Molecular Cloning and Expression of SmIrV1, a Schistosoma mansoni Antigen with Similarity to Calnexin, Calreticulin, and OvRall

(Received for publication, November 30, 1992)

Thomas R. Hawn†, Timothy D. Tom, and Mette Strand

From the Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Schistosomiasis is a tropical parasitic disease estimated to affect 200–300 million people worldwide with over 200,000 deaths annually. Although safe and efficacious drugs exist to kill schistosomes, the expense and delivery of such chemotherapy is compromised in many countries because of economic reasons and problems of reinfestation. Evidence for age-dependent natural immunity in humans (1) and induced protection of up to 91% by vaccination in mice (2, 3) indicates that successful control of this organism may be most effectively achieved through a vaccine.

The life cycle of Schistosoma mansoni, one of three major species of schistosomes, includes infection by cercaria which penetrate the host’s skin, transform into schistosomula, enter the venous system, migrate through the lungs, and reside in the mesenteric veins as adults. The target of protective immunity is probably the lung stage schistosomula which is most susceptible to elimination in normally infected as well as chronically infected mice.

*This work was supported in part by Grant AI-19217 from the National Institutes of Health and by the John D. and Catherine T. MacArthur Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

†Supported by the Medical Scientist Training Program Grant 5T32GM07039–18.

vaccinated mice (4). Protection induced in mice by vaccination with irradiated cercaria can be partially transferred to naïve mice with sera or IgG resulting in up to 50% of the original protection (5). On the basis of the importance of the humoral immune response, we have previously identified antigens that are uniquely or more strongly recognized by sera of immunized mice compared with chronically infected unvaccinated mice (6). Antibodies were raised against these candidate vaccine antigens and used to screen an adult worm cDNA expression library. Out of 20,000 plaques screened, five clones were identified representing two genes, SmIrV1 and SmIrV5 (IrV = Irradiated Vaccine). The gene product of cDNA clone SmIrV5 has similarity to myosin and induces protection levels of 42–83% in mice (7). By comparison, other antigens, including glutathione S-transferase (8) and parasomyosin (9, 10) have shown levels of protection in the range of 39–43% and 24–53%, respectively. Since complete protection was not obtained with SmIrV5 alone and in an effort to more fully characterize potential targets of the protective immune response, we selected SmIrV1 as the next candidate for immunoprophylactic analysis.

We report here the molecular cloning and sequencing of the SmIrV1 gene which has a deduced amino acid sequence of 582 residues and is similar to three proteins: 1) calreticulin, an intracellular protein hypothesized to be involved with calcium sequestration (11, 12) as well as having been identified as an autoantigen (13); 2) calnexin, a phosphoprotein hypothesized to be involved with the endoplasmic reticulum (ER) (14) and a molecular chaperone involved in folding and assembly of proteins as they transit the ER (15–17); and 3) OvRall, a surface protein found on a larval stage of Onchocerca volvulus, a filarial nematode that is the causative agent of river blindness in the tropics (18). In addition to sequence analysis, we have expressed a fragment of SmIrV1 and demonstrated that it reacts with sera from vaccinated as well as chronically infected mice.

EXPERIMENTAL PROCEDURES

Parasites—Worm pairs of S. mansoni (Puerto Rican strain) were obtained from female mice (strain CD-1, Charles River Breeding Laboratories) infected 7 weeks earlier with 250 cercaria by tail immersion (19). The worm pairs were obtained by hepatic portal perfusion and then immediately frozen in liquid nitrogen until use.

Isolation and Identification of SmIrV1 cDNA Clones—Two different adult worm libraries were screened. 1) an oligo(dT)-primed Agt11 library was screened as described previously (7) with a 1:75 dilution of an antisemir (anti-IrV) raised in a rabbit against a subset of S. mansoni glycoproteins that show unique or enhanced immunogenicity in mice vaccinated twice with radiation-attenuated cercaria (6). 2) An oligo(dT)-primed library constructed in UNI-ZAP XR (Stratagene) according to the manufacturer's instructions was screened with

