Invited review

Technological advances and genomics in metazoan parasites

D.P. Knox*

Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik, Midlothian, Scotland EH26 0PZ, UK

Received 25 August 2003; received in revised form 24 October 2003; accepted 31 October 2003

Abstract

Molecular biology has provided the means to identify parasite proteins, to define their function, patterns of expression and the means to produce them in quantity for subsequent functional analyses. Whole genome and expressed sequence tag programmes, and the parallel development of powerful bioinformatics tools, allow the execution of genome-wide between stage or species comparisons and meaningful gene-expression profiling. The latter can be undertaken with several new technologies such as DNA microarray and serial analysis of gene expression. Proteome analysis has come to the fore in recent years providing a crucial link between the gene and its protein product. RNA interference and ballistic gene transfer are exciting developments which can provide the means to precisely define the function of individual genes and, of importance in devising novel parasite control strategies, the effect that gene knockdown will have on parasite survival.

Keywords: Parasite; Helminth; Gene expression; Microarray; RNA interference; Proteomics

1. Introduction

Metazoan parasites (nematodes, cestodes and fluke) cause a wide variety of debilitating, and frequently fatal disease syndromes in man and animals. Chemotherapy with a variety of different drugs provides the foundation for current control but is being undermined by the development of drug resistance in the target parasites. An additional problem is that the precise mode of action of many of these chemotherapeutics remains undefined. The search for novel drugs continues but with diminished impetus because of the cost of development, licensing requirements and marketing along with concerns about drug resistance and, in the case of food animals, drug residues in meat and dairy products. Control by vaccination has been a long-held aim but the reality is that there are no licensed vaccines for any human parasitic infection. Slightly more progress has been made in countering parasites affecting domestic animals exemplified by the use of irradiated larval vaccines to control lungworm (Dictyocaulus spp.) in cattle and sheep and the recent exciting progress made in developing effective recombinant vaccines against the cattle tick Boophilus microplus (Willadsen, 1997) and the sheep cestode Taenia ovis (Rickard et al., 1995) but the genetic complexity of the metazoan parasite which allows it to evade the host immune response for sufficient time to complete the life cycle. It is also due, probably in equal part, to the variety of the host immune response which, again, is under complex genetic control. Early thoughts at the start of the molecular parasitology revolution were that vaccine production would simply require the identification of parasite proteins which interact with the immune system and the genes encoding them. These genes would then be expressed in Escherichia coli and the resultant recombinant protein would be the basis of a vaccine. This was almost borne out by the vaccination strategies developed with success against B. microplus (Willadsen, 1997) and T. ovis (Rickard et al., 1995) and, recently, Echinococcus granulosus (Lightowlers et al., 2002) but has been frustratingly elusive in most other metazoans. Therefore, there is now a realisation that logical vaccine production requires a clear insight into the host–parasite relationship at the molecular level. In the last decade, the tools have started to become available to make this approach a feasible reality. Technological advances in sequencing technologies and associated bioinformatics have led to the initiation of whole genome and expressed sequence tag (EST) sequencing projects and these advances have been rapidly followed by spectacular advances in methodologies which allow gene expression profiling to be
monitored in a cost-effective manner. Parallel developments in protein analysis (proteomics) give us the potential to identify the critical elements of the host–parasite relationship and, from this knowledge, devise new means for parasite control.

2. Parasite genome and expressed sequence tag sequencing initiatives

A large number of parasite sequencing initiatives are now underway building on the impetus gained by the completion of the Caenorhabditis elegans genome sequencing project (The C. elegans Sequencing Consortium, 1998). This programme as well as the bioinformatic tools developed for analysis, an ongoing process, provides the foundation stone for gene discovery in metazoan parasites. Sequence data can be derived using different approaches but can be placed into three basic categories (Tarleton and Kissinger, 2001). These are complete or nearly complete genome sequences, genome-survey sequence (GSS) tags generated by sequencing random clones or bacterial artificial chromosome (BAC) ends and, finally, ESTs which are generated from mRNAs expressed at the different stages of the parasite life cycle. The status of the various metazoan parasite-sequencing initiatives can be obtained by simple searches on the world-wide web. It is immediately evident that the numbers are large and increasing daily and this raw sequence data need to be stored, annotated and processed in a manner which can be accessed and interrogated by the wider research community. The rapid progress in acquiring parasite sequence data has been paralleled in advances in data management and interrogation with easy access made possible because of the development of the world-wide web. Most sequencing projects permit sequence-downloads and the data are often pre-processed with sequences clustered on the basis of similarity. Usually, the project website has an on-site BLAST capability, will provide translations of the predicted open reading frame (ORF) and some provide links to other web-based tools. Some permit text-based searching. The major websites are listed by Parkinson et al. (2003) and many useful links are found in http://www.nematode.net/, www.nematodes.org and http://www.tigr.org/.

