Infectivity and insertional mutagenesis of endogenous retrovirus in autoimmune NZB and B/W mice

Gabriele B. Beck-Engeser, Tomasz Ahrends, Gero Knittel, Rafael Wabl, Mirjam Metzner, Dan Eilat and Matthias Wabl

Correspondence
Matthias Wabl
mutator@ucsf.edu

1Department of Microbiology and Immunology, University of California, San Francisco, CA 94143-0414, USA
2Department of Medicine, Hadassah University Hospital and The Hebrew University Faculty of Medicine, Jerusalem 91120, Israel

Murine leukaemia virus has been suggested to contribute to both autoimmune disease and leukaemia in the NZB mouse and in the (NZB × NZW) F1 (abbreviated B/W) mouse. However, with apparently only xenotropic but no ecotropic virus constitutively expressed in these mice, few mechanisms could explain the aetiology of either disease in either mouse strain. Because pseudotyped and/or inducible ecotropic virus may play a role, we surveyed the ability of murine leukaemia virus in NZB, NZW and B/W mice to infect and form a provirus. From the spleen of NZB mice, we isolated circular cDNA of xenotropic and polytropic virus, which indicates ongoing infection by these viruses. From a B/W lymphoma, we isolated and determined the complete sequence of a putative ecotropic NZW virus. From B/W mice, we recovered de novo endogenous retroviral integration sites (tags) from the hyperproliferating cells of the spleen and the peritoneum. The tagged genes seemed to be selected to aid cellular proliferation, as several of them are known cancer genes. The insertions are consistent with the idea that endogenous retrovirus contributes to B-cell hyperproliferation and progression to lymphoma in B/W mice.

INTRODUCTION

NZB and (NZB × NZW) F1 (abbreviated B/W) mice suffer not only from autoimmune disease (haemolytic anaemia and lupus, respectively), but also from hyperproliferation of lymphocytes and, eventually, from lymphoma and leukaemia. Male B/W mice succumb to autoimmune disease later than female B/W mice, and thus live long enough to acquire lymphoma. Female B/W mice that have been cured of lupus (Wofsy & Seaman, 1987) do the same. Interestingly, the number of different B-cell clones found in the peritoneum of B/W mice decreases with age, and monoclonal or biclonality is common by 6 months (Tarlinton et al., 1988). This seems to be driven by antigen (Tarlinton et al., 1988), but may also be helped by tumorigenic mutations.

NZB is a mouse strain with two striking virologic characteristics: high titres of xenotropic murine leukaemia virus (XMV), and high synthesis rate of the envelope protein gp70. To link the virologic, immunological and tumorigenic characteristics of NZB, investigators hypothesized that XMV constitutes an ‘autoimmune virus’ (Levy, 1974). Although this retroviral hypothesis of autoimmune disease in NZB mice generated considerable enthusiasm when XMV was discovered, it was never proved and was later dismissed. This was mainly due to experiments with crosses of the NZB and the SWR mouse strains, in which autoimmunity in the progeny did not depend on the presence of replication-competent virus (Datta et al., 1978a). Meanwhile, endogenous retroelements are known to be adjuvants in the immune response to T-cell-independent type 2 antigens (Zeng et al., 2014), and they are implicated in monogenic (Beck-Engeser et al., 2011; Stetson et al., 2008) and polygenic (‘spontaneous’) (Beck-Engeser et al., 2009) autoimmune disease. Also, importantly, in the NZB × SWR mice, virus titres and high-level expression of NZB-like serum gp70 did correlate with lymphoma incidence (Datta et al., 1978b); but because the prototypic NZB virus is xenotropic, there seemed to be a limited number of effects it could have in lymphomagenesis.

While proliferation may be helped by binding of retroviral envelope protein to Toll-like receptor (TLR)-4 (Rassa et al., 2002), tumorigenesis via insertional mutagenesis by retrovirus requires infection. Generally, four types of
endogenous retroviruses are distinguished by their host range, which is
dictated by the sequence of the envelope (env) glycoprotein
gp70: ecotropic (infectious in the mouse; making up less
than 10 % of the endogenous retroviruses); xenotropic
(XMV; infectious in species other than the mouse); and
polytropic (PMV; infectious in the mouse and other
species), including modified polytropic (MPMV; polytrop-
ic with a deletion in gp70, which may or may not affect
host range) (Stoye & Coffin, 1987). The sequences of retro-
viral origin make up almost 10 % of the mouse genome.
The genomic sites of XMs, PMVs and MPMVs are
defined in common inbred strains (Frankel et al., 1990;
Jern et al., 2007) and in NZB mice (Frankel et al., 1992).
Although few of the endogenous viruses are infectious,
and fewer still are replication competent, a large variety
of viruses can be produced in individual mice through
somatic recombination. The viral diversity could presum-
ably also lead to pseudotyping and phenotypic mixing.
Pseudotyping and recombinant viruses may mediate inser-
tional mutagenesis in the NZB mice, which constitutively
only express xenotropic but no ecotropic virus.

The germ line-encoded virus sequences vary between mouse
strains. Because the genomes of the NZB and NZW mice
have not yet been sequenced, the complement of endogen-
ous retroviruses in these genomes is only partially known.
The NZB mouse can express two distinct groups of replica-
tion-competent xenotropic viruses: the NZB-X1 which is
spontaneously produced, and the NZB-X2 which is induced
upon treatment of fibroblasts with iododeoxyuridine (Elder
et al., 1980). In addition to these xenotropic viruses, the
NZW mouse produces an inducible ecotropic virus (Chatto-
padhyay et al., 1980; Stephenson et al., 1975); however,
neither type has been sequenced.

