Group II Introns Designed to Insert into Therapeutically Relevant DNA Target Sites in Human Cells

Huatao Guo *et al.*

Science **289**, 452 (2000);
DOI: 10.1126/science.289.5478.452

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Group II Introns Designed to Insert into Therapeutically Relevant DNA Target Sites in Human Cells

Huatao Guo,[1] Michael Karberg,[1] Meredith Long,[2] J. P. Jones III,[2] Bruce Sullenger,[2] Alan M. Lambowitz[1]*

Mobile group II intron RNAs insert directly into DNA target sites and are then reverse-transcribed into genomic DNA by the associated intron-encoded protein. Target site recognition involves modifiable base-pairing interactions between the intron RNA and a >14-nucleotide region of the DNA target site, as well as fixed interactions between the protein and flanking regions. Here, we developed a highly efficient *Escherichia coli* genetic assay to determine detailed target site recognition rules for the *Lactococcus lactis* group II intron Ll.LtrB and to select introns that insert into desired target sites. Using human immunodeficiency virus–type 1 (HIV-1) proviral DNA and the human *CCR5* gene as examples, we show that group II introns can be retargeted to insert efficiently into virtually any target DNA and that the retargeted introns retain activity in human cells. This work provides the practical basis for potential applications of targeted group II introns in genetic engineering, functional genomics, and gene therapy.

Group II introns are catalytic RNAs that function as mobile genetic elements by inserting directly into target sites in double-stranded DNA (*1, 2*). This mobility is mediated by a multifunctional intron-encoded protein (IEP) that has reverse transcriptase (RT), RNA splicing (maturase), and DNA endonuclease activities (*2–5*). After translation, the protein promotes RNA splicing, presumably by facilitating formation of the catalytically active intron RNA structure. It then remains associated with the excised intron to form a ribonucleoprotein (RNP) complex, which has DNA endonuclease/integrase activity. In homing, the major mobility pathway, the excised intron RNA in this complex reverse-splices into a specific target site in double-stranded DNA (*6–8*). The associated IEP then cleaves the opposite strand in the 3′ exon of the DNA target, 9 or 10 nucleotides (nt) downstream of the intron insertion site, and uses the 3′ end of the cleaved strand as a primer to reverse-transcribe the inserted intron RNA. The resulting cDNA copy of the intron is incorporated into the recipient DNA primarily by recombination mechanisms in yeast mitochondria (*7*) and by repair mechanisms in bacteria (*8*). Homing frequencies approach 100% for both fungal mitochondrial and bacterial introns (*7, 8*).

To initiate mobility, the intron-encoded RNP complex uses both its RNA and protein components to recognize specific sequences in its DNA target site (*9–11*). For the well-studied *Lactococcus lactis* Ll.LtrB intron, the DNA target site extends from position −26 in the 5′ exon (E1) to position +9 in the 3′ exon (E2; positions numbered from the intron insertion site) (Fig. 1A) (*11*). A 14-nt region of the DNA target site (E1 −13 to E2 +1) is recognized primarily by base pairing with the intron RNA. This region includes short sequence elements denoted IBS2, IBS1, and δ′, which are complementary to intron sequences EBS2, EBS1, and δ (IBS and EBS refer to intron and exon binding sites, respectively) (Fig. 1, A and B). These same sequence elements are involved in base-pairing interactions required for RNA splicing (*1*). The regions of the DNA target site flanking the IBS and δ′ sequences are recognized by the IEP. The protein first recognizes a small number of nucleotide residues in the distal 5′ exon region (E1 −26 to −11) and appears to cause local DNA unwinding, enabling the intron to form base pairs with the IBS and δ′ sequences for reverse splicing. Antisense-strand cleavage occurs after reverse splicing and requires additional interactions between the protein and 3′ exon. The finding that at least a 14-nt region of the DNA target site is recognized by base pairing with the intron RNA raises the possibility that group II introns can be retargeted to recognize any 14-nt DNA sequence, juxtaposed to the fixed posi-

[1]Institute for Cellular and Molecular Biology, Department of Chemistry and Biochemistry, and Section of Molecular Genetics and Microbiology, School of Biological Sciences, University of Texas, Austin, TX 78712, USA. [2]Center for Genetic and Cellular Therapies, Department of Surgery, Duke University Medical Center, Box 2601, Durham, NC 27710, USA.

