

Identification of Parasitic Genes by Computational Methods

A. Bhattacharya, S. Bhattacharya, A. Joshi, S. Ramachandran and R. Ramaswamy

A number of parasite genome projects are under way, and large amounts of nucleotide sequence data are becoming available for analysis. There is an urgent need for development of theoretical tools to analyze the genome data, including identification of protein-coding sequences. The majority of the methods developed to date require prior information about the genome before accurate predictions can be made. Because such information is not available for many parasites, these methods cannot be directly applied. In this article, Alok Bhattacharya and colleagues describe some of the gene-prediction methods commonly in use, and a new method, GeneScan, that they have developed for the analysis of parasite genomes.

The nucleotide sequences of the entire genomes of several organisms are already available, and data are currently being generated for many others. The full nucleotide sequence of an organism can lead to a more complete understanding of the organism, either through direct analysis, or via the so-called 'functional genomics approach'¹. In parasitic species, such analysis may be useful to identify new targets for development of vaccines and chemotherapeutic agents, and for understanding specific mechanisms of pathogenesis – eg. pathogenicity islands have been identified in a number of organisms, such as *Vibrio cholerae*².

Several parasite genomes are currently being sequenced and analyzed, including *Leishmania major*, *Plasmodium falciparum*, *Brugia malayi* and *Trypanosoma cruzi*. Expressed sequence tag (EST) analysis³ and complete nucleotide sequencing of these genomes are being pursued as well as physical mapping (more details can be found on the parasite genome website: <http://www.ebi.ac.uk/parasites/parasite-genome.html>). Once the genome sequences are available, the major task is to identify and characterize protein-coding and regulatory regions, and to predict the probable biological functions of the encoded proteins.

Many computational methods have been developed to address the complex problem of identifying possible coding regions in genomic DNA sequences. Commonly used programs employ a range of mathematical techniques, such as neural nets, Markov chain analysis, hidden Markov models, dynamic programming and linguistic analysis. The resulting algorithms predict the coding properties of unknown sequences with varying degrees of success.

The application of standard gene-identification algorithms (often developed for microbial and vertebrate genes) to parasite genomes may be limited because of

insufficient information available to 'train' programs based on the detection of standard patterns. The considerable variation in genetic organization and gene sequences among parasitic organisms, even among closely related genera (as compared with vertebrates), and the limited number of genes characterized so far from some parasites make it difficult to define consensus splice sites, transcription regulatory sequences, and transcription start and stop sites. A secondary issue is that some parasitic organisms have AT-rich genomes and the performance of many gene identification methods shows a noticeable deterioration with decreased GC content⁴. This may be due to coding regions being generally more GC-rich than noncoding ones even in genomes that are AT-rich (as in some parasites). Because many of the methods are based on discriminative power of oligonucleotide frequency derived measures, GC content *per se*, rather than some intrinsic biological constraint, influences results of gene-finding programs.

Different methods for gene identification

The major goal of gene-identification methods (Box 1) is to locate all the genes in a query genomic sequence. Ideally, the programs should not only find the genes, but also accurately predict the boundaries, such as introns and exons, translational start and stop sites and, if possible, promoter sites. It would also be desirable if the same method (with minimal modification or additional data) could be used for analysis of any genomic sequence regardless of its characteristics. No method available has all these capabilities.

The majority of gene-finding methods have been based on a consensus approach requiring an existing database for training⁵. The so-called 'signal-based methods' use signatures, such as locating start and stop codons and consensus sequences for transcription start sites, promoters, splice sites and polyadenylation signals to locate a gene in a genomic sequence. The content-based methods, however, look for bias in the codon usage, oligonucleotide frequencies and other similar indicators. Both signal- and content-based methods have drawbacks, mainly due to variation in the signals and content among different genes of an organism (which may be a major problem in parasitic organisms), as among different organisms. Some methods require the calculation of several statistical measures, the analysis of which may be organism-specific, and may, therefore, yield ambiguous results. By themselves, signal and content sensors cannot find genes accurately because statistical signals may be weak. Alternatively, there may be some correlation between content and signal. Recently developed methods such as GenLang combine both sensors and so have improved gene-finding capabilities.