To be able to characterize insertional mutagenesis in these
mouse models, we describe here the virus types that are
produced by NZB, NZW and B/W mice. We sequenced
the gene fragments encoding the env proteins recovered
from Mus dunni cells infected with plasma of the various
mouse strains and identified three types of xenotropic
virus, plus an additional variant xenotropic virus. Because
we also recovered polytropic virus from all mice, we inves-
tigated whether pseudotyped virus might be infectious to
NZB cells. This is indeed the case, and allows for the possi-
bility that insertional mutagenesis generates the lymph-
omas of NZB mice late in life.

We also determined the complete sequence of a putative
ecotropic NZW virus, which we isolated from a lymphoma
derived from a B/W mouse. Therefore, we investigated
whether insertional mutagenesis plays a role in hyperpro-
liferation and lymphoma formation in B/W mice. Indeed,
we found new retroviral integration events in endogenously
activated splenic B-cells, in peritoneal B-cells of B/W mice
and in the lymphoma from which we had isolated the eco-
tropic virus mentioned above. This supports the view that
murine leukaemia virus (MLV) contributes to B lymphoid
hyperplasia in these mice.

RESULTS

MLV gp70 envelope protein subunit expression on
lymphocytes

Retroviral env gp70 is encoded as part of a complete retro-
viral genome(s) (Lerner et al., 1976) and expressed on the
cell membrane of lymphoid and epithelial cells of all
mouse strains. However, mere expression of a viral protein
does not necessarily imply the presence of infectious or
replication-competent viral particles. Serum gp70, for
example, is produced by the liver of all mouse strains
(Hara et al., 1981) as an acute-phase protein (Hara et al.,
1982). However, most of the gp70 that is expressed on
lymphocytes and macrophages does indeed come from com-
plete viral particles. This can be concluded from the increase
in gp70 expression in TLR-7- and RAG-deficient mice
(Young et al., 2012; Yu et al., 2012), an effect most easily
explained by the spread of replication-competent virus.

To compare viral expression in the NZB, NZW and B/W
mice with that in AKR and C57BL/6 mice, we stained
cells from brain, thymus, spleen and peritoneum of various
mouse strains with the 83A25 antibody, which has broad
specificity to gp70. In C57BL/6 mice, no gp70 is expressed
in the brain, but is expressed in thymus, peritoneum and,
to a smaller extent, spleen (Fig. 1a, b). gp70 expression is
much higher in the AKR mouse, which expresses the pro-
totypic leukaemia virus Akv. In AKR, virtually all thymo-
cytes and most splenocytes express MLV, whereas only a
fraction of the peritoneal cells do so (Fig. 1a, b). NZB,
NZW and B/W mice show gp70 expression patterns similar
to those in the AKR mouse (Fig. 1a, b).

Viruses in plasma of NZB, NZW and B/W mice

Our flow data showed that gp70 is expressed in great
abundance on thymus, spleen and peritoneal cells. To
characterize infectious or replication-competent viruses,
we infected M. dunni cells with the plasma of NZB,
NZW and B/W mice. Apart from an endogenous MLV,
which is expressed only when induced (Bonham et al.,
1997) (and the sequence of which is known), M. dunni
cells express no other retroviral sequences and can be
infected by xenotropic, polytropic and ecotropic viruses.
We isolated DNA from these cells infected with plasma,
and amplified and sequenced proviral DNA for the
proline-rich region (PRR) of the gp70 env protein of pro-
viral MLVs. This region varies according to virus type, but
the flanking amino acid sequences are invariant, so that
two (to account for the codon variation) common PCR
primer pairs cover the complete spectrum of MLV
types. One primer pair covers the ecotropic Akv-type
viruses (Akv primers), and the other mainly covers the
xenotropic and polytropic viruses (NZB primers).
As expected, Akv primers did not amplify any DNA
from the supernatant of M. dunni cells infected with
plasma from NZB, NZW and B/W mice, whereas the
NZB primers yielded a 280 bp band. This may represent xenotropic and polytropic viruses (Fig. 2a). Owing to a 27 bp deletion in the PRR (Stoye & Coffin, 1987), the amplicon of MPMV runs slightly below that of the other viruses. We detected sequence-confirmed MPMV in the DNA from cells infected with plasma from young NOD mice (Fig. 2a), but not in the plasma of NZB, NZW and B/W mice. This indicates that these latter mice do not produce viral particles with genomes encoding env of MPMV.

We then sequenced the PRR amplicons without subcloning them, which allowed us to estimate the heterogeneity of viral genomes in the plasma. Fig. 2(b) shows the chromatograms of a segment in which the xenotropic sequences differ most from the polytropic sequences. Virus produced from M. dunni cells containing the proviral xenotropic NZB-X2 genome (NZB9-1) yielded a ‘clean’ chromatogram, whereas the chromatograms from NZB, NZW and B/W mice showed the dominant xenotropic sequence superimposed on polytropic viral sequences. In the B/W mouse, xenotropic and polytropic viruses are present at the same ratio (Fig. 2b). Owing to the additional and predominant expression of an Akv-type virus (see below), the chromatogram of the B/W-derived NYC lymphoma is much more complex (Fig. 2b). To identify the various viruses that infected the M. dunni cells, we sequenced bacterial subclones of amplicons covering the complete env of MPMV.