EST sequencing projects are being undertaken for most of the important metazoan parasites of man and animals because they provide a potentially cost-effective and rapid means to elucidate parasite biology and to discover novel pathways which may be exploited for future control. The aim is to produce >300 000 ESTs from a range of human, animal and plant parasites with an estimated coverage of 30–50% of the genes of each organism (www.nematode.net, www.nematodes.org; Brooks and Isaac, 2002; Parkinson et al., 2003). Of course, this means that 50–70% of the genes will not be represented, a deficiency which can only be overcome by generating genomic sequence data (updates on genome projects can be viewed at e.g. www.ebi.ac.uk/parasites/parasite-genome.html), the latter being undertaken for Schistosoma mansoni (TIGR, Rockville, MD) Onchocerca volvulus (Steve Williams, Smith College, Northampton, MA) and Brugia malayi (Mark Blaxter, University of Edinburgh). Genome sequence projects have also commenced for Strongyloides stercoralis, Echinococcus spp., Echinostoma paranaense and Fasciola hepatica (Parkinson et al., 2003). It should be noted that relatively small EST datasets can be highly informative and can result in the discovery of many new genes of immediate relevance to the parasite and its lifecycle (Hoekstra et al., 2000; Maizels et al., 2000; Grant and Viney, 2001; Kenyon et al., 2003).

Gene expression can be examined in silico in species where EST projects are being conducted on several life cycle stages such as S. mansoni, Haemonchus contortus, B. malayi and O. volvulus. EST sequencing has drawbacks including redundancy in data as a result of over-expressed genes, sequencing large numbers of housekeeping genes which are not developmentally regulated and the under-representation of rare transcripts. Simply comparing the relative abundance of a transcript in the different life cycle stages provides useful information about shifting patterns of gene expression associated with development. These changes will also reflect the adaptation of the parasite to its environment within the host including requirements for invasion of the host, feeding, reproduction and evasion of host defences (Gasser and Newton, 2000; Boag et al., 2001). However, where mRNA is harvested from parasites in previously uninfected host donor animals, the parasite may not be expressing genes which may encode peptides/ proteins which allow it to manipulate or evade specific elements of the host anti-parasite immune response. Such proteins, if they exist, may be critical targets for novel parasite control strategies. This is exemplified by the discovery of ESTs encoding several novel immunomodulatory proteins from the filarial nematode B. malayi including a macrophage migration inhibitory factor (MIF; Blaxter et al., 1996), a transforming growth factor β homologue (Gomez-Escobar et al., 1998, 2000) and serpin-like serine proteinase inhibitors (Zang et al., 1999). All possess biological activity against the predicted target pathway and, intriguingly, homologues of all are found in C. elegans.

With prior knowledge of parasite biology, it is possible to examine patterns of gene expression associated with a particular life cycle characteristic, for example, blood-feeding. Extensive evidence suggests that this is mediated by a cascade of proteases. Aspartyl proteases may cleave the globin peptide chain resulting in the disruption of normal structure and exposing the peptide to the activity of cysteine proteases and cathepsin Ds which can then break the peptides into smaller fragments. These can be split into individual aminocids by the action of metalloproteases, dipeptidyl peptidases and aminopeptidases (Tort et al., 1999). For example, sequences encoding these enzymes are
prevalent in the ESTs from 11-day-old and adult *H. contortus*, both blood feeding stages, but less abundant or absent in the third larval stage ESTs and in EST datasets from the related ovine abomasal parasite *Teladorsagia circumcincta* which is not an obligate blood feeder (Nembase at www.nematodes.org.uk).

The availability of EST or genomic sequence databases allows the immediate analysis of specific individual genes of interest without the need for probes, such as antibody or oligonucleotides, for cDNA library screening. Full-length coding sequences can be rapidly isolated from cDNA using PCR and primers derived from EST or genomic sequence.

Current predictions are that about 70% of parasitic nematode genes will have homologues in *C. elegans*, a percent similarity which makes this free-living nematode an invaluable tool for defining gene function. Conversely, this also predicts that 30% of nematode genes will have no homologues in the databases and these may include parasite-specific genes which would be a source of targets for parasite-specific vaccine and drug targets (Dalton et al., 2003). Comparative gene analysis is widely employed to identify homologues, orthologues and paralogues, particularly comparisons with *C. elegans* (Grant and Viney, 2001). These authors identified three areas where caution is required. First, gene predictions from the *C. elegans* genome sequence can be inaccurate, second, ascribing homology depends on the stringency of the bioinformatic conditions chosen and, third, sequence orthologues are not necessarily functional orthologues.

The challenge now is to interrogate the databases in a manner which will rapidly identify potential new drug targets and vaccine candidates. This issue has been addressed in a recent review (Dalton et al., 2003) and the authors suggested that vaccine candidates could be identified in an initial screen for proteins secreted by the parasite or those likely to be expressed on the surface or as transmembrane proteins. These proteins are likely to be pivotal to the parasitic way of life through their molecular interaction with host cells and molecules (Dalton et al., 2003). Many of these molecules can be identified from EST data using signal sequence prediction algorithms (Dalton et al., 2003) or, where large scale EST data is unavailable, using signal sequence trapping which exploits the presence of N-terminal signal sequences on ES and membrane anchored molecules (Tashiro et al., 1997; Smyth et al., 2003). Other possibilities in database mining include text-based searching based on gene ontology (Ashburner et al., 2000), in particular the GO consortium (http://www.geneontology.org) which provides updates of linkages from key model organisms and pathogens, the entries being categorised under molecular function, biological process and cellular component. At a more fundamental level, the Parasite-Genome site (http://www.ebi.ac.uk/parasites/parasite-genome.html) provides tools to BLAST 18 different parasite databases at one location as well as translations of ORFs, motif searches, key word searches and links to many useful sites (Tarleton and Kissinger, 2001).