Correlation methods and GeneScan

One feature of all the methods described in Box 1 is that they are context dependent. In each case, a sample set is required, and information obtained from the

Alok Bhattacharya and Amita Joshi are at the School of Life Sciences, Sudha Bhattacharya is at the School of Environmental Sciences and Ram Ramaswamy is at the School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110067, India. Srinivasan Ramachandran is at the Center for Biochemical Technology, New Delhi 110 007, India. **Tel: +91 11 610 7676, Fax: +91 11 616 5886, e-mail: alok@jnuniv.ernet.in**

Box 1. Some Commonly used Gene-finding Methods

Sequence similarity approach

The inherent assumption in this approach⁹ is that, if a query sequence is homologous to another protein or gene, then it is a gene. This may of course include pseudogenes or cryptic genes, but excludes noncoding DNA. One of the main advantages of this method is that it can also identify rRNA and tRNA genes and assign putative functionality. The major problems, at least with regard to parasites, arise from the lack of representative databases (which are searched for homologues) and the difficulty of defining similarity quantitatively: about 40–60% of parasite genes are not significantly similar to the genes present in the currently available databases.

GeneMark

This algorithm has been crucial in discovering a large number of genes in the several completed genome projects, such as *Saccharomyces cerevisiae*¹⁰, *Mycoplasma genitalium*¹¹ and *Haemophilus influenzae*¹². In this technique¹³, differences in oligonucleotide frequencies between protein-coding and non-coding sequences have been exploited for identification of protein-coding sequences (<http://genemark.biology.gatech.edu/GeneMark>). A Markov chain model has been used to calculate probabilities of oligonucleotides, taking into account correlations between nucleotide frequencies in different positions of the sequence. Because correlation between nucleotides differs in coding and noncoding sequences, the corresponding Markov models also differ. Correct reading frames are predicted using a phased or non-homogenous Markov model, which produces distinct models corresponding to the six possible reading frames. GeneMark is an extremely successful and sophisticated algorithm that is being used for annotation in many (mostly prokaryote) genome projects. For each sequence, GeneMark determines the set of all possible open reading frames (ORFs) and the *a priori* probability that the ORFs are coding, on the basis of which a complete set of predictions is made. A hidden Markov model version, GeneMark.hmm, has recently been developed¹⁴.

GLIMMER

This also uses Markov models (interpolated Markov models or IMMs) to identify the coding regions and distinguish them from noncoding DNA¹⁵. The GLIMMER-M version is being used for analysis of the *Plasmodium* genome. The IMM approach (<http://www.tigr.org/~salzberg/glimmer-nar.pdf>) is based on a combination of Markov models from first through eighth order, weighting each model according to its predictive power. GLIMMER's IMM is a three-periodic non-homogeneous Markov model. GLIMMER has been tested on many whole bacterial genomes, including *Haemophilus influenzae*, *Escherichia coli* and *Helicobacter pylori*. The GLIMMER system consists of two programs: the first is the training program, which takes an input set of sequences and builds and outputs the IMM for them. These sequences can be complete genes or just partial ORFs. For a new genome, this training data can consist of those genes with strong database hits as well as very long ORFs that are statistically almost certain to be genes. The second program uses this IMM to identify putative genes in an entire genome.

GeneId

This is a hierarchical rule-based system for identifying probable protein-coding genes: it first locates specific signals in a sequence, such as potential splice sites, start and stop codons, and then assembles these into potential first, internal and last exons¹⁶ (<http://www.imimes/GeneIdentification/GeneId>). Exons are evaluated according to a number of characteristics related to coding and splicing, and only likely exons are kept that meet certain minimum statistical criteria as compared with a sample set. Mutually exchangeable exons (normally in the same frame) are put together in groups. Only the top 15 ranking first and last exon classes and the top 35 ranking internal exon groups from each sequence are kept and assembled into potential gene models, with ORFs ranked according to quality of the exons they contain. GeneId also compares predicted exons to a network of protein databases (using BLAST). Predicted exons with matches in the database are marked by

the locus name of the match, which increases prediction accuracy. However, the number of exons are vastly overpredicted by GeneId and if the sequence contains frameshift errors in exons this may affect the quality of the prediction in the current implementation. The sample sets that have been used in GeneId consist of the first, internal and terminal exons from mammalian, rodent and vertebrate groups of GenBank (<http://www.ncbi.nlm.nih.gov>) and, therefore, predictions will be biased towards the genomes of higher eukaryotes.

