Nucleotide sequence of EV1, a British isolate of maedi–visna virus

David R. Sargan,* Ian D. Bennet, Christina Cousens, Douglas J. Roy, Barbara A. Blacklaws, Robert G. Dalziel, Neil J. Watt and Ian McConnell

Department of Veterinary Pathology, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, U.K.

We have isolated a maedi–visna-like virus from the peripheral blood mononuclear cells of a British sheep displaying symptoms of arthritis and pneumonia. After brief passage in fibroblasts this virus (designated EV1) was used to infect choroid plexus cells. cDNA clones of the virus were prepared from these cells and sequenced. Gaps between non-overlapping clones were filled using gene amplification by the polymerase chain reaction. The genome structure is similar to that described for visna virus strain 1514, and differs from that described for visna virus strain SA-OMVV in not having a W reading frame. Overall the genome differs by about 20% between each of these strains, but there is fivefold variation in the amount of divergence of derived amino acid sequences of different open reading frames. Two sequenced EV1 clones each contain only one copy of the 43 bp repeat, with paired AP-1 sites, which is a feature of other ruminant lentiviral long terminal repeats (LTRs). However, analysis of viral DNA in infected cells by gene amplification shows that LTRs with two repeats do occur, albeit at a relatively low frequency.

Introduction

The ovine lentivirus maedi–visna virus causes diseases characterized by incubation periods of months to years, slow progressive symptomologies and chronic and severe inflammatory pathologies. Visna and maedi represent two extremes of presentation of a complex of diseases thought to be caused by a single virus type (i.e. visna virus), and mediated at least in part by the genetic background of the host (Haase, 1986; Cheevers & McGuire, 1988). However, it is likely that some of the observed clinical differences are also mediated by genetic differences between visna virus strains (Narayan & Cork, 1985). In all of these diseases, the major productively infected cell lineage is the monocyte/macrophage (Narayan & Cork, 1985). In all of these diseases, the major productively infected cell lineage is the monocyte/macrophage (Narayan et al., 1982; Gendelman et al., 1985), but there may also be infection of neural and other cell types, leading to a spectrum of symptoms associated with these tissue distributions. Visna disease is associated with locomotive dysfunction, paralysis and wasting. The major sites of inflammation include the choroid plexus and meninges of the brain and sites in the spinal cord, as well as lymphoid tissue. Maedi is characterized by pneumonia caused by massive inflammatory infiltrates of mononuclear cells, and smooth muscle hyperplasia in the lungs. In both diseases there are pronounced lymphadenopathies, and variable degrees of arthritis and mastitis. Three sequences of maedi–visna viruses have been published. Two are from closely related strains of the virus derived originally from the same in vivo isolate from a sheep with visna disease, strains 1514 (Sonigo et al., 1985) and LV1 (Braun et al., 1987). The other, strain SA-OMVV, was derived from an epithelial tumour cell line isolated from a sheep with pneumonia associated with jaagsiekte (Coetzee et al., 1976; Querat et al., 1990). These viruses had been maintained in tissue culture for many years prior to cloning.

Here we present the complete genome sequence of a British visna virus isolated from a sheep which had arthritic and maedi-like lesions. This virus was passaged minimally in tissue culture prior to cloning by cDNA and polymerase chain reaction (PCR) methods. Comparison with other published sequences shows that sequence divergence has occurred most rapidly at sites in proteins which are predicted to be on their surfaces, consistent with selection by the humoral component of the immune system.

Methods

Virus isolation. Virus was isolated by cocultivation of peripheral blood mononuclear cells of a sheep with an acute maedi-like disease with an ovine skin cell line (derived by explant and cultivation of a skin biopsy from an uninfected sheep) until syncytia appeared (12 days). Extracellular virus was passaged twice more through these cells (m.o.i. <0.1 TCID₅₀/cell) and harvested at 12 days post-infection (p.i.). The resulting virus line was named EV1. Another sample of skin cells was infected with this virus, and used as a source of RNA for cloning. Poly(A)⁺ RNA extracted at 7 days p.i. was used as a template for
cDNA synthesis by the method of Gubler & Hoffman (1983). Double-stranded cDNA was cloned into λgt10. The resulting library was screened using the long SacI fragment of a genomic clone of visna virus strain 1514 (Molineaux & Clements, 1983). Positive clones (162) were picked for further analysis. Sequencing of 20 of these clones, and restriction analysis of the rest, showed that they did not contain the whole viral genome. The remainder of the genome was cloned by PCR (Scharf et al., 1986) using the conditions of Ohara et al. (1989), with extrachromosomal DNA purified from infected cells by the method of Hirt (1967) as substrate, and using primers derived from sequenced regions of the EV1 virus (Fig. 1).