While these sequencing initiatives are invaluable, the parallel development of molecular markers is an equally important goal (Grant and Viney, 2001). Markers can be used to map important life history traits leading to positional cloning and subsequent gene discovery using ‘forward genetics’. Grant and Viney (2001) provide a thoughtful discussion of the approaches available and their application. Mapping is undertaken using, amongst other techniques, polymorphic ESTs, microsatellites, amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD) and, perhaps the most abundant and versatile of all polymorphisms, single nucleotide polymorphisms (SNPs; Grant and Viney, 2001; Brooks, 2003).

The explosion in the amount of sequence data available through genome and EST projects has been matched by the development of high throughput technologies to monitor patterns of gene expression. Before discussing these, it is pertinent to discuss the impact of the polymerase chain reaction (PCR) in molecular parasitology.

### 3. Polymerase chain reaction

#### 3.1. Applications of PCR

The PCR has become an essential tool in all aspects of molecular parasitology. Since the initial description of the technique (Mullis et al., 1986; Saiki et al., 1986), it has been used to directly isolate specific genes from cDNA pools or gene libraries, for classification, for population genetics and studies of genetic polymorphism, for diagnostic purposes and so on. The applications of PCR have been discussed in detail in a number of recent reviews (e.g. McKeand, 1998; Gasser, 2001; Prichard and Tait, 2001). This technique has revolutionised molecular parasitology.

The review of Gasser (2001) is recommended to the reader because it provides a comprehensive insight into the variety of PCR-based methods for the identification of parasites and the study of genetic variability. To summarise, specific target sequences can be selected on the basis of prior knowledge from related organisms and can be used to differentiate closely related nematode species. The internal transcribed spacers (ITS) of ribosomal DNA are species-specific and provide a DNA region which allows for high sensitivity and specificity for DNA-based assays for parasitic helminths (Gasser et al., 1997; Prichard and Tait, 2001). Schröder et al. (1999) used this approach to differentiate between *Cooperia*, *Haemonchus*, *Trichostrongylus*, *Nematodirus* and *Ostertagia* genera from as little material as a single egg. Zarlenga et al. (1998) used ITS-based PCR to accurately estimate the proportion of *Ostertagia ostertagi* eggs in a mixed faecal population with a sensitivity equivalent to 0.05 eggs. Newton et al. (1998) used a technique termed PCR-linked restriction
3.1. DNA microarrays

The accumulation of large amounts of sequence information has led to the development of high throughput methodologies which can be applied at the genome scale to define patterns of gene expression. The most widely applied method to date is the use of DNA arrays where the known gene set is gridded onto a solid support medium in an ordered manner and probed with mRNA/cDNA from different parasite life cycle stages of interest. Technological advances in robotics allow thousands of genes to be spotted onto glass microscope slides. Nucleic acid probes can carry fluorescent labels (Cy3 and Cy5 dyes) which allow the direct analysis of differential gene expression with the aid of sophisticated image analysis software. Experiments can be repeated in a highly reproducible manner which allows statistical validation of the output.
Arrays can be produced by several methods, the method chosen reflecting the level of funding available to a project. ‘DNA chip’ technology is where oligonucleotides representing gene-specific sequences are synthesised directly on a flat glass surface. Although they are ‘state of the art’, they are very expensive and require dedicated equipment to monitor the output. Microarrays are high-density arrays of gene-specific PCR products gridded onto a glass surface. The PCR products can be prepared in any laboratory and spotted using robotics in house or contracted out. Dedicated array readers can be purchased but arrays can be read using standard fluorescent or light microscopes depending on the detection system employed to visualise hybridisation. Finally, and more within the financial reach of the ordinary laboratory, high density arrays of PCR products can be gridded on nylon membranes using standard robotic facilities. These can be interrogated with chemiluminescent, colorimetric or isotopic probes and the output monitored using fluorimeters, densitometers or phosphoimagers.

The global analysis of gene expression using microarray is established in C. elegans (Kim et al., 2001; Murphy et al., 2003). A recent paper described the global analysis of dauer gene expression in C. elegans and identified 1984 genes that show significant expression changes during the transition from the dauer to the non-dauer state (Wang and Kim, 2003). Genes identified included those encoding transcription factors, components of signalling pathways and those involved in metabolic pathways important for dauer survival and longevity. This analysis has clear relevance for understanding the mechanisms involved in larval arrest associated with several parasitic nematode infections such as ostereagiosis in cattle.

Analysing the molecular events leading to larval arrest is hampered by the difficulties associated with obtaining sufficient material. One possibility is to use laser capture microdissection. This approach has been applied to isolate liver stages of Plasmodium falciparum from human hepatocyte cultures and the quality of the parasite RNA recovered confirmed by RT-PCR (Semblat et al., 2002). cDNA microarrays are being used to unravel the pathways involved in sexual maturation and egg production of schistosomes in the bloodstream, the most important pathological consequences of infection being associated with this process (Hoffmann et al., 2002). The array comprised 576 ESTs from two different developmental stages of the parasite and analyses identified 12 new female and four new male-associated transcripts in the mature adult parasite. Clearly, the development of array projects to analyse patterns of gene expression associated with the different stages of a parasite life cycle is a priority if the expanding EST and genome sequence datasets are to be exploited to their full potential.

