Current Millennium Biotechniques for Biomedical Research on Parasites and Host-parasite Interactions

Antonio RL Teixeira[+], Augusto Simões-Barbosa, Eric Faudry, Silene P Lozzi, Enrique R Argañaraz, Marian D'Souza-Ault, Jaime M Santana

Laboratório de Pesquisa Multidisciplinar da Doença de Chagas, Faculdade de Medicina, Universidade de Brasília, Caixa Postal 04536, 70919-970 Brasília, DF, Brasil

The development of biotechnology in the last three decades has generated the feeling that the newest scientific achievements will deliver high standard quality of life through abundance of food and means for successfully combating diseases. Where the new biotechnologies give access to genetic information, there is a common belief that physiological and pathological processes result from subtle modifications of gene expression. Trustfully, modern genetics has produced genetic maps, physical maps and complete nucleotide sequences from 141 viruses, 51 organelles, two eubacteria, one archeon and one eukaryote (Saccharomices cerevisiae)*. In addition, during the Centennial Commemoration of the Oswaldo Cruz Institute the nearly complete human genome map was proudly announced, whereas the latest Brazilian key stone contribution to science was the publication of the* Shillela fastidiosa *genomic sequence highlythed on a Nature cover issue. There exists a belief among the populace that further scientific accomplishments will rapidly lead to new drugs and methodological approaches to cure genetic diseases and other incurable ailments. Yet, much evidence has been accumulated, showing that a large information gap exists between the knowledge of genome sequence and our knowledge of genome function. Now that many genome maps are available, people wish to know what are we going to do with them. Certainly, all these scientific accomplishments will shed light on many more secrets of life. Nevertheless, parsimony in the weekly announcements of promising scientific achievements is necessary. We also need many more creative experimental biologists to discover new, as yet un-envisaged biotechnological approaches, and the basic resource needed for carrying out mile stone research necessary for leading us to that "promised land" often proclaimed by the mass media.*

Key words: genomics - proteomics - transcriptomics - Chagas disease - leishmaniasis - post-genomics

CAN THE LATEST SCIENTIFIC ACHIEVEMENTS IMPROVE QUALITY OF LIFE OF PEOPLE SUFFERING FROM INFECTIOUS DISEASES?

We all agree that recent advances in biomedical technology and genomics have opened up exciting possibilities for the development of vaccines, chemo- and immuno-therapeutics, and new tools for control of transmission of major tropical diseases. Also, it has been considered that infectious diseases represent a tremendous public health burden and continual threat to people living in tropical and subtropical regions of the world. However, the economists in the administration of public or private institutions have curtailed financial support to scientific research in infectious diseases, highly prevalent in the tropics arguing that those diseases

will be wiped out solely by social development. This sort of pragmatic econometrics gained much influence among policy makers, regardless of a lack of an ethical consideration of the needs of those individuals already affected by infectious diseases or at continuous risk of contracting the infections. Furthermore, recent annual reports from international organizations showing certain deteriorating healthy condition in the tropical regions of the world strongly advise policy makers that unhealthy life conditions preclude optimal individual productivity and socioeconomic development.

In this respect, an international organization report reminds the reader that 20% of the world's population carries at least one helminthic parasite infection, and that 10% are infected with those parasites causing malaria, trypanosomiasis, leishmaniasis, schistosomiasis or filariasis. Greater than 70% of deaths in children less than five years of age and greater than 40% of deaths in developing countries are due to infectious and parasitic diseases. Enteric infections, caused by viruses, bacteria and protozoa, including mycobacterial infections such as leprosy, are responsible for immea-

surable global morbidity and mortality. Childhood mortality rates in tropical developing countries are five times higher than those in industrialized countries. Furthermore, the morbidity and mortality of several tropical diseases in endemic areas are increasing, due to changing ecological patterns leading to an increase in incidence and prevalence (e.g., malaria in Amazônia), geographic spread (e.g., dengue) in the Americas, drug resistance, and/or failure of currently available control measures (Teixeira et al. 2000).

What can the latest genome sequence information do for fostering scientific research that would lead to more effective treatment, prevention and control of tropical infectious diseases? The answer to this question addresses the reader for countless highly significant examples showing how have scientific knowledge and technologies improved quality of life in the tropics? Basically, every tropical disease cited in the above paragraph would serve for showing the important practical benefits of past and current research. In very pragmatic language, the money and human lives saved by fostering public health in Brazil with the aid of scientific knowledge generated in our national institutions certainly exceed by several folds the investment made. Nonetheless, new pathogens have been discovered and the increased biological diversity of several pathogenic species has been elucidated in correlation with increasing microbe drug resistance, virulence and/or pathogenicity. These threatening factors represent a clear need for sustained research support in biomedical sciences and require skillful introduction of the latest scientific achievements aiming at a better quality of life in the next decades. In the following paragraphs we will depict the biological research needed to be carried out in parallel with the latest scientific achievements, in order to develop new tools for treatment, prevention and control of infectious tropical diseases.