1 The abbreviations used are: IrV, irradiated vaccine; ER, endoplasmic reticulum; GST, glutathione S-transferase; PBS, phosphate-buffered saline; PAGE, polyacrylamide gel electrophoresis.
 probes as described by Sambrook (20). Plaques were transferred onto nitrocellulose filters (Millipore), immobilized by UV cross-linking (Stratallinker, Stratagene), prehybridized for 1-2 h at 55 °C in a fresh hybridization solution with 1 × 10^6 cpm/ml of probe, washed two times for 10 min at 25 °C with 2 × SSC (standard sodium citrate) and 0.1% SDS, washed two times for 20 min at 65 °C with 0.1 × SSC and 0.1% SDS, and exposed to x-ray film overnight at −70 °C. Clones were plaque-purified and subcloned into phagemids (pBluescript II SK−, Stratagene) before sequencing.

**Northern Blot Hybridization**—Adult worm mRNA was isolated, electrophoresed, and transferred to a nylon membrane (Zeta-Probe, Bio-Rad) as described previously (21). The blot was processed as for the plaque screening described above except the prehybridization and hybridization was at 42 °C and the second wash step was at 50 °C.

**DNA Sequencing**—DNA sequencing was performed by the dyeode chain termination method of Sanger et al. (22) using [α-32P]dATP (Amersham Corp., 800 Ci/mmol) and a Sequenase kit (United States Biochemical Corp.). Both strands of the total coding region were sequenced.

**Computer Analyses**—The GenBank™ and the Protein Identification Resource Data Banks were searched using the software developed by Genetic Computer Group, University of Wisconsin, Madison, WI.

**Expression of Glutathione S-Transferase (GST) Fusion Proteins**—A blunt-end ligation was used to subclone SmIrVlc into pGEX-2T (Amrad Corp.) (Fig. 1). This 1.37-kilobase fragment with 976 coding base pairs was generated from SmIrV1-17 by cleavage with Bsu361 at an internal site and EcoRI.

**Isolation and Identification of cDNA Clones Encoding SmIrV1**—From an 11.1 cDNA expression library constructed from adult worm mRNA, a 1.4-kilobase clone called SmIrV1-1 was identified which encoded a protein product that reacted with a serum directed against schistosome proteins that exhibit unique or enhanced immunogenicity in vaccinated mice (6, 7). In order to obtain a full-length clone, a probe encompassing a 492-base pair HindIII/Xbal restriction fragment called SmIrV1ala was used to screen a UNI-ZAP XR adult worm cDNA library (Fig. 1). Out of 150,000 plaques screened, seven positive clones were obtained. One of these, designated SmIrV1-17, contained the largest insert which was 1.8 kilobases. An additional screen was performed with SmIrVib, a 409-base pair fragment from the 5′ end of SmIrV1-17 generated by cleavage with Bsu361 at an internal site and EcoRI from the pBluescript multicloning site. Out of 250,000 clones screened, 11 positive clones were obtained, four of which had insert sizes of 2.2 kilobases. In order to insure that a full-length clone was obtained, a final screen was performed with SmIrVld, a 447-base pair fragment from the 5′ end of one of these clones generated by cleavage with ClaI at an internal site and EcoRI from the pBluescript multicloning site. Out of 250,000 plaques screened, three positive clones were obtained, all of which had inserts of 2.2 kilobases. One of these, clone 65 (SmIrV1-65), was used for further analysis.

**Northern Blot Hybridization**—We examined the expression of SmIrV1 mRNA in the adult worm stage using SmIrVib as a probe and detected a single mRNA species of 2.5-2.6 kilobases (Fig. 2).

**Deduced Amino Acid Sequence—SmIrV1-65** contained an open reading frame that spanned 1796 nucleotides with the first initiation codon (ATG) at base pair 51 (Fig. 3). The nucleotides around this start codon (ACCAGT) fulfill Kozak's criteria for a strong ribosomal initiation site (ACCATGG) (25). The termination codon (TAA) was followed by 391 base pairs of untranslated codons, and a polyadenylation signal (AAATAA) was positioned 15 base pairs from the poly(A) tail. A putative signal sequence was characterized by a basic residue (amino acid 5) followed by a stretch of hydrophobic residues (7-17), then several polar residues (19-22), and finally small residues at positions −3 and −1 (23 and 25 or 24 and 26) in relation to the cleavage site. The deduced amino acid sequence of SmIrV1-65 contained the entire coding region: a strong ribosomal binding initiation site, a putative signal sequence, and seven clones with a size of 2.2 kilobases.

