Proteomic analysis of *Glossina pallidipes* salivary gland hypertrophy virus virions for immune intervention in tsetse fly colonies

Henry M. Kariithi,1,2† Ikbal Agah Incé,1,3† Sjef Boeren,4 Jacques Vervoort,4 Max Bergoin,5 Monique M. van Oers,1 Adly M. M. Abd-Alla2 and Just M. Vlak†

1Laboratory of Virology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
2Insect Pest Control Laboratory, Joint FAO/IAEA Programme of Nuclear Techniques in Food and Agriculture, Vienna, Austria
3Department of Biology, Faculty of Arts and Sciences, Giresun University, 28049 Giresun, Turkey
4Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands
5Laboratoire de Pathologie Comparée, Université Montpellier 2, Place Eugene Bataillon, 34095 Montpellier, France

Many species of tsetse flies (Diptera: Glossinidae) can be infected by a virus that causes salivary gland hypertrophy (SGH). The genomes of viruses isolated from *Glossina pallidipes* (GpSGHV) and *Musca domestica* (MdSGHV) have recently been sequenced. Tsetse flies with SGH have reduced fecundity and fertility which cause a serious problem for mass rearing in the frame of sterile insect technique (SIT) programmes to control and eradicate tsetse populations in the wild. A potential intervention strategy to mitigate viral infections in fly colonies is neutralizing of the GpSGHV infection with specific antibodies against virion proteins. Two major GpSGHV virion proteins of about 130 and 50 kDa, respectively, were identified by Western analysis using a polyclonal rabbit antibody raised against whole GpSHGV virions. The proteome of GpSGHV, containing the antigens responsible for the immune-response, was investigated by liquid chromatography tandem mass spectrometry and 61 virion proteins were identified by comparison with the genome sequence. Specific antibodies were produced in rabbits against seven candidate proteins, including the ORF10/C-terminal fragment, ORF47 and ORF96 as well as proteins involved in peroral infectivity PIF-1 (ORF102), PIF-2 (ORF53), PIF-3 (ORF76) and P74 (ORF1). Antiserum against ORF10 specifically reacted to the 130 kDa protein in a Western blot analysis and to the envelope protein of GpSGHV, detected by using immunogold-electron microscopy. This result suggests that immune intervention of viral infections in colonies of *G. pallidipes* is a realistic option.

**INTRODUCTION**

Tsetse flies (Diptera: Glossinidae) are the vectors of sleeping sickness in humans (or human African trypanosomosis, HAT) and the livestock disease nagana (or African animal trypanosomosis, AAT) (Steelman, 1976; WHO, 2001). The continuing presence and advancement of tsetse flies prevents the development of sustainable and profitable livestock production systems, thus greatly influencing food production, natural-resource utilization and human settlement in almost two thirds of sub-Sahara Africa (Jordan, 1986). There are no effective vaccines against trypanosomosis and the disease is mainly managed by the prophylactic and curative treatment with trypanocidal drugs. However, there have been reports of development of resistance to the available trypanocidal drugs (Aksoy & Rio, 2005). It is generally accepted that vector control remains the most effective way of managing the disease and the sterile insect technique (SIT) that relies upon the sequential release of sterile male flies into the wild has proven to be a robust technique for use in an area-wide integrated pest management (AW-IPM) approach (Hendrichs et al., 2007; Vreysen et al., 2000). Mating of the sterile males with virgin
female wild flies produces no offspring, which leads to a reduction in tsetse fly population density. *Glossina austeni* has been successfully eradicated from the Island of Unguja, United Republic of Tanzania using insecticide impregnated targets, insecticide ‘pour on’ on livestock and the release of sterile insects, and efforts are being made to do the same in the Southern Rift Valley of Ethiopia (Feldmann, 2005).

Sterile males for AW-IPM programmes with an SIT component are produced in mass rearing facilities and sterilized with ionizing radiation (usually $^{60}\text{Co}$ or $^{137}\text{Ce}$). However, the production of *Glossina pallidipes* flies is hampered by the fact that the fly colonies are contaminated by a salivary gland hypertrophy virus (GpSGHV) (Ellis & Maudlin, 1987; Jaenson, 1978; Jura et al., 1987, 1989), which affects the productivity and fecundity of these colonies. The low productivity of these colonies makes the rearing very cumbersome and often leads to colony collapse.

GpSGHV is a rod-shaped, double-stranded, circular DNA virus with a genome of 190 kbp and averaging 70 × 640 nm in size (Fig. 1) (Abd-Alla et al., 2008; Odindo et al., 1986). The presence of the virus in the salivary glands of male and female flies explains the hypertrophied appearance (hyperplasia) of these glands. The virus has also been associated with testicular degeneration and ovarian abnormalities (Kokwaro et al., 1990; Sang et al., 1998, 1999). Although it is not yet clear how the fly gets infected with the virus and how exactly the virus affects the mating and feeding behaviour of the fly, it has been demonstrated recently that in mass-rearing facilities the virus is released from the infected fly with saliva upon blood ingestion and transmitted through the blood to other flies (Abd-Alla et al., 2010). In nature, the virus is probably transmitted vertically from mother to offspring, either trans-ovum or through infected milk glands (Jura et al., 1989; Sang et al., 1996, 1998) and the infection is largely asymptomatic. Similar viruses have been described from the bulb fly *Merodon equestris* (Amargier et al., 1979) and the house fly *Musca domestica* (Coker et al., 1993).