Sequencing and sequence analysis. Inserts from positive clones from the λgt10 library or amplification products from PCR were gel-purified on agarose gels and cloned into the phagemid pTZ18R (Pharmacia). Overlapping subclones were prepared using both Sau3AI and the exonuclease Bal 31, and sequenced on both strands (Fig. 1). All parts of the genome were sequenced at least twice on each strand using at least two (and on average four) independent subclones. Single-stranded and double-stranded forms of the phagemid were prepared for sequencing by the dideoxynucleotide chain termination method (Sanger et al., 1977). Sequences were assembled and analysed using version 6.2 of the University of Wisconsin Genetics Computing Group package (UWGCG; Devereux et al., 1984).

Results and Discussion
The cloning strategy and genome structure of EV1 are shown in Fig. 1. The complete DNA sequence is shown in Fig. 2. The virus genome is 9203 bases long and shows a similar gene arrangement to the 1514 and SA-OMVV strains of visna virus. Different open reading frames (ORFs) have from 65% (rev) to 90% (gag) amino acid identity with each of these strains (Table 1). Overall the identity values for the nucleotide sequence are 81% with the published sequence of strain 1514, and 82% with that of SA-OMVV.

Parts of the virus were cloned using PCR which has a significant error rate of 1 per 1500 to 4000 bases amplified on most templates, including lentivirus templates (Keohavong & Thilly, 1989; Meyerhans et al., 1989; Perrin & Gilliland, 1990). It has been suggested that presentation of a PCR-generated sequence is valid if the intrinsic misincorporation rate during cloning is significantly lower than the variation within the locus under study (Meyerhans et al., 1990). In any lentivirus population the replicating virus exists as a cohort of quasi-species. In natural infections in vivo of a single individual, these quasi-species differ by up to 4% in sequence (Balfe et al., 1990). In the most highly variable positions in the genome (the structural sequences of the rev gene), two EV1 cDNA clones overlapping by 516 bases show 10 single base changes and one reciprocal insertion and deletion (2-3% differences). Elsewhere in the env gene, a PCR clone overlapping neighbouring cDNA clones by 550 bases differs from them by five point mutations. Four sequenced copies of a short PCR-generated fragment of the pol gene (positions 3905 to 4096) are identical to each other and to cDNA clones which they overlap by 155 bases. Since overlapping cDNA clones in some parts of the genome show greater differences than overlapping PCR clones in other parts, it is likely that in many parts of the genome differences between individual viruses in
Fig. 2. Sequence of the EVI virus genome. Sequence of the main ORFs is shown, as are the splice donor and acceptor sites used in the subcloned into pTZ18R and sequenced at least twice per subclone from an average of four independent subclones.
Table 1. Sequence divergence of various ORFs between different strains of visna virus

<table>
<thead>
<tr>
<th></th>
<th>Dissimilarity (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>pol</td>
<td>env</td>
</tr>
<tr>
<td></td>
<td>gag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) Amino acid sequence</td>
<td></td>
<td>pro</td>
<td>rt</td>
</tr>
<tr>
<td>EV1/1514</td>
<td>13-9</td>
<td>13-5</td>
<td>20-9</td>
</tr>
<tr>
<td>EV1/SA-OMVV</td>
<td>18-1</td>
<td>13-5</td>
<td>13-5</td>
</tr>
<tr>
<td>1514/SA-OMVV</td>
<td>12-2</td>
<td>20-9</td>
<td>20-9</td>
</tr>
<tr>
<td>(ii) DNA sequence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EV1/1514</td>
<td>13-2</td>
<td>18-5</td>
<td>18-2</td>
</tr>
<tr>
<td>EV1/SA-OMVV</td>
<td>14-4</td>
<td>13-8</td>
<td>17-9</td>
</tr>
<tr>
<td>1514/SA-OMVV</td>
<td>13-0</td>
<td>17-3</td>
<td>20-8</td>
</tr>
</tbody>
</table>

* Percent dissimilarity in amino acid or genomic sequence between the major protein products of pairs of strains is shown.

the EV1 population are greater than those introduced by PCR-based cloning.

