for an additional 1 hour at 30°C. The postincubation step increased the efficiency of co-transformation by about 10 fold. Cells were pelleted, resuspended in TE, and plated on selective medium [F. M. Ausubel, *et al.*, Eds., *Current Protocols in Molecular Biology* (Wiley-Interscience, New York, 1991)].
8. G. V. Kalpana and S. P. Goff, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10593 (1993).
9. The *INI1* cDNA from positive clone pD2.1 was inserted into plasmid pGEX2TK to form pGVK10. The GST-Ini1 fusion protein encoded by pGVK10 was affinity purified on glutathione-agarose beads (G-beads) according to standard procedures [D. B. Smith and K. S. Johnson, *Gene* **67**, 31 (1988)]. The immobilized GST-Ini1 was used as an affinity matrix for binding of IN. The HIV-1 IN expressed in *Escherichia coli* from the T7 promoter was solubilized with buffer containing 1 M NaCl. The IN lysates were adjusted to 0.2 to 1 M NaCl, as indicated in Fig. 2, and then incubated with G-beads alone, G-beads with GST alone, or G-beads with GST-Ini1. The beads were washed extensively with buffer Y [50 mM tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, 0.5% NP-40, and 1 mM phenylmethylsulfonyl fluoride (PMSF)], and the bound proteins were separated by SDS–polyacrylamide gel electrophoresis (PAGE) and subjected to immunoblot analysis with polyclonal antibodies specific for HIV-1 IN.
10. Overlapping cDNA clones encoding full-length Ini1 were obtained from screening two phage libraries: λZAPII-HeLa cDNA library (Stratagene) and λgt11-HeLa cDNA library (gift of A. Efstratiadis and S. Zeitlin).
11. M. Kozak, *J. Biol. Chem.* **266**, 19867 (1991).
12. L. Neigeborn and M. Carlson, *Genetics* **108**, 845 (1984); E. Abrams, L. Neigeborn, M. Carlson, *Mol. Cell. Biol.* **6**, 3643 (1986).
13. B. C. Laurent, M. A. Treitel, M. Carlson, *Mol. Cell. Biol.* **10**, 5616 (1990).
14. ———, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2687 (1991); B. C. Laurent and M. Carlson, *Genes Dev.* **6**, 1707 (1992); S. K. Yoshinaga, C. L. Peterson, I. Herskowitz, K. R. Yamamoto, *Science* **258**, 1598 (1992); C. L. Peterson and I. Herskowitz, *Cell* **68**, 573 (1992); M. Carlson and B. C. Laurent, *Curr. Opin. Cell Biol.* **6**, 396 (1994).
15. P. A. Khavari, C. L. Peterson, J. W. Tamkun, D. B. Mendel, G. R. Crabtree, *Nature* **366**, 170 (1993); C. Muchardt and M. Yaniv, *EMBO J.* **12**, 4279 (1993).
16. A full-length cDNA clone lacking only the first five codons was inserted into pGEX2TK, and the full-length Ini1 fusion protein (GST-Ini1F) was isolated with the use of G-beads and eluted with 20 mM glutathione. The protein was dialyzed against a large volume of storage buffer [25 mM Hepes (pH 7.2), 50 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol, 20% glycerol, 1 mM PMSF, 1 μg/ml each of pepstatin, aprotinin, and leupeptin] and stored at −70°C. Recombinant HIV-1 IN protein was isolated from bacterial cultures carrying plasmid pINC6H essentially as described [M. Drelich, R. Wilhelm, J. Mous, *Virology* **188**, 459 (1992)], with minor modifications. Integrase-joining activity assays were performed in a total volume of 30 μl and contained 1 ng of a double-stranded DNA oligonucleotide from the HIV-1 U5 terminus, consisting of one strand labeled at the 5′ end, representing the already-processed substrate (sequence 5′-GGATCCGGAAAATCTCTAGCA-3′), and its unlabeled complement with extra CA dinucleotide overhang at the 5′ end; 10 ng of pBluescript DNA as target; and ~15 ng of IN. The reactions were stopped by addition of EDTA to 50 mM final concentration. The products were treated with proteinase K and SDS and analyzed by electrophoresis on a 1% agarose gel. The gel was dried and exposed to autoradiography to monitor transfer of the oligonucleotide to the relaxed circular and linear target DNA.
17. B. R. Cairns *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1950 (1994); C. L. Peterson, A. Dingwall, M. P. Scott, *ibid.*, p. 2905.
18. J. N. Hirschhorn, S. A. Brown, C. D. Clark. F. Winston, *Genes Dev.* **6**, 2288 (1992); W. Kruger and I. Herskowitz, *Mol. Cell. Biol.* **11**, 4135 (1991); F. Winston and M. Carlson, *Trends Genet.* **8**, 387 (1992).
19. H. Kwon, A. N. Imbalzano, P. A. Khavari, R. E. Kingston, M. R. Green, *Nature* **370**, 477 (1994); A. N. Imbalzano, H. Kwon, M. R. Green, R. E. Kingston, *ibid.*, p. 481; J. Cote, J. Quinn, J. L. Workman, C. L. Peterson, *Science* **265**, 53 (1994).
20. W. Wang, G. R. Crabtree, G. V. Kalpana, S. P. Goff, data not shown.
21. J. Luban, K. L. Bossolt, E. K. Franke, G. V. Kalpana, S. P. Goff, *Cell* **73**, 1067 (1993).
22. Single-letter 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.
23. K. Gorski *et al.*, *Cell* **47**, 767 (1986).

24. We thank M. Carlson, X. Yang, and S. Silverstein for helpful suggestions; S. Fields and J. Luban for plasmids and advice; W. Osheroff, R. Swanstrom, and S. LeGrice for IN plasmids; B. Nilson and D. Helland for IN antibodies; S. Zeitlin and A. Efstratiadis for the λgt11 cDNA library; and S. Amin for technical assistance. Supported by PHS award U01 AI 24845 to S.P.G. and by an award from the American Foundation for AIDS research (AmFAR) to G.V.K. S.P.G. is an Investigator of the Howard Hughes Medical Institute. G.V.K. is a Helen Hay Whitney Fellow. W.W. is a Damon Runyon Fellow.

26 August 1994; accepted 2 November 1994