The entire GpSGHV genome has been sequenced (Abd-Alla et al., 2008) and 160 ORFs have been identified. Putative functions could be assigned to only very few ORFs by blasting against databases. Most notable is the presence of homologues of the *per os* infectivity factor (PIF) genes of baculoviruses. These factors are involved in the oral transmission of baculoviruses from insect to insect. Their presence in GpSGHV suggests a similar mechanism of infection, which is in fact compatible with GpSGHV transmission via the blood meal.

One strategy to control the GpSGHV infections of *G. pallidipes* flies in mass-rearing facilities is to prevent horizontal GpSGHV transmission by immune-complexing the virus in the blood meal and/or in the saliva. In this paper, we report the proteomic analysis of the GpSGHV proteome and the identification of the immune-responsive, virus-encoded proteins for the development of antibodies to be used in immune intervention in order to prevent SGHV infections in *G. pallidipes* colonies.

**RESULTS AND DISCUSSION**

**Analysis of the GpSGHV proteome**

The recent sequencing of the Uganda isolate of GpSGHV (Abd-Alla et al., 2008, 2009) has provided information that greatly facilitated the assignment of ORFs for virion proteins in the genome. The major disadvantage in our study is the lack of a cell culture system for the production of GpSGHV virions which made it rather difficult to purify a large amount of GpSGHV virions with high quality. GpSGHV virions were therefore, purified from hypertrophied salivary gland cells dissected from infected tsetse flies. In addition, to minimize the disadvantage of losing the envelope protein as observed after purification of GpSGHV by using sucrose gradients as described previously (Abd-Alla et al., 2007), the virus particles were purified over a Nycodenz gradient, resulting in only a single viral band at a density of 1.153 g cm$^{-3}$ (Fig. 1a). The integrity of the virions after Nycodenz purification was checked by transmission electron microscopy (TEM), but they appeared to be fragile (Fig. 1b, c). It is possible that some host proteins have been co-purified, but also that they are intricately associated with virions, such as actin (Lanier & Volkman, 1998; Wang et al., 2010). Separation of the purified GpSGHV proteins by gradient SDS-PAGE revealed at least 35 proteins ranging in size from 10 to >130 kDa (Fig. 2a). The most abundant proteins run at about 43–50 kDa (multiple bands) and 130 kDa.

**Fig. 1.** Purification of GpSGHV virions used for MS analysis. Virions from hypertrophied salivary glands were purified by Nycodenz gradient centrifugation (a) the arrow indicates a single distinct band that was selected after centrifugation for 1 h at 150 000 g. Micrographs show a negatively stained ultrathin section (b) and a suspension of the isolated GpSGHV (c).
The gel lane was divided into seven slices containing proteins with a molecular mass ranging from lower than 26, 26–34, 34–43, 43–55, 55–95, 95–130 to 130 kDa, respectively (Fig. 2b), and protein extracts of each of the slices were subjected to liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. A decoy database strategy was used (Elias & Gygi, 2007) which, after applying the appropriate filters, resulted in 90 protein hits: 61 viral proteins, 28 contaminants and one decoy hit giving a false discovery rate of 1.1 %. The LC-MS/MS method allowed the detection of GpSGHV proteins that were present in relatively low quantities. Fifty-seven GpSGHV proteins were identified with two or more peptides (Table 1). Manual verification of the four peptides with a single identified peptide not only revealed a sufficient overlap between virtual and measured MS/MS spectra, but also showed that no other peptides present in the database can explain the measured spectra.

The ORFs corresponding to 61 virion proteins were superposed on the physical map of GpSGHV (Fig. 3). The virion protein genes were almost equally distributed over both strands of the genome. Within the genome two segments encoding ORF11–26 and ORF114–139 were devoid of virion protein genes. In contrast, other segments (ORF78–113) were densely populated with virion protein genes (22 of 29, 76 %). Also ORF62, a giant ORF was identified as encoding a constituent of GpSGHV virions representing a protein of about 511 kDa. Such a large virion protein is not unusual for large dsDNA viruses, as white spot syndrome virus has a 664 kDa virion protein that is a major nucleocapsid protein (Leu et al., 2005; van Hulten et al., 2001). Whereas the 511 kDa protein is probably a minor component (1.8 % peptide coverage), another large protein of 127 kDa (ORF10) was found in high abundance (with 16.9 % coverage) and probably represents the 130 kDa protein seen in SDS-PAGE (Fig. 2a).