GRAIL

This is a suite of tools (<http://compbio.ornl.gov>) designed to provide analysis and putative annotation of DNA sequences both interactively and through the use of automated computation¹⁷. The coding recognition portion of the system uses an artificial neural network that combines a series of coding prediction algorithms based on different coding measures such as the frame bias, Fickett's measure¹⁸, fractal dimension, coding hexamers and others. There are three basic versions – GRAIL 1, GRAIL 1a and GRAIL 2 – which differ in the sizes of the variable window used and in whether additional information, such as splice junctions, translation start and stop sites, and noncoding scores of 60-base regions on either side of a putative exon, is employed. All three versions have been trained to recognize coding regions in human DNA sequences, although they also work well on a number of other organisms, particularly other mammals.

GENEFINDER

This algorithm¹⁹ uses statistical criteria (primarily log likelihood ratios, or LLRs) to identify likely genes within a given genomic sequence. LLRs assess the hypothesis that a test sequence is a site of a given type (essentially splice sites or start or stop sites) against the hypothesis that the test sequence is 'random'. In the LLR approach, these frequencies are estimated from the known sites together with the test site, so as to obtain the maximum likelihood under the hypothesis that the test sequence and known sites are drawn from the same family. Candidate genes are evaluated on the basis of scores that reflect their splice site, translation start site, coding potential LLRs, and intron sizes. A dynamic programming algorithm is used to find the set of non-overlapping candidate genes (on a given strand) having the highest total score (among all such sets). In addition to the highest-scoring list of candidate genes, GENEFINDER also gives a list of high-scoring exons – that is, those whose normalized scores exceed the user-definable parameters. This method is currently being used for annotation in the *Caenorhabditis elegans* genome project. GENEFINDER has some drawbacks in terms of identification of overlapping coding regions and the inability to predict with sequences containing errors.

TESTCODE

The asymmetry in codon usage gives rise to compositional variations in coding versus noncoding regions. This observation can be used to develop methods for identification of coding regions. In the technique TESTCODE¹⁸ a total of eight parameters – four positional and four content-based (one for each nucleotide) – are used to judge whether a sequence is coding or not. For a test sequence one calculates the eight parameters and obtains the probabilities (π_i) and weight (w_i) for each of them. The sum of all π_i s and w_i s is evaluated to get the TESTCODE indicator. This method cannot find exact boundaries by itself but it is well suited to be used along with other methods.

Other methods

Additional methods have been developed based on Linguistic analysis (GeneLang²⁰), artificial neural networks and dynamic programming (GeneParser²¹), and hidden Markov models (Genescan²²). The need for training of these algorithms on a set of specific indicators makes them well suited to the study of particular species. Wentian Li of the Rockefeller University at <http://linkage.rockefeller.edu/wli/gene/> maintains an excellent compilation of links to genomic databases and computational tools. Many of the programs have been compared against each other using a common test data set; an example of one such comparison can be seen in the paper of Burslet and Guigo²³.

sample set is used to determine the coding potential of the query sequence. In recent years, attempts have been made to develop methods for gene prediction that do not require prior information. Many of these methods are based on analysis of correlations within DNA sequences – that is, analysis of the probabilities of finding nucleotides separated by a given distance. In general, noncoding sequences appear to have long-range correlation, whereas coding sequences show a striking short-range correlation^{6,7}, and this observation can be used to develop techniques for detection of probable genes. These existing correlations embedded in the DNA sequence may not be apparent upon superficial examination but can be studied by means of several computational tools, one of which is the discrete Fourier transform⁸.