**Viral long terminal repeat (LTR)**

The LTR of the virus is 362 bases in length and is flanked by a 5' primer-binding site complementary to the 3' sequence of tRNA\(^{\text{Y}}\) (positions 152 to 167) and by a polypurine tract (positions 8896 to 8911). Inside these priming sites are the short inverted repeats (IRs) GAC (U5) and CTG (U3). The U3 LTR region containing the viral promoter is unusually short (210 bases). Like to the promoters of 1514, SA-OMVV and caprine arthritis-encephalitis virus (CAEV), it contains potential AP-1-binding sites as well as an AP-4-binding site upstream of the TATA box. These sites have been implicated in viral transcription and in activation of the promoter by tat (Gabuzda et al., 1989; Hess et al., 1989). However, EV1 contains only one copy of the 43 bp repeat containing two potential AP-1 sites which occurs twice in these other promoters (Fig. 3a). Two independently derived cDNA clones of the EV1 LTR had this structure, making it unlikely that one repeat unit had been lost as an artefact of cloning. To see whether the population included viruses with two 43 bp repeat units, we performed PCR on extrachromosomal DNA populations. Using a 5' primer with two independent 3' primers spanning the LTR and a variety of amplification conditions, we found that the major amplified product corresponded to molecules containing one repeat unit, but that a small proportion of the product (about 5%) was about 40 bases longer than expected from our EV1 sequence as judged by agarose gel electrophoresis (Fig. 3b). Polyacrylamide gel electrophoresis showed that the longer fragment band actually contained at least three products differing in length by only a few bases and sequencing showed that the lower band contained only one 43 bp repeat. The group of longer fragments did not yield a clean sequence in the region of the 43 bp repeat units, suggesting heterogeneity. Preliminary sequence data suggest that this group contains products with two repeat units, but different products may have recombined at a number of different sites within, and on the 3' side of, the repeat units. As the primers used in isolating these LTR variants bind at unique sites, primer binding is not the cause of the variations seen. Furthermore, the variants are found when a range of Mg\(^{2+}\) concentrations (1.5 to 5 mM) are used for PCR amplification, so that they are unlikely to arise as artefacts of changes in the processivity of the enzyme. We are currently cloning these variants to obtain a full sequence and to study their functional significance.

The R region is 82 bp long and contains the poly(A) signal AATAAA at positions 9186 to 9191. In common with other lentivirus R regions, it encodes an RNA which is capable of folding into complex secondary structures including a conserved stable hairpin at the 5' end (bases 1 to 53 for EV1; standard free energy of formation, \(AG^\circ = -92\) kJ/mol; Freier et al., 1986). In the case of visna viruses, the functional significance of this is not clear, as the presumptive tat protein appears to interact with the viral promoter upstream of the cap site.

The U5 region of EV1 contains several short insertions and deletions when compared with the equivalent regions of visna virus 1514, SA-OMVV or CAEV. A cluster of seven T residues (positions 97 to 103) occupies the same position as an AT-rich region found in both other visna virus isolates, CAEV and other lentiviruses.
The gag gene

The gag ORF encodes a polypeptide of 51K which contains consensus aspartate protease cleavage sites suggesting mature gag proteins of 16.8K (p16), 24.7K (p25), and 9.6K (p14). The 25K gag protein (together with the reverse transcriptase core domain of pol) is the most highly conserved of any of the viral proteins (differing at about 4% of residues between each of the three visna virus strains), whereas p14 is much more variable, differing at 14% of residues between EV1 and 1514, and containing five amino acid insertions in EV1 and SA-OMVV relative to 1514 at positions 1775 to 1789 (EV1 sequence). These two insertions, which differ from each other in sequence, are followed by a further five nucleotides of unrelated sequence in all three visna virus strains immediately preceding sequences which are probably involved in ribosome slippage in the generation of the gag/pol precursor (see below). A subline of the EV1 virus, maintained in culture for some 8 months after cloning, had acquired a further amino acid change in this region (data not shown). Thus, the region of gag around the ribosomal slippage site may be a mutational hotspot.

