Pharmacogenomics and Computational Genomics: An Appraisal

P. K. TRIPATHI*, SHALINI TRIPATHI AND N. K. JAIN[1]
Rajarshi Rananjay Sinh College of Pharmacy, Amethi-227 405, [1]Department of Pharmaceutical Sciences, Dr. H. S. Gour Vishwavidyalaya, Sagar-470 003, India.

Pharmacogenomics deals with the interactions of individual genetic constitution with drug therapy. It is very likely that pharmacogenetic tests will make up a significant proportion of total molecular biology testing in future. Therefore, this article emphasizes the applications of pharmacogenomics, and computational genome analysis in drug therapy.

Patients sometimes experience adverse drug reactions (ADR) leading to deterioration of their underlying condition in clinical practice. Physiology of individual patient can vary, so the response of an individual to drug therapy may also be highly variable. This is a major clinical problem, since this inter-individual variability is until now only partly predictable. Besides these medical problems, cost-benefit calculations for a given pharmacological therapy will be affected significantly. This is exemplified by a recent US study, which estimates that over 100,000 patients die every year from ADRs[1].

There are many factors such as age, sex, nutritional status, kidney and liver function, concomitant diseases and medications, and the disease that affects drug responses. In recent years, it has become clear that genetic factors significantly modify drug responses. These factors should be evident for the physician.

The first examples include hemolysis in patients with glucose-6-phosphate dehydrogenase deficiency when administered antimalarial drugs and severe prolonged muscle relaxation with suxamethonium in patients with cholinesterase deficiency. This has been supplemented by numerous polymorphisms in drug-metabolizing enzymes, cellular receptors, transporters and plasma proteins. It is important to note that most of these polymorphisms or defects do not manifest themselves as a phenotype without a pharmacological challenge. Thus, the genetic constitution of an individual is extremely relevant for both efficacy and safety of a given drug regimen and this is the central topic of pharmacogenomics. Drug development and pretreatment genetic analysis of patients will be two major practical aspects in pharmacogenomics. Highly specialized drug research laboratories will achieve the first goal, while the second goal has to be achieved by clinical laboratories[2,3].

Drug development:

There are two pharmacogenomic approaches. First, for most therapies a correct diagnosis is mandatory for a satisfactory therapeutic result. This is more relevant because many diseases may be caused by different genetic defects or be significantly affected by the genetic background of an individual. Thus, phenotypically similar disease states may have quite different underlying pathobiochemical mechanisms e.g., the therapy of M3-AML (promyelocytic leukemia) with retinoids. A small subgroup of patients carries an unusual chromosomal rearrangement-t(11;17)(q23;q21), leading to a PLZF/RARA rather than the typical PML/RARA fusion gene[4]. These patients are retinoid-resistant and will perhaps need other treatments. Thus, a better classification and understanding of disease mechanisms will be the basis for a targeted development of new drugs.

Second approach, the response of an individual to a specific therapy may depend on the genes interacting with drug metabolism and/or action. The first genes shown to affect outcome of therapies coded for enzymes that are involved in the metabolism of drugs. These are very obvious targets for pharmacogenomic studies. The best-investigated examples were the cytochrome P450

*For correspondence
E-mail: tripathi.pushpendra@rediffmail.com

enzymes and N-acetyltransferase[5]. In the cytochrome P450 system, there are examples for drug toxicity related to poor metabolism or even complete lack of metabolism, as well as for reduced drug efficacy due to ultra rapid metabolism. It is estimated that more than one half of all currently used drugs are metabolized by P450 enzymes, and that CYP3A4 accounts for roughly 50% of these, followed by CYP2C6 (20%), and CYP2C9 and CYP2C19 (15%)[5]. CYP2C6, CYP2C9, CYP2 C 19, and CYP2A6 have been shown to be functionally polymorphic[6]. At present, drug development tries to avoid substances whose metabolic pathways are significantly influenced by polymorphisms in P450 enzymes. However, with the more widespread availability of clinical tests, this may no longer be necessary. While genetic polymorphisms may significantly affect the metabolism of drugs via P450 enzymes, it should be kept in mind that drug-drug interactions in genetically normal individuals play a similarly important role, as can be seen for example in the severely increased risk for toxicity if statins modified by CYP3A4 are co-administered with drugs inhibiting CYP3A4, as for instance, mibefradil.

In the long run, genes coding for receptors or signaling molecules involved in the pathophysiology of disease will be perhaps of broader relevance. G-protein-coupled receptors are a good example to illustrate this point. Several receptor genes from this family including the ß$_1$- and ß$_2$-adrenergic receptors[7,8] the cholecystokinin$_2$ (CCK$_2$) receptor[9,10] and Mu opioid receptor[11] have been shown to harbor polymorphisms that affect ligand affinity. This is of importance because it is conceivable to develop ligands for the receptor that may or may not be affected by these polymorphisms. As a consequence, the presence of a polymorphism, which affects the binding properties of the natural ligand, would not interfere with the action of the synthetic ligand. Knowledge of such polymorphisms enables the design of drugs that are effective in all patients rather than drugs that are effective only in patients with a certain genetic constellation. If a receptor polymorphism causes disease, this approach would constitute a causal therapeutic intervention. An even farther-reaching option would be to silence constitutively active receptors by ligands specifically designed to bind only to these mutant receptors and inactivate them[12]. Such ligand-independent receptors underlie several human diseases, e.g., thyroid adenoma[13], precocious puberty[14], and Jansen's metaphyseal chondrodysplasia[15]. Overall, pharmacogenomic approaches offer interesting perspectives for molecular design and development of more specific drugs with significant benefits to patients[16].