*To whom correspondence should be addressed. E-mail: lambowitz@mail.utexas.edu

that the small fraction of HDV1 molecules that assume the HDV fold are highly active. Because U73C affects the fold and not the core of the ribozyme, the intersection molecules that assume the HDV fold must also be highly active.
23. For the ligase ribozyme, this search was informed by an experiment in which active ligase variants were selected from a large pool of minor variants of intersection sequences. A pool of more than 10^{15} variants was based on two sequences that were similar to that shown in Fig. 2A and also had detectable ligation and cleavage activity: 5′-Ggactcccat tagactgggc cGCCTCCTCG CGGCgggagt tGGGccaggg aggtaagccc tttctggGCt AAgggccca and 5′-Ggactcccat tagactgggc cGCCTCCTCG CGGCgggagt tGGGccaggg aggtaagccc tttctggGCt AAggccca (lowercase letters indicate positions mutagenized such that they each had a 10% probability of changing to one of the three other nucleotides; uppercase letters indicate positions not mutagenized because of their importance for HDV self-cleavage activity). Selection for ligation activity was performed as described (*35*). Pool activity reached that of the prototype ribozyme by round three, at which point the pool was cloned and sequenced. Each of the 20 isolates had a different nucleotide sequence, indicating that there are many highly active sequences in the vicinity of the intersection sequence. Sequences of active clones informed the design of LIG1, LIG2, and LIG4.
24. M. A. Tanner and T. R. Cech, *RNA* **2**, 74 (1996).
25. Supplemental data showing the predicted secondary structures of each construct (Fig. 3) and explaining the ligation activity of truncated ribozymes (Fig. 2B) are available at *Science* Online at www.sciencemag.org/feature/data/1050240.shl.
26. J. Maynard Smith, *Nature* **225**, 563 (1970).
27. The Intersection Theorem (*5*) states that there exists at least one intersection sequence for every pair of RNA secondary structures (usually as suboptimal conformations).
28. S. Ohno, *Evolution by Gene Duplication* (Springer-Verlag, New York, 1970).
29. D. L. Minor Jr. and P. S. Kim, *Nature* **380**, 730 (1996).
30. M. H. J. Cordes, N. P. Walsh, C. J. McKnight, R. T. Sauer, *Science* **284**, 325 (1999).
31. X. Ye *et al.*, *Chem. Biol.* **6**, 657 (1999).
32. Although protein enzyme networks probably do not intersect, a single gene can give rise to alternative amino acid sequences having different folds through the use of overlapping reading frames. This mechanism for expression (and perhaps emergence) of multiple protein folds and functions appears prevalent in viruses, where there is a high cost to carrying nonfunctional sequences.
33. S. Dalal, S. Balasubramanian, L. Regan, *Folding Des.* **2**, R71 (1997).
34. F. E. Cohen, *J. Mol. Biol.* **293**, 313 (1999).
35. E. H. Ekland and D. P. Bartel, *Nucleic Acids Res.* **23**, 3231 (1995).
36. A. R. Ferre-D'Amare, K. Zhou, J. A. Doudna, *Nature* **395**, 567 (1998).
37. Ribozyme activities were determined as described (*10*, *14*). Because no antisense oligonucleotide was used to protect the cleavage site when transcribing the HDV variants, all cleavage rates are reported relative to the rate of the similarly prepared prototype, 0.1 min^{-1} (*14*). All except seven ribozymes were transcribed from synthetic oligonucleotide templates. The others (LIG P, LIG2, LIG1, INT, HDV1, HDV2, and HDV P) were transcribed from polymerase chain reaction (PCR)–amplified DNA that had been generated from plasmid DNA template and appropriate primers. This removed the small amount of sequence heterogeneity observed with long synthetic templates, eliminating concern that sequence heterogeneity might increase the apparent activity of inefficient ribozymes.
38. We thank W. Johnston for assistance with in vitro selection (*23*) and helpful discussions and N. Lau for assistance with Web fig. 1. We thank P. Kim as well as M. Lawrence, P. Unrau, M. Glasner, N. Bergman, and others in the lab for comments on the manuscript. Supported by the NSF/Alfred P. Sloan Foundation (E.A.S.) and the NIH (D.P.B.).

8 March 2000; accepted 18 May 2000

tions recognized by the IEP. By using crude target site recognition rules deduced from biochemical experiments, the Ll.LtrB intron could in fact be retargeted to specific sites in a plasmid-borne *E. coli thyA* gene. At best, however, these retargeted introns were very inefficient, presumably reflecting the cumulative effect of multiple changes from the normal target site sequence and/or additional constraints that must be satisfied for optimal base-pairing interactions (*11*).

To determine more detailed target site recognition rules, it is necessary to test a large number of different nucleotide combinations. For this purpose, we developed a new *E. coli* genetic assay (Fig. 1C) in which a modified Ll.LtrB intron containing a phage T7 promoter near its 3′ end is expressed from a T7lac promoter in a chloramphenicol-resistant (Cam^R) donor plasmid (pACD-LtrB) (*12*). A compatible ampicillin-resistant (Amp^R) recipient plasmid (pUCR-LtrB/Tet) contains the Ll.LtrB target site (ligated E1-E2 sequence) inserted upstream of a promoterless tetracycline resistance (tet^R) gene (*13*), so that movement of the intron into the target site activates the expression of that gene. To assay mobility, we cotransformed the donor and recipient plasmids into an *E. coli* (DE3) strain, which contains an isopropyl-β-D-thiogalactopyranoside (IPTG)–inducible T7 RNA polymerase (*14*). After induction with 2 mM IPTG, cells cotransformed with the wild-type donor, and recipient plasmids gave 10 to 40% Amp^RTet^R colonies indicative of mobility events, compared to 0.001% for cells transformed with the Amp^R recipient plasmid alone. Correct integration of the intron was confirmed by DNA sequencing of 10 mobility events.

As expected, efficient mobility was abolished by mutations that delete a large segment of the intron open reading frame (ΔORF), inhibit the RT (YAAA) or DNA endonuclease (ΔZn or ΔConZn) activities of the IEP, or inhibit the ribozyme activity of the intron RNA (ΔD5) (*15*). Further, in experiments using a "twintron" construct, in which a self-splicing group I intron (the *td* intron) was inserted into the group II intron (*8*), 95% of the mobility products had spliced the *td* intron; this result confirmed that mobility occurs through an RNA intermediate (*16*). Deletion analysis showed that a target site extending from positions −25 to +9 was sufficient for maximal mobility, whereas further 5′ deletions to −13 reduced mobility by a factor of ~7000 and 3′ deletions to +4 reduced mobility by a factor of ~240.

Use of the donor plasmid pACD-ΔORF+ORF1 led to a marked increase in mobility frequency (Fig. 1D) (*17*). This plasmid has a large deletion in the "loop" of intron domain IV, which removes most of the LtrA ORF, and expresses the LtrA protein separately from a position downstream of the 3′ exon. This configuration gave very high mobility frequencies (~70% Tet^R colonies), even without IPTG induction to stimulate donor plasmid transcription, and the frequencies increased to 100% with a low concentration of IPTG (100 μM). The increased mobility frequencies appear to be due to greater resistance of the ΔORF intron to nucleolytic cleavage in domain IV rather than increased expression of the LtrA protein (*18*).

The very high mobility frequencies enabled us to determine detailed target site rec-