INFORMATION FROM GENE MAPPING AND SEQUENCING

A long list of review papers on gene mapping and sequencing advances is available in the literature (Schuler et al. 1996, Weinstein 1998, O'Brien et al. 1999, Evans & Relling 1999, Wolf & Matzke 1999, Brent 2000) with information from sequencing projects for virus, bacteria, yeast, flies, mice, and humans. Unfortunately, not much is known about the genomes of microorganisms causing tropical diseases, and, therefore, most of the discussion in the following paragraphs will be based in inferences from the various genomes from which complete or almost complete sequencing have been accomplished, particularly the human genome (Schuler et al. 1996).

The word genomics was selected to designate the study of information developed by the generation and analysis of genes and genomes through systematic DNA sequencing projects, and very recently the term functional genomics has been used to designate all information about what genes do, with the understanding that the information was produced systematically. The current literature includes the terms: proteomics and transcriptomics, where "omics" means all the work systematically produced on an industrial scale. Below we will describe some early findings that led to the rise of the importance of genomics and systematic gene sequence mapping in modern biological sciences, which are equally important for the studies of infectious diseases and host-parasite relationships.

DNA polymorphism and linkage information - Previous research discoveries leading to the concept that genomics could be of major added value stemmed from the finding that DNA polymorphism (RFLPs: restriction fragment linkage polymorphisms, or SNPS: single nucleotide polymorphisms) could be associated to the Mendelian rules of inheritance. This discovery is of great importance for the biological sciences (Botstein et al. 1980). Such a correlation led to development of markers linked to inherited traits (alleles) for isolation of genes responsible for a genetic disease. The isolation of the genes responsible for Huntington's chorea (Gusella et al. 1983) was the first breakthrough in modern genetics and this observation generated the human medical genetics strategy for isolation of individual disease genes (Fraser et al. 1998, Winzler et al. 1999).

Genome organization - The genome is organized by pairing of four bases: CG and AT, namely guanine, cytosine, adenine and thymine. The systematic sequencing of these structures has revealed an array of features that have been taken as some kind of previous selection rather than evolutionary accident, regardless that the functional significance of a sequence is sometimes unclear. The two examples demonstrate this point. The short chromosomes of *Saccharomices cerevisiae* show higher recombination rates than larger ones, and ORF's (opened reading frames) with higher than average CG content at silent positions tend to be on those smaller chromosomes (Bradnam et al. 1999). On the other hand, AT-rich areas near centromeres show lower rates of meiotic recombination than the rest of the genome (Goffeau et al. 1996). Brent (2000) considers these finding "all fascinating, but so far uninterpreted". However, the *Treponema pallidum* complete genome sequence and mapping has suggested the information might help the formulation of a medium for growth of the agent of syphilis (Fraser et al. 1998).

Genome sequence and inferred protein - Encoded effector molecules are complement of sequences from genomic DNA, expressed sequence tags (EST) and full-length cDNA. This information generated the belief that having more sequences lead to more knowledge of a gene's function, because if that information were reported for one organism, similar sequences from other organisms may have similar gene functions. Truthfully, we lack knowledge of the function of genomic protein complements. This assertion is substantiated by the fact that most sequences are of bacterial genomes and most of those refer to microbial metabolism than to other aspects of biology (Brent 2000). Also, the depicted frameworks of biological function revealed by coding sequences are evocative but not often conclusive. Between 15 and 40% of the proteins encoded by any genome do not have a function readily.

Gene regulation - Automatic sequencing projects have made possible the large-scale identification of sites of regulatory protein action. Nucleotides conserved in noncoding regions between some pairs of organisms allow identification of functional sites. However, a comprehensive understanding of how those sites work together with proteins to regulate gene expression awaits further technological development (Venkatesh et al. 1997).

Phylogeny and evolution - It has been speculated that vertebrate genomes may represent a quadruplication of an ancestral metazoan genome that gave rise to helminths and flies (Spring 1997). An attribute of organismic DNA sequences has been considered to demonstrate that some changes have led to speciation and existing phylogeny. Therefore, analysis of DNA sequences has led to readjustments in phylogenies once based solely on morphology or paleontological information. These observations made possible the use of spliceosomal introns as markers to define phylogenies (Barbujani 1997, Jin et al. 1999). In addition, DNA sequences demonstrate numerous instances of horizontal gene transfer among different lower organisms species (Teixeira et al. 1991, 1994, Argañaraz 1996, Jain et al. 1999, Simões-Barbosa et al. 1999). Altogether, this information may lead to a much more complete understanding of the natural history of living species, as we incorporate knowledge from molecular evolution.