One cautionary note is required though. ESTs represent the genes being expressed at the time the parasite is harvested. Transiently expressed genes which may influence the immediate environment within the host, for example genes expressed to initiate or downregulate anti-parasite host immune responses, may not be represented in the EST dataset. This problem will be overcome when whole genome sequencing projects are complete and the full complement of the parasites genes can be arrayed with certainty. An alternative approach, where genome sequence is unlikely to be available in the short term, is the application of ‘shotgun’ genomic DNA microarrays where random inserts from a genomic DNA library are arrayed and gene expression monitored by differential hybridisation and sequencing of relevant clones. Using this, large differences in gene expression were identified between the blood stage trophozoite form of the malaria parasite, P. falciparum, and the sexual stage gametocyte form (Hayward et al., 2000).

4.2. Serial analysis of gene expression

Serial analysis of gene expression (SAGE) provides a sensitive and quantitative measure of gene expression in an organism (Velculescu et al., 1995, 2000). This approach samples short (14 bp) tags from the mRNA population of interest which, when concatenated and sequenced, contain sufficient information to identify the source transcript using standard BLAST searches. The frequency of each tag in the SAGE library accurately reflects the abundance of the transcript. The technique has been applied to analysing changes in gene expression associated with developmental arrest and longevity in C. elegans (Jones et al., 2001) and to monitor gene expression in P. falciparum (Patankar et al., 2001; Munasinghe et al., 2001). Recently, SAGE has been applied to examine gene expression in adult H. contortus and preliminary analysis of ~2000 tags indicates a pattern of gene expression corresponding to that predicted by EST analysis (Skuce and Knox, unpublished). Importantly, gene-specific PCR primers based on tag sequences were used to successfully amplify the precise transcript from adult worm cDNA, even with the capability of discriminating members of closely related gene families. These observations indicate that SAGE can be applied to parasites where database sequence information is relatively limited and that it may not be necessary to sequence tens of thousands of tags to gain representative information on levels of gene expression at a given time point.

5. Proteomics

The analyses outlined above provide a profile of the genes being expressed at a specific point in the life cycle and are now complemented by sophisticated techniques to characterise the actual proteins synthesised (proteomics). Proteomics is a systematic study of the structure function and control of the many and diverse proteins expressed in health and disease. The key to this emerging field is the availability of technologies which allow proteome-wide analyses in contrast to traditional reductionist biochemical
analysis of single proteins. Proteomics and other complementary analyses are core components of ‘systems biology’ that seeks to comprehensively describe biological systems by the integration of diverse types of biological data (Patterson and Aebersold, 2003).

Proteins define the biological phenotype of the organism and are the primary targets of most therapeutic agents (Barrett et al., 2000), the aforementioned review providing an excellent account of the possibilities arising from the use of proteomics in parasite biology. Importantly, proteomics provides information on post-translational modifications such as glycosylation (Packer and Harrison, 1998) which are vital for correct protein function and are increasingly recognised as contributing to immunogenicity, an important consideration in vaccine development.

Two key technologies are at the core of proteomics (Barrett et al., 2000). Firstly, two-dimensional polyacrylamide electrophoresis with powerful image analysis software and, secondly, mass spectrometry with database searching. The outcome of gel fractionation is now so reproducible it is possible to produce proteome maps from different points in the parasite lifecycle and compare these to identify and quantify specific protein spots. The reproducibility, now routine, allows protein spots to be quantified, a capability which creates an almost three-dimensional picture of protein expression patterns. Comparisons are aided by a variety of labelling techniques including the use of Cy-dyes, lectins or spot identification after immunoblot probing.

Proteomics is likely to have a major impact on drug discovery and validation. In the first instance, this type of analysis can be used to confirm that a novel drug does affect the target protein in vivo (Page et al., 1999). Drug treatment will usually cause the up- or downregulation of particular biochemical pathways and definition of these will help to define the mode of drug action. The same approach can be used to define mechanisms of resistance to existing drugs (Barrett et al., 2000), a vitally important research area given used to define mechanisms of resistance to existing drugs and are increasingly recognised as contributing to immunogenicity, an important consideration in vaccine development.

6. Genetic transformation

The technologies described above provide in-depth information on the gene complement of an organism and the nature of the protein products encoded but they do not wholly address the question of function. This is crucial when it comes to selecting potential targets for novel parasite control strategies. Gene knockout technologies provide a means to address this.

6.1. RNA interference

RNA interference (RNAi) has been applied to the study of gene function in a variety of divergent organisms (Hannon, 2002). Double-stranded RNA (dsRNA), corresponding to a particular gene, is introduced into the target organism and this results in specific interference of the targeted gene at the post-transcriptional level (Tabara et al., 1998). The RNAi effect spreads throughout the worm and is heritable to at least the F1 generation. RNAi was rapidly adapted for the study of gene function in C. elegans and applied to the systematic study of gene function in specific chromosomes (Fraser et al., 2000) and in genome-wide analyses (Maeda et al., 2001). High-throughput RNAi became feasible when the ability to induce RNAi by soaking worms in dsRNA solutions was demonstrated (Tabara et al., 1998) with the efficiency and reproducibility of the procedure being enhanced by the addition of spermidine (Maeda et al., 2001). The latter study examined 2479 non-redundant cDNA clones representing approximately 13% of predicted genes in C. elegans. Phenotypic effects such as embryonic lethality, morphological abnormalities and arrested larval development were identified for about 27% of the genes tested. A recent report described the outcome of RNAi on approximately 86% of the 19 427 genes predicted in C. elegans. Mutant phenotypes were
identified for 1722 genes, two-thirds of which had not previously been associated with a phenotype (Kamath et al., 2003). RNAi can be applied to the identification of novel drug targets (O’Neil et al., 2001) and to the analysis of genes associated with specific life cycle events such as embryogenesis (Plano et al., 2000) or metabolic processes such as fat regulation (Ashrafì et al., 2003).