**Gene homology and domain analyses**

Homology searches with the identified GpSGHV proteins performed against GenBank/EMBL, Swisprot and PIR databases revealed that of the 61 proteins identified, four were unique to GpSGHV (encoded by ORFs 2, 39, 47 and 49). Neither the nucleotide sequences nor the deduced protein sequences of these genes shared significant homology to other genes/proteins identified so far. In addition, among the GpSGHV proteins identified by LC-MS/MS, five proteins had homologues in baculoviruses and nudiviruses (ORFs 1, 53, 76, 102 and 110) and one had homology to a nimavirus ORF (ORF36) (Table 1).

Among the GpSGHV virion proteins identified (Table 1) there are homologues of all four baculovirus per os infectivity factors (PIF-1, PIF-2, PIF-3 and PIF-0/P74). These factors, encoded by ORFs 1, 53, 76 and 102, are thought to be involved in the oral infection process of baculovirus occlusion-derived virions by binding to midgut epithelial cells (Slack & Arif, 2007). PIF-1, PIF-2, and P74 have been shown to mediate specific binding of ODVs to midgut cells, suggesting that they are directly involved in the virus–host-cell interaction at the initial step of infection (Kikhno et al., 2002; Ohkawa et al., 2005; Slack et al., 2010). Although PIF-3 appears to be an occlusion-derived virus (ODV)-associated protein (Li et al., 2007), it does not appear to be involved in specific binding and its function is not known yet. The PIF proteins are absolutely essential for baculovirus oral infectivity (Kikhno et al., 2002; Kuzio et al., 1989; Peng et al., 2010; Pijlman et al., 2003; Slack et al., 2010; Song et al., 2008). The presence of these PIF proteins in the GpSGHV proteome is a strong indication that they might have a similar function in GpSGHV infection following uptake via a blood meal.

**Comparative proteomics of GpSGHV and MdSGHV**

Recently, a limited proteomic analysis has been carried out for an SGHV from the house fly *Musca domestica* (MdSGHV), the genome of which is also entirely sequenced (Garcia-Maruniak et al., 2008). This virus is highly related to GpSGHV and is proposed to be accommodated in a newly proposed virus family Hytrosaviridae (Abd-Alla et al., 2009;
Table 1. The GpSGHV virion proteins identified by MS in order of descending molecular mass (511–10 kDa)