The deduced primary amino acid sequence has a high degree of hydrophilicity, an isoelectric point (pl) of 5.0, a predicted molecular mass of 66 kDa, and can be divided into three distinct regions: N term (1-246), the middle proline rich “P” region (247-392), and the C term (393-582) (Fig. 1). The P region is proline (18.5%, normal (n) = 5.2) and tryptophan (5.5, n = 1.4) rich and highly charged (D, E; K; 11.6, 8.9, and 11.0% respectively, n = 5.3, 6.2, 5.9) with an isoelectric point of 4.5. In addition, it contains two sequences which are repeated four times with variation: FXXIDPDAXKPE DWDE (repeat sequence one) and GWXXXPXIXNPXX X (repeat sequence two). Chou-Fasman predictions indicate that this region is rich in β-turn secondary structure. Furthermore, two PEST sequences are present spanning residues 278-297 and 297-319 with scores of 9.8 and 5.6. PEST sequences are defined by a score greater than zero and range from −50 to +55. They have been found in a number of proteins with short half-lives and are speculated to be part of a signal for proteolytic degradation (26). The C-terminal region of SmIrV1 also contains a PEST sequence spanning residues 550-566 with a score of 24.8. This region is acidic (D and E; 7.3 and 12.0%, respectively) with an isoelectric point of 4.7, a potential N-linked glycosylation site at residue 544, and a hydrophobic stretch of residues from 465 to 482. Although the hydrophathy values suggest that these residues may represent a transmembrane spanning region, this possibility seems unlikely due to the β-strand structure predicted by the Chou-Fasman algorithm (27). Furthermore, this region contains two α-helix-breaking proline residues and does not contain the requisite 22 amino acids for a transmembrane spanning helix.

**Comparison to Sequences in the Protein Identification Resource Data Base**—Comparison of the SmIrV1 sequence with a Protein Data Bank revealed 41-53% identity with three proteins (Table IA): 1) calreticulin (calregulin, HACBP, CRP55, 55 kDa by SDS-PAGE, 396 amino acids) (11, 12), 2) calnexin (cdpp90, p88, IP90, 30 kDa by SDS-PAGE, 573 amino acids) (14-16), and 3) OvRall (42 kDa by SDS-PAGE) (18). The middle P region contains the greatest similarity among all of the proteins with 59-71% identity and includes the two sets of sequences repeated three times in calreticulin and OvRall and four times in calnexin and SmIrV1 (Table I, B and C, and Fig. 4). A consensus sequence was generated for each repeat sequence and used to search the Protein Identifi-
**S. mansoni Antigen**

\[ A \]

\[
\begin{array}{cccccccc}
-1.1 & -0.5 & 0 & 0.5 & 1 & 1.5 & 2 & 2.5 \\
361 & 361 & 361 & 361 & 361 & 361 & 361 & 361 \\
\end{array}
\]

\[ \text{clone 65} \]

\[ \text{clone 17} \]

\[ \text{clone 1} \]

**FIG. 1.** Schematic representation of SmIrV1 cDNA and deduced amino acid sequence. A, restriction enzyme map with sites of cleavage indicated. Relevant fragments for cloning and expression are depicted. The coding region is represented by an open box, whereas the segments sequenced are indicated by arrows. B, schematic representation of SmIrV1 protein:

- \( n \) = N-terminal domain from residues 1 to 246;
- \( P \) = P domain from residues 247-392; C = C-terminal domain from residues 393-582;
- ss = putative signal sequence;
- \( rS1 = \text{repeat sequence PXXIXDP(D,E)(A,D)XKP(E,D)DWD(E,D)} \);
- \( rS2 = \text{repeat sequence GXWXXPXIXNPXYX} \).