Genetic polymorphisms - The latest advancements in systematically generated DNA information will foster population geneticists' research towards determining the full array of polymorphisms within species. In fact, polymorphism data have already allowed reconstruction of the migration of different human populations during prehistory (Jin et al. 1999) and may shed light on other population-related anthropological issues (Owens & King 1988). This area of scientific advancement combined with very powerful statistical methods (Pearson & Lipman 1988, Autschul et al. 1990) may give important information about genetics of susceptibility and resistance to infectious diseases. Such experimental biological research based in systematic DNA sequencing will represent an important basis for constructing "predictive medicine" in this century.

Comparative genome hybridization - This biotechnique can be performed using an ordered array of DNA from various sources at different intervals along the genome. The data are obtained by comparing hybridization patterns of differentially labeled DNA sequences taken from different stages of cell cycle/organization. The application of this technique in various disease conditions will shed light on the genomic differences between healthy and sick cells, and apoptosis (Kallioniemi et al. 1992, Lockhart & Winzler 2000).

RNA genomics - The level of expression of large populations of mRNAs can be detected by hybridizing to surface immobilized arrays of nucleic acids. Lately, the use of non-proprietary arrays from polymerase chain reaction (PCR) products were developed, which greatly increased the speed of researching gene expression and producing important results (Liang & Pardee 1992, DeRisi et al. 1997). These studies have suggested new regulatory relationships between transcription factors and transduction pathways. A large number of techniques can define different functional mRNA classes and current knowledge on cell biology will be greatly increased from gene expression monitoring.

Protein genomics - O'Farrel (1975) produced the first large-scale surveys using 2-dimensional gels for separation of a complex mixture of proteins by isoelectric focusing and differences in molecular weight. The protein separated in an individual spot (Lozzi 1998, Rabiloud et al. 1998, Simões-Barbosa et al. 2000) can be identified by mass spectrometric determination of molecular weight and sequence (Mann et al. 1993). The combination of mass spectrometric proteolytic peptide mapping and genome database searches provides a rapid, sensitive, and potentially accurate means for identifying proteins (Li et al. 1999, Wilkins et al. 1999). Database search algorithms detect the matching between proteolytic peptide masses from a mass spectrometry map and theoretical proteolytic peptide mass of the proteins in a genome database (Eriksson et al. 2000). The number of sample masses that matches is used to compute a score for each protein, and the protein that yields the best score is assumed as the identification re-

sult. These studies have been complemented by insertional mutagenesis experiments aiming at detection of subcellular localization of a significant portion of the proteins expressed in an organism. The tagged proteins are localized by immunofluorescent light microscopy. Furthermore, interactions among binary protein products (Fields & Song 1999) have been tested by two-hybrid experiments to determine the equilibrium dissociation constant of the binary interactions. The data are now stored in searchable databases of various types.

BIOTECHNOLOGICAL APPROACHES USED FOR INFECTIOUS DISEASE RESEARCH

Based on the description of the most relevant biotechnological applications and on various examples of functional genome research, some of which were discussed above, scientists have conducted research on infectious diseases affecting people in the tropical region of the world, aiming at disclosure of host-parasite interactions and pathogenetic mechanisms associated with tissue lesions. We will briefly describe how such advances are currently used in biomedical and ecological research as illustrated by Chagas disease and leishmaniases.

Proteomics - Currently, comparative proteomes of amastigote and trypomastigote forms of *Trypanosoma cruzi* reveals state specific proteins related to survival of the parasite in its intracellular stage. The differential analysis of gene expression showed that *T. cruzi* gene expression is controlled mostly at post-transcriptional level (Teixeira 1998).

High resolution 2-dimensional electrophoresis followed by mass spectrometry has been used (Lozzi et al. 1998, Faudry et al. 1999, Lozzi 1999) to identify subunits of apyrases in salivary glands of triatomines, which are important for the successful feeding and transmission of the *T. cruzi* infection. The spots of individual proteins in the nylon membrane were subjected to tryptic digestion for mass spectrometric analysis. This study showed that a polypeptidic oligomer encloses the apyrase activity in *Triatoma infestans* salivary gland secretion. Also, we have recently utilized 2-dimensional electrophoresis for identifying differentially expressed proteins in macrophages that were transfected with a sequence of minicircle of kDNA from *T. cruzi* (Simões-Barbosa et al. 1999, 2000). This study revealed marked differences in gene expression among uninfected and kDNA transfected macrophages.