We set out to differentiate the xenotropic viruses according to Lamont et al. (1991) and displayed in Baudino et al. (2008) into four subgroups, Xeno-I through Xeno-IV. However, the original grouping was done with viruses derived from an AKR cell line, and the sequences displayed in Baudino et al. (2008) were from three different strains: the sequences of Xeno-I (comprising viruses X-1 and X-2, which differ at the C terminus of gp70) are from NZB mice; Xeno-II, Xeno-III and Xeno-IV are from C57BL/6; and PMV and MPMV are from newborn HRS/J mice. Therefore, the sequences determined in our study differ:

**Viruses in NZB mice**

The plasma of NZB contained xenotropic and polytropic but not ecotropic (Jenkins et al., 1982) virus (Table 1). We set out to differentiate the xenotropic viruses according to Lamont et al. (1991) and displayed in Baudino et al. (2008) into four subgroups, Xeno-I through Xeno-IV. However, the original grouping was done with viruses derived from an AKR cell line, and the sequences displayed in Baudino et al. (2008) were from three different strains: the sequences of Xeno-I (comprising viruses X-1 and X-2, which differ at the C terminus of gp70) are from NZB mice; Xeno-II, Xeno-III and Xeno-IV are from C57BL/6; and PMV and MPMV are from newborn HRS/J mice. Therefore, the sequences determined in our study differ.

---

**Fig. 1.** Expression of MLV gp70 in cells from various mouse strains. Flow cytometry histograms and dot plots of cells from C57BL/6, AKR, NZB, NZW and B/W mice, and an NYC lymphoma. (a) Flow cytometry histograms of gp70 expression in brain, thymus, spleen and peritoneum. y-axis, cell number; x-axis, fluorescence intensity on a logarithmic scale, in arbitrary units. Filled curve, first plus second antibody; empty curve, second antibody only. (b) Quantitative data from flow cytometry as in (a), representing individual mice. y-axis, mean fluorescence intensity; x-axis, mouse strains.
Endogenous retrovirus in autoimmune mice

Xenotropic MLVs can infect *M. dunni* cells, and these viruses probably helped the non-replication-competent polytropic viruses that we recovered. However, xenotropic virus cannot help polytropic virus infect NZB cells. Rather, xenotropic virus would need to be pseudotyped with polytropic env to obtain the enabling gp70 sequence. To test this hypothesis, we made use of a by-product of the retroviral replication cycle: a fraction of the linear retroviral cDNA molecule is covalently circularized by host factors – reactions that prevent subsequent integration of the retroelement DNA. In this process, ligation of the two ends forms a circle, in which the two LTRs that flank the linear cDNA are joined (the so-called 2-LTR circles) (Fig. 3a), or homologous recombination between the LTRs results in formation of a 1-LTR circle. While the circular forms are dead-end products, they are a reliable and reproducible measure of recent retroviral infection events. We thus prepared and sequenced the fragment containing the breakpoints (sites of ligation) of retroviral circular DNA from NZB spleen cells. The fact that we could isolate 2-LTR circles indicated that endogenous MLV can infect NZB cells. Because we identified LTR sequences from both xenotropic and polytropic virus (Fig. 3b), we conclude that xenotropic virus is pseudotyped or has recombined with polytropic virus. This would be similar to polytropic MuLV being pseudotyped within ecotropic virions, when cells were co-infected with both types of viruses (Rosenke et al., 2012).

**Viruses in NZW mice**

As with NZW mouse, we recovered polytropic and xenotropic, but no ecotropic virus from NZW plasma. As in NZB, the env of the polytropic virus of NZW starts with a methionine codon and has a perfect ORF. In addition, we could distinguish an XMV variant, named XMV-5, by the HHPPP motif in place of the HPPPS, and the SLVDGAY motif instead of the NLVEGAY motif in the PRR. Also, the hypervariable region VRA of gp70 (GenBank accession number KM282109) was quite different from the other known XMVs.

In the NZW mouse strain an ecotropic virus can be induced (Stephenson et al., 1975; Chattopadhyay et al., 1980) However, this virus has not been well characterized. We cloned a virus from the NYC lymphoma (designated NZW-1) and compared its sequence (GenBank accession number KP087798.1) to that of the ecotropic AKV (Herr, 1984). Altogether there are 12 amino acid exchanges between the two viruses, but within the gp70 there are...
Table 1. Virus types determined from sequences of proviruses integrated in infected *M. dunni* cells

*M. dunni* cells were infected with plasma from NZB mice and supernatant of the LNC lymphoma cell line from NZB; with plasma from NZW and B/W mice; and from supernatant of the NYC lymphoma derived from a B/W mouse. XMV-1c, Constitutively expressed Xeno-I (X1) from NZB; XMV-1i, inducible Xeno-I (X2) from NZB. PMV, Polytropic virus; MPMV, modified polytropic virus; NZW-1, MLV isolated from NYC lymphoma of B/W mice. Numbers in the table body represent the number of isolated gp70 sequences encoded by the respective virus type. –, Not found; ND, not done.

<table>
<thead>
<tr>
<th>Mouse strain</th>
<th>XMV</th>
<th>XMV-1c</th>
<th>XMV-1i</th>
<th>XMV-2</th>
<th>XMV-3</th>
<th>XMV-4</th>
<th>XMV-5</th>
<th>PMV</th>
<th>MPMV</th>
<th>NZW-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZB 8 mo</td>
<td>11</td>
<td>–</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NZB 14 mo</td>
<td>6</td>
<td>–</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LNC</td>
<td>23</td>
<td>1</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NZW 1.5 mo</td>
<td>8</td>
<td>–</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NZW 6 mo</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>B/W 3 mo</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B/W 12 mo</td>
<td>5</td>
<td>–</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NYC</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
</tbody>
</table>

---

**Fig. 3.** (a) Schematic of circularization of late reverse transcription (RT) cDNA product of endogenous MLV. Empty and filled arrows show PCR primers used to amplify the part of circular DNA that contains the site where the two LTR ends of the linear sequence are joined to form a circle. (b) Upper panel, 2-LTR sequences amplified from NZB spleen cells. Sequences are from a 0.5 kb amplicon that contains the 2-LTR junction (X); boxed, inverted repeat sequences; shaded, nucleotide differences; dots indicate nucleotides missing, as compared to prototype sequence. Lower panel, frequencies of U5 sequences with nucleotide replacements as compared to the consensus sequence; dashes indicate nucleotide identity.