**FIG. 2.** Northern blot hybridization with 3 µg of adult worm mRNA probed with \( \alpha^{32P} \)-labeled SmIrV1b as described under "Experimental Procedures." Molecular size standards in kilobases are indicated.
**S. mansoni Antigen**

**FIG. 3.** cDNA and deduced amino acid sequence of SmIrV1. The putative signal sequence is in **boldface lettering**, repeat sequence one is **underlined**, repeat sequence two is **underlined** and **boldface**, the hydrophobic stretch of residues is **boxed**, and the putative polyadenylation signal is **double underlined**. The GenBank accession number is L08641.
S. mansoni Antigen

TABLE I
Comparison of SmIrV1 by region with calreticulin, OvRall, and calnexin

A. sequence identity comparison of domains N, P, and C. B and C, comparison of repeat sequences one and two. The repetitive regions from all four proteins were aligned and suggested consensus sequences were derived. The number of residues matching the consensus sequence is recorded out of a total of 14.

<table>
<thead>
<tr>
<th>Domain</th>
<th>N</th>
<th>P</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Sequences compared</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SmIrV1:calreticulin</td>
<td>38</td>
<td>60</td>
<td>35</td>
<td>41</td>
</tr>
<tr>
<td>SmIrV1:OvRall</td>
<td>40</td>
<td>59</td>
<td>26</td>
<td>41</td>
</tr>
<tr>
<td>SmIrV1:calnexin</td>
<td>52</td>
<td>71</td>
<td>37</td>
<td>53</td>
</tr>
</tbody>
</table>

B. Protein

<table>
<thead>
<tr>
<th>Consensus</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PXXIXD</td>
<td>PDKPEDEWDE</td>
</tr>
<tr>
<td>E</td>
<td>D</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
</tr>
</tbody>
</table>

SmIrV1
251–267
PKIEIDDPEDKKPSDWDDE
268–284
REKIVDTNAAKPDWDDE
287–303
PATTEDSEAVKPSGWLD
306–322
PEMTDPAAVPKKDWDDR
Calnexin
254–270
SREIPEDEQKPEDWDDE
271–287
RPKIPDPDAVRPDWDWNE
290–306
PAKIPDEEATKPDGWLD
309–325
PEYVPDPDAEKPEDWDDE
Calreticulin
187–203
PKKKIKDPADSKPEDWDDE
205–221
PEHIPDPDAEKPEDWDDE
222–238
PEHIPDPDAEKPEDWDDE
OvRall
101–117
PKKKIKDPADSKPEDWDDE
118–134
PEHIPDPDAEKPEDWDDE
135–151
PEHIPDPDAEKPEDWDDE
150–151
PEHIPDPDAEKPEDWDDE

C. Protein

<table>
<thead>
<tr>
<th>Consensus</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>GXWXXP</td>
<td>PXXKNPXYX</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

SmIrV1
326–339
GEWVAPQINNPKCA
345–368
GKWKRIPIPNPKYK
359–372
GKWSPAIPNPNNYNK
373–386
GIWTPRKiPnPYPF
Calnexin
329–342
GEWEAPQIANPKCE
348–361
GVWQRPMIDNPNYK
362–375
GKWKPMPIDNPNYQ
376–389
GKWKPMPIDNPDDF
Calreticulin
242–255
GEWEPVIPQNPYK
256–269
GEWKPRQIDNPDYK
270–283
GTWHPETIDNPEYS
OvRall
205–218
GEWEPMPVMDNPEYK
219–232
GEWKPKNNKNPAYK
233–246
GKWHPETIEIPDYT

DISCUSSION

We report here the molecular cloning, sequencing, and expression of the SmIrV1 protein, an antigen selected for study based on its enhanced immunogenicity in mice with protective immunity against schistosomiasis. Comparison of the deduced amino acid sequence of SmIrV1 to sequences in a Protein Data Bank revealed significant similarity to calreticulin, calnexin, and OvRall. Calreticulin has been implicated as having a calcium-sequestering function for intracellular signaling requirements such as the inositol triphosphate-mediated response (11, 30). Calnexin also has been shown to bind calcium and has been hypothesized to participate in the regulation of calcium's effect on protein translocation (31) as part of a complex with the signal sequence receptor (SSRα...
FIG. 4. Multiple sequence alignment of calreticulin, OvRall, calnexin, and SmIrV1. Regions of identity among all four sequences are boxed.