The env gene

The env ORF encodes a polypeptide of 114.7K. A putative proteolytic cleavage signal (RKRR) at position 7982 separates the mature extracellular protein from the transmembrane segment to give proteins of 76.6K and 38.1K prior to glycosylation or other modifications. There are 29 putative glycosylation sites in the env ORF of EV1, of which 25 are shared with visna virus 1514, whereas the other four are found among clustered sites, and are shifted by only a few amino acids from three equivalent sites in strain 1514. The transmembrane segment differs by about 20% from both 1514 and SA-OMVV, but putative fusion and transmembrane do-
mains of this protein and sequences which may interact with the major extracellular glycoprotein (by analogy with the similar HIV transmembrane glycoprotein) are all well conserved (Komalski et al., 1987; Felser et al., 1989; Freed et al., 1990). The extracellular envelope glycoproteins (gp110) of all three visna viruses show areas of highly conserved sequence as well as areas of high variability (Fig. 4). The first exon of the rev gene is coincident with the N terminus of env. The rev gene is the most rapidly evolving throughout the lentivirus group; rev protein sequences show more inter-strain variation than tat protein sequences, although the underlying genomic variation is similar for both gene products (Table 1). The rev gene also varies more than gp41 in those parts of the genome which encode both. Thus, non-silent mutations are accepted more readily by rev than by tat or gp41. If there is selection for rev variants, this may drive rapid evolution in the N terminus of the env protein. Elsewhere, variable residues are grouped at sites which are largely in hydrophilic and possibly external positions in the env protein (Hopp & Woods, 1981; Emini et al., 1985) consistent with variations being fixed by selection of the immune system. Selection acting upon antigenic variants of visna virus env is well documented (Narayan et al., 1978, 1981; Scott et al., 1979). Two possible hypervariable domains exist towards the carboxyl end of the extracellular portion of env, distinguished by the insertion and deletion of short repeated sequence elements as well as by point mutation, a mechanism important in the generation of diversity in HIV hypervariable regions (Starich et al., 1986; Balfé et al., 1990). These domains occur at residues 579 to 607 and 649 to 664 in EV1 env. The latter is immediately prior to the likely protease cleavage site for env maturation and is the major site of variation between the closely related strains 1514 and LV1 (Sonigo et al., 1985; Braun et al., 1987). The former is between conserved cysteine residues and includes a predicted β-turn; it may be analogous to the third hypervariable loop of HIV (Javaherian et al., 1986). When this env sequence is compared with that of CAEV (Saltarelli et al., 1990), variable sites are found in the same regions as seen with other visna virus isolates, though CAEV is almost twice as dissimilar to the visna virus group as they are to each other.

**Short open reading frames**

ORFs encoding vif- and tat-like proteins of 28-2K and 11-1K respectively, are present. A further ORF has been identified in cDNA clones (Fig. 5), and contains a first exon which shares the methionine and first 45 amino acids of the env gene spliced to a 3′ ORF at position 8589, similar to the structure reported for other visna viruses (Mazarin et al., 1988; Querat et al., 1990). The putative protein encoded by this reading frame is 18.5K and is the least conserved protein amongst sequenced visna virus isolates (approximately 35% difference between each virus). Nonetheless it shares some sequence features with the rev protein of HIV-1 and HIV-2, including grouped basic residues, and a sequence GLEKLTLE with homology to the sequence PLERLTLD known to be required for rev function in HIV-1 and conserved between sequenced HIVs and SIVs are shown underlined. Amino acids encoded by the polypurine stop codons are largely in hydrophilic and possibly external positions in the Env protein. Elsewhere, variable residues are grouped at sites which are largely in hydrophilic and possibly external positions in the env protein (Hopp & Woods, 1981; Emini et al., 1985) consistent with variations being fixed by selection of the immune system. Selection acting upon antigenic variants of visna virus env is well documented (Narayan et al., 1978, 1981; Scott et al., 1979). Two possible hypervariable domains exist towards the carboxyl end of the extracellular portion of env, distinguished by the insertion and deletion of short repeated sequence elements as well as by point mutation, a mechanism important in the generation of diversity in HIV hypervariable regions (Starich et al., 1986; Balfé et al., 1990). These domains occur at residues 579 to 607 and 649 to 664 in EV1 env. The latter is immediately prior to the likely protease cleavage site for env maturation and is the major site of variation between the closely related strains 1514 and LV1 (Sonigo et al., 1985; Braun et al., 1987). The former is between conserved cysteine residues and includes a predicted β-turn; it may be analogous to the third hypervariable loop of HIV (Javaherian et al., 1986). When this env sequence is compared with that of CAEV (Saltarelli et al., 1990), variable sites are found in the same regions as seen with other visna virus isolates, though CAEV is almost twice as dissimilar to the visna virus group as they are to each other.