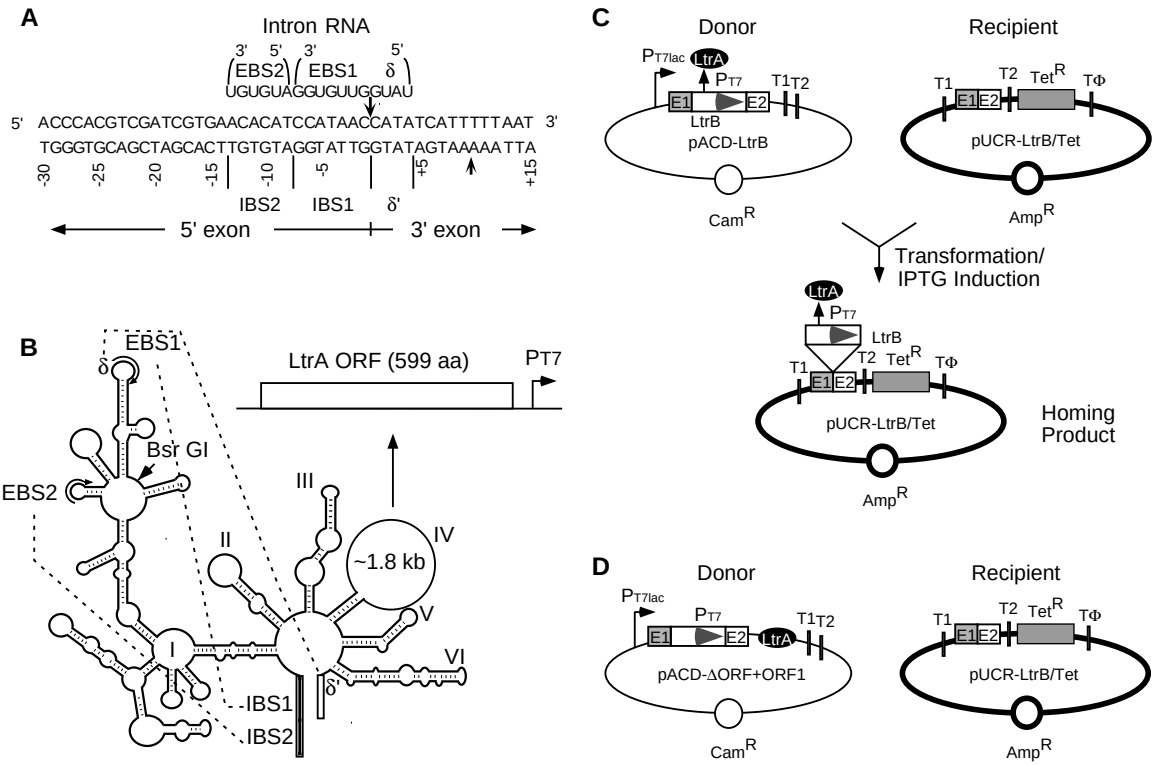

Fig. 1. *Escherichia coli* genetic assay based on the Ll.LtrB intron for analyzing group II intron–DNA target site interactions. (**A**) Natural Ll.LtrB DNA target sequence from position −30 to +15 and base-pairing interactions with the intron RNA. Sequence elements IBS2 and IBS1 in the 5′ exon and δ′ in the 3′ exon of the DNA target are recognized primarily by base pairing with sequence elements EBS2, EBS1, and δ located in domain I of the intron RNA. The intron insertion site in the top (sense) strand and the endonuclease cleavage site in the bottom (antisense) strand are indicated by arrows. (**B**) Schematic of the Ll.LtrB intron showing base-pairing interactions EBS1-IBS1, EBS2-IBS2, and δ-δ′ between the intron and flanking exons. The inset shows the location of the LtrA ORF and the T7 promoter introduced into intron domain IV in donor plasmids. (**C**) Genetic assay. The donor plasmid pACD-LtrB is a Cam^R pACYC184 derivative containing the full-length Ll.LtrB intron and flanking exons, with a phage T7 promoter inserted downstream of the LtrA ORF in intron domain IV (*12*). The intron and flanking exon sequences (E1 and E2) are cloned behind a T7lac promoter, and *E. coli rrn*B T1 and T2 transcription terminators are positioned downstream of the intron. The recipient pUCR-LtrB/Tet is a compatible Amp^R plasmid with an Ll.LtrB target sequence (ligated *ltrB* exons E1 and E2) cloned upstream of a promoterless tet^R gene (*13*). An *E. coli rrn*B T1 transcription terminator, which terminates both *E. coli* and T7 RNA polymerase, is inserted upstream of the target site, and an *rrn*B T2 terminator, which terminates *E. coli* but not T7 RNA polymerase, is inserted between the target site and the tet^R gene. A phage T7 TΦ terminator is inserted downstream of the tet^R gene to terminate T7 RNA polymerase. Movement of the intron carrying the phage T7 promoter into the DNA target site activates expression of the tet^R gene. (**D**) Mobility assay using pACD-ΔORF+ORF1. This plasmid has a deletion in the loop of intron domain IV, which removes most of the LtrA ORF, and the LtrA protein is expressed separately from a position downstream of the 3′ exon. This configuration gives higher mobility frequencies approaching 100%.

ognition rules for the wild-type Ll.LtrB intron. We did so by performing mobility assays with recipient plasmids in which positions −30 to +15 of the Ll.LtrB target site were partially randomized (30% "doped" with non–wild-type nucleotide residues) (*19*). The data showed that the wild-type nucleotide residues between positions −24 and +7 were selected to different degrees. In the protein-recognition regions, the most critical positions were readily identified as G −21 and T +5, which were shown previously to be stringently required for reverse splicing into the DNA target site and antisense-strand cleavage, respectively (*11*). In the region of the DNA target site potentially recognized by base pairing with the intron RNA, the data showed strong or moderate selection against nucleotide substitutions at positions −13 to −8 in IBS2, −6 to −1 in IBS1, and +1 to +4 in the δ′ region. Comparison of the number of potential base pairs in selected target sites and the original recipient pool showed some selection for base pairing at each position between −13 and +4, except for −7, which instead showed clear selection against base pairing (Fig. 2A). Over the 16 positions potentially recognized by base pairing, 99% of the selected target sites have 13 or more potential base pairs, and none has less than 12 potential base pairs (Fig. 2B). Because δ-δ′ is not essential for RNA splicing in vitro or in vivo (*20*), the extended δ-δ′ interaction, potentially involving positions +1 to +4, is presumably required primarily for reverse splicing into DNA.

As an initial test of the targeting rules, we designed group II introns targeted to different regions of HIV-1_{LAI} provirus and the human gene encoding the CCR5 chemokine receptor. The latter is required together with CD4 for infection of macrophages by HIV-1, and it has been shown that individuals homozygous for *CCR5* mutations are resistant to HIV-1 infection while having no other pathologies (*21, 22*). Consequently, disabling *CCR5* has been considered a means to block HIV-1 infection and the progression of acquired immunodeficiency syndrome.