The analysis of a large set of coding and noncoding sequences has revealed that the Fourier transform of coding sequences has a distinct peak (Fig. 1) at frequency $f = 1/3$ (which implies a periodicity of three), and that this peak is noticeably absent from noncoding sequences. This peak in the Fourier spectrum corresponds to a three-base periodicity in the correlation, and has been found in genes from a wide range of organisms, from *Mycoplasma* to humans, and is independent of any sequence feature including genomic base composition and minor sequencing errors.

Three-base periodicity is the basis of the gene prediction method GeneScan⁸. In order to detect this, the query sequence is converted into a digital signal and its Fourier spectrum is computed; the quantitative measure used is the signal:noise ratio of the power-spectral peak at frequency $f = 1/3$ – that is, the intensity of this peak divided by the average of the total spectrum. This measure, denoted P_N , N being the length of the gene, distinguishes likely coding sequences ($P_N > 4$) from likely noncoding sequences ($P_N < 4$). A few genes also display other peaks in addition to the prominent peak at $f = 1/3$; an example is the serine-rich protein of *Entamoeba histolytica* (Fig. 1b). The significance of the additional periodicities is not very clearly understood at present. No non-protein-coding sequences, such as those encoding rRNA, show a significant peak at frequency $f = 1/3$ (Fig. 1c). The program is available for distribution and can be accessed at <http://202.41.10.146/GS.html>.

From our analysis of a large number of sequences including those from parasites we do find a small number (<2%) of genes that do not display the usual pattern of a coding region. The majority of these genes encode ribosomal proteins or proteins that have repeated motifs. Several of these proteins are small (around 12 kDa or less), leading to greater noise in the Fourier spectrum of genes because of the statistical nature of the analysis. However, this may not be the only reason why some genes lack a three-base periodicity.

A few examples of GeneScan patterns of genomic sequences from different parasites are shown in Fig. 2. The GeneScan pattern generated matches very well with the annotated data. In addition to detecting the annotated regions, GeneScan can also identify potential coding regions not identified by other methods. There are 79 genes predicted in *Leishmania* chromosome 1 using the programs TESTCODE, CODONPREFERENCE and sequence similarity approaches (eg. using BLAST). GeneScan identified all the 79 potential genes, and in