It is worthy to note that RNAi knocks down gene function but is not the same as a gene knockout. It is transitory and the phenotype can depend on the half-life of the protein product of the gene. If no phenotype is observed, this may reflect genetic redundancy, for example where the target gene is a member of a gene family. Finally, different tissues show differing susceptibility to the technique with the nervous system being particularly refractory.

Many parasite genes will not have precise homologues in *C. elegans* and may be species-specific and will require specific gene-knockout in the parasite of interest to accurately ascribe function. Reproducible RNAi effects can be produced in the free-living stages of *H. contortus* but not, as yet, in parasitic stages. This is an area of urgent research effort given impetus with the demonstration that acetylcholinesterase expression in the rodent intestinal nematode can be suppressed in vivo by soaking infective L3 in the RNA solution prior to infection (Hussein et al., 2002).

6.2. Analysis of parasite gene function in *C. elegans*

The studies in *C. elegans* have immediate applicability to parasitic nematodes where the genes of interest have high identity and it may be inferred that function is conserved. In fact the latter can be confirmed by using the parasite gene to functionally rescue the RNAi phenotype in *C. elegans* (Britton and Murray, 2002; Boag et al., 2003). This approach provides a powerful tool for the analysis of gene function in parasites when it is combined with the expanding databases carrying information on gene mutants in *C. elegans* such as the Caenorhabditis Genetics Centre (http://biosci.umn.edu/CGC/CGChomepage.htm) or the *C. elegans* gene knock-out project at University of Colombia (http://www.zoology.ubc.ca/kogenomics/kowebpge.html; Brooks and Isaac, 2002). Transformation of *C. elegans* with parasite genes confirmed the genetic association of β-tubulin genes with benzimidazole resistance (Kwa et al., 1995) and has been used to study the function and location of particular parasite gene promotors (Qin et al., 1998; Britton et al., 1999).

6.3. Transfection

Transfection (transgenesis) provides another tool to study gene function in metazoans. The gene of interest can be introduced by microinjection into the gonads, an approach which has formed the basis of many of the classical genetic studies in *C. elegans* (Mello and Fire, 1995), or by ballistic gene transfer where DNA is introduced into the target organism coated on the surface of a solid support. While providing the opportunity to conduct experiments paralleling the many in *C. elegans*, transfection could be used to introduce potentially lethal genes into a parasite population, a means of gene knockout which would help define gene function and may have applications in parasite control.

*C. elegans* has been transfected with the promotors from parasite genes and function confirmed using reporter (*gfp* or *lacZ*) transgenes (Qin et al., 1998; Britton et al., 1999). The promotors directed the expected spatial expression pattern but temporal regulation differed. However, the transgenes can be modified to dissect the promoter elements which determine spatial and temporal expression (Brooks and Isaac, 2002). *Gfp*-reporter gene constructs were prepared using promotors for each of two housekeeping genes, actin-2 and an ATPase, from *S. stercoralis* and injected into the gonad of free-living females of the same parasite (Lok and Massey, 2002). Forty-one percent of injected females survived the procedure and both reporter gene constructs were transiently expressed in the maternal gonad or in intrauterine embryos. Expression persisted in embryonating eggs deposited by free-living females and continued to be observed as late as the three-fold stage of embryonic development. Embryonic expression of the transgene suggests that heritable transformation, in this parasite at least, may be a feasible proposition. Transient transfection in adult female *B. malayi* has been reported with the exogenous DNA also being taken up in the developing embryos, possibly forming semi-stable heritable extra chromosomal arrays by analogy with *C. elegans* (Higazi et al., 2002). The authors proposed that it might be possible to isolate transfected microfilaria and use these to infect mosquitoes and then jirds to generate transected parasite lines.

Ballistic DNA transfer utilises tungsten, gold or glass microparticles with DNA chemically linked to the surface. The particles are introduced into the target using a ‘gene-gun’ at high pressure and under vacuum. The technique has been used with success to transform the nematodes *C. elegans* and *Litonomoides sigmodontis*, the latter often serving as a model for human pathogenic lymphatic filariae (Jackstadt et al., 1999), and *B. malayi* (Higazi et al., 2002). Of crucial importance, the technique offers the possibility that heritable transformation may be possible in parasitic nematodes. Transient expression of reporter gene constructs, introduced by ballistic transfer, has been reported in *S. mansoni* (Wippersteg et al., 2002). Adult male and sporocysts stages of *S. mansoni* were successfully transformed with a heat shock protein-green fluorescent protein (*gfp*) construct with *gfp* expression being visible along the surface of the adults and also within sporocysts (Wippersteg et al., 2002).

Transfected microorganisms (bacteria and viruses) have been employed as a means to deliver antigen to a specific
infection site using replicating vectors such as *Salmonella* or *Vaccinia* (Chatfield et al., 1995) or, as described in a recent report, construction of a recombinant orf virus that expresses an *E. granulosus* vaccine antigen (Marsland et al., 2003).