Protein structure determination - Advances in genomics research allowed scientists to rapidly obtain protein sequences. Protein structure determination is considered important for elucidation of protein functions related to key biological activities.

Furthermore, genomics research has suggested the targets for recombinant technology approaches aiming at production of proteins of interest. However, further determination of the structure of recombinant or non-recombinant proteins is fundamental for engineering and designing drugs required for treatment of infectious diseases, and for improvement of current therapeutic drugs available.

When the protein is abundant in a pure form, its structure can be determined experimentally by X-ray crystallography or nuclear magnetic resonance. In the case of the apyrase of *T. infestans* we have undertaken experiments for obtaining the recombinant polypeptides for subsequent studies (Naiva et al. 1989). After obtaining the recombinant proteins we shall determine the structures (Weber et al. 1989). The calculation of minimum free energy from different adopted polypeptide conformations might predict protein folding by computational and theoretical physics. However, this prediction *ab initio* may not be achieved due to the complexity of the protein folding and, therefore, the protein structure will be determined by comparative studies of homologous structure, refined with mathematical tools (Hunter 1997). The *T. infestans* sequence will be first aligned with sequences of homologous proteins deposited in a database, and a 3-dimensional model might be proposed by fitness of compared structures (Langen et al. 1997, Edgard et al. 2000). Secondary and tertiary structures might be predicted also by an existing database (Eriksson et al. 2000). Several algorithms might be used for statistical calculation of the ability of each residue to form a specific secondary structure, or it can be inferred by comparison with portions of known proteins. Energy molecular dynamics calculations based on minimum free energy conformation are then used to refine the theoretical model. Experimental refinement can be further achieved by site-directed mutagenesis in that sequence related to the protein function (DeRisi et al. 1997, Eriksson et al. 2000).

Application of protein structure determination - Knowledge of protein structure is required for understanding function and interaction of protein with ligands. Our study has led to investigations on modifications of the protein, its ligands, or both, that could modify its function. Determination of protein structure is considered a major step towards scientific breakthroughs in major diseases: drug design can be targeted to synthesis of new inhibitors or agonists while protein engineering fosters construction of more stable proteins with increased activity. In this research area, docking algorithms have been used to study interactions between macromolecules (protein-protein or protein-DNA), but these studies are currently applied principally for

examining interactions between small molecules-macromolecules, such as in drug design. For example, protein structure modeling and docking was used to design parasite protease inhibitors (Quemeneur et al. 1998). These types of studies require international collaboration of research workers in the cutting edge of protein biochemistry.

RNA genomics - Differential display of mRNA - The biotechniques for manipulation of genes and their products have given us important information for understanding the phenotype derived from gene regulatory and expression mechanisms. The cDNA subtraction techniques are based in the knowledge that differentially expressed transcript will not hybridize, thus maintaining a single strand form. In the first version of the method, single and double DNA/RNA strand can be separated by affinity chromatography in hydroxiapatite columns. This technique has not been used because it has low sensitivity for detection of transcripts of low representation. This was followed by the development of the DNA microarray technique discussed above, and other PCR-based techniques for analyzing gene expression were quickly developed. These techniques require lower quantities of mRNA and show high sensitivity. Two different adapters are added to the tester cDNA, which are hybridized with excess of the driver cDNA. A second hybridization performed among the products from the first hybridization experiment might yield differential transcripts possessing the two different adapters, which are exponentially amplified by PCR. A pan-like structure suppresses PCR when the template shows same adapter at both ends (suppression PCR). A variant to this suppression consists of adding oligo-(dT) 30-latex particles or biotin/streptavidin affinity to separate subtracted cDNAs.

The technique of differential display of mRNA (DD), which is based in RT-PCR (Liang & Pardee 1992, Liang et al. 1994, Zhang et al. 1996) has allowed us to analyze hundreds of transcripts in a single experiment. The DD technique utilizes an oligo dT anchored primer in the reverse transcription reaction. One-base anchored primer can secure three possibilities: oligo dTA, oligo dTC and oligo dTG. Therefore, one third of the mRNA population is retrotranscribed at the first base immediately before the poly A tail. Combining this cDNA with the same one-base oligo dT anchored primer plus an arbitrary primer of 13 mer results in PCR amplification of several transcripts, which can be analyzed on a sequencing gel. Many different combinations of anchored primers and arbitrary primers can be used to screen thousands of transcripts.