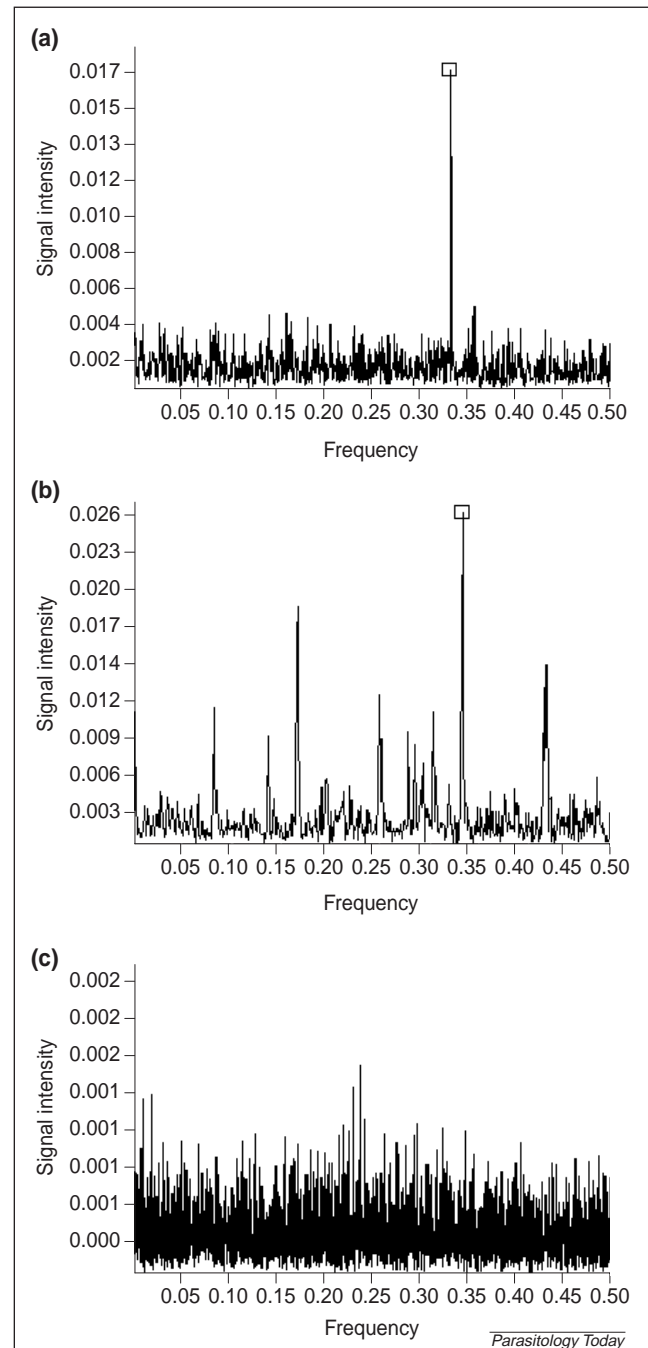


Fig. 1. Power spectra for the entire DNA sequence of two fairly typical coding sequences and a noncoding DNA sequence from the protozoan parasite *Entamoeba histolytica*. The actin gene (accession number M16338) shows a single peak at $f = 1/3$, with signal:noise ratio $P_N > 4$ (a). The spectrum of a serine-rich protein (M80910) shows multiple peaks with the largest peak at $f = 1/3$ and $P_N > 4$ (b). Ribosomal RNA sequence (X65163) shows a flat Fourier spectrum with no distinct periodicity ($P_N < 4$ at $f = 1/3$) (c).

addition, it predicted another 12 genes, all of which display open reading frames (ORFs) of 145–339 amino acids. The sense strand of only one of these 12 ORFs is in the opposite orientation to the neighboring genes, raising doubts about its transcription. It is likely that these genes are atypical (as compared with other *Leishmania* genes) in terms of codon usage, or they may lack consensus start and/or stop sites and other features normally displayed by other *Leishmania* genes and,