While transfection is a useful technology for the analysis of gene function and expression, it has other potential and vital practical applications. Given that several studies, such as those mentioned above, have demonstrated that parasitic nematode promoters are functionally active in *C. elegans*, this raises the possibility that *C. elegans* or cell lines derived from this organism can be used to express parasite proteins of interest as vaccine candidates (e.g. Redmond et al., 2001) with the advantage that the protein products will carry more appropriate post-translational modifications than those attainable from yeast, insect or mammalian cells. Conformational epitopes could include glycan and are likely to contribute to the protection. Studies (Newlands and Knox, unpublished) show the > 50% of the IgG responses in lambs protected against *H. contortus* challenge by immunisation with proteins isolated from the parasite intestine were directed at the glycan component and it is becoming increasingly recognised that glycans themselves may be antigenic targets for immune intervention. Nematodes have unique carbohydrate structures which can be species-, stage- or tissue-specific (Haslam et al., 1996, 1998, 2001).

7. Protein structure and function prediction–drug design and vaccine development

Modelling protein structures can identify unique features of the parasite protein which can be exploited for drug and vaccine design. McGrath et al. (1995) described the crystal structure of cruzain, the major cysteine protease present in *Trypanosoma cruzi*, the protozoan parasite which cause Chagas’ disease in man. The structure was modelled using papain as a reference template with cruzain complexed to a potent inhibitor, the inhibitor greatly reducing parasitaemia in a cell culture system (McGrath et al., 1995). Wasilewski et al. (1996) showed that cysteine protease inhibitors blocked schistosome haemoglobin digestion in vitro and reduced worm and egg production in vivo. In addition, Lim et al. (1999) showed that invasion of human skin by schistosome cercariae was blocked by topical administration of specific inhibitors.

As the databases expand, in silico predictions of protein structures directly from amino acid sequences will become an increasingly powerful tool for the identification of novel drug and vaccine therapies. Comparative modelling of protein structures is a predictive technique based on the structures of other proteins that have been confirmed experimentally. The accuracy of the prediction depends on selecting the correct template, aligning the query sequence to this and building non-conserved surface loops (Contreras-Moreira et al., 2003). The accuracy of the prediction can be enhanced using new bioinformatic tools such as in silico protein recombination. This is a genetic algorithm with crossover and mutation which exploits the variability of templates and sequence alignments to produce optimised models by artificial selection (Contreras-Moreira et al., 2003). Aligning this information with databases describing chemical interactions with proteins will accelerate the rate of discovery of compounds which may inhibit protein function and hence harm the invading organism. Accurately predicting what parts of a protein are exposed on the surface and may be accessible to immune attack will become an increasingly powerful tool in vaccine development and may help to overcome some of the problems of correct recombinant protein folding encountered at present (see Dalton et al., 2003).

8. Recombinant protein expression

Much of the preceding discussion has focused on the identification of genes and their protein products which are critical at the host–parasite interface, often with the longer term aim of evaluating their potential as drug targets or vaccine candidates. Ideally, the recombinant protein will have the same functional properties as the native form and, in the case of vaccine candidates, correct post-translational modification. The latter affects the immunogenicity of the protein and there is increasing evidence to suggest it affects its value as a vaccine component. Several of the major vaccine candidates under evaluation for the control of *H. contortus* are glycosylated, the glycan being highly immunogenic (Haslam et al., 2001) and bacterially expressed versions are ineffective protective antigens, presumably because they lack the appropriate post-translational modifications (Knox et al., 2003). On the other hand, the only two metaozoan parasite vaccines to reach the commercialisation stage are expressed in *E. coli* (Wiladsen, 1995; Lightowlers, 2002). Functionally active recombinant proteins, for example those with enzyme activity, are often expressed in yeast or baculovirus (Dalton et al., 2003; Williamson et al., 2003), systems which glycosylate the protein product but not in the same manner as helmint parasites with the possibility that inappropriate post-translational modifications might mask protective epitopes or affect immunogenicity. Recent reports bring a new perspective to this issue. Humoral responses are correlated to the expression of protective immunity in humans with schistosomiasis infection (Nyame et al., 2003) and in sheep immunised against *H. contortus* with ES (Vervelde et al., 2003) or gut antigens (Knox et al., 2003). A high proportion of the humoral response is directed to the glycan component and a monoclonal antibody to a glycan epitope, in the presence of complement, killed schistosomula in vitro (Nyame et al., 2003). Therefore, the nature of the glycan on a recombinant protein could well be crucial to achieve protection. The ability to express parasite antigens in
the correct manner still requires considerable research towards the development of appropriate expression systems, be these parasite cell lines or transgenic close relatives of the target organism such as *C. elegans* in the case of parasitic nematodes. Another possibility receiving increasing attention is the use phage antibody display technology to select high affinity antibodies against a range of antigens (Griffiths and Duncan, 1998), an approach which can also be applied to conformational epitopes and carbohydrates (Deng et al., 1995). In the case of the latter, peptides can be selected which bind to most carbohydrate-specific antibodies. These peptides could then be incorporated into vaccines against parasite antigens where glycan epitopes are thought to be important for the stimulation of protective immunity.

Choosing the appropriate means of expressing recombinant proteins in cellular systems can be difficult. Until recently, the choice of systems was limited but there is now a bewildering variety of options available for recombinant protein expression, including bacteria, yeasts, insect cells, mammalian cell lines and plants. Production of a particular protein in sufficient quantities in any one system is not guaranteed and failure may be due to a variety of reasons. For example, the recombinant protein may be toxic to the host cells or the level of expression may be restricted due to the codon bias of the host cells differing from that in the organism of origin. Furthermore, it is one thing to produce a recombinant protein in bulk but quite another to produce one with the appropriate immunogenic properties as discussed above. In addition, the commercial reality is that the process must be relatively simple and amenable to inexpensive scale-up. If not, the vaccine has little chance of being a commercial success.