<table>
<thead>
<tr>
<th>ORF</th>
<th>GenBank accession no.</th>
<th>Predicted molecular mass (kDa)</th>
<th>Protein coverage (% aa)</th>
<th>No. of peptides identified</th>
<th>pI</th>
<th>Characteristics of deduced proteins</th>
<th>Conserved domain(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>62</td>
<td>YP_001687010</td>
<td>511.81</td>
<td>1.8</td>
<td>7</td>
<td>4.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>YP_001686993</td>
<td>200.98</td>
<td>4.1</td>
<td>7</td>
<td>5.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>YP_001686986</td>
<td>136.46</td>
<td>42.20</td>
<td>130</td>
<td>5.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>YP_001686958</td>
<td>126.96</td>
<td>81</td>
<td>305</td>
<td>5.60</td>
<td>Major protein</td>
<td>TM</td>
</tr>
<tr>
<td>40</td>
<td>YP_001686988</td>
<td>104.065</td>
<td>3.9</td>
<td>3</td>
<td>9.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>YP_001687031</td>
<td>81.54</td>
<td>13</td>
<td>9</td>
<td>5.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>YP_001686949</td>
<td>81.35</td>
<td>13.8</td>
<td>12</td>
<td>4.73</td>
<td>p74 (bac) Md039</td>
<td>TM</td>
</tr>
<tr>
<td>104</td>
<td>YP_001687052</td>
<td>77.84</td>
<td>7</td>
<td>6</td>
<td>7.30</td>
<td></td>
<td>TM, coiled-coil region</td>
</tr>
<tr>
<td>88</td>
<td>YP_001687036</td>
<td>77.71</td>
<td>15.3</td>
<td>11</td>
<td>7.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>YP_001687050</td>
<td>76.07</td>
<td>12.4</td>
<td>9</td>
<td>5.02</td>
<td>PIF-1 (bac) Md090</td>
<td>TM</td>
</tr>
<tr>
<td>71</td>
<td>YP_001687019</td>
<td>71.95</td>
<td>2</td>
<td>1</td>
<td>6.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>YP_001687034</td>
<td>70.13</td>
<td>29.4</td>
<td>28</td>
<td>7.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>YP_001687012</td>
<td>69.98</td>
<td>43.7</td>
<td>42</td>
<td>8.95</td>
<td>ATPase/Helicase (Md097)</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>YP_001687056</td>
<td>63.90</td>
<td>30.5</td>
<td>15</td>
<td>5.34</td>
<td>Cell division 48-like protein (Md033) (Md084)</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>YP_001686994</td>
<td>61.50</td>
<td>37.7</td>
<td>24</td>
<td>6.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>YP_001687055</td>
<td>59.53</td>
<td>45.1</td>
<td>31</td>
<td>5.99</td>
<td>Cell division 48-like protein (Md33)</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>YP_001687009</td>
<td>57.40</td>
<td>38.3</td>
<td>14</td>
<td>8.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>YP_001687054</td>
<td>55.05</td>
<td>31.3</td>
<td>15</td>
<td>5.31</td>
<td></td>
<td>Coiled-coil region</td>
</tr>
<tr>
<td>27</td>
<td>YP_001686975</td>
<td>53.03</td>
<td>5.5</td>
<td>3</td>
<td>6.12</td>
<td>Chitinase Chit1 precursor-like protein</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>YP_001687018</td>
<td>50.89</td>
<td>23.9</td>
<td>13</td>
<td>9.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>YP_001686989</td>
<td>48.74</td>
<td>24.9</td>
<td>15</td>
<td>6.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>YP_001687088</td>
<td>48.41</td>
<td>10</td>
<td>5</td>
<td>9.26</td>
<td></td>
<td>Coiled-coil region</td>
</tr>
<tr>
<td>47</td>
<td>YP_001686995</td>
<td>47.16</td>
<td>61.9</td>
<td>98</td>
<td>4.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>YP_001687045</td>
<td>44.37</td>
<td>62.9</td>
<td>83</td>
<td>8.61</td>
<td></td>
<td>(Md025)</td>
</tr>
<tr>
<td>96</td>
<td>YP_001687044</td>
<td>43.50</td>
<td>77.2</td>
<td>170</td>
<td>6.57</td>
<td>Major protein (Md025)</td>
<td>TM</td>
</tr>
<tr>
<td>44</td>
<td>YP_001686992</td>
<td>42.81</td>
<td>20.3</td>
<td>8</td>
<td>8.84</td>
<td></td>
<td>SP</td>
</tr>
<tr>
<td>34</td>
<td>YP_001686982</td>
<td>41.15</td>
<td>7.7</td>
<td>3</td>
<td>9.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>YP_001686955</td>
<td>41.02</td>
<td>15.3</td>
<td>5</td>
<td>9.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>YP_001686981</td>
<td>40.98</td>
<td>10.3</td>
<td>3</td>
<td>9.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>YP_001687001</td>
<td>40.21</td>
<td>27.8</td>
<td>11</td>
<td>5.05</td>
<td>PIF-2 (bac) (Md071)</td>
<td></td>
</tr>
<tr>
<td>154</td>
<td>YP_001687102</td>
<td>40.07</td>
<td>8</td>
<td>4</td>
<td>7.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>YP_001686950</td>
<td>38.64</td>
<td>31.8</td>
<td>17</td>
<td>8.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>YP_001687041</td>
<td>38.51</td>
<td>57.8</td>
<td>39</td>
<td>9.53</td>
<td></td>
<td>(Md022)</td>
</tr>
<tr>
<td>39</td>
<td>YP_001686987</td>
<td>37.64</td>
<td>4.8</td>
<td>1</td>
<td>4.47</td>
<td></td>
<td>TM, SP</td>
</tr>
<tr>
<td>52</td>
<td>YP_001687000</td>
<td>36.67</td>
<td>6.5</td>
<td>7</td>
<td>8.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>YP_001687053</td>
<td>34.75</td>
<td>7.9</td>
<td>2</td>
<td>9.82</td>
<td></td>
<td>Coiled-coil region</td>
</tr>
<tr>
<td>31</td>
<td>YP_001686979</td>
<td>33.58</td>
<td>19.3</td>
<td>5</td>
<td>9.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>YP_001687061</td>
<td>33.10</td>
<td>34.5</td>
<td>11</td>
<td>8.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>YP_001686998</td>
<td>32.72</td>
<td>29.2</td>
<td>45</td>
<td>6.33</td>
<td></td>
<td>(Md086)</td>
</tr>
<tr>
<td>94</td>
<td>YP_001687042</td>
<td>32.68</td>
<td>57.5</td>
<td>34</td>
<td>9.28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Garcia-Maruniak et al., 2009). GpSGHV has 42 ORFs with homologues in MdSGHV (Garcia-Maruniak et al., 2009). Comparative analysis of the GpSGHV virion proteins with all MdSGHV ORFs (Garcia-Maruniak et al., 2008) showed that 33 of the 61 identified GpSGHV virion proteins were homologous to 29 MdSGHV ORFs (Table 2). The difference in ORF numbers (33 compared with 29) follows from the notion that four MdSGHV ORFs have pairs of homologues in GpSGHV (see below). Of the 29 MdSGHV ORFs homologous to the GpSGHV proteome ORFs, only 13 were actually identified in the MdSGHV proteome (Garcia-Maruniak et al., 2008).