Consensus: A TCTTGGTGGAG GTGCTTCCCA TATGATTGAA CTACCACCT GCAGGCTTTTC TCA
only three, at positions 33, 45 and 147 (counting from the start methionine; Table S1, available in the online Supplementary Material). If, in analogy to the Moloney MLV gp70, we consider the segment of amino acids 81 to 163 (nt 6026–6271 of the viral RNA genome) to encompass the hypervariable region A (Bae et al., 1997), then only the E in position 147 in NZW-1, instead of the G in AKV, might influence receptor binding. In the PRR of the gp70 ('diagnostic region', nt 6574-6842 in Table S1), there was no difference. We therefore consider the NZW-1 virus to be most likely ecotropic. At positions 109 and 110 of gag p30, the virus has amino acids QR, which indicates an N-tropism (DesGroseillers & Jolicoeur, 1983; Kozak & Chakraborti, 1996). NZB and NZW mice were typed Fv1n (Baliji et al., 2010) and, more specifically, Fv1nr (W. P. Rowe and J. W. Hartley, cited in Steeves & Lilly, 1977), which restrict some but not all N-tropic viruses (Kozak, 1985; Steeves & Lilly, 1977). Therefore, once induced, NZW-1 may be able to spread in both NZW and B/W mice. However, if it is restricted (Stevens et al., 2004), it may need some mutational/recombinational event to modify the capsid protein.

**De novo insertions in B/W spleen cells**

In light of likely ongoing retroviral infections in the NZB mouse and also, presumably, in the NZW mouse, MLV may be partially responsible for the hyperproliferation of spleen cells in the B/W mouse. Ecotropic MLV of NZW, or recombinant/pseudotyped MLV from either NZB or NZW, may cause insertional mutagenesis, which would, in turn, drive increased cell division.

In tumours, at least the first transformative event (e.g. a retroviral insertion) is clonal, and therefore the causative mutation is likely to be present in stoichiometric amounts. However, it is not known whether hyperproliferation in the B/W spleen also tends toward monoclonality. Therefore, to recover de novo retroviral insertion sites, we investigated single activated cells. By fusing spleen cells to generate hybridomas, we (i) selected the activated (dividing) cells, since these cells fuse preferentially. Subsequently, we (ii) separated the bulk of activated cells into single cell clones. To identify endogenous MLV integration sites, we analysed 80 hybridomas from the spleens of two 7-month-old, sick, non-immunized B/W mice and found a total of six de novo insertions in five hybridomas (Table 2 and Fig. 4). In addition to the names of the genes, Table 2 lists the location of the insertions in or near the genes and the layout of the insertions. A separate column gives the number of additional insertions at a given locus in lymphomas recovered by us, or in lymphocytic/myeloid tumours reported by others (Akagi et al., 2004; Sokol et al., 2014; Uren et al., 2008). Because the insertions were absent from the BALB/c-generated fusion

<table>
<thead>
<tr>
<th>Cells</th>
<th>Gene</th>
<th>Location</th>
<th>Layout</th>
<th>Additional inserts</th>
</tr>
</thead>
<tbody>
<tr>
<td>B/W hybridoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Lyzl4</td>
<td>Chr9 : 121461423</td>
<td>T+G− 25538</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Lsamp</td>
<td>Chr16 : 40063012</td>
<td>G+T− Intron 1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Jarid1b</td>
<td>Chr1 : 136518121</td>
<td>G+T+ Intron 21</td>
<td>2</td>
</tr>
<tr>
<td>52</td>
<td>Rsbn1l</td>
<td>Chr5 : 20422593</td>
<td>G−T− Intron 3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>Sema4b</td>
<td>Chr7 : 87355314</td>
<td>G+T+ Intron 2</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>Lrba</td>
<td>Chr3 : 86328110</td>
<td>G+T− Intron 27</td>
<td>7</td>
</tr>
<tr>
<td>B/W peritoneal cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>297</td>
<td>Isg2012</td>
<td>Chr3 : 87736045</td>
<td>G+T+ Exon 2</td>
<td>0</td>
</tr>
<tr>
<td>247</td>
<td>Trio</td>
<td>Chr15 : 27663180</td>
<td>G−T− Intron 10</td>
<td>12</td>
</tr>
<tr>
<td>243</td>
<td>Popdc3</td>
<td>Chr10 : 45040457</td>
<td>G+T+ 1814</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Nsmaf</td>
<td>Chr4 : 6368259</td>
<td>G−T− Intron 1</td>
<td>6</td>
</tr>
<tr>
<td>B/W lymphoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NYC</td>
<td>Prex1</td>
<td>Chr2 : 166558404</td>
<td>G−T+ 19122</td>
<td>24</td>
</tr>
<tr>
<td>NZW mouse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germ line</td>
<td></td>
<td>Chr8 : 118995834</td>
<td>T−</td>
<td>0</td>
</tr>
</tbody>
</table>
Fig. 4. De novo endogenous MLV integration sites in splenic B-cells of B/W mice. Screen prints of a customized version of the UCSC Genome Browser (mouse genome assembly mm9). ‘YourSeq’, Genomic sequence tag recovered in a chimeric sequence containing retrovirus LTR. Numbers at the top show nucleotide position at the particular chromosome. Bright green
partner NS0/1, and from the genomes of both the NZB and NZW mice, the insertions likely arose de novo in the B/W spleen cells. However, in the NS0/1 fusion partner, we identified other integration sites, which contributed to the tumorigenesis of the myeloma-derived cells (Knittel et al., 2014). Therefore, we considered the possibility that by generating hybridomas, we were studying subclones of the fusion partner, rather than the spleen cells themselves. These subclones would contain de novo insertions, which, however, should not be recovered in the bulk culture of NS0/1. Of 105 subclones tested, none contained any of the six insertions listed in Table 2.