Analysis of the native protein and the gene encoding it provides important information to inform expression strategies, e.g. is the protein secreted, is it glycosylated, are the protective epitopes dependent only on secondary structure or is it necessary to obtain tertiary conformation, and does the protein possess some measurable activity (e.g. enzymatic) that can be employed to assess whether secondary/tertiary structure has been obtained? All of these factors can be used to select the best pro- or eukaryotic expression system. Our understanding of protein expression in the cell is constantly improving and this, in turn, is leading to the development of better systems. For additional discussions and references on the types of expression systems available, the advantages/disadvantages of each type and on means of optimising expression, readers should also consult Andersen and Krummen (2002), Gilbert and Albala (2002) and Dalton et al. (2003).

### 9. DNA vaccination

DNA vaccines are based on plasmid or viral vectors which contain the inserted gene or cDNA of interest under the control of a strong promoter. DNA vaccines can be administered by intramuscular, epidermal or mucosal routes and the route of administration influences the type of immune response induced. Intramuscular immunisation appears to favour Th1-type responses whereas, intradermal administration, using a gene-gun to propel plasmid DNA-coated gold particles into the epidermis, promotes Th2-type responses (Robinson and Torres, 1997; Feltquate, 1998). Intraspleen delivery has also been described (Can0 et al., 2001). The DNA is taken up by host cells and the protein product is expressed. This protein is then recognised as foreign by the immune system of the host, inducing an appropriate immune response. They eliminate the requirement for the expression and purification of recombinant proteins. The DNA vaccines are also stable, not requiring refrigerated storage, a great advantage in field applications. Recent developments in this field have been the subject of comprehensive review (Alarcon et al., 1999; Watts and Kennedy, 1999).

Rodents have been vaccinated with DNA encoding antigens from *Schistosoma japonicum* (e.g. Waine et al., 1997; Zhou et al., 2000) and *S. mansoni* (Dupre et al., 1999; Da’dara et al., 2002), the immunisation regimes stimulating protective immune response with reductions in worm and egg numbers. Recently, DNA vaccine constructs expressing candidate protective antigens, Sj28GST and Sj23, have been evaluated in the field in sheep, waterbuffalo (Shi et al., 2001) and cattle (Shi et al., 2002) with partial protection against challenge infection being recorded in both studies.

Humoral responses in mice following vaccination with DNA-encoding GSTs from *F. hepatica* have been evaluated with a ‘cytoplasmic’ construct, given by intramuscular injection, generated a T-helper cell (Th)-1 type response whilst intradermal injection of an ‘extracellular’ construct elicited a Th-2 type response (Smooker et al., 1999). These observations were confirmed and extended using DNA vaccines encoding a fatty acid binding protein from *Fasciola gigantica* and cathepsin L from *F. hepatica* (Smooker et al., 2001). Moreover, rats were successfully protected against experimental challenge with *F. hepatica* following intramuscular injection of cDNA encoding a cysteine protease from adult *F. hepatica* (Kofia et al., 2000).

This approach has been extensively tested for vaccination of sheep against *T. ovis*. Both plasmid DNA and recombinant ovine adenovirus constructs encoding the 45 W antigen (Rothel et al., 1997a,b) were evaluated and the latter induced some protective immunity against challenge. DNA vaccines encoding 45 W, 18 k and 16 k host-protective antigens were given by intramuscular injection with antigen-specific antibody response only being detected to 45 W (Drew et al., 2000). Immune responses to DNA vaccines can be modulated by co-delivery of cytokine genes (Scheerlinck et al., 2002; Siddiqui et al., 2003) and efficacy can be improved by DNA-prime and protein-boost strategies (e.g. Rothel et al., 1997a;b; Scheerlinck et al., 2002; Siddiqui et al., 2003). Direct injection of partial expression
libraries from Mycobacterium pulmonis induced protection against challenge (Barry et al., 1995) and mice were partially protected against Plasmodium chabaudi infection by a similar approach (Rainczuk et al., 2003).

10. Concluding remarks

The above discussion summarises the variety of molecular tools which are now available for studying parasite biology, genetic diversity and host–parasite interactions. The challenge now is to exploit these technologies, and develop them further as required, to address the specific question being asked. The major questions of parasitism will be addressed with multidisciplinary approaches. To date, studies on the molecular basis of parasitism, including host responses, have been largely restricted to looking at single genes or gene families and proteins in isolation, the latter with relatively insensitive techniques being applied to quantification, if at all. We can now contemplate the global analysis of gene expression by, e.g. microarray and correlate changes with alterations in the proteome and metabolome map of the parasite subjected to the same experimental conditions. All these parameters can be quantified to enable statistical analysis. Experiments of this type will inform about the pathways involved in a process with the real bonus of providing information on the selected and related pathways which may or may not have been thought to be relevant intuitively. Gene expression in indicated pathways can be confirmed and the analysis refined using quantitative real-time PCR. To illustrate this broad approach, with the possible exception of the benzimidazoles, we still do not understand the mode of action of many of the anthelmintics in current use. Gene expression and proteome comparisons in normal, drug-treated, drug-susceptible and drug-resistant worms would be expected to provide an insight into the mode of drug action and the principal proteins involved as well as changes involved in the expression of resistance. These changes may be directly as a result of altered gene expression levels or as an effect of the drug blocking a particular protein function. These analyses would be complemented by analyses of the products of metabolism (metabolome) where, for example, accumulation of a metabolite may indicate that the anthelmintic is interfering with a particular metabolic pathway.