The homology of the virion proteins and the presence of a number of proteins shared between GpSGHV and MdSGHV virions suggests that the two viruses have further properties in common, such as their virion structure and assembly mechanisms, and gives further credence to the

<table>
<thead>
<tr>
<th>ORF no.</th>
<th>GenBank accession no.</th>
<th>Predicted molecular mass (kDa)</th>
<th>Protein coverage (% aa)</th>
<th>No. of peptides identified</th>
<th>pI</th>
<th>Characteristics of deduced proteins</th>
<th>Conserved domain(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>YP_001687020</td>
<td>31.75</td>
<td>27.9</td>
<td>7</td>
<td>8.32</td>
<td>(Md102)</td>
<td>TM</td>
</tr>
<tr>
<td>8</td>
<td>YP_001686956</td>
<td>31.58</td>
<td>31.4</td>
<td>9</td>
<td>9.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>YP_001687039</td>
<td>31.47</td>
<td>7.1</td>
<td>2</td>
<td>7.96</td>
<td></td>
<td>TM, SP</td>
</tr>
<tr>
<td>67</td>
<td>YP_001687015</td>
<td>30.99</td>
<td>47.1</td>
<td>30</td>
<td>4.55</td>
<td></td>
<td>TM, SP</td>
</tr>
<tr>
<td>69</td>
<td>YP_001687017</td>
<td>30.89</td>
<td>41.9</td>
<td>49</td>
<td>7.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>YP_001686980</td>
<td>30.08</td>
<td>13.9</td>
<td>3</td>
<td>4.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>YP_001687033</td>
<td>30.07</td>
<td>40</td>
<td>15</td>
<td>9.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>YP_001687026</td>
<td>27.83</td>
<td>9.3</td>
<td>3</td>
<td>7.93</td>
<td></td>
<td>TM</td>
</tr>
<tr>
<td>146</td>
<td>YP_001687094</td>
<td>26.67</td>
<td>9</td>
<td>2</td>
<td>9.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>YP_001687032</td>
<td>25.75</td>
<td>28.8</td>
<td>8</td>
<td>9.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>YP_001687024</td>
<td>24.39</td>
<td>9.5</td>
<td>2</td>
<td>4.34</td>
<td>PIF-3 (bac)</td>
<td>TM</td>
</tr>
<tr>
<td>110</td>
<td>YP_001687058</td>
<td>23.62</td>
<td>27.9</td>
<td>4</td>
<td>7.71</td>
<td>mp-nase (bac)</td>
<td>TM</td>
</tr>
<tr>
<td>112</td>
<td>YP_001687060</td>
<td>19.05</td>
<td>19.2</td>
<td>8</td>
<td>4.64</td>
<td></td>
<td>TM</td>
</tr>
<tr>
<td>82</td>
<td>YP_001687030</td>
<td>18.67</td>
<td>23.3</td>
<td>3</td>
<td>9.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>YP_001686991</td>
<td>16.89</td>
<td>69.4</td>
<td>28</td>
<td>9.91</td>
<td>(Md072)</td>
<td>TM</td>
</tr>
<tr>
<td>30</td>
<td>YP_001686978</td>
<td>15.91</td>
<td>10.5</td>
<td>1</td>
<td>10.17</td>
<td>Thymidylate synthase (nima)</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>YP_001686984</td>
<td>13.79</td>
<td>12.2</td>
<td>2</td>
<td>5.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>YP_001687046</td>
<td>13.52</td>
<td>51.3</td>
<td>10</td>
<td>10.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>YP_001687016</td>
<td>12.64</td>
<td>37</td>
<td>6</td>
<td>7.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>YP_001687049</td>
<td>12.33</td>
<td>62.3</td>
<td>37</td>
<td>9.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>YP_001686997</td>
<td>10.00</td>
<td>12.2</td>
<td>1</td>
<td>8.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Positioning of the 61 virion (structural) proteins encoded by GpSGHV on the genomic map of GpSGHV (Abd-Alla et al., 2008). The arrows indicate the positions and the direction of transcription for the ORFs.
Table 2. Proteins represented in the proteome of GpSGHV with homologue genes in MdSGHV genome

The MdSGHV homologues that were also identified in the MdSGHV proteome are in bold. MdSGHV proteome data (Garcia-Maruniak et al., 2008, 2009). Four GpSGHV ORFs share a homologue in MdSGHV. ORF length is given in amino acids.

<table>
<thead>
<tr>
<th>GpSGHV ORF</th>
<th>MdSGHV ORF</th>
<th>GpSGHV ORF</th>
<th>MdSGHV ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF Length</td>
<td>ORF Length</td>
<td>ORF Length</td>
<td>ORF Length</td>
</tr>
<tr>
<td>1 (p74)</td>
<td>696</td>
<td>39 (p74)</td>
<td>707</td>
</tr>
<tr>
<td>30</td>
<td>133</td>
<td>82</td>
<td>359</td>
</tr>
<tr>
<td>31</td>
<td>285</td>
<td>74</td>
<td>698</td>
</tr>
<tr>
<td>32</td>
<td>259</td>
<td>70</td>
<td>967</td>
</tr>
<tr>
<td>33</td>
<td>348</td>
<td>55</td>
<td>416</td>
</tr>
<tr>
<td>40</td>
<td>901</td>
<td>72</td>
<td>136</td>
</tr>
<tr>
<td>41</td>
<td>413</td>
<td>73</td>
<td>390</td>
</tr>
<tr>
<td>42</td>
<td>144</td>
<td>83</td>
<td>1780</td>
</tr>
<tr>
<td>53</td>
<td>359</td>
<td>86</td>
<td>381</td>
</tr>
<tr>
<td>61</td>
<td>494</td>
<td>97</td>
<td>595</td>
</tr>
<tr>
<td>64</td>
<td>595</td>
<td>94</td>
<td>336</td>
</tr>
<tr>
<td>68</td>
<td>327</td>
<td>90</td>
<td>672</td>
</tr>
</tbody>
</table>