Consistently, we have observed a high rate of false positives, which is the main problem with the DD technique. However, being aware that the DD provides the 3' end of the transcript (usually the non-coding region), an entire transcript can be searched in a cDNA library using the RT-PCR 5' RACE. Normalization of cDNA libraries increases the discriminating power for the identification of differentially expressed genes and allows identification of poorly expressed sequences in microorganisms causing many tropical diseases (Soares et al. 1994, Bonaldo et al. 1996, Simões-Barbosa 2000).

Gene sequencing and post-genomics analysis - Biological research is being dramatically improved by two factors: the massive increase in the amount of DNA sequence information and the development of new technologies to study its signification. New types of experiments are possible and will help us to understand how cells and biological systems work. At a time that many organisms have had their genomes completely sequenced, including much of the human one, interdisciplinary research is necessary to understand what those genes do and how they interact among themselves and with the environment, what goes wrong in disease, how parasites and hosts interact and how to develop a drug. Postgenomics analysis will revolutionize the way biology is studied. The goal of functional genomics is to understand how molecular components work together to comprise functioning biological systems.

The current trend for automatic production of genome sequences has generated an incredibly large amount of data. An almost equally large amount of these data and information require post-genomics analysis by experienced experimental biologists. Information on genomics is being used experimentally with new technologies such as systematic analysis of protein-protein interactions, systematic gene knockouts, double-strand RNA, genome-wide messenger RNA expression profiling by using oligonucleotide chips, serial analysis of gene expression, green fluorescent protein, and DNA microarrays (Hieter et al. 1999, Lockhart & Winzler 2000). However, the new technologies are not enough. It is necessary to change the way biologists think.

Examplifying such analytical approach, the complete genome mapping of *Borrelia burgdorferi*, a spirochete causing Lyme disease, a tick-borne zoonosis, allowed identification of putative gene sequences for proteases that may play a role in the metabolism, growth and parasitic-host infectivity. In our laboratory, we are employing an in-gel activity-mass spectrometry approach to identify protease gene function in this spirochete. Then,

using both natural and fluorogenic substrates Santana and co-workers identified some proteolytic activities, and gel enzymography using gelatin co-polymerized with acrylamide revealed two major proteolytic activities (unpublished data). Furthermore, inhibition patterns of both proteases showed that they belong to the serine-protease class. Maximal specific activities were found at neutral pH. On-going research will require purification of the proteases from the spirochete cell free extracts by a combination of ion exchange and hydrophobic chromatographic procedures using FPLC and HPLC. The purified enzymes will be submitted to mass spectroscopy analyses to determine their fingerprinting and peptide sequence-tags. Secondly, taking advantage that the *B. burgdoferi* genome is known (Fraser et al. 1997), the genes encoding each protease will be amplified by PCR, cloned and expressed. Thereafter, the recombinant proteases will be subjected to protein structure analyses and enzymatic activity characterization aiming at developing specific inhibitors both by drug design and combinatorial chemistry techniques.

Biological and molecular characterizations of kinetoplastid flagellates causing cutaneous and visceral leishmaniasis and Chagas disease - Non-pathogenic and pathogenic *Leishmania* spp. may be undistinguishable from each other and *cruzi*-like protozoa are morphologically undistinguishable from the virulent trypanosomatids (Teixeira 1987). In addition to knowing the extent of biodiversity in the taxa, and that these protozoan flagellates are phylogenetically related, these parasites share constitutional genes and expressed proteins and antigenic determinants. These are strong reasons for pursuing phylogenetic classifications (Fernandes et al. 1998, 1999, Briones et al. 1999), phenotyping and genotyping of these infectious agents. Therefore, isoenzymic analysis of freshly isolated kinetoplastid flagellates has detected diverse isoenzymic profiles. Also, template DNA from parasitic forms isolated from vertebrate and invertebrate hosts can be amplified by PCR with primers derived from DNA sequences of archetype protozoan flagellates (Teixeira et al. 2000). Furthermore, DNA sequences from known kinetoplastid flagellates have been used as a probe for *in situ* hybridization characterization of wild flagellates. The positive hybridization further shows the specificity of the PCR amplification products used for genotyping protozoan flagellates isolated from vertebrate and invertebrate hosts.

The advances in molecular biology also allows separation between virulent, pathogenic kinetoplastid protozoa belonging to the subgenus *Schizotrypanum*, which have been divided in sylvatic and domestic groups with basis on phyloge-netic classification. This biotechnique consists in using primers either for the mini-exon gene or for the rDNA sequence of kinetoplastids (Souto et al. 1996). In addition to phylogenetic classification of different protozoan flagellate isolates, this biotechnique shows further molecular differences between the non-pathogenic flagellates (*T. rangeli, T. hastatus* and other *cruzi*-like flagellates) and the virulent *T. cruzi* and *Leishmania* spp.