It is important to avoid performing a global analysis of gene expression experiment and then focusing attention on a limited gene set which just happens to be the laboratories’ main interest for the last decade! The outcome must be viewed with a ‘neutral eye’. Array experiments must be analysed within strict statistical constraints and each of maybe hundreds of up-regulated genes given the same degree of attention to avoid the output of the experiment being corrupted by individual biases. Of course, this requires potentially a considerable commitment of resources to conduct an experiment with appropriate neutrality. Of course, this requires potentially a considerable commitment of resources to conduct an experiment with appropriate neutrality. Caution is required in the interpretation of these data sets. For the metazoan parasites where sequence data are available, it is usually in the form of ESTs and, these datasets do not represent even 50% of the genome! Also, as mentioned above, the ESTs only represent the genes being expressed at a given time point in the parasite lifecycle. Therefore, many important genes may not be represented. This leads to the obvious conclusion that we must continue to press for the funding required to generate whole genome sequence data for the major metazoan parasites.

Given the present author’s interests, the availability of the technologies described above provides an enormous opportunity to analyse the molecular events at the host–parasite interface as infection progresses and host anti-parasite immunity develops. mRNA can be isolated from the parasite as infection progresses and the host immune response develops. The latter can be monitored by conventional immunological techniques and with analyses of immune gene arrays. A recent study used microarray to analyse the responses of dendritic cells and macrophages to bacterial, protozoan and helminth parasites. The responses induced by the individual pathogens showed major differences in the functionally related gene profiles associated with each infectious agent (Chaussabel et al., 2003). These authors concluded that the association of characteristic functional gene clusters with each infectious agent is consistent with the concept that antigen-presenting cells have pre-wired signalling patterns for use in the response to different pathogens. Experiments of this type combined with the parallel analysis of gene expression in the parasite would allow us to look at the host–parasite interaction in a new light. Is the conventional Th2-like response to helminths actually initiated deliberately by the parasite? This approach will throw up surprises which may become targets for immuno- or chemotherapy. Gene expression can be confirmed by proteomics and a variety of conventional immunological techniques used to identify parasite proteins which are recognised by and may modulate host immune responses. Gene knockout can then be used to eliminate the protein product of the gene and to define the effect of this on parasite survival, in theory a ‘perfect protection trial’. In conventional protection trials, more and more finely refined protein fractions are tested in a series of trials. These are expensive, time-consuming and often not wholly conclusive in that it is always possible that a very minor protein component is responsible for the protective responses stimulated by vaccination with an essentially, but not completely, pure parasite protein. Also, the outcome of the trial can be critically influenced by the choice of adjuvant. In addition, the purification process needed may alter the properties of the target protein and affect immunogenicity.
Gene knockout technologies provide the exciting possibility of defining genes, and by extrapolation the proteins encoded, which are critical for parasite survival without the need for protein purification and the difficulties this imposes where material is scarce. Provided that RNAi persists, or can be manipulated to persist, through the entire parasite lifecycle, this technique could become the approach of choice for the initial screening of candidate molecules. Assuming a high level of gene knockdown in the target, in vivo RNAi should provide convincing proof that the protein product is a valid target before going down the road of native and recombinant protein testing. However, in assessing the outcome of these RNAi experiments, we need to expand the range of phenotype measurements to include those (egg output, egg viability, worm weight, sex ratio) commonly used in determining vaccine efficacy to date. Proteome comparisons between normal and RNAi-treated worms may provide a rapid focus on modulated protein expression in response to the gene knock down and, thus, information on compensatory gene expression which may aid parasite survival. One of the questions that is often posed concerning vaccine targets is functional redundancy within gene families. Can the parasite up-regulate the expression of a related gene to compensate for loss of specific gene function? There is some evidence in the literature to support this view but the question can now be addressed using RNAi and array techniques allied with real-time PCR experiments.

Underpinning these studies will be an expanding requirement for bio-informatics to store data, to access it, to identify genes and proteins, to analyse patterns of gene expression and so on. As the bioinformatic tools develop, interspecies comparisons should enable predictive biology to be undertaken, a process which should accelerate technology transfer across the field of metazoan biology. The techniques available are also highly reproducible within and between laboratories and should provide us with the capability to incorporate data from geographically distinct laboratories into common databases. The resultant analytical possibilities should enhance the power of individual laboratory experiments and effectively ‘globalise’ research efforts on a scale not possible to date. In the example above where I proposed that correlation of gene expression in the parasite with host immune gene expression might identify novel interactions where the parasite may direct the immune response, such a proposition would be supported if similar patterns of gene expression were observed in quite distinct parasites in the face of a developing common immune response. This comparison could be performed, e.g. between flatworms and gastrointestinal nematodes.

Of course, these brief concluding remarks will probably not include many peoples’ ideas on technology exploitation. All the better! The more ideas there are, the better our prospects of exploiting the genomic revolution. The tools are there, it is up to us to apply them!

Acknowledgements

I am indebted to the ASP for providing the opportunity to attend the Darwin meeting. I also acknowledge the Scottish Executive Environment and Rural Affairs Department for supporting my research.

References