For targeting, the HIV-1_{LAI} and *CCR5* DNA sequences were scanned for the best matches to the fixed positions recognized by the IEP, and the intron RNA was then modified to form base pairs with the adjacent sequences for the EBS-IBS and δ-δ′ interactions (positions −11 to −8 and −6 to +3 or +4). The data from the initial selection experiment with partially randomized DNA target sites (*19*) were used to obtain a quantitative measure of the ability to substitute a nucleotide residue at a particular position. This "mutability value" was calculated by comparing the ratios (R) of mutant to wild-type nucleotide residues in active target sites and the initial recipient pool, using the expression $[(R_{mut/wt})_{active}/(R_{mut/wt})_{pool}] - 1$ (*23*). To select target sites in the HIV-1_{LAI} and *CCR5* DNAs, we initially used a mutability value of −0.6 as a lower limit for nucleotide substitutions at positions recognized by the IEP (taken as −30 to −12 and +4 to +7 in initial experiments). The resulting search sequence, 5′-N_7(G,T,C)N(G,A)(A,T,C)-N_2(G,A)(A,T,C)(G,A,C)N_2CN_{11}↓N_3(G,A,C)-T(A,T,C)N, where N represents any of the four bases and the downward arrow indicates the intron insertion site, gave 18 matches in HIV-1_{LAI} and two in *CCR5*. The HIV-1 target sites were ordered by using successively more stringent cutoff values, and introns targeted to the two best sites along with the two *CCR5* sites were tested for their ability to insert into HIV-1_{LAI} and *CCR5* DNA targets in the *E. coli* genetic assay (Fig. 3) (*24*). Also tested was HIV1-4069s, in which position −12 did not meet the −0.6 cutoff, to determine whether the GC base pair at this position, which was highly conserved in the initial selection experiment (*19*), could be replaced with a compensatory AT base pair. Because the retargeted introns have modified EBS and δ sequences, complementary IBS and +1 sequences were introduced into the donor plasmid to ensure efficient splicing.

All of the retargeted introns inserted at precisely the correct positions in the HIV-1 and *CCR5* target sites, as confirmed by sequencing multiple events. Two introns, HIV1-4021s and HIV1-4069s, inserted at high frequencies (>60% after IPTG induction); the remaining three introns (CCR5-1019s, CCR5-759a, and HIV1-3994s) inserted at lower frequencies (0.16 to 10.6%). The two most efficient introns have compensatory changes at EBS-IBS and δ-δ′ positions −12, −6, or +1, where the wild-type nucleotide was strongly conserved in the initial selection (*19*); this finding indicates that protein recognition at these positions is not essential for efficient integration. The less efficient introns presumably have deleterious combinations of nucleotides that are not readily predicted at this stage.

To alleviate the necessity of predicting such deleterious combinations, we developed an alternate, selection-based approach in which the desired DNA target site is simply cloned in the recipient vector upstream of the promoterless *tet*^{R} gene, and introns that insert into that site are selected from a combinatorial library having randomized target site recognition sequences (EBS and δ) (*25, 26*). The corresponding IBS sequences in the 5′ exon of the donor plasmid library were also randomized to eliminate selection for the wild-type EBS sequences during RNA splicing in vivo. Although the requirement for base pairing between the two sets of randomized sequences in unspliced precursor RNA reduces the complexity of the spliced intron pool, the approach was successful because of the very high integration efficiency in this system.

Single transformations with the combinatorial library yielded 13 introns that inserted at different positions in the HIV-1_{LAI} and *CCR5* target sites. The introns were retested individually and were shown to integrate into their target sites at frequencies ranging from <0.001% to 53%. Data for a subset of the selected introns are summarized in Fig. 3. Most of the efficient introns insert into target sites having the G −21 and T +5 residues, which were found to be critically required for protein recognition [see above and (*11*)]. However, introns HIV1-54a/9186a and CCR5-24s deviate at these positions but still insert at frequencies of 3 to 5%, possibly reflecting partial compensation by other target site nucleotide residues. The most efficient *CCR5* target site has a disfavored nucleotide residue at position −16, which excluded it from the initial computer search for potential target sites. Two of the selected HIV-1 introns have mismatches in EBS-IBS and δ-δ′ interactions, and their integration efficiencies increased substantially when these were "cor-

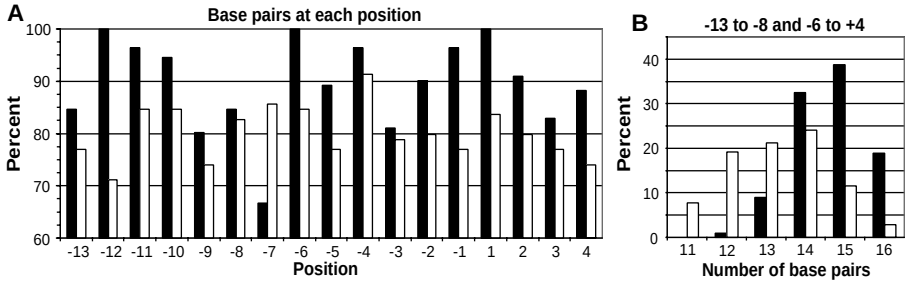

Fig. 2. Base-pairing requirements for different positions of the DNA target site. A mobility assay was performed with the wild-type donor plasmid pACD-ΔORF+ORF2 (*17*) and a recipient plasmid pool in which DNA target site positions between −30 and +15 were partially randomized ("doped") to contain 70% of the wild-type nucleotide and 10% each of the three mutant nucleotides (*19*). The number of potential base pairs with the intron RNA was compared in active target sites (black bars) and the original recipient pool (white bars). (**A**) Percentage of target sites having a potential base pair with the intron RNA at each position between −13 and +4. (**B**) Percentage of target sites having the indicated number of potential base pairs over the indicated interval. Selection for base pairing in the active target sites is evident at all positions except −7.

rected." Such correction can be effected routinely with the use of appropriate polymerase chain reaction (PCR) primers in the process of recloning the selected intron into the donor plasmid. Although the selected introns have a range of mobility frequencies, it should be relatively straightforward to enrich for the most efficient introns by carrying out multiple rounds of selection.