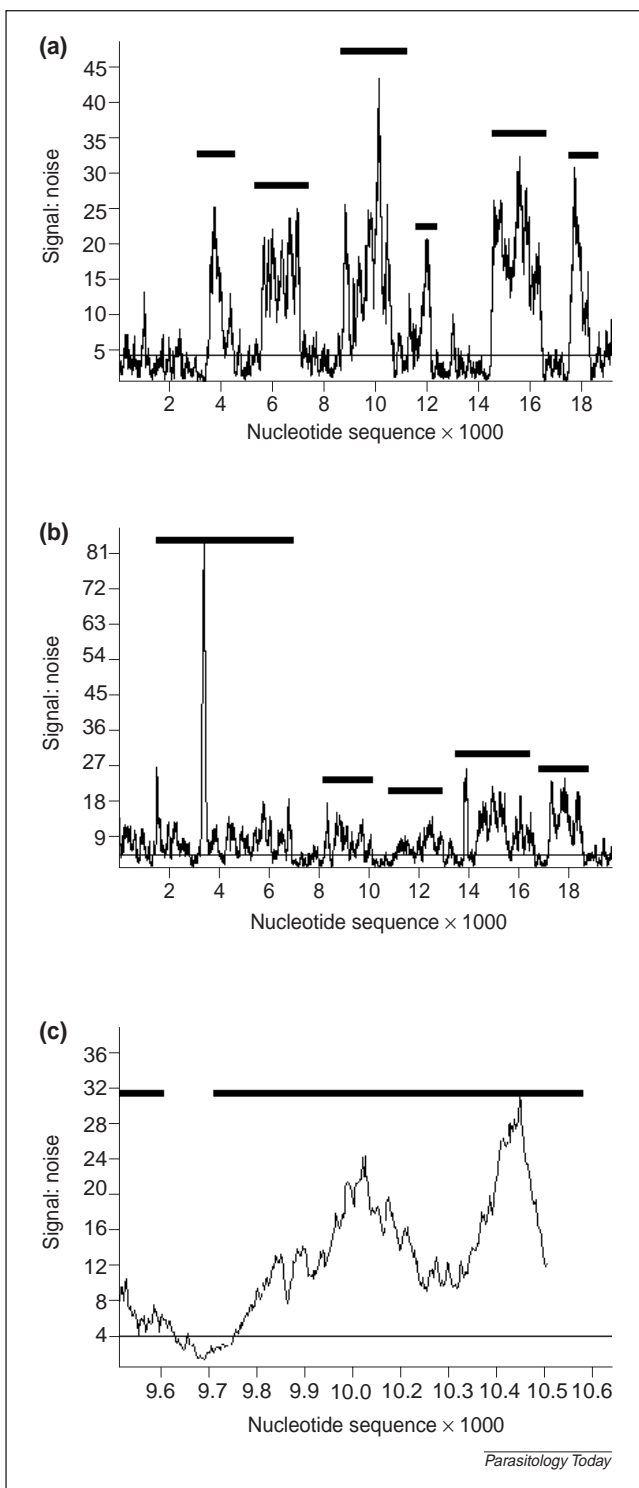


Fig. 2. GeneScan analysis of selected parasite genomic sequences. The window size and the cut-off value of P_N for this analysis was 200 and 4 respectively (see Ref. 8 for details). The profile of a segment of *Leishmania* chromosome I (position 1–20 000, accession number AE001274) is shown in (a). Locations of the predicted and/or experimentally observed genes as mentioned in the database are shown by solid lines. The GeneScan pattern shows a number of regions where P_N is above the cut-off value, namely 4, which is indicative of coding regions. Power spectra of predicted coding regions show a peak at $f = 1/3$ much greater than 4 (not shown here). GeneScan can identify all five genes predicted for the first 20 000 nucleotides of a cosmid of *Trypanosoma cruzi* (AF052832) (b). The spectrum of the genomic fragments encoding *Plasmodium falciparum* chromosome 2 gene PFB1040w (AF001433) identifies the intron known to span nucleotide positions 9580 to 9727 (the scan dips below the cut-off value of 4) (c).

thus, were not picked up in the original annotation. These putative genes also do not have any homologs in the databases, suggesting that they may be unique to the parasite. A table of these and other similar predicted genes in a number of genomes is available at <http://202.41.10.146/newgenes.html>.

Identification of parasitic genes containing introns may be more difficult as ORF analysis cannot be carried out, and there is insufficient information (especially for some parasites) available on consensus splice donor and acceptor sites necessary for prediction by other techniques. When present, introns in parasite genes are short (typically <100 nucleotides), and therefore, it is difficult to carry out many statistical calculations with a high degree of accuracy. Fourier-based methods may also not be very useful in identifying small introns as these are also based on statistical parameters and the noise level increases with decreasing window length. In general, the window length is kept at about 200 nucleotides or less while analyzing introns. The pattern obtained with the genomic sequence of a fragment of *P. falciparum* chromosome 1 shows the presence of an intron (Fig. 2c; the scan dips below $P_N = 4$ for a length of nucleotides), which is consistent with the annotation. Our results with a number of intron-containing genes of different protozoa show that about 60% of the introns can be reliably identified.

Conclusion

Genome projects of many parasites have been initiated and it is clear that rapid analysis of sequence information by computational means is essential for the success of these projects. Although there are a number of gene identification tools available, these require prior information for optimum accuracy and need to be modified for different organisms, and, as such, may not be directly applicable to parasite sequences. We describe a Fourier-based method GeneScan and show that it can be used for analysis of parasite genes. GeneScan, in its current form, has certain advantages: it is independent of GC content, it does not require prior information about the genomic structure of the organism and it can tolerate sequencing errors. Inevitably, it has a few limitations – mainly its inability to detect the boundaries of exons, high noise in small coding regions and absence of positive signals in a small group of proteins. Approaches that use results of analyses from a multiplicity of programs, and judiciously interpret this information, are likely to be most successful. GeneScan, being an *ab initio* method, can play a vital role in the first step of such an approach.

Acknowledgements

AJ acknowledges the award of a Junior Research fellowship by the University Grants Commission. We thank Shirish Tiwari, Colin Gros and Unitech Software for collaboration in the development of GeneScan. The authors acknowledge the help provided by the staff of the Bioinformatics Center, Jawaharlal Nehru University during the preparation of the manuscript.