Fig. 4 shows a graphic representation of the proviral insertions recovered from the spleen cells (‘YourSeq’). Of these, Lyz14, Rsbn11 and Lsamp are not known to be cancer genes. However, Lrba has seven insertions in other mouse tumours and thus can be considered a mouse cancer gene. In human patients, mutations in the LRBA gene cause a syndrome of autoimmunity, lymphoproliferation and humoral immune deficiency, by substantially increasing degradation of the immune response inhibitor CTLA-4 (Lo et al., 2015). Sema4b is a known cancer gene. Eleven integrations into the Sema4b locus have been published, and in our lymphoma screen, we found four additional insertions (Fig. 4). Of these insertions, five are 5' to the transcription start site, four are in intron 1, five are in intron 2, and one is in intron 4. This pattern suggests constitutive expression of the gene. Jarid1b (also known as lysine [K]-specific demethylase 5B) is also a likely cancer gene. It demethylates Lys-4 of histone H3, thereby playing a central role in histone code. Lysine-specific demethylases are known tumour suppressors. We also found an insertion in Fbx110 (also known as lysine (K)-specific demethylase 2B, human orthologue KDM2B) in a plasmacytoma (Knittel et al., 2014).

**De novo insertions in B/W peritoneal cells**

The number of different B-cell clones found in the peritoneum of B/W mice decreases with age, and monoc- or biclonality of their membrane IgM is common by 6 months. Many clones from different mice show similar rearrangements at both the Ig heavy and light chain loci (Stall et al., 1988). This suggests that the monoclonal peritoneal cells are precursors of the lymphomas suffered by B/W mice. As we had done previously for spleen cell hybridomas, we used anchored PCR to recover proviral insertions in peritoneal cells. Fig. 5 represents these insertions, and Table 2 lists the four de novo MLV integrations recovered from 17 3-month-old B/W mice. All insertions disrupt genes (counting the insertion in Popdc3 as a promoter insertion), which indicates selection of the genetic event in this cell population.

Two of the four genes, Trio and Nsmaf, were also targeted in tumours in other studies (Table 2). This shows that these genes are common integration sites, and thus can be considered oncogenes. Trio, or triple functional domain (PTPRF interacting) protein, is a rho family GDP/GTP exchange factor. Nsmaf, a.k.a. FAN, or neutral sphingomyelinase activation-associated factor, encodes a WD-repeat protein that binds the cytoplasmic sphingomyelinase activation domain of the 55 kDa TNF receptor. The protein is required for TNF-mediated activation of neutral sphingomyelinase, and may play a role in regulating TNF-induced cellular responses such as inflammation (Montfort et al., 2009). Nsmaf is expressed in B- and T-cells and lymphoid organs (UCSC Genome Browser). Unlike Trio and Nsmaf, there are no reports that Isg20l2 or Popdc3 may act as an oncogene.

**Passenger integrations versus drivers of tumorigenesis**

Because genes represent only 2% of the genome and, by and large, MLV integration into the genome is random, integrations are less likely to occur in or near genes than in intervening sequences. However, there is some preference of pseudotyped MLV to integrate into reference sequence (RefSeq) genes in HeLa cells: 34.0% of the integrations landed in RefSeq genes, while in a set of computersimulated random integrations, only 22.4% did (Wu et al., 2003). Although transcription start sites were preferred, the association of integration sites with gene activity was weak (Mitchell et al., 2004). In contrast to integrations in short-term cell culture, where one expects little selection, tumours *ex vivo* have undergone strong selection for cell growth. In mouse end-stage tumours, irrespective of the distance to transcription start sites, the majority of integrations colocalized with common epigenetic enhancer markers in murine lymphoid tissues (Akagi et al., 2004; Sokol et al., 2014; Uren et al., 2008). Of 6132 tags we analysed previously in T lymphomas induced by exogenous MLV, 46.4% were in genes. Of 833 tags in B lymphomas, 53.2% were in genes.

Of the ten insertions shown in Table 2, nine were within genes, defined as the unit spanning 2 kb upstream of the transcription start site to the end of the 3'UTR (i.e. 90.0%, with central confidence limits of 58.7–97.7%, at 0.95 confidence). Intriguingly, this percentage of insertions
into a gene is even (statistically) higher than the 53.2 % in B lymphomas generated by exogenous MLV. Several mutations in Table 2 also represent common integration sites, as catalogued by the published retroviral tags and our own B and T lymphoma screens. Therefore, we interpret integrations into genes of the B/W cells here as an indication of selection and as drivers of tumorigenesis.

**Insertion in a B lymphoma of the B/W mouse**

The retroviral insertions in some of the quasimonoclonal populations of peritoneal B-cells in B/W mice suggested that lymphomas from the B/W mice may be driven by insertional mutagenesis. We thus investigated the NYC cell line, which was adapted from a tumour of a B/W mouse that was treated with anti-CD4 antibodies and thus did not develop glomerulonephritis (Wofsy & Seaman, 1987). Because it expresses low levels of the characteristic surface marker CD5 (Fig. 6a), NYC is likely derived from a peritoneal B1 cell. The growth of cells of this line depends on the presence of the antigen receptor, which can be co-precipitated with MLV protein p30 (Jäck et al., 1992). Furthermore, the 2-LTR circles present in these cells indicate continuous MLV infection (Beck-Engeser et al., 2009).