Identification of immunodominant proteins

To identify those GpSGHV virion proteins that give an immune response in rabbits, antiserum was raised against purified GpSGHV virions and tested against GpSGHV proteins in a Western blot analysis (Fig. 4). The antiserum reacted with several GpSGHV proteins, more specifically to a 130 kDa protein, to proteins in the 50 kDa range and to an array of proteins in the 30 kDa range. On the basis of the proteomic LC-MS/MS and the Western blot analyses (Table 1 and Fig. 4) ORF10 (approx. 127 kDa) and six ORFs with molecular sizes ranging between 43 and 50 kDa (ORFs 41, 47, 70, 96, 97 and 140) were tested as candidates for the observed immunodominant proteins in GpSGHV (Fig. 4). These ORFs were expressed in Escherichia coli BL21 cells. Due to its large size, ORF10 was split into two fragments (N- and C-terminal fragments with calculated molecular sizes of approx. 69 and 66 kDa, respectively). In addition, due to their assumed important role in oral infectivity and potential target for immune intervention, all four PIF genes were selected as candidates for protein expression in bacteria. ORF1 (p74), ORF41, ORF70, ORF97 and the N-terminal fragment of ORF10 could not be successfully expressed in E. coli. Of the six proteins that were successfully made and that reacted with the rabbit antiserum against GpSGHV virions (PIF-1, PIF-2, PIF-3, ORF47, ORF96 and the C-terminal fragment of ORF10), two ORFs, ORF10 (C-terminal) and ORF96, were selected to generate mono-specific polyclonal antibodies based on the fact that these two ORFs were found to be the major viral proteins (Table 1). Antibodies raised against the proteins encoded by ORF10 (C-terminal) and ORF96 were
tested against GpSGHV and homogenates of hypertrophied salivary glands of infected *G. pallipides* flies (Fig. 5a). As expected these antibodies reacted against the 130 and 50 kDa bands, respectively.

**Immunolocalization of SGHV proteins**

TEM using the antibody against the C-terminal fragment of ORF10 gave an indication that the protein is likely to be a component of the envelope of the virus [Fig. 5b(ii)]. Immunolocalization studies of ORF96 using the specific rabbit antibody against this protein did not give conclusive evidence whether this protein is part of the virus envelope as there were no gold particles observed on the virions under the conditions used in the TEM studies. This could mean that either the ORF96 protein could have hidden epitopes or the antiserum was not suited for immuno-EM. Furthermore, it could be seen in the TEM that only a few GpSGHV virion rods remained entirely intact after Nycodenz preparation. This confirms the fragile nature of GpSGHV and further work needs to be done to study its stability under different conditions such as temperature, and especially the effects of virus handling on its infectivity to the tsetse fly. It is to be noted that a high density of gold particles was observed when antiserum against ORF10 was used, most likely on debris of the viral envelope. In the control experiments (preimmune serum), no gold particles could be seen for either the GpSGHV virion or the nucleocapsid. These studies may be direct evidence that the ORF10 could be involved functionally in the formation and/or the assembly of the GpSGHV envelope.

**Conclusions**

Current proteomic analysis of GpSGHV allowed us to determine a total of 61 proteins. The identities of the proteins within the virion proteome revealed many candidates which provide a basis for further studies focusing on the virulence and pathogenesis of GpSGHV as well as on mechanisms of virus infection and transmission in tsetse flies. Comparison of the 61 identified GpSGHV ORFs with dsDNA viruses of other virus families showed only a few homologies (5), more specifically with baculoviruses and nudiviruses. These involve the PIF proteins that are essential for oral infectivity of baculoviruses in insects. Whether they are also functional PIF proteins remains to be investigated. The proteomic data also clearly indicated that GpSGHV has a total of 28 structural ORFs that encode proteins not encoded in MdSGHV. Twenty-nine GpSGHV ORFs have homologues in MdSGHV, and so far 13 of these have also been identified in the MdSGHV proteome. In addition, MdSGHV has a number of virion proteins that do not have homologues in GpSGHV. Therefore, this analysis further
supports the placement of these two members of the newly proposed family *Hybraviridae* into two separate genera (proposed names *Glossinavirus* and *Muscaivirus*) (Abd-Alla et al., 2009). The role of these virion proteins in virion structure and infectivity will be the subject of future investigations.

Proteomic and immunolocalization data indicated that the ORF10 protein is abundant and present on the virion envelope. In the light of these findings, the ORF10 protein would be a good target for studies to mitigate infections of tsetse colonies by the SGHV. The approach would be to supplement the blood meal with the ORF10 antibody or by immunizing cattle who are the blood donors in tsetse fly rearing systems.