CONCLUDING REMARKS

The description of the most relevant biotechnological applications in parallel with systematic genomic sequencing and functional genome research has made possible the use of the scientific knowledge for the study of parasites and host-parasites interactions, for infectious diseases highly prevalent in the tropics. In the case of Chagas disease, the 2-dimensional associated mass spectrometry allowed us to identify oligomeric forms of apyrases in salivary glands of triatomines, which are important for successful feeding and transmission of *T. cruzi*. Post-genomic analysis of sequence mapping of *B. burgdoferi* allowed identification of specific proteolytic activities in the cell-free extract of this infectious agent. Comparative proteomics of amastigote and trypomastigote forms of *T. cruzi* under investigation has revealed some stage-specific proteins related to the survival of the parasite in the cytoplasm of the host cell. Furthermore, differential analysis of gene expression in trypanosomatids revealed that these protozoa have their gene expression mostly regulated at the post-transcriptional level.

We have used 2-dimension electrophoresis for identification of proteins differentially expressed in macrophages that were transfected with minicircle kDNA sequence form *T. cruzi* as a result of natural infection. Also, the differential display of mRNA was used for differential analysis of gene expression among uninfected and *T. cruzi*-kDNA transfected macrophages. The results from these experiments have shown that, as a consequence of integration of kDNA minicircle sequence into the genome of the host-cell alteration of gene expression can occur. The analysis of thousands of transcripts from the control (uninfected) and from kDNA sequence transfected macrophages resulted in the characterization of genes that were over and under expressed. Further genomic analysis has shown that blocking of the gene expression resulted from integration of a chimeric insertion sequence. Segment of *T. cruzi* kDNA minicircle sequence covalently linked into a host cell LINE-1 transposable element was associated with gene shut off. We believe that these findings may correlate with pathogenic mechanisms for diseases showing variable patterns of clini-

cal manifestations. Many more biotechniques and experimental biological research are required for unraveling the molecular bases of autoimmune diseases such as Chagas disease. Moreover, biomedical science needs be geared towards interdisciplinary approaches, and re-writing the academic curricula to strengthen the use of mathematics and computer science.

In conclusion, we described previous review articles depicting the front wave of experimental biological research carried out in parallel with systematically produced data and information in genomics. We believe that the potential impact of this increase in knowledge and capability during this millennium depends on further development of new biotechniques, in parallel with statistics and computer-based computation methods. Future genomic biology will probably spur development of computer based analytical frameworks.

We believe that in this 21st century, genomic discoveries will lead to meaningful knowledge necessary for improving medical practice and public health worldwide. This very optimistic appraisal of scientific production in developed countries, however, is far from receiving significant contributions from scientists in the field of tropical diseases. Currently, most scientists in underdeveloped countries in the tropics worry about whether or not the water supply will be cut or rationed, or whether the electric power cuts will destroy their biological stocks. These basic issues must be addressed in order for there to be an equal opportunity for Brazilian scientists to fully compete for ever more scarce research funding. This is an important issue for discussion with policy makers in the current Brazilian system for science and technology.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990. Basic local alignment search tool. *J Mol Biol 215*: 403-410.

Argañaraz ER 1996. *Integração de Seqüências de Minicírculo de kDNA de Trypanosoma cruzi em Elemento L1 no Genoma da Célula Hospedeira,* PhD Thesis, Universidade de Brasília, 96 pp.

Barbujani G, Gagagni A, Minch E, Cavallier-Sforza LC 1997. An apportionment of human DNA diversity. *Proc Natl Acd Sci USA 94*: 4516-4519.

Bonaldo MF, Lennon G, Soares MB 1996. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res 6*: 191-806.

Botstein D, White RL, Skolnick M, Davis RW 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet 32*: 314-331.

Bradmam KR, Seoighe C, Sharp PM, Wolfe KH 1999. G + C content variation along and among *Saccharomyces cerevisiae* chromosomes. *Mol Biol Evol 16*: 666-675.

Brent R 2000. Genomic biology. *Cell 100*: 169-183.

Briones MR, Souto RP, Stolf BS, Zingales B 1999. The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. *Mol Biochem Parasitol 104*: 219-232.

DeRisi JL, Iyer VR, Brown PO 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science 278*: 680-686.