In this cell line, we identified an insertion into Prex1 (Fig. 6b). Prex1 is a bona fide proto-oncogene; various other studies list 23 MLV insertions that presumably cause constitutive (over)expression of the gene and, in some cases, truncation. Prex1 functions as a RAC guanine nucleotide exchange factor and is synergistically activated by

---

**Fig. 5.** De novo endogenous MLV integration sites in peritoneal B-cells of B/W mice. Screen prints of a customized version of the UCSC Genome Browser depicting (a) Popdc3; (b) Trio; (c) Nsmaf; and (d) lsg2012 locus in the mouse. Layout as in Fig. 4; PicoAKV2, insertions recovered from our (different) large B lymphoma screen; the insertions recovered in this study are marked by ‘YourSeq’.
phosphatidylinositol-3,4,5-triphosphate and the β, γ subunits of heterotrimeric G protein (Welch et al., 2002). According to the UCSC Genome Browser, the protein is strongly expressed in brain, trachea, thymus, bone marrow, and B- and T-cells. Prex1 is also overexpressed in prostate cancer (Knight-Krajewski et al., 2004) and breast cancer (Sosa et al., 2010). In addition, patients whose tumours express Prex1 are more likely to develop metastasis, when compared with those whose tumours do not express Prex1 (Sosa et al., 2010).
De novo integrations ought to increase expression of both MLV and the target gene. Taqman analysis of the steady-state mRNA levels confirmed that NYC expresses large amounts of MLV mRNA – 1 or 2 orders of magnitude more than plasmacytomas or LPS-activated spleen cells, which are known for their high-level expression (Moroni & Schumann, 1975) (Fig. 6c). Stimulation of splenic B-cells also increases expression of Prex1 by an order of magnitude, to the level of that in peritoneal cells (Fig. 6c). NYC cells express more than twice as much Prex1 mRNA than do peritoneal, LPS-activated spleen cells or testicular cells (Fig. 6c). This is likely because Prex1 overexpression in NYC cells drives their proliferation. However, this high expression level is not a general prerequisite for proliferation, as, for example, NIH3T3 cells do not express Prex1, and the strongly proliferating thymocytes express only a little (Fig. 6c).

DISCUSSION

Proliferative disease by insertional mutagenesis due to endogenous retrovirus activity is an established fact and a major cause of spontaneous mouse tumours in general. Integration is potentially mutagenic on two counts: it may disrupt a vital region of the host genome, or bring powerful regulatory functions of the virus to bear on the expression of host genes. Because the prototypic NZB virus is xenotropic, there seemed to be a limited number of effects it could have in lymphomagenesis. However, as we showed here, xenotropic virus replicates in NZB cells – presumably pseudotyped by polytropic virus. Therefore, recombinant virus could be generated and spread. This could happen at a higher frequency in NZW mice, once the Akv-like (putative) ecotropic NZW-1 virus is induced. However, we did not find the NZW-1 virus in the plasma of NZW or B/W mice and thus do not know how it is induced. Because we cloned the NZW-1 virus from a B/W lymphoma, we also do not know whether the virus is recombinant or encoded in the NZW germ line.

The experimental insertion frequency (6 of 80) was not high enough to conclude that this mutagenesis contributes in an important way to hyperproliferation in the spleen. Perhaps the splinkerette method identifies only a small fraction of inserts, and we are underestimating the frequency. However, in 17 mice, the yield from quasi-monoclonal cells from the peritoneum was better. We thus suggest that hyperproliferation is the initial step of lymphomagenesis in the peritoneum of these mice. There, the tendency toward monoclonality strongly points to a preleukaemic stage. The initiating step may be antigenic activation of a cell that will grow into a clone, followed by insertional mutagenesis mediated by MLV. MLV then activates a proto-oncogene (or deactivates a tumour suppressor) and causes hyperproliferation – more aggressive growth and larger cell numbers, from which new oncogenic events can be selected. This would lock the cell into continuous proliferation (Fig. S1a). The peritoneal cell-derived NYC lymphoma with its insertion in the Prex1 gene may be an example of this sequence of events. Alternatively, the retroviral insertion may happen earlier in ontogeny, such as in a pro-B-cell. However, in a proto-oncogene, the insertion may have no effect until the B-cell reaches the mature stage and is activated by antigen. In that case, many B-cells of different specificities would contain the first mutation, and these would only hyperproliferate after antigenic activation.

Mouse tumours characterized in B/W mice share at least three characteristics with human chronic lymphocytic leukaemia (Stall et al., 1988): they apparently are driven, in part, by (self) antigen (Dühren-von Minden et al., 2012; Jäck et al., 1992; Tarlinton et al., 1988), they express the surface marker CD5 (Tarlinton et al., 1988) and they have the particularly puzzling feature of an indolent phase of tumour evolution followed by a rapid diffuse expansion of a large B-cell lymphoma (Richter’s syndrome). The underlying cause of this transition has been difficult to pinpoint. We suggest that, at least in the B/W mouse, the cause is an additional insertional mutagenesis event.

The ability of MLV to cause insertional mutagenesis in the B/W mice may also provide a mechanistic link between autoimmunity and cancer. Analogous to the B/W mice, which suffer from autoimmune disease and lymphadenopathy and eventually lymphoma (Stall et al., 1988), human systemic lupus erythematosus patients have a well-documented increased risk of non-Hodgkin’s lymphoma (Bernatsky et al., 2005). As another example, BALB/c mice injected with mineral oil suffer from inflammation and develop both a lupus-like disease (Satoh et al., 2003) and plasmacytomas, the latter of which (in addition to translocations) are driven by de novo insertions of endogenous MLV (Knittel et al., 2014). The increased susceptibility to cancer in autoimmune humans and mice could be solely due to the large number of proliferating cells, which are the targets of mutagenesis. In that case, the tumour formation would be an epiphenomenon of the autoimmune disease with unknown aetiology. But we hypothesize that similarly to in lymphoma formation, in autoimmunity an immature B-cell may not be tolerized to self-antigen because the retroelement insertion blocks the relevant pathway (Fig. S1b). Because a single mutation in a stem cell may jeopardize the tolerization of a larger lymphocyte population, whereas cancer generally requires accumulation of several mutations within the same cell, autoimmune disease will precede cancer. For cancer, the genome must also allow repeated productive viral recombination to overcome the superinfection barrier – a requirement that is hindered by certain haplotypes (Datta et al., 1983).