**METHODS**

**Preparation of virus particles and analysis by LC-MS/MS.**

GpSGHV virions were purified from hypopharyngeal salivary glands collected from a *G. pallidipes* colony maintained at the Entomology Unit of the FAO/IAEA Agriculture & Biotechnology Laboratory, Seibersdorf, Austria. Hypopharyngeal salivary glands were collected from dissected flies, homogenized in Tris buffer (50 mM, pH 7.8) and clarified twice by centrifugation for 10 min at 3000 g. The supernatant was layered onto a 10–60% linear Nycodenz gradient and centrifuged for 1 h at 27 000 g. The viral band was taken and washed in Tris buffer and centrifuged for 1 h at 150 000 g. The viral pellet was resuspended in Tris buffer.

The purified virus particles were solubilized in 2× concentrated Laemmli buffer, and fractionated by SDS-PAGE (12%). Fermentas PageRuler Prestained Marker proteins were used. The gel was stained with colloidal blue and the gel lane containing the virion proteins was excised into seven contiguous sections spanning the complete gel lane based on a comparison with molecular markers.

In-gel proteinigestions and peptide extractions were performed at 25 °C according to a method described previously (Ince et al., 2010). The peptides resulting from this digestion were analysed by LC-MS/MS by injecting 18 μl sample onto a 0.10×32 mm Prontosil 300-5-C18H pre-concentration column (Bishop) at a flow rate of 6μl min⁻¹ for 5 min. Peptides were eluted from the pre-concentration column onto a 0.10×200 mm Prontosil analytical column 300-3-C18H (Bishop) with an acetonitrile (ACN; HPLC grade) gradient at a flow rate of 0.5 μl min⁻¹ for 50 min. The gradient consisted of 10–35% (v/v) ACN increased in water with 1 mM formic acid (H₂O) in 50 min. As a subsequent cleaning step, the ACN concentration was increased to 80% (v/v) in 3 min (with 20% water and 1 mM formic acid (H₂O) in both the ACN and water). Between the pre-concentration and analytical column, an electroospray potential of 3.5 kV was applied directly to the eluent via a solid 0.5 mm platinum electrode fitted into a P875 Upchurch microT. Full-scan-positive mode Fourier transform mass spectra (FTMS) were measured between mass-to-charge ratios of 380 and 1400 with a LTQ-Orbitrap spectrometer (Thermo electron).

Tandem mass spectrometry (MS/MS) scans of the four most abundant doubly and triply charged peaks in the FTMS scan were recorded in a data-dependent mode in the linear trap (MS/MS threshold=10,000). All MS/MS spectra obtained with each run were analysed with BioWorks 3.1.1 software (Thermo Fisher Scientific). A maximum of a single differential modification allowed per peptide was set for oxidation of methionines and de-amidation of asparagine and glutamine residues. Carbamidomethylation of cysteines was set as a fixed modification. Trypsin specificity was set to fully enzymic and a maximum of three missed cleavages with monoisotopic precursor and fragment ions. The mass tolerance for peptide precursor ions was set to 10 p.p.m. and for MS/MS fragment ions to 0.5 Da.

A GpSGHV protein database was used for the analysis (EF568108; created February 25, 2008; downloaded from www.ncbi.nlm.nih.gov/sites/entrez) after adding a list of commonly observed contaminants like: BSA (P02769, BSA precursor), trypsin (P00760, bovine), trypsin (P00761, porcine), keratin K22E (P35908, human), keratin K1C9 (P35527, human), keratin K2C1 (P04264, human) and keratin K1C (P35527, human). A decoy database was created by adding the reversed sequences using the program SequenceReverser from the MaxQuant package (Cox & Mann, 2008).

To identify the proteins in the GpSGHV virions, the spectra obtained from the LC-MS/MS were searched against the GpSGHV ORF database using BioWorks 3.3.1. The peptide identifications obtained were filtered in BioWorks with the following filter criteria: Da≥0.08, Xcorr≥1.5 for charge state 2+, Xcorr≥3.3 for charge state 3+ and Xcorr≥3.5 for charge state 4+ (Peng et al., 2003). Only those proteins that showed a BioWorks score factor larger then 0.9 were considered.

**Detection of immunodominant protein candidates.** Following separation by 12% SDS-PAGE, proteins of purified GpSGHV particles were transferred to Immobilon-P membranes. Membranes were blocked overnight by incubation with 1% low-fat milk and 0.05% Tween 20 in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄ and 1.76 mM KH₂PO₄, pH 7.4) at room temperature. Membranes were washed once with 0.2% low-fat milk in PBS-Tween 20 for 5 min, incubated with rabbit primary antibody (anti-GpSGHV; diluted 1:5000, see below) at room temperature for 30 min, washed three times with PBS-Tween 20, and further incubated for 30 min in goat anti-rabbit IgG-alkaline phosphatase conjugate (Promega) diluted 1:3000. Blots were washed three times with alkaline phosphatase buffer (0.1 M Tris/HCl, 5 mM MgCl₂, pH 9.5) and stained with 1% of NBT/BCIP in alkaline phosphatase buffer.