**Short open reading frames**

ORFs encoding vif- and tat-like proteins of 28-2K and 11-1K respectively, are present. A further ORF has been identified in cDNA clones (Fig. 5), and contains a first exon which shares the methionine and first 45 amino acids of the env gene spliced to a 3′ ORF at position 8589, similar to the structure reported for other visna viruses (Mazarin et al., 1988; Querat et al., 1990). The putative protein encoded by this reading frame is 18.5K and is the least conserved protein amongst sequenced visna virus isolates (approximately 35% difference between each virus). Nonetheless it shares some sequence features with the rev protein of HIV-1 and HIV-2, including grouped basic residues, and a sequence GLEKLTLE with homology to the sequence PLERLTLD known to be required for rev function in HIV-1 and conserved between sequenced HIVs and SIVs are shown underlined. Amino acids encoded by the polypurine stop codons are largely in hydrophilic and possibly external positions in the Env protein. Elsewhere, variable residues are grouped at sites which are largely in hydrophilic and possibly external positions in the env protein (Hopp & Woods, 1981; Emini et al., 1985) consistent with variations being fixed by selection of the immune system. Selection acting upon antigenic variants of visna virus env is well documented (Narayan et al., 1978, 1981; Scott et al., 1979). Two possible hypervariable domains exist towards the carboxyl end of the extracellular portion of env, distinguished by the insertion and deletion of short repeated sequence elements as well as by point mutation, a mechanism important in the generation of diversity in HIV hypervariable regions (Starich et al., 1986; Balfé et al., 1990). These domains occur at residues 579 to 607 and 649 to 664 in EV1 env. The latter is immediately prior to the likely protease cleavage site for env maturation and is the major site of variation between the closely related strains 1514 and LV1 (Sonigo et al., 1985; Braun et al., 1987). The former is between conserved cysteine residues and includes a predicted β-turn; it may be analogous to the third hypervariable loop of HIV (Javaherian et al., 1986). When this env sequence is compared with that of CAEV (Saltarelli et al., 1990), variable sites are found in the same regions as seen with other visna virus isolates, though CAEV is almost twice as dissimilar to the visna virus group as they are to each other.

**Short open reading frames**

ORFs encoding vif- and tat-like proteins of 28-2K and 11-1K respectively, are present. A further ORF has been identified in cDNA clones (Fig. 5), and contains a first exon which shares the methionine and first 45 amino acids of the env gene spliced to a 3′ ORF at position 8589, similar to the structure reported for other visna viruses (Mazarin et al., 1988; Querat et al., 1990). The putative protein encoded by this reading frame is 18.5K and is the least conserved protein amongst sequenced visna virus isolates (approximately 35% difference between each virus). Nonetheless it shares some sequence features with the rev protein of HIV-1 and HIV-2, including grouped basic residues, and a sequence GLEKLTLE with homology to the sequence PLERLTLD known to be required for rev function in HIV-1 and conserved between sequenced HIVs and SIVs are shown underlined. Amino acids encoded by the polypurine stop codons are largely in hydrophilic and possibly external positions in the Env protein. Elsewhere, variable residues are grouped at sites which are largely in hydrophilic and possibly external positions in the env protein (Hopp & Woods, 1981; Emini et al., 1985) consistent with variations being fixed by selection of the immune system. Selection acting upon antigenic variants of visna virus env is well documented (Narayan et al., 1978, 1981; Scott et al., 1979). Two possible hypervariable domains exist towards the carboxyl end of the extracellular portion of env, distinguished by the insertion and deletion of short repeated sequence elements as well as by point mutation, a mechanism important in the generation of diversity in HIV hypervariable regions (Starich et al., 1986; Balfé et al., 1990). These domains occur at residues 579 to 607 and 649 to 664 in EV1 env. The latter is immediately prior to the likely protease cleavage site for env maturation and is the major site of variation between the closely related strains 1514 and LV1 (Sonigo et al., 1985; Braun et al., 1987). The former is between conserved cysteine residues and includes a predicted β-turn; it may be analogous to the third hypervariable loop of HIV (Javaherian et al., 1986). When this env sequence is compared with that of CAEV (Saltarelli et al., 1990), variable sites are found in the same regions as seen with other visna virus isolates, though CAEV is almost twice as dissimilar to the visna virus group as they are to each other.
for rev function in HIV-1, and is conserved between the human and simian immunodeficiency viruses (SIV) (Perkins et al., 1989; Venkatesh et al., 1990). This protein has been termed rev in visna virus by Mazarin et al. (1988). The same cDNA could also encode a protein with a short first exon linked to the carboxyl end of the env transmembrane glycoprotein, as reported for visna virus 1514 (Mazarin et al., 1988). The W′ ORF reported for the SA-OMVV isolate is greatly truncated at the N terminus if present, although there is a methionine residue preceding an ORF of 36 amino acids immediately 5′ to tat, which shows amino acid homology with the carboxyl end of the W′ ORF. This very short ORF is also present in strain 1514 virus.