In the autoimmune-prone mouse strain MRL/lpr, a connection between insertional mutation and autoimmunity is evident. The lpr mutation is caused by the insertion of a retrotransposon into the Fas gene, which results in
aberrant splicing of the Fas message. This, in turn, results in proliferative disease, which accelerates the lupus-like disease (Drappa et al., 1993). Lack of tolerance also happens in non-autoimmune mice that (i) have a mutation leading to hyperresponsive external calcium entry in B-cells (Yu et al., 2005), (ii) are overexpressing B-cell activators, including BAFF (Mackay et al., 1999), or (iii) are deficient (knockout) in inhibitors, including CD22 (Nitschke et al., 1997; O’Keefe et al., 1996; Otipoby et al., 1996; Sato et al., 1996), Lyn (Hibbs et al., 1995; Lamagna et al., 2013, 2014) and FcγRIIB (Bolland & Ravetch, 2000). Similarly, we suggest that insertional mutagenesis may accelerate lupus in B/W mice. However, unlike the germ line insertions introduced by the experimenter, or that in the MRL/lpr mouse, the insertions would be generated somatically, and when accumulating, they would lead to lymphoma.

METHODS

Flow cytometry. The following antibodies were used: FITC-coupled goat anti-mouse IgM (μ chain specific; Southern Biotechnology), PE-coupled anti-CD5 (rat IgG2a; Pharmingen) and rat monoclonal antibody 83A25 specific to MLV gp70 (IgG 2a) plus APC-coupled goat anti-rat Ig (Pharmingen).

2-LTR sequences. Circular DNA was isolated (von Schwedler et al., 1990) and retroviral 2-LTR circle sequences were amplified. The following primers were used: ‘NZB primers A’: 5’-CAG GTG TGG TTT GCC CCT AGT CTT-3’ (U3rev) and 5’-ATC CCA ATC GTG TGC TCG CTT-3’ (U5forw); and ‘NZB’ primers B: 5’-TAC CTT TAT TAT AGG CCT GGG G-3’ (U3rev) and 5’-GGG CCA CTC CTA CCA TAG ACT G-3’ (R’c’forw). The 50 μl PCR mixture contained 2 μg of circular DNA, 0.2 μM each of primer, 200 μM dNTP, 1.5 mM MgCl2 and 1 U of recombinant Taq DNA polymerase (Invitrogen) in 1 X PCR buffer (final concentration). Cycle conditions: 94°C for 5 s; 35 cycles of 60°C for 1 min and 72°C for 1 min; and finally, 72°C for 5 min. PCR-amplified products were gel purified using a Qiaquick gel extraction kit (Qiagen), cloned into a pCR2.1 TOPO vector using a TOPO TA cloning kit (Invitrogen), and sequenced with M13 sequencing primers.

PCR amplification of ecotropic and polytropic PRR and full-length polytropic env. M. dunni cells were infected using plasma from mice or cell culture supernatant. Genomic DNA was isolated 10 days after infection using a DNA Blood and Tissue kit (Qiagen). The ecotropic proline-rich region (PRR) was amplified using primers specific for the ecotropic virus AKV gp70 280F CGGGTCCCAATAGGGCCCCAACCCKGTGGCATTTG. The env was amplified using gp70 specific forward (876gp70F ACGGCCGCTACATGCAAACCGCG) and reverse (876gp70R GGGGTACTGCGACTGCTTGT) primers. The PRR of the polytropic gp70 was sequenced using primers specific for the PRR domain gp70 280F CGGGTCCCAATGGGCTCAATCGG and gp70 280R GGGGCTCAACCCCTCGTACTGAG. There were no bands in the non-infected M. dunni cells.

Recovery of de novo MLV integration sites. De novo MLV integration sites were recovered using the splinkerette approach (Mikkers et al., 2002), described in detail in Knittel et al. (2014). Cellular DNA was restriction-digested with the enzymes Apol, Nsp1 or BstY1 and ligated to a dsDNA oligonucleotide with a stable hairpin (splinkerette). Primers (given below) were such that they amplified a chimeric sequence containing part of the retroelement LTR and genomic sequences upstream of the integration site. Further validation of the de novo integration site included amplification of a chimeric fragment with a genomic 5’ primer and a 3’ LTR primer. Finally, because the databases provide the genome of the C57Bl/6 mouse but not that of the NZB or NZW mouse, DNA fragments of the NZB and NZW mice spanning the retroelement insertion site were sequenced. Re-sequencing excluded the possibility that the retroelement insertions were in the germ line of these mouse strains.