**Selection of ORF, PCR amplification and gene cloning.** Candidate ORFs for protein expression were selected based on the molecular masses of LC-MS/MS identified proteins and the immunoblot analyses. Viral DNA was extracted from purified virus as reported previously (Abd-Alla et al., 2007), and approximately 5 ng of the DNA was used as template. PCR amplifications were performed with HF Phusion *Taq* DNA polymerase (Finnzymes), using the reaction mixture recommended by the supplier. The primers were designed to amplify the hydrophilic regions of the selected ORFs (Supplementary Table S1, available in JGV Online) and were used at a final concentration of 0.2 mM. The PCR conditions were 98 °C for 30 s; 98 °C for 10 s, 59 °C for 20 s and 72 °C for 30 s/kbp for 25–30 cycles; and finally 72 °C for 5 min. The PCR products were individually inserted into pET11a/blunt cloning vector (CloneJET PCR Cloning kit; Fermentas). The resulting recombinant plasmids were purified with homemade GF/F columns as described in Borodina et al. (2003) and the inserts were sequenced to confirm the sequences. The inserted DNA fragments were recloned into pET28a (+) (Sambrook et al., 1989) at the multiple cloning site.

**GpSGHV proteins: expression, purification and production of antisera.** *E. coli* BL21(DE3) cells were transformed with the pET28-derived plasmids to express the cloned genes according to the pET system manual (Novagen). The bacteria expressing the viral genes were sonicated in Laemmli sample buffer (Laemmli, 1970) and purified using preparative SDS-PAGE (model 491 Prep Cell, Bio-Rad Laboratories) according to manufacturer’s instructions. Ten micro-litres of each recombinant protein was analysed on 12% SDS-PAGE.
followed by silver staining according to standard protocols. The purity and quantity were verified with Coomassie blue staining and with Western blot analysis, using specific immune serum directed against His-tag. To reduce the amount of SDS in the samples, each protein fraction was concentrated with Centriprep YM-10 centrifugal membranes (Amicon bioseparations).

Antisera were prepared against the purified proteins (proteins encoded by ORF96 and the C-terminal fragment of ORF10) by injecting rabbits with 0.4–0.8 mg of the recombinant protein emulsified in Freund’s incomplete adjuvant. Two booster injections were given at 2-week intervals. Another two antisera were prepared against the P74 protein (ORF1) using the synthetic oligopeptides LYEHSKDEDGVYHRA-C (aa 114–128) and C-SEENKIASIDDKEQF (aa 612–626) (Pacific Immunology Corp). Polyclonal antibody against the whole GpSGHV particle was collected from a rabbit that was used for several months to feed tsetse flies in CIRAD, France.

**EM and immunolocalization of authentic viral proteins.** Aliquots (5 µl) of GpSGHV virion suspension were adsorbed onto carbon-coated and ionized-nickel grids (400 mesh) for 5 min at room temperature and treated for negative staining with phosphotungstic acid or for immunogold labelling. For the latter, the grids were then blocked with blocking buffer (5 % BSA, 5 % normal serum, 0.1 % cold water skin gelatin, 10 mM phosphate buffer, 150 mM NaCl, pH 7.4) for 30 min, and incubated with primary antibody or pre-immune rabbit serum (1 : 20 dilution in incubation buffer) for 1.5 h at room temperature and treated for negative staining with phosphotungstic acid or for immunogold labelling. For the latter, the grids were then coated and ionized-nickel grids (400 mesh) for 5 min at room temperature and treated for negative staining with phosphotungstic acid or for immunogold labelling. For the latter, the grids were then blocked with blocking buffer (5 % BSA, 5 % normal serum, 0.1 % cold water skin gelatin, 10 mM phosphate buffer, 150 mM NaCl, pH 7.4) for 30 min, and incubated with primary antibody or pre-immune rabbit serum (1 : 20 dilution in incubation buffer) for 1.5 h at room temperature. After incubation and several washes, the grids were incubated with goat anti-rabbit secondary antibody conjugated with gold particles (10 nm diameter; 1 : 20 dilution in incubation buffer) for 45 min at room temperature. The grids were washed extensively with incubation buffer to remove excess salt, and negatively stained with 2 % sodium phosphotungstate (pH 6.5) for 5–10 s. The specimens were examined with a transmission electron microscope (100 kV EM, JEM-1011; JEOL).

**ACKNOWLEDGEMENTS**

The authors would like to thank Wageningen University and Research Centre, The Netherlands for awarding a Master of Science grant to Mr Henry M. Karithi to carry out these studies. We are indebted to the Entomology Unit, FAO/IAEA Agriculture and Biotechnology Laboratories, Seibersdorf Austria for providing the virus used in this research. All proteomic LC-MS/MS measurements were done at Biqualys Advanced Analysis Company, the Netherlands (www.biqualys.nl). We would like to thank Marc Vreyesen for reviewing the manuscript.

**REFERENCES**