Calcium binding studies with calreticulin have demonstrated one high affinity site located in the proline-rich region and numerous low affinity sites in the acidic C terminus (32). The location of the high affinity binding site suggests a possible role in calcium binding for the repeat sequences PXXIXDPDAXKPEDWDE and GXWXXF'XIXNPXYX which are found with minor variations four times in SmIrV1 and calnexin and threes times in calreticulin and OvRall. Almost the entire P domain is composed of these two sequences, which is the most highly conserved domain within this family of proteins. Since the P domain does not have similarity to known calcium-binding proteins such as those containing an EF-hand motif (28) or annexins (29), it may represent a new motif for calcium binding. Besides the high affinity site, calreticulin contains 20 low affinity sites in its acidic tail where calcium may be chelated by the long stretches of acidic residues (12). Calnexin and SmIrV1 could potentially have similar low affinity sites but would probably bind less calcium due to their higher isoelectric points. In contrast, this putative calcium-sequestering region has been abrogated in OvRall which contains a long stretch of basic residues.

Besides its role in calcium binding and protein translocation, calnexin has been identified as a possible chaperone molecule which aids in the assembly of various proteins in the ER including the T-cell receptor, membrane-bound immunoglobulin (15), and class I histocompatibility molecules (16). Immunofluorescent studies on calnexin demonstrated an ER and nuclear envelope localization with no evidence of surface staining. Wada et al. (14) have predicted a topology of calnexin as a type I transmembrane protein with the N-terminal portion within the lumen of the ER. However, this hydrophobic region in calnexin, similar to the stretch in SmIrV1, is predicted to have a β-strand secondary structure which is inconsistent with the requisite 22-amino acid α-helix structure for a transmembrane-spanning region of a type I integral membrane protein (27).

Similar to calnexin, calreticulin has also only been found in a number of intracellular locations, including the cytoplasm (33), nuclear envelope, nucleoli-like structures, ER, sarcoplasmic reticulum (34), and inositol trisphosphate-sensitive calciosomes (35). Besides calcium sequestration, the function of calreticulin remains unknown. Burns et al. (36) have shown that its mRNA and protein expression levels are increased in activated T-lymphocytes, whereas Rojiani et al. (33) have shown its association with a highly conserved sequence found in the α subunit of integrins. In addition, it has been implicated as an autoantigen as it is recognized by antibodies of and SSRβ (14).
patients with systemic lupus erythematosus and Sjogrens disease (37). The intracellular localizations of calreticulin and calnexin and their proposed functions do not provide a model for SmIrV1’s potential role as a vaccine candidate. However, the surface localization of OvRall provides a precedent for a nonintracellular location for this family of proteins. OvRall was identified by injection of larvae of O. volvulus into rabbits, a nonpermissive host. These sera were used to screen a cDNA library, to identify OvRall, and to show that it immunolocalizes to the surface of larvae (18). Furthermore, sera of patients with onchocerciasis reacted with calreticulin and raised the possibility that autoimmune reactions may contribute to the pathogenesis of this disease (37) (38). However, recent preliminary studies suggest that there is not a correlation between antibody titers to OvRall and development of disease. If a similar larval surface localization exists for SmIrV1, it would correlate with the hypothesis that the surface of schistosomules are targets of elimination in immune mice (39) (40).

Epitopes which may be the targets of a protective immune response are present in SmIrV1c which is recognized by sera from vaccinated as well as chronically infected mice. The binding of chronic sera to Schistosoma japonicum GST (Sj26) may be due to cross-reactivity. This protein has been identified previously by its enhanced immunogenicity with sera from mice resistant to S. japonicum (41). Reactivity of Sj26 with chronic sera of mice infected with S. mansoni has not been reported previously and may be a result of the increased sensitivity of the enhanced chemiluminescence detection technique. However, cross-reactivity between schistosome species has been found for the larger 28-kDa isoenzyme of GST (42).

Further experiments will include production of anti-SmIrV1 antibodies for localization studies and analysis of stage-specific expression of the native antigen. In addition, the immunoprophylactic potential of SmIrV1 will be investigated by vaccination trials in mice.

Acknowledgments—We thank Drs. John P. Dalton and Scott Roberts for initial studies on SmIrV1, Alvin M. George for expert technical assistance, and Dr. Martin Rechsteiner for help in the PEST sequence analysis.

REFERENCES