The oligonucleotides for the splinkerettes were, for Apol, 5’-CGAAGTACACGGAGATTGCTAGGGAGACCCGTTGCAATGAGATCTG- TGTGCACTAAGTG-3’ and 5’-AATTCACATGTCGACACAGTC- TCTCTGATTTTTTTTCAAAAAA-3”; for Nsp1, 5’-CGAAGTACACGGAGATTGCTAGGGAGACCCGTTGCAATGAGATCTG- TGTGCACTAAGTG-3’ and 5’-GATCACCAGTCTGATCCCTCACGTCGACACACTGCTCATAAATTTTTTTTCAAAAAA-3”; for BstY1, 5’-CGAAGATACACGGAGATTGCTAGGGAGACCCGTTGCAATGAGATCTG- TGTGCACTAAGTG-3’ and 5’-GATCACCAGTCTGATCCCTCACGTCGACACACTGCTCATAAATTTTTTTTCAAAAAA-3”. The samples were then digested with EcoRV restriction nuclease to prevent amplification of internal viral fragments. The ligated fragments were amplified in three PCR rounds using nested primers. First primer set: forward primer binding to the splinkerette, CGAAGAAGTAAC- CCAGCTCTCTGTACTTCC (LTR), with PCR conditions activation at 95°C for 2 min; 27 cycles of 95°C for 20 s, 60°C for 30 s, 72°C for 2 min 15 s; final elogation at 72°C for 5 min. Second primer set: GTGGCCTCAGACTGTTGCTGCACGACACACTGCTCATAAATTTTTTTTCAAAAAA-3’ and 5’-CCAGCTCTCTGTACTTCC (splinkerette) and CCAGCTCTCTGTACTTCC (LTR), with PCR conditions being the same. Third-round adaptor binding primers contained the restriction sites and were therefore specific for each enzyme: Apol, ATGGACTGTTGCTGACACATGTTGGAATT(T/C); Nsp1, CTGGGTCTCGACACATGTTGGAATT(C/T); BstY1, GACTGGTGCTGACACATGTTGGAATT(C/T). The LTR primer was CTGAGAACATCCGCTCTGTGGTAC or CCAGCTCTCTGTACTTCC (splinkerette) and CCAGCTCTCTGTACTTCC (LTR), with PCR conditions being the same. Third-round adaptor binding primers contained the restriction sites and were therefore specific for each enzyme: Apol, ATGGACTGTTGCTGACACATGTTGGAATT(T/C); Nsp1, CTGGGTCTCGACACATGTTGGAATT(C/T); BstY1, GACTGGTGCTGACACATGTTGGAATT(C/T). The LTR primer was CTGAGAACATCCGCTCTGTGGTAC or CCAGCTCTCTGTACTTCC (splinkerette) and CCAGCTCTCTGTACTTCC (LTR), with PCR conditions being the same. Third-round adaptor binding primers contained the restriction sites and were therefore specific for each enzyme: Apol, ATGGACTGTTGCTGACACATGTTGGAATT(T/C); Nsp1, CTGGGTCTCGACACATGTTGGAATT(C/T); BstY1, GACTGGTGCTGACACATGTTGGAATT(C/T).

Provlral insert validation. To validate the integration sites, PCR was run with genomic up- and downstream primers of the presumed integration site, as well as with a genomic forward primer and a proviral reverse primer. As the viral primer, either CTGAGAACATCCGCTCTGTGGTAC or CCAGCTCTCTGTACTTCC (splinkerette) and CCAGCTCTCTGTACTTCC (LTR), with PCR conditions being the same. Third-round adaptor binding primers contained the restriction sites and were therefore specific for each enzyme: Apol, ATGGACTGTTGCTGACACATGTTGGAATT(T/C); Nsp1, CTGGGTCTCGACACATGTTGGAATT(C/T); BstY1, GACTGGTGCTGACACATGTTGGAATT(C/T). The LTR primer was CTGAGAACATCCGCTCTGTGGTAC or CCAGCTCTCTGTACTTCC (splinkerette) and CCAGCTCTCTGTACTTCC (LTR), with PCR conditions being the same. Third-round adaptor binding primers contained the restriction sites and were therefore specific for each enzyme: Apol, ATGGACTGTTGCTGACACATGTTGGAATT(T/C); Nsp1, CTGGGTCTCGACACATGTTGGAATT(C/T); BstY1, GACTGGTGCTGACACATGTTGGAATT(C/T). The LTR primer was CTGAGAACATCCGCTCTGTGGTAC or CCAGCTCTCTGTACTTCC (splinkerette) and CCAGCTCTCTGTACTTCC (LTR), with PCR conditions being the same. Third-round adaptor binding primers contained the restriction sites and were therefore specific for each enzyme: Apol, ATGGACTGTTGCTGACACATGTTGGAATT(T/C); Nsp1, CTGGGTCTCGACACATGTTGGAATT(C/T); BstY1, GACTGGTGCTGACACATGTTGGAATT(C/T).

Quantitative real-time reverse transcriptase (RT)-PCR. Total mRNA was isolated using an RNeasy MiniKit (Qiagen). After digestion with DNase I (Ambion), the mRNA was reverse-transcribed into single-stranded cDNA using a SuperScript III First Strand Synthesis System for RT-PCR (Invitrogen), followed by quantitative PCR with an Applied Biosystems 7300 real-time PCR cycier. A gene-specific forward and reverse primer (Applied Biosystems) and a gene-specific FAM-labelled MGB probe (Applied Biosystems) were used. MLV, MLVfor AGGCGGTTGAGAGACATCC, MLVrev AGGCGGTTGAGAGACATCC, MLV probe CCCCACCGTGCCCAACCCT (Lötscher et al., 2007); Prex1 (Applied Biosystems); mouse 18S rRNA (Applied Biosystems) was used as endogenous control. Quantitative real-time PCR data were analysed according to the comparative C_T method.

ACKNOWLEDGEMENTS

We thank Leonard Evans for rat monoclonal antibody 83A25 to MLV gp70 and the hybridoma producing it; Dusty Miller for the NZB9-1 plasmid and the Mus dunni cell line with and without this plasmid
integrated; Elizabeth Raveché for the LNC tumour cell line; Christine Kozak for discussions; and Mary McKenney for editing the manuscript. Supported by grants from the NIH (R01AI041570) and the Luspus Research Institute to M. W., and from the United States–Israel Binational Science Foundation to D. E. and M. W. The authors have no financial conflicts of interest.

REFERENCES


Drappa, J., Brot, N. & Elkon, K. B. (1993). The Fas protein is expressed at high levels on CD4+CD8+ thymocytes and activated mature lymphocytes in normal mice but not in the lupus-prone strain, MRL lpr/lpr. Proc Natl Acad Sci U S A 90, 10340–10344.