References

- 1 Brownstein, M.J. *et al.* (1998) Functional genomics, in *Trends Guide to Bioinformatics* (Brenner, S. *et al.*, eds), pp 27–29, Elsevier Science
- 2 Karaolis, D.K. *et al.* (1998) A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strain. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3134–3139

- 3 Marra, M.A. *et al.* (1998) Expressed sequence tags-ESTblishing bridges between genomes. *Trends Genet.* 14, 4–7
- 4 Guigo, R. and Fickett, J.W. (1995) Distinctive sequence features in protein coding, genic non-coding, and intergenic human DNA. *J. Mol. Biol.* 253, 51–60
- 5 Haussler, D. (1998) Computational genefinding, in *Trends Guide to Bioinformatics* (Brenner, S. *et al.*, eds), pp 12–15, Elsevier Science
- 6 Voss, R.F. (1992) Evolution of long-range fractal correlations and $1/f$ noise in DNA base sequences. *Phys. Rev. Lett.* 68, 3805–3808
- 7 Ossadnik, S.M. *et al.* (1994) Correlation approach to identify coding regions in DNA sequences. *Biophys. J.* 67, 64–70
- 8 Tiwari, S. *et al.* (1997) Prediction of probable genes by Fourier analysis of genomic sequences. *Bioinformatics* 13, 262–270
- 9 Gish, W. and States, D.J. (1993) Identification of protein coding regions by database similarity search. *Nat. Genet.* 3, 266–272
- 10 Tettelin, H. *et al.* (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome VII. *Nature* 387, 81–84
- 11 Fraser, C.M. *et al.* (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397–403
- 12 Fleischmann, R.D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512
- 13 Borodovsky, M.Y. *et al.* (1986) Statistical patterns in primary structures of the functional regions of the genome in *Escherichia coli*. III. *Molek. Biol.* 120, 1390–1398
- 14 Lukashin, A.V. and Borodovsky, M. (1998) GeneMark.hmm: new solution for gene finding. *Nucleic Acids Res.* 26, 1107–1115
- 15 Salzberg, S.L. *et al.* (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 26, 544–548
- 16 Guigo, R. *et al.* (1992) Prediction of gene structure. *J. Mol. Biol.* 226, 141–157
- 17 Uberbacher, E.C. and Mural, R.J. (1991) Locating protein-coding regions in human DNA sequences by a multisensor neural network approach. *Proc. Natl. Acad. Sci. U. S. A.* 88, 11261–11265
- 18 Fickett, J.W. (1982) Recognition of protein coding regions in genomic DNA. *Nucleic Acids Res.* 10, 5303–5318
- 19 Solovyev, V.V. and Salamov, A.A. (1997) The Gene-Finder computer tools for analysis of human and model organisms genome sequences, in *Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology* (Rawling, C. *et al.*, eds), pp 294–302, AAAI Press
- 20 Dong, S. and Searls, D.B. (1991) Gene structure prediction by linguistic methods. *Genomics* 23, 540–551
- 21 Snyder, E.E. and Stormo, G.D. (1995) Identification of protein coding regions in genomic DNA. *J. Mol. Biol.* 248, 1–18
- 22 Burge, C. and Karlin, S. (1997) Prediction of complete gene structure in human genomic DNA. *J. Mol. Biol.* 268, 78–94
- 23 Burset, M. and Guigo, R. (1996) Evaluation of gene structure prediction programs. *Genomics* 34, 353–367

Letters

No NO Production During Human *Toxoplasma* Infection

I would like to comment on the recent article by Miller *et al.*¹. Certainly, there is now increased interest in the study of virulence among isolates of *Toxoplasma gondii*, which was, for some time, a neglected aspect of this protozoan infection. We have suggested hypotheses relating to this topic². Miller *et al.*¹ postulate that the 70 kDa heat shock protein (HSP70) is an innate factor of mice-virulent strains and that nitric oxide (NO) is an important inducer of this protein (which could be related to stage conversion of the parasite). Miller *et al.*¹ argue that mice are the ideal model host in which to study the immune system. However, there is strong evidence of species-specific differences between human and murine effector mechanisms against *Toxoplasma*.

