This project is examining the mechanisms of protein cross linking induced by phenol oxidase mediated processes in the schistosome eggshell. Synthetic peptides are being used as models for this process. Physical and chemical techniques are being used to examine their properties. Keywords: molecule-molecule interactions; crosslinking (chemistry); (RT).
It is clear that the eggshells of schistosomes are very sophisticated structures and all of their components are highly evolved. From the results obtained to date we stand to learn some very interesting lessons in protein structure and function from this project, many of them unexpected.

The eggshells of schistosomes are made up almost entirely of protein with possibly a small amount of carbohydrate (1,2). These proteins are extensively cross-linked to form a structure which is very well adapted to its functional roles, namely the protection of the developing embryo during its passage through the hosts tissues and its ability to readily split during hatching. The eggshell is very protease resistant and is almost completely insoluble in commonly employed agents such as SDS and urea.

Recent molecular cloning and DNA sequencing of the genes which encode the eggshell proteins reveals that they are glycine, tyrosine and lysine rich polypeptides with unusual sequence properties. One protein in which we are particularly interested is made up of a large number of repeats of a pentapeptide with the sequence Gly-Tyr-Asp-Lys-Tyr. Examination of the DNA sequence shows that the amino acid sequence of the repeats is very strongly conserved whilst the DNA sequence of the repeats is not. Thus we conclude that the amino acid sequence is under strong selective pressure that prevents substitutions etc. altering the protein repeat sequence.

This led us to hypothesize that the repeat must confer some special property on the protein and we favored the idea that a particular secondary structure was adopted by the repeat. Our main approach to this problem has been to synthesize a series of synthetic peptides containing various numbers of repeats. These are shown in Figure 1. The repetitive nature of this protein means that we can hope to discover the structure of the entire polypeptide by understanding the structures of small parts made of several repeats.

<table>
<thead>
<tr>
<th>Name</th>
<th>Amino acid sequence.</th>
<th>Length (no. of res.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F4-1</td>
<td>GYDKY</td>
<td>5</td>
</tr>
<tr>
<td>F4-2</td>
<td>GYDKGYDKY</td>
<td>10</td>
</tr>
<tr>
<td>F4-3</td>
<td>GYDKGYDKGYDKY</td>
<td>15</td>
</tr>
<tr>
<td>F4-3Cys</td>
<td>GYDKGYDKGYDKGYDKY</td>
<td>16</td>
</tr>
<tr>
<td>F4-6</td>
<td>GYDKGYDKGYDKGYDKGYDKY</td>
<td>30</td>
</tr>
</tbody>
</table>

Figure 1 Synthetic peptides synthesized. The names of the peptides are based on the name of the cDNA clone (F4) whose sequence they are derived from. Peptides F4-1, F4-2, F4-3, and F4-3Cys were all synthesized using t-BOC protected amino acids. F4-6 was synthesized using F-MOC protected amino acids. All were synthesized using an Applied Biosystems Peptide synthesizer.

Our main intention was to investigate the potential secondary structure of the repeats using various physical techniques. However we have found the peptides to have unique and quite unexpected chemical properties that we are just beginning to characterize in more detail.

During the process of purifying and characterizing the F4-6 peptide we observed several
spontaneous effects. Whilst dissolving the peptide (30mg) in 1ml of 0.3% acetic acid before running it on a 90cm Sephadex G50 gel permeation column we noticed that the white peptide gave a brown solution as it dissolved. After running it over the G50 column various fractions were analyzed by reverse phase chromatography on a 25cm Lichrosphere C-18 column using an acetonitrile gradient in 0.1% TFA. As figure 2 shows the peptide has been altered to give a series of distinct derivatives which elute earlier from the column and are presumably less hydrophobic/more hydrophilic in nature (3).

The brown coloration which appeared in the solution whilst dissolving the peptide suggested that we may be seeing oxidation of the peptide at tyrosine residues (in an analogous way to the formation of "dopachrome" from the oxidation of DOPA).

When the solutions were rigorously degassed and the procedure repeated under nitrogen no derivatives were observed confirming our guess that oxygen was apparently oxidizing the peptide. We have tried to find some kind of precedent for this degree of spontaneous oxidation of a tyrosine containing synthetic peptide but have been unable to find similar examples. Counter examples can be given such as the enkephalin peptides which contain tyrosine residues which are essentially stable for years. N-terminal tyrosines have been reported to be somewhat less stable (5% oxidation in a year) but when present internally (as all the tyrosines in the F4 peptides are) they are stable (R.V. Lewis pers. comm. We are thus forced to consider the possibility that this peptide has some unusual properties that render it easily oxidized. This observation and suggestion needs following up in much greater detail before we can draw firm conclusions.

Perhaps this observation should not be too surprising since we believe the purpose of this protein to be in the production of cross-links subsequent to the oxidation of tyrosine or DOPA residues by phenol-oxidase. Perhaps we are characterizing a protein that has evolved to be easily oxidized, what one could refer to as "substrate evolution".

These observations have opened up a new set of possibilities in our thinking about the biological "purpose" of the repeats. We think that these properties deserve following up in parallel with our structural studies. Some of the questions that we should be able to answer using the peptide approach are.

1) What is the minimum length of peptide that exhibits this property.
2) What is the chemical change that has occurred. For example has tyrosine been converted to DOPA or to quinone and are cross-links induced by these changes.
3) What are the pH optima etc. for this change.
4) By synthesizing additional peptides with variant sequences we can hope to elucidate which features of the sequence are necessary for the effect. The high degree of sequence conservation might suggest all of them are.

Structural studies.

Preliminary CD spectra of the F4-6 peptide suggest that there is no appreciable amount of either alpha helix or beta sheet under the conditions tried so far. This means that the structure of the peptide (should it adopt one) is not one of the two easily identified structural motifs which confirms our tentative conclusions drawn from an inspection of the peptide sequence. This in no way precludes the possibility of definite structure in the peptide and much more data will be required to make any definitive statements. Our molecular modeling studies of peptides are in an early stage and it is too early to draw any useful conclusions.

Eggshell structure.

FTIR studies on the eggshell using a microscope attachment and a polarized beam suggest some small degree of anisotropy of the protein molecules composing the eggshell although we have been unable to confirm earlier reports of longitudinal fibers in the eggshell observed by polarizing light microscopy. (4) However, numerical analysis of eggshell split orientation upon hatching show that the eggshell splits longitudinally the majority of the time (>70%) in natural hatching, whereas bursting of the eggshell by osmotic stress produces breaks with a quite different distribution of split orientation. This suggests that hatching is more than
just a swelling effect rupturing the eggshell.

The phenol oxidase gene.

We have made some progress towards identifying the schistosome phenol oxidase gene. We have constructed genomic DNA libraries in the EMBL-3 vector system and we are in the process of screening these libraries with a cDNA clone of the mouse tyrosinase gene (the gene from melanoma (this gene encodes the copper containing enzyme rather than the one expressed in adrenal tissue which does not.)

References.

Figure 2. Spontaneous oxidation of F4 peptides. The left panel shows the absorbance profile of the F4-6 peptide fractionated on sephadex G-50 in 0.3% acetic acid. The right panel shows the numbered fractions from the sephadex column analyzed by C-18 RPHPLC. The shaded peak is the major peak observed when the peptide is dissolved and fractionated under anaerobic conditions presumably representing the unmodified peptide. The fractions run identically on the C-18 column after reduction with sodium borohydride.