Supporting Online Material for

Recombination of Retrotransposon and Exogenous RNA Virus Results in Nonretroviral cDNA Integration

Markus B. Geuking,* Jacqueline Weber, Marie Dewannieux, Elieser Gorelik, Thierry Heidmann, Hans Hengartner, Rolf M. Zinkernagel,* Lars Hangartner*

*To whom correspondence should be addressed. E-mail: hangartner.lars@virology.uzh.ch (L.H.); geuking@mcmaster.ca (M.B.G.); rolf.zinkernagel@usz.ch (R.M.Z.)

Published 16 January 2008, Science 323, 393 (2009)
DOI: 10.1126/science.1167375

This PDF file includes:

Materials and Methods
SOM Text
Figs. S1 to S3
Tables S1 and S2
References
Recombination of retrotransposon and exogenous RNA virus results in non-retroviral cDNA integration

Markus B. Geuking, Jacqueline Weber, Marie Dewannieux, Elieser Gorelik, Thierry Heidmann, Hans Hengartner, Rolf M. Zinkernagel, and Lars Hangartner

Materials and Methods

Cell lines and virus
MC57G (ATCC CRL-2295) is a methylcholanthrene-induced fibrosarcoma cell line of C57BL/6 origin, Vero (ATCC CCL-81) is a green monkey kidney cell line, HeLa (ATCC CCL-2) is a human-derived tumor cell line. MDCK (ATCC CCL-34) is a canine cell line, and CHO is a Chinese hamster-derived cell line. LCMV strain WE was originally provided by F. Lehmann-Grube, Hamburg, Germany, and was grown on L929 cells (ATCC CCL-1) for 48 hours in minimal essential medium containing 5% FCS after infection with an initial multiplicity of infection of 0.01.

Mice
C57BL/6 and RAG1−/− mice were bred and maintained at the Biologisches Zentrallabor of the University of Zurich. Mice were infected with 200pfu LCMV-WE intravenously. All animals were used in accordance with the guidelines of the Kanton Zurich and the Swiss Veterinary Office animal protection laws.

Transfections and IAP-containing plasmids
The various IAP element-containing plasmids 213P12, 440N1, 262J21, 92L23, and 92L23Neo were constructed and kindly provided by M. Dewannieux (1). The plasmids containing the IAP mutants 440N1ΔRT, 440N1LinkerStopRT, and 440N1ΔGag were also provided by M. Dewannieux. The Human Immunodeficiency Virus 1 (HIV-1) Gag-Pol plasmid pCMVΔR8 containing all HIV-1 genes except the envelope gene (2) was kindly provided by Ekaterina Safroneeva. The control plasmid pEGFP-N3 was purchased from Clontech. For transfection of LCMV-infected cells, 5x10^5 or 10^5 cells were seeded into individual 6-well or 24-well dishes, respectively. LCMV-infected samples were infected at a multiplicity of infection of 0.001 on the same day. Cells were incubated overnight at 37°C and 5% CO₂. Medium was then replaced with 2ml of DMEM or IMDM devoid of serum and antibiotics followed by incubation at 37°C and 5% CO₂ for at least one hour. Lipofectamin liposomes for transfection of the various tested plasmids were prepared according to the manufacturer’s instructions in OptiMEM medium (Invitrogen). The serum- and
antibiotic-free medium was then replaced with the transfection mixtures, and cells were incubated for at least five hours at 37°C and 5% CO₂. Transfection mixes were then replaced by 2ml of DMEM or IMDM containing 10% FCS and tissue culture antibiotics, and incubated at 37°C and 5% CO₂. Transfection efficiency was determined using EGFP-transfections. Three to four days later, DNA was extracted using a protocol involving SDS-mediated lysis, Proteinase K digestion, salting out of proteins and SDS with NaCl, isopropanol precipitation, washing of the nucleic acid with 70% EtOH and dissolving the DNA in Tris-EDTA buffer after air drying. 100 to 200 ng genomic DNA from transfected cells was then used as template for PCR. Infection of cells with LCMV-WE did not alter transfection efficiency and transfection did not interfere with LCMV-replication as determined by measuring the viral titer in the supernatant of transfected and non-transfected HeLa cells.

**Generation of MC57.18 and MC57.23**

MC57G cells were infected with LCMV-WE at a multiplicity of infection of 0.001 and cultured for three days before they were divided into pools of 100 cells. Following expansion in 96 well plates, pools were screened by nested PCR for the presence of cGP sequences. Primers RC1 (5’ GAG CTC TGC AGC AAG GAT CAT CC 3’) and DM1 (5’ GGA TTC TAT CCA ATA AAA GGA TGG 3’) were used for the first round of amplification and primers WEU76 (5’ GCC ATG GGT CAG ATT GTG ACA ATG 3’) and RCM (5’ GGT ACT GAT AGC TTG TTT GGC TGC ACC 3’) were used for the second round of amplification. The cycling conditions for the two rounds were (4min 94°C; (45sec 94°C, 45sec 50°C, 45sec 72°C)x35, 10min 72°C) and (4min 94°C; (30sec 94°C, 30sec 60°C, 30sec 72°C)x28, 10min 72°C), respectively. Positive pools were then subcloned, expanded, re-screened and subcloned again using limiting dilution. Two independent stable cGP positive cell lines (MC57.18, and MC57.23) were obtained. The same nested PCR specific for GP was used to test for cGP formation in IAP-transfectants.