Edgar PF, Douglas JE, Cooper GJS, Dean B, Kydd R, Faull RLM 2000. Comparative proteome analysis of the hippocampus implicates chromosome 6q in schizophrenia. *Moll Psychiatry 5*: 85-90.

Eriksson J, Chat BT, Fenyo D 2000. A statistical basis for testing the significance of mass spectrometric protein identification results. *Analythical Chem 72*: 999-1005.

Evans WE, Relling MV 1999. Pharmacogenomics: Translating functional genomics into rational therapeutics. *Science 286*: 487-491.

Faudry E, Lozzi SP, Santana JM, Sousa MV, Ricart CAO, Rocha PS, Macedo TC, Teixeira ARL 1999. *Triatoma infestans* saliva apyrase is a 480 kDa oligomer. Simpósio Internacional Doença de Chagas, Rio de Janeiro.

Fernandes O, Catanho MP, Segura I, Labrada LA, Derré R, Saravia N, Degrave W 1999. Minicircle variable probes for characterization of *Leishmania (Viannia)* species. *J Parasitol 85*: 563-568.

Fernandes O, Sturm NR, Derré R, Campbell DA 1998. The mni-exon gene: a genetic marker for zymodeme III of *Trypanosoma cruzi*. *Mol Biochem Parasitol 95*: 129-133.

Fields S, Song O 1999. A novel genetic system to detect protein-protein interactions. *Nature 340*: 245-246.

Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, Owen W, Ketchum KA, Dodson R, Hickey EK 1997. Genomic sequence of a Lime disease spirochaete, *Borrelia burgdorferi*. *Nature 390*: 580-586.

Fraser CM, Norris SJ, Weinstock GM, White O, Granger G, Sutton RD, Gwinn M, Hickey EK, Cleyton R, Ketchum KA 1998. Complete genome sequence of *Treponema pallidum*, the syphylis spirochete. *Science 281*: 375-387.

Goffeau A, Barrel BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Hohnston M 1996. Life with 6000 genes. *Science 274*: 546-563.

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature 306*: 234-238.

Hieter P 1999. Functional genomics: what do yeast proteins do? *Nature 402*: 362-363.

Hunter WN 1997. A structure-based approach to drug discovery; crystallography and implications for the development of antiparasite drugs. *Parasitology 114*: 17-29.

Jain R, Rivera MC, Lake JL 1999. Horizontal gene transfer among genomes: the complexity, hypothesis. *Proc Natl Acd Sci USA 96*: 3801-3806.

Jin L, Underhill PA, Doctor V, Davis RW, Shen P, Cavalli-Sforza LL, Oefner PJ 1999. Distribution of haplotypes from a chromosome 21 region distinguishes multiples prehistoric human migrations. *Proc Natl Acd Sci USA 96*: 3796-3800.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science 258*: 818-821.

Langen H, Gray C, Roder D, Juranville JF, Takacs B, Fountoulakis M 1997. From genome to proteome: protein map of *Haemophylus influenzae. Electrophoresis 18*: 1184-1192.

Li XP, Pleissner KP, Regitz-Zagrosek V, Jungblut PR 1999. A two-dimensional electrophoresis database of rat heart proteins. *Electrophoresis 20*: 891-897.

Liang P, Pardee AB 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science 257*: 967-971.

Liang P, Zhu W, Zhang X, Guo Z, O´Connell RP, Averboukh L, Wang F, Pardee AB 1994. Differential display using one-base anchored oligo-dT primers. *Nucl Ac Res 22*: 5763-5764.

Lockhart DJ, Winzler EA 2000. Genomics, gene expression and DNA arrays. *Nature 405*: 827-836.

Lozzi SP 1998. *Caracterização Bioquímica e Funcional de Proteínas com Atividade Apirásica na Saliva de T. infestans,* PhD Thesis, Universidade de Brasília, 118 pp.

Lozzi SP, Faudry EC, Nunes CA, Santana JM, Felix CR, Gréllier P, Ricart CAO, Sousa MV, Teixeira ARL 1998. Detection, purification and mass spectrometry identification of the apyrase of *Triatoma infestans. Mem Inst Oswaldo Cruz 93* (Suppl. 1).

Mann M, Hojrup P, Roepstorff P 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrometry 22*: 338-345.

Naiva MA, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature 337*: 615-620.

O' Brien SJ, Menotti-Raymond M, Murphy WJ, Nash WG, Wienberg J, Stanyon R, Copeland NG, Jenkins NA, Womack JE, Graves JAM 1999. The promise of comparative genomics. *Science 286*: 458-481.

O'Farrel PH 1975. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem 250*: 4007-4021.