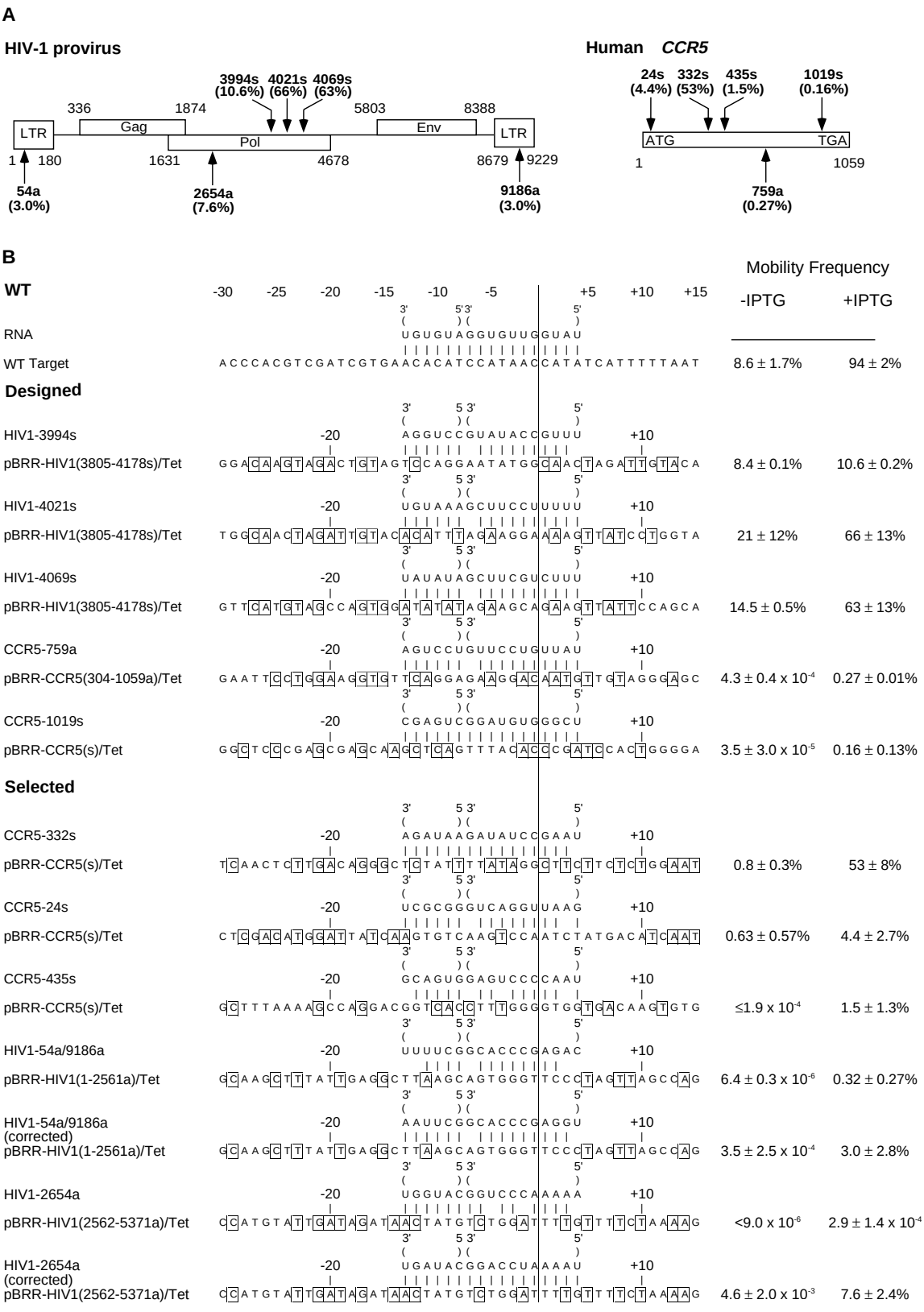

Fig. 3. Design and selection of group II introns that insert into specific DNA target sites. (**A**) Maps showing group II intron insertion sites in the HIV-1 provirus and human *CCR5* gene. Insertion sites in the top (sense) and bottom (antisense) strands are indicated by arrows above and below the target DNA, respectively. Introns are identified by the position number in the target site (HIV-1 sequence, GenBank accession number K02013; *CCR5* sequence, GenBank accession number AF031237), followed by "s" or "a" indicating the sense or antisense strand, respectively. The numbers in parentheses indicate mobility frequencies in the *E. coli* genetic assay in the presence of 100 μM IPTG. The intron HIV1-54/9186a has integration sites in each long terminal repeat (LTR). (**B**) DNA target site sequences and base-pairing interactions for designed and selected introns. The wild-type Ll.LtrB target site and base-pairing interactions are shown above for comparison. Nucleotide residues in the HIV-1 and *CCR5* target sites that match the wild-type sequence are boxed. Mobility frequencies in the presence or absence of 100 μM IPTG (mean ± SD for at least two experiments) are shown to the right. Mobility events were confirmed by sequencing a region extending from a position downstream of the intron's EBS sequences through the 5′ junction with the target DNA, using primer LtrBA2 (complementary to intron positions 301 to 326). For determination of mobility frequencies, the selected introns were reconstructed in the donor plasmid by PCR and tested in the *E. coli* genetic assay with the indicated recipient plasmids containing the HIV-1 or *CCR5* target sites (*26*). The selected introns HIV1-54a/9186a and 2654a have mismatches in the EBS-IBS and δ-δ′ interactions, and their mobility frequencies increased substantially when these were "corrected."

Ultimately, we hope to impede HIV-1 replication in patients by disrupting the HIV-1 coreceptor *CCR5* gene or HIV-1 proviral DNA through targeted intron insertion into DNA. To determine whether group II intron RNPs can function in a human cellular environment, we cotransfected 293 embryonic kidney cells or CEM T cells with plasmids containing either *CCR5* or HIV-1 DNA targets and RNP particles containing the retargeted introns (CCR5-332s or HIV1-4069s), which had been packaged separately into liposomes (*27*). PCR analysis of DNA isolated from the transfected cells gave products expected for integration of the introns into the DNA target sites (Fig. 4A) (*28*). By contrast, such products were not detected with DNA from mock-transfected cells, from the transfection mix incubated without cells, or from cells that were transfected separately with either target DNA or RNPs and mixed before DNA extraction. Restriction enzyme analysis and sequencing confirmed that the retargeted introns had inserted at the correct locations in the *CCR5* and HIV-1 DNAs (Fig. 4, A and B).