**Inverse PCR to identify cGP-flanking sequences in clone MC57.18 and MC57.23**

15µg of genomic DNA from clone MC57.18 or MC57.23 were digested with NcoI (not present in the GP gene). 1.125µg of NcoI fragments were self-ligated in a total volume of 400µl favoring intra- over intermolecular ligation. The cGP-flanking sequences were amplified by long-range PCR (TaqPlus® Long, Stratagene) using the inverse primer pair WE333r (5’ AGT CGT TGG TGA AAG TTG GCT CCA G 3’) and WE590f (5’ CTT GTC ATC TTC GGA CCC ACA GAG C 3’) and the high-salt buffer provided by the manufacturer. In order to obtain extended sequence information of the cGP-flanking region beyond the NcoI site in clone MC57.18, a long-range PCR on genomic DNA was performed using the forward primer
WE924f (5’ GGT TAT TGC CTG ACC AAA TGG A 3’) and the reverse primer LTR3’ (5’ TGT TAT TAG ACG CGT TCT CAC G 3’). All PCR products were subcloned into pGEM-T (Promega) and sequenced.

**Nested PCRs to detect GP/IAP and NP/IAP hybrid DNA**

The nested PCR detecting NP/IAP hybrid sequences consisted of the following two primer pairs: WE2180r (5’ AGT ATT CAC ACG GCA TGG A 3’) x psenv1f (5’ CGG GAG CAG GTC AGG GTT G 3’) and WE2139r (5’ CAG GCT GGG CTG ACC TCA T 3’) x psenv2f (5’ AGG TCA GGG TTG CTC TGG G 3’). The cycling condition for the first and second amplification was (4min 94°C; (1min 94°C, 1min 50°C, 1min 72°C)x35, 10min 72°C). To detect GP/IAP hybrid sequences the primer pairs WE(HPI)1619f (5’ CGG TTC ATG TCC AAA GCC AC 3’) x IAPLTRfwd1 (5’ GCT GCA GCC AAT CAG GGA G 3’) and WE(HPI)1652f (5’ CAA GGG GAT CTG TAG T 3’) x IAPLTRfwd2 (5’ AAG ATT CTG GTC TGT GGT G 3’) were used. The cycling conditions for the first and second amplification were (4min 94°C; (30sec 94°C, 45sec 60°C, 2min30sec 72°C)x35, 10min 72°C). PCR products were directly sequenced or subcloned into pGEM-T (Promega) and sequenced using primers M13fwd and M13rev. Sequences were analyzed using Sequencher (Gene Codes Corporation, Ann Arbor, MI, U.S.A.) or LaserGene software (DNASTAR, Inc., Madison, WI, U.S.A.).
Supporting text 1

**Attempts to obtain 3’ flanking sequence in MC57.18 reveal potential large deletion on chromosome 10**

In order to obtain 3’ flanking sequences in MC57.18, different PCRs, including long-range and nested PCR, using a forward primer within the cGP/IAP and a reverse primer within the putative chromosome 10 sequences were performed (Table S1 and Table S2). These PCRs span a region of 143 to 1946bp downstream of the cLCMV/IAP integration site on chromosome 10. None of the PCRs yielded a product, indicating the presence of a large chromosomal deletion that might have been generated during the integration process. Together with the smaller deletions shown in Fig. 1 this indicated that integration of the cLCMV/IAP recombination product might have occurred at sites of DNA damage.

Supporting text 2

**Potential integration mechanism**

It is important to note that the sequence of the IAP primer-binding site of 92L23 differs from tRNA^Phe^ in two nucleotides (fig. S2). The nucleotides present in several cLCMV/IAP recombinants indicated that these sequences were actually tRNA^Phe^- and not IAP primer-binding site-derived (Fig. 2). This finding strongly suggested that these recombination events involved a faulty jump of a tRNA^Phe^-containing IAP reverse transcription intermediate and/or faulty reverse transcription of the tRNA primer.