It has been shown that NO does not have a significant role in human monocytes or macrophages³. Neither competitive inhibition of L-arginine metabolism (NMA) nor depletion of L-arginine (arginase) altered intracellular antimicrobial activity against *Toxoplasma gondii*, *Chlamydia psittaci*, or *Leishmania donovani* in humans. By contrast, NMA and arginase readily reversed the antimicrobial effect of mouse peritoneal macrophages stimulated either *in vitro* or *in vivo* by interferon (IFN)- γ . Activated mouse cells could also be induced to release enhanced levels of nitrite (an effect not seen in human cells). These results suggest that the

arginase-dependent generation of NO is a species-restricted macrophage mechanism, which is unlikely to participate in the intracellular antimicrobial activity of IFN- γ -stimulated human mononuclear phagocytes. Instead, IFN- γ readily induced normal monocyte-derived macrophages (MDM) to express indoleamine 2,3-dioxygenase (IDO) activity and stimulated MDM, alveolar macrophages, and oxidatively deficient chronic granulomatous disease MDM to degrade tryptophan⁴. All IFN- γ -activated, tryptophan-degrading macrophages killed or inhibited *T. gondii*, *C. psittaci*, and *L. donovani*. In addition, IFN- γ -treated mouse macrophages showed neither IDO activity nor tryptophan degradation, but killed *T. gondii* and *L. donovani*. These results suggest that tryptophan depletion contributes to the oxygen-independent antimicrobial effects of the activated human macrophage.

While NO cannot be activating HSP70 in the human immune response to *T. gondii*, tryptophan depletion would produce enough 'respiratory' stress to induce stage conversion. This merits further study.

Finally, I have obtained evidence of the existence of NO production not only in the host cell, but also in the parasite (Gomez-Marín, J.E., PhD thesis, Université de Reims, 1997). *Toxoplasma*, like other protozoans such as *Tetrahymena*⁵, *Trypanosoma*⁶ and *Plasmodium*⁷, has its own

constitutive NO synthase (producing 2–6 μM of nitrites), which could be essential in intracellular signaling. The NO defensive mechanism, where levels of nitrites can reach 120 μM or more, is possibly too toxic for humans, who live for many years, but survivable for mice, which live for only a few months.

References

- 1 Miller, C.M.D. *et al.* (1999) Cytokines, nitric oxide, heat-shock proteins and virulence in *Toxoplasma*. *Parasitol. Today* 15, 418–422
- 2 Gómez-Marín, J.E. *et al.* (1997) Does human toxoplasmosis involve an imbalance in T1/T2 cytokines? *Medical Hypotheses*, 48, 161–169
- 3 Murray, H.W. and Teitelbaum, R.F. (1992) L-arginine-dependent reactive nitrogen intermediates and the antimicrobial effect of activated human mononuclear phagocytes. *J. Infect. Dis.* 165, 513–517
- 4 Murray, H.W. *et al.* (1989) Role of tryptophan degradation in respiratory burst-independent antimicrobial activity of γ interferon-stimulated human macrophages. *Infect. Immun.* 57, 845–849
- 5 Christensen, S.T. (1996) Cell death, survival and proliferation in *Tetrahymena thermophila*. Effects of insulin, sodium nitroprusside, 8-Bromo cyclic GMP, NG-methyl-L-arginine and methylene blue. *Cell Bio. Int.* 10, 653–666
- 6 Paveto, C. *et al.* (1995) The nitric oxide transduction pathway in *Trypanosoma cruzi*. *J. Biol. Chem.* 270, 16576–16579
- 7 Ghigo, D. *et al.* (1995) Erythrocyte stages of *Plasmodium falciparum* exhibit a high nitric oxide synthase (NOS) activity and release an NOS-inducing soluble factor. *J. Exp. Med.* 182, 677–688

Jorge E. Gómez-Marín

Unidad de Parasitología, Facultad de Medicina
Universidad Nacional de Colombia
Santafé de Bogotá DC
Colombia