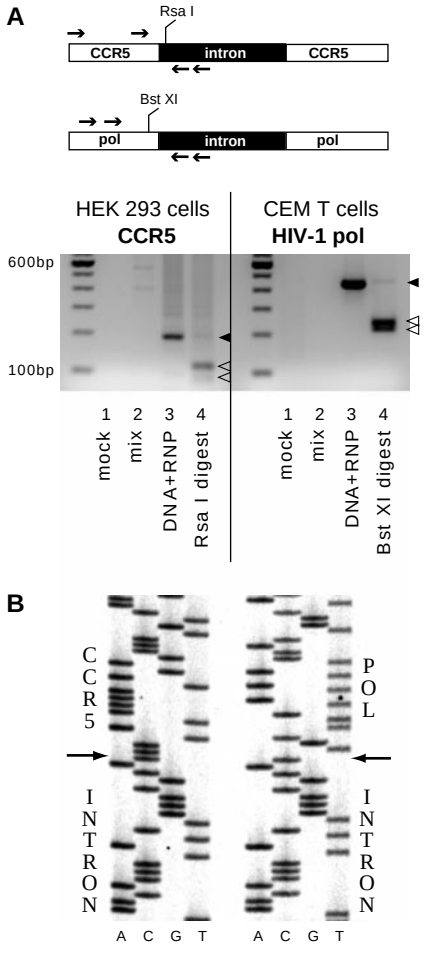

Fig. 4. Intron insertion into *CCR5* and HIV-1 *pol* genes in human cell lines. (**A**) PCR amplification of integration junctions. Integration events into the *CCR5* and HIV-1 *pol* genes are shown schematically at the top. Primer sites are designated by arrows, and restriction sites are labeled. PCR products were analyzed in a 2% agarose gel. Solid arrowheads indicate PCR products corresponding to 5′ integration junctions in the *CCR5* (left) and HIV-1 *pol* (right) genes. Open arrowheads indicate the restriction fragments of integration products. The left lane shows molecular mass markers. (**B**) Representative sequencing gels for intron integration into the *CCR5* (left) and HIV-1 *pol* (right) genes. The arrows denote integration junctions.

Our results establish that group II introns can be retargeted to insert efficiently into desired DNA target sites and that the intron RNP particles retain activity in human cells. Because the number of obligatory fixed positions is relatively small (<5), potential target sites should exist in most genes, and the overall target sequence should be unique even in large genomes. On the basis of the present results, target site recognition appears to be sufficiently malleable to obtain group II introns that insert into any desired gene. Attractive features of this system for genetic manipulation are the wide host range of the Ll.LtrB intron, an insertion mechanism that is independent of host-cell recombination functions, and the ability to readily introduce additional genetic markers into intron domain IV (*5, 8*). Introns inserted in the antisense orientation cannot splice and yield unconditional disruptions, whereas those inserted in the sense orientation could be used for conditional disruptions by linking intron splicing to synthesis of the LtrA protein under the control of an inducible promoter. Intron libraries with randomized target site recognition sequences would generate insertions at random chromosomal locations. In other work, we have used group II introns for efficient chromosomal gene disruption in bacteria (*29*), and experiments are in progress to determine whether eukaryotic chromosomal genes can be targeted similarly by optimizing the introduction or intracellular expression of group II intron RNP particles. Such methods would greatly facilitate analysis of eukaryotic organisms that lack efficient homologous recombination systems (flies, worms, mice, plants, and human cells) and potentially have direct therapeutic applications.

References and Notes

1. F. Michel and J. L. Ferat, *Annu. Rev. Biochem.* **64**, 435 (1995).
2. A. M. Lambowitz, M. G. Caprara, S. Zimmerly, P. S. Perlman, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 1999), pp. 451–485.
3. S. Zimmerly, H. Guo, P. S. Perlman, A. M. Lambowitz, *Cell* **82**, 545 (1995).
4. S. Zimmerly *et al.*, *Cell* **83**, 529 (1995).
5. M. Matsuura *et al.*, *Genes Dev.* **11**, 2910 (1997).
6. J. Yang, S. Zimmerly, P. S. Perlman, A. M. Lambowitz, *Nature* **381**, 332 (1996).
7. R. Eskes, J. Yang, A. M. Lambowitz, P. S. Perlman, *Cell* **88**, 865 (1997).
8. B. Cousineau *et al.*, *Cell* **94**, 451 (1998).
9. H. Guo, S. Zimmerly, P. S. Perlman, A. M. Lambowitz, *EMBO J.* **16**, 6835 (1997).
10. J. Yang, G. Mohr, P. S. Perlman, A. M. Lambowitz, *J. Mol. Biol.* **282**, 505 (1998).
11. G. Mohr, D. Smith, M. Belfort, A. M. Lambowitz, *Genes Dev.* **14**, 559 (2000).
12. Donor plasmids contain the Ll.LtrB intron or its derivatives with a phage T7 promoter (annealed oligonucleotides 5′-TCGACTAATACGACTCACTATAGGG and 5′-TCGACCCTATAGTGAGTCGTATTAG) inserted into an engineered Sal I site immediately downstream of the stop codon of the LtrA ORF in intron domain IV (Fig. 1, B and C) (*8*). The wild-type donor plasmid, pACD-LtrB, contains the full-length Ll.LtrB intron, flanked by 177 nt of exon 1 (E1) and 90 nt of exon 2 (E2), cloned downstream of the T7lac promoter in a pACYC184 derivative carrying a *cam*^{R} gene. The LtrA ORF was derived from plasmid pLIP3, where it had been modified via silent mutations to introduce a series of unique restriction sites for cassette mutagenesis (H. Ma and A. M. Lambowitz, unpublished data). *Escherichia coli rrn*B T1 and T2 transcription terminators were inserted downstream of the intron between the Bam HI and Hinc II (position 3211) sites of pACYC184.
13. Recipient plasmid pUCR-LtrB/Tet is a pUC19 derivative containing the Ll.LtrB target site [ligated E1-E2 sequence from position −178 to +91; D. A. Mills, D. A. Manias, L. L. McKay, G. M. Dunny, *J. Bacteriol.* **179**, 6107 (1997)] cloned upstream of a promoterless *tet*^{R} gene (Fig. 1C). The Ll.LtrB target site was inserted in a 35–base pair (bp) polylinker in the previously constructed vector pUCR-Tet. This vector contains an *E. coli rrn*B T1 transcription terminator, which terminates both *E. coli* and T7 RNA polymerase, upstream of the polylinker, and an *rrn*B T2 terminator, which terminates *E. coli* but not T7 RNA polymerase, between the polylinker and the *tet*^{R} gene [A. Orosz, I. Boros, P. Venetianer, *Eur. J. Biochem.* **201**, 653 (1991); L. Hartvig and J. Christiansen, *EMBO J.* **15**, 4767 (1996)]. The *tet*^{R} gene is preceded by a T7 TΦ terminator to terminate T7 RNA polymerase [F. W. Studier, A. H. Rosenberg, J. J. Dunn, J. W. Dubendorff, *Methods Enzymol.* **185**, 60 (1990)]. For the nucleotide sequence of the target site region of pUCR-LtrB/Tet, see *Science* Online (www.sciencemag.org/feature/data/1050641.shl). pBRR-Tet is a lower copy number vector derived from pUCR-Tet by swapping a 1.9-kb Alw NI–Bam HI fragment containing the pUC19 replication origin for the corresponding fragment of pBR322. This swap necessitated reinserting a T7 TΦ terminator at the Sty I site downstream of the *tet*^{R} gene.
14. For mobility assays, *E. coli* HMS174(DE3) (Novagen, Madison, WI) was cotransformed with donor and recipient plasmids and grown in LB medium with chloramphenicol (25 μg/ml) and ampicillin (100 μg/ml) overnight (16 to 20 hours, 37°C). A 50-μl sample of the overnight culture was inoculated into 10 ml of LB with the indicated antibiotics and grown to an optical density of 0.2 to 0.3 at 600 nm; at that point, 250 μl was inoculated into 5 ml of fresh LB without antibiotics and induced with 100 μM or 2 mM IPTG for 1 hour at 37°C. Cells were harvested by centrifugation, washed, resuspended in 10 ml of ice-cold LB, and plated on LB plus ampicillin in the presence or absence of tetracycline (25 μg/ml). Mobility frequencies were calculated as the proportion of Tet^R colonies. The length of induction and IPTG concentration were critical parameters, because overinduction leads to loss of plasmids carrying T7 promoters as well as decreased cell viability [F. W. Studier and B. A. Moffatt, *J. Mol. Biol.* **189**, 113 (1986)] and possibly to toxic effects of the Ll.LtrB intron.
15. pACD-ΔORF has amino acids 40 to 572 of LtrA replaced with TR, and pACD-YAAA has the conserved YADD in the RT domain changed to YAAA (*5*) (where A = Ala, D = Asp, R = Arg, T = Thr, and Y = Tyr). pACD-ΔD5 is deleted for intron domain V (positions 2398 to 2429). pACD-ΔConZn and pACD-ΔZn express truncated LtrA proteins lacking part or all of the Zn domain (positions 2200 to 2205 and 2065 to 2070, respectively, changed to consecutive stop codons