Owens K, King MC 1988. Genomic views of human history. *Science 286*: 451-453.

Pearson WR, Lipman DJ 1988. Improved tools for biological sequence comparison. *Proc Natl Acd Sci Usa 85*: 2444-2448.

Quemeneur E, Moutier M, Charbonnier JB, Menez A 1998. Engineering cyclophilin into a proline specific endopeptidase. *Nature 391*: 301-304.

Rabilloud T, Kieffer S, Procaccio V, Louwagie M, Courchesne PL, Patterson SD, Martinez P, Garin J, Lunardi J 1998. Two dimensional electrophoresis of human placental mitochondria and protein identification by mass spectrometry. Toward a human mitochondrial proteome. *Electrophoresis 19*: 1006-1014.

Schuler GD, Bogushi MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek E 1996. A gene map of the human genome. *Science 274*: 540-567.

Simões-Barbosa AM 2000. *Transferência Horizontal de Seqüência de Minicírculo de kDNA de Trypanosoma cruzi para Transposon LINE-1 e Alteração da Expressão do Gene p9 na Célula Hospedeira*, PhD Thesis, Universidade de Brasília.

Simões Barbosa A, Barros AM, Nitz N, Arganaraz ER, Teixeira ARL 1999. Integration of *Trypanosoma cruzi* kDNA minicircle sequence in the host genome may be associated with autoimmune serum factors in Chagas disease patients. *Mem Inst Oswaldo Cruz 94*: 249-252.

Simões-Barbosa A, Santana JM, Teixeira ARL 2000. Solubilization of delipidated macrophage membrane proteins for analysis by two-dimensional electrophoresis. *Electrophoresis 21*: 641-644.

Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L, Efstratiadis A 1994. Construction and characterization of a normalized cDNA library. *Proc Natl Acd Sci USA 91*: 9228-9232.

Souto RP, Fernandes O, Macedo AM, Campbell DA, Zingales B 1996. DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi. Mol Biochem Parasitol 83*: 141-152.

Spring J 1997. Vertebrate evolution by interspecific hybridization – are we polyploid? *FEBS Lett 400*: 2-8.

Teixeira ARL 1987. The stercorarian trypanosomes. In EJL Soulsby, *Immune Responses in Parasitic Infections: Immunology, Immunopathology, and Immunoprophylaxys*, *CRC Press.*, Boca Raton, Fl., UI: 8510912.

Teixeira ARL, Argañaraz ER, Freitas Jr L, Lacava ZGM, Santana JM, Luna H 1994. Possible integration of *Trypanosoma cruzi* kDNA minicircle into the host cell genome by infection. *Mut Res 305*: 171-175.

Teixeira ARL, Lacava Z, Santana JM, Luna H 1991. Inserção de DNA de *Trypanosoma cruzi* no genoma da célula hospedeira de mamífero por meio de infecção. *Rev Soc Bras Med Trop 24*: 55-58.

Teixeira ARL, Monteiro PS, Rebelo JM, Argañaraz ER, Vieira D, Lauria-Pires L, Nascimento R, Vexenat AC, Silva AR, Ault SK, Costa JM 2000. Emerging Chagas disease: trophic network and cycle of transmission of *Trypanosoma cruzi* from palm-trees in the Amazon. *Emerging Inf Dis* (in press).

Teixeira SMR 1998. Control of gene expression in Trypanosomatiae. *Braz J Med Biol Res 31*: 1503-1516.

Venkatesh B, Si-Hoe SL, Murphy D, Brenner S 1997.

Transgenic rats reveal functional conservation of regulatory controls between the Fugu isotocin and rat oxitocin genes. *Proc Natl Acd Sci USA 94*: 12462-12466.

Weber IT, Miller M, Jaskolski M, Leis J, Skalka AM, Wlodawer A 1989. Molecular modeling of the HIV-1 protease and its substrate binding site. *Science 243*: 928-931.

Weinstein JN 1998. Fishing expeditions. *Science 282*: 687.

Wilkins MR, Gasteiger E, Gooley AA, Herbert BR, Molloy MP, Binz PA, Ou KL, Sanchez JC, Bairock A, Williams KL, Hochstrasser DF 1999. High-throughput mass spectrometric discovery of protein pos-translational modifications. *J Mol Biol 289*: 645-657.

Winzler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science 285*: 901-906.

Wolf AP, Matzke MA 1999. Epigenetics: regulation through repression. *Science 286*: 481-486.

Zhang H, Zhang R, Liang P 1996. Differential screening of gene expression difference enriched by differential display. *Nucl Ac Res 24*: 2454-2455.