The recombination events demonstrated here, in particular those observed in HG4 cells, displayed typical features of RT copy-choice recombination. These include a strong-stop/pausing of RT in the tRNA primer (3), a hairpin structure in the acceptor RNA (4) (fig. S3), and short stretches of up to 4nt sequence homology (3, 5, 6), as found in more than 60% of the identified recombination products. Short additions, deletions and mis-incorporations, like those found here, have also been demonstrated to accompany RT strand transfer (6). *In vitro* (MC57G cells) and *in vivo* in mice, however, the recombination mechanism appears to be less canonical.

Supporting text 3

**Evolutionary considerations**

Precisely how common such recombination events occur between different viral families remains unknown. The case described here probably reflects co-evolution between host (*Mus musculus*), endogenous retrotransposon (IAP), and its highly adapted exogenous virus (LCMV). Similar recombination and integration, both of which have previously been considered as impossible, may also occur for other host/virus combinations and have to be taken into account when assessing the safety of RNA viruses to be used in gene therapy, and in the phylogenetic and evolutionary interpretation of genome sequence data.
Supplementary Fig. S1. Details of the homology break-points in MC57.18 and MC57.23.

(A) Alignment to determine 5' homology break-points between LCMV sRNA and cGP sequences from MC57.18 and MC57.23. (B) Alignment to determine 3' homology break-points between LCMV sRNA and cGP sequences from MC57.18 and MC57.23. (C) Alignment to compare homologies of MC57.23 with mouse chromosome 7 and IAP element 92L23 sequences. Interestingly, sequences homologous to the direct IAP LTR repeats (shaded) could be found at this position on chromosome 7.
Supplementary Fig. S2. The sequence of the IAP 92L23 primer binding site and tRNA\textsuperscript{Phe} differ in two nucleotides.

tRNA\textsuperscript{Phe} hybridized to the IAP 92L23 genomic RNA is depicted. Bold black letters indicate LTR-derived sequences, bold gray letters indicate the direct repeat at the end of the IAP LTR. tRNA\textsuperscript{Phe} is depicted in red with the signature di-nucleotide in bold. The bridging di-nucleotide between the IAP LTR and the primer binding site is shown in orange.
Supplementary Fig. S3. Homology break-points in relation to secondary structures.

The secondary structure of the LCMV intergenic RNA loop (A) and tRNA^{Phe} (B) was determined using the online mfold server provided by the Rensselaer Polytechnic Institute (http://www.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi). The secondary structure of the 5’ untranslated region of LCMV (C) was predicted manually, based on mfold results. Break-points of homology to LCMV (A and C) or tRNA^{Phe} (B) found in sequences of the indicated samples are shown by an arrow. Blue letters indicate the stop codon of the LCMV GP and NP open reading frames.
Supporting tables

**Table S1. PCR approaches to amplify 3’ flanking chromosome 10 sequences in MC57.18.**

<table>
<thead>
<tr>
<th>Primer pair used for PCR</th>
<th>Expected Product Length (expected # of bp on chromosome 10)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conventional PCR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAPfwd2 x chr10rev2</td>
<td>1860bp (1352bp)</td>
<td>NOT successful</td>
</tr>
<tr>
<td>WE924f x chr10rechtsrev</td>
<td>3413bp (795bp)</td>
<td>NOT successful</td>
</tr>
<tr>
<td>WE924f x chr10rev2</td>
<td>3970bp (1352bp)</td>
<td>NOT successful</td>
</tr>
<tr>
<td><strong>Long range PCR (TaqPlusLong)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WE924f x chr10rev</td>
<td>4564bp (1946bp)</td>
<td>NOT successful</td>
</tr>
<tr>
<td><strong>Nested PCR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st IAPfwd2 x chr10rev2</td>
<td>1860bp (1352bp)</td>
<td>NOT successful</td>
</tr>
<tr>
<td>2nd IAPfwd1 x chr10rev1</td>
<td>468bp (143bp)</td>
<td>NOT successful</td>
</tr>
</tbody>
</table>

**Table S2. Primer sequences for primers used in Table S1.**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forward Primers</strong></td>
<td></td>
</tr>
<tr>
<td>WE924f</td>
<td>GGTTATTGCCTGACCAAATGGA</td>
</tr>
<tr>
<td>IAPfwd1</td>
<td>ATTCGCCGTTACAAGAT</td>
</tr>
<tr>
<td>IAPfwd2</td>
<td>GGGAGCAGGTCAGGGTGT</td>
</tr>
<tr>
<td><strong>Reverse Primers</strong></td>
<td></td>
</tr>
<tr>
<td>chr10rev</td>
<td>AAGCCCACACCCACAGTG</td>
</tr>
<tr>
<td>chr10rechtsrev</td>
<td>CCCAGCAACAGTCATAGCA</td>
</tr>
<tr>
<td>chr10rev1</td>
<td>CTTCTCCACTTTGCGCA</td>
</tr>
<tr>
<td>chr10rev2</td>
<td>TGGGCACTTCTGTCTTCTG</td>
</tr>
</tbody>
</table>

**Supporting references**