TGA-TAA). We confirmed by SDS–polyacrylamide gel electrophoresis (PAGE) that the mutant donor plasmids express wild-type levels of LtrA protein.
16. In 29 of 39 mobility products, the group I intron spliced using the normal 5′ and 3′ splice sites, and eight used cryptic 5′ splice sites at LI.LtrB positions 1810 (six clones), 1985 (one clone), and 2185 (one clone).
17. pACD-ΔORF+ORF1 (Fig. 1D) is a derivative of pACD-ΔORF (*15*). The donor intron has a deletion in the loop of intron domain IV, which removes most of the LtrA ORF, but retains regions of domain IV that are required for binding the LtrA protein [H. Wank, J. San Filippo, R. N. Singh, M. Matsuura, A. M. Lambowitz, *Mol. Cell* **4**, 239 (1999)]. An intact LtrA ORF was cloned downstream of the 3′ exon in the same plasmid (1.9-kb Sca I–Sal I fragment of pACD-LtrB filled in with Klenow polymerase inserted into the Fsp I site). pACD-ΔORF+ORF2 is a derivative of pACD-ΔORF+ORF1 that contains shorter exons (26 bp of E1 preceded by a Hind III site and 11 bp of E2 followed by a Pst I site) to facilitate manipulation of the IBS and δ′ sequences.
18. Reverse splicing assays showed that RNP particles from cells expressing pACD-ΔORF+ORF1 (ΔORF intron) insert intact intron RNA, whereas those from cells expressing pACD-LtrB (full-length intron) insert partially degraded intron RNA (*5*). SDS-PAGE showed that pACD-ΔORF+ORF1 and pACD-LtrB express similar levels of LtrA protein.
19. For supplementary data, see *Science* Online (www.sciencemag.org/feature/data/1050641.shl).
20. G. Mohr, X. Cui, A. M. Lambowitz, unpublished data.
21. M. Dean *et al.*, *Science* **273**, 1856 (1996).
22. R. Liu *et al.*, *Cell* **86**, 367 (1996).
23. S. J. Deminoff, J. Tarnow, G. M. Santangelo, *Genetics* **141**, 1263 (1995).
24. Recipient plasmids containing HIV-1 and *CCR5* target sites were constructed by inserting the indicated segments of the target DNA into the Xho I site of the polylinker of pBRR-Tet (*13*) by blunt-end ligation after filling in sites with T4 DNA polymerase. Target DNAs were inserted in both the sense and antisense orientations relative to the *tet*^{R} gene. HIV-1 and *CCR5* target DNAs were derived from pLAI3 [K. Peden, M. Emerman, L. Montagnier, *Virology* **185**, 661 (1991)] and pCMV-CKR5 [P. Bieniasz *et al*., *EMBO J.* **16**, 2599 (1997)], respectively. For targeting experiments, the donor plasmid pool was electroporated into cells containing the desired recipient plasmid.
25. The combinatorial intron library pACD-DL was constructed in pACD-ΔORF+ORF2 (*17*) by inserting random sequences in place of EBS and δ positions (EBS2, −13 to −8; EBS1, −6 to −1; and δ, +1 to +4) that potentially form base pairs with the DNA target site. The corresponding 5′ exon positions in IBS1 and IBS2 of the precursor RNA were also randomized to provide nucleotide combinations that could form base pairs with the randomized EBS sequences for RNA splicing. The randomized sequences were introduced in a two-step PCR, in which the first step used primers that contain the randomized sequences to generate two partially overlapping PCR products (positions −26 to 222 and 202 to 318). The two PCR products were mixed and amplified with external primers, which introduce Hind III and Bsr GI sites, and then cloned between the corresponding sites of pACD-ΔORF+ORF2. The ligation products were electroporated into *E. coli* DH10B, and the library was amplified in LB medium containing chloramphenicol.
26. For selection experiments, the donor plasmid library containing ~2×10^8 different introns was transformed into *E. coli* HMS174(DE3) containing recipient plasmids with HIV-1 or *CCR5* DNAs inserted upstream of the promoterless *tet*^{R} gene [pBRR-HIV1(1-2561a)/Tet, pBRR-HIV1(2562-5371a)/Tet, or pBRR-CCR5(s)/Tet) (*24*)]. About 10^8 Amp^{R}Cam^R transformants were obtained. The frequency of Tet^R colonies was 2.5×10^{-6}, with 40% having mobility events. The integrated introns were then reconstructed in the pACD-ΔORF+ORF2 donor plasmid (*17*) by PCR amplifying a segment extending from position −13 in the target site to intron position 326 downstream of the EBS sequences, by using the 3′ primer LtrBA2 in combination with a target site specific 5′ primer. The latter has a 5′ Hind III site, followed by wild-type 5′ exon positions −26 to −14, and the target site junction sequence from −13 to +18 in the intron. The PCR products were digested with Hind III and Bsr GI, and swapped for the corresponding segment of pACD-ΔORF+ORF2. Mobility frequencies for the selected introns were determined in the *E. coli* genetic assay using recipient plasmids containing the HIV-1 or *CCR5* DNA target sites (*24*).
27. Human embryonic kidney 293 cells (10^6) were transfected with the *CCR5* cDNA plasmid pT7CCR5 (1 μg) and/or RNP particles containing intron CCR5-332s (1.2 μg) in 12 μl of DMRIE-C (Gibco-BRL) for 4 hours. CEM T cells (3×10^5) were transfected with pLAIgag-pol (0.5 μg) and/or RNP particles containing intron HIV1-4069s (0.5 μg) in 12 μl of DMRIE-C for 5 hours. In both cases, DNA was incubated with half the total lipid for 15 to 30 min, and RNPs were incubated with the remaining half just before mixing and transferring to cells, to minimize the possibility that reverse splicing occurs extracellularly. DNA was isolated from transfected cells after 24 hours using a QIAamp DNA Blood Kit (Qiagen). The RNP particles used in the transfections were reconstituted with in vitro spliced intron RNA and purified IEP, as described [R. Saldanha *et al.*, *Biochemistry* **38**, 9069 (1999)].
28. Nested PCR (total of 60 cycles) was carried out on DNA samples treated with ribonuclease. Upstream primers were specific to target DNA sequences (*CCR5* primer, 5′GCGCGGATCCATGGATTATCAAGTGTCAAG; nested primer, 5′-CCGAAGCTTGTCAACTCTTGACAGGGCTC) (HIV-1 *pol* primer, GGGGGGATCCACACAAAGGAATTGGAGG; nested primer, GGGGGGATCCGTCAGTGCTGGAATCAGG). Downstream primers were specific to intron sequences (primer, 5′-GGCCCTTTCGTTTCGTTC; nested primer, 5′-GGCCCGATTGTCTTTAGGTA). PCR products containing the integrated CCR5 intron were cut with Rsa I at a site located in the intron, and those containing the integrated HIV-1 pol intron were cut with Bst XI at a site located in the HIV-1 *pol* sequence. PCR products were cloned by standard techniques and sequenced using a Thermo Sequenase Radiolabeled Cycle Sequencing kit (USB).
29. M. Karberg, H. Guo, R. Coon, A. M. Lambowitz, in preparation.
30. We thank J. Edwards for technical assistance and M. Belfort (Wadsworth Center, Albany, NY) for comments on the manuscript. Supported by NIH grants GM37949 (A.M.L.) and AI40981 (B.S.).