Visna virus contains ORFs with some homology to the regulatory genes vir, tat and rev of HIV-1. It is unclear whether nef or other HIV regulatory proteins have visna virus equivalents. We have previously suggested that the rev ORF of visna virus 1514 might also have nef activity (Sargan & Bennet, 1989). The nef gene product of HIV has a cytoplasmic location (Franchini et al., 1986; Guy et al., 1990); recently it has been shown that visna virus strain 1514 rev protein may also be primarily cytoplasmic (Mazarin et al., 1990), in contrast to HIV rev, which is predominantly located in the nucleus (Cullen et al., 1988; Felber et al., 1989). In HIV it has been reported that the sequence KEKGGLEG is essential for GTPase action (Samuel et al., 1987; Guy et al., 1990), although the presence and significance of this activity has been disputed (Kaminchik et al., 1990). A related sequence occurs at a more carboxyl location in 1514, but in EV1 and SA-OMVV the same sequence is separated from the rev ORF by a single stop codon (sequence KRKGGLSG, positions 8900 to 8923 in EV1). In all visna virus strains a further sequence constraint is imposed by the polyuridine tract which underlies the first four residues of this sequence, but only in the rev ORF does it give rise to the KRKGGLSG peptide. HIV in tissue culture often produces a non-functional or smaller nef protein, sometimes by introduction of a stop codon (Sodroski et al., 1986; Guy et al., 1990; Laurent et al., 1990; Zweig et al., 1990). If the visna virus 1514 rev protein does have nef function, it is possible that a similar event has happened in EV1 and SA-OMVV.

**Other sequence features**

A second polyuridine tract, exactly repeating that which flanks U3, is found at positions 4757 to 4772. A similar feature has been reported in other visna viruses. The conservation of this feature, which probably functions as an alternative RNase H-resistant primer site for second DNA strand synthesis, suggests that the gapped DNA structure is functionally important in these viruses.

**Evolution of EV1**

EV1 is slightly more different from either 1514 or SA-OMVV than the latter two are from each other. The known history of the Icelandic sheep flock suggests that 1514 and SA-OMVV have undergone divergent evolution for at least 42 years (1933 to 1975) (Querat et al., 1990), and 1514 has diverged from EV1 for at least 55 years, assuming divergence prior to visna virus being imported to Iceland in 1933. EV1 is almost equally diverged from 1514 and SA-OMVV, suggesting that divergence occurred at about the same time. There is thus a possibility of a common origin for these strains at about this period. These three viruses are somewhat more diverged than any other species-specific lentivirus subgroup.

We wish to thank Drs J. E. Clements and O. Narayan for visna virus 1514 DNA clones and Dr M. Dawson for blood from a visna virus-infected sheep. Part of the computing analysis was performed on the SERC VAX at Daresbury Laboratory. The work was supported by a Wellcome Trust Project Development Award. C.C. is in receipt of an MRC studentship, B.A.B. is supported by MRC ADP grant number G890079. R.G.D. is supported by a Wellcome Trust Lectureship.

**References**