21 March 2000; accepted 18 May 2000

A Neural Basis for General Intelligence

John Duncan,[1]* Rüdiger J. Seitz,[2] Jonathan Kolodny,[1] Daniel Bor,[1] Hans Herzog,[3] Ayesha Ahmed,[1] Fiona N. Newell,[1] Hazel Emslie[1]

Universal positive correlations between different cognitive tests motivate the concept of "general intelligence" or Spearman's *g*. Here the neural basis for *g* is investigated by means of positron emission tomography. Spatial, verbal, and perceptuo-motor tasks with high-*g* involvement are compared with matched low-*g* control tasks. In contrast to the common view that *g* reflects a broad sample of major cognitive functions, high-*g* tasks do not show diffuse recruitment of multiple brain regions. Instead they are associated with selective recruitment of lateral frontal cortex in one or both hemispheres. Despite very different task content in the three high-*g*–low-*g* contrasts, lateral frontal recruitment is markedly similar in each case. Many previous experiments have shown these same frontal regions to be recruited by a broad range of different cognitive demands. The results suggest that "general intelligence" derives from a specific frontal system important in the control of diverse forms of behavior.

As discovered by Spearman (*1*) early in the last century, measures of performance or success in diverse cognitive tests show a pattern of almost universal positive correlation: To some extent at least, the same people tend to perform well in very different tasks. To explain this result, Spearman put forward the hypothesis of a general or *g* factor making some contribution to success in diverse forms of cognitive activity. People with high *g* scores will be those usually performing well, leading to the interpretation of *g* as "general intelligence." Factor analysis can be used to show which tasks are most correlated with *g* and are thus the best general intelligence measures; often, these turn out to be tests of novel problem solving such as Raven's Progressive Matrices (*2*). An alternative hypothesis, originally proposed by Thomson (*3*), has also received detailed consideration. According to this hypothesis, any task receives contributions from a large set of component factors or information-processing functions. Universal positive correlation arises not for any common reason, but simply because any two tasks are likely to share at least some components. The "*g* factor" measured by standard intelligence tests is now interpreted as the average efficiency of the total set of cognitive functions (*4*); as Thomson showed, tasks with high apparent *g* correlations will be those sampling the total set

[1]Medical Research Council Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK. [2]Department of Neurology, Heinrich-Heine-University Düsseldorf, Moorenstrasse 5, D-40225 Düsseldorf, Germany. [3]Institute of Medicine, Research Center Jülich, D-52407 Jülich, Germany.

*To whom correspondence should be addressed. E-mail: john.duncan@mrc-cbu.cam.ac.uk