Pharmacogenomics as diagnostic tool in clinical therapeutics:

Much more relevant for laboratory medicine is the analysis of genetic polymorphisms in patients before therapy. With the expansion of our knowledge about gene-drug interactions the number of diagnostic tests will increase rapidly. The typical questions here relate to the risk for ADRs, and potential therapy failures.

At present, the most widely used tests apply to the detection of patients that might experience severe or even life-threatening toxicity from a certain drug. Usually, this is associated to deficient activity of an enzyme that is involved in the metabolism and/or inactivation of a drug. The consequence is prolonged and exaggerated drug action. Probably one of the most widespread assays is the analysis of thiopurine-methyltransferase (TPMT) activity. The polymorphism of TPMT has been described about 25 years ago by Weinshilboum and Sladek[17], its association with azathioprine toxicity was detected only a little over 15 years ago[18], and broad interest in the pharmacogenetic trait has developed only in the mid 90s. The rationale whether a genetic or a biochemical test should be used to analyse TPMT is summarised in the paragraph on methodology, because it may be taken as a paradigm for most if not all pharmacogenetic tests. The major considerations relate to sensitivity, specificity and cost effectiveness. TPMT also nicely exemplifies that one of the key criteria for demand of a specific test by the clinician is the risk for so far unpredictable serious adverse events.

Analysis of P450 polymorphic enzymes is less routine in clinical practice, even though approximately one quarter of all drugs are metabolised by the polymorphic CYP2D6 and significantly decreased or absent activity is present in more than 5% of Caucasians. Decreased activity of CYP2D6 is associated among others with cardiotoxicity of tricyclic antidepressants, proarrhythmic effects of antiarrhythmic drugs. Decreased activity of CYP2C9 can lead to bleeding under warfarin or tolbutamide therapy. Increased effects of diazepam may accompany defects in CYP2C19. A recent study suggests that a poor-metaboliser status for CYP2C19 (CYP2C19*2) may be associated with an increased risk for ventricular tachycardias and torsade de pointes-arrhythmias under treatment with terodiline, an anticholinergic agent with Ca-blocking activity[19]. Terodiline has been implicated to cause QT prolongation and life-threatening arrhythmias. If this association is confirmed, one might be able to reduce the incidence of such potentially lethal complications by genetic tests of patients before treatment.

One of the major areas in which polymorphic P450 enzymes are relevant, are psychiatric diseases. Especially, polymorphisms in CYP2D6 may significantly affect drug levels of antipsychotic and antidepressant medications. Even though this could be predicted by genetic analysis, the usual approach here is not testing of pharmacogenetic polymorphisms but rather regular determinations of drug levels. This will not only provide plasma levels in relation to drug dose, but also permit monitoring of compliance with the prescribed drug[20]. Altered metabolism of substances by P450 enzymes can be predicted by analysis of the metabolism of test substances or by genotyping. While genotyping never achieves 100% sensitivity and does not take into account drug-drug interactions, it is much easier to perform, is not confounded by any underlying disease, and carries no risk of adverse reactions to the test substances[21]. At present, there is no general consensus which patients should be tested for P450 polymorphisms. Candidates are patients with poor therapeutic responses despite optimal dosing, or patients with a relative that had ADRs to one of the drugs metabolised by a polymorphic P450 enzyme. As outlined above, in the future more drugs may be developed that are affected by P450 polymorphisms and require testing of a patient before institution of therapy. In this case, laboratories must be prepared to perform routine tests for P450 polymorphisms.

Another major problem for clinical utility and acceptance of a pharmacogenetic test are impracticability. At present, few convincing examples for short therapeutic trials are available. It has been shown that a common polymorphism in the gene for cholesteryl ester transfer protein (CETP) is associated with the effect of pravastatin on the progression of coronary atherosclerosis[22]. However, the reported association of the apolipoprotein E polymorphism with the response to treatment with tacrine[23] is a good example for a spurious genetic association. It is a good reminder of the general problems observed with association studies that are well known from many trials[24].

Besides these strictly clinical considerations, the classic diagnostic characteristics of the test applied will be important, i.e., its specificity, sensitivity and predictive value of a positive and/or negative test. The latter is obviously related to the prevalence of the polymorphism or gene defect in the population. These properties of a diagnostic test will determine its cost effectiveness. And finally, it will be relevant for assay development whether the therapeutic regimen in question is common or rare.

Our knowledge regarding the relevance of genetic polymorphisms for pharmacotherapy is still very limited. To improve this, it will be mandatory to accompany clinical drug trials with genotype analysis of the probands. This can be either performed by analyzing candidate genes of potential relevance or by genome-wide single nucleotide polymorphism (SNP) analysis. The candidates identified by this approach can be sequenced and analyzed for polymorphic markers that can be included in clinical trials. At current prices, the costs for a genome-wide SNP analysis in a drug trial have been estimated to be probably far above 10 US $ million. Thus, genome-wide analyses will be the exception in drug trials as long as the technology does not improve significantly. One improvement could be allele discrimination by identifying haplotypes. This has been impossible with the current PCR-based techniques for SNP detection, but would reduce the number of SNPs needed significantly[25].

As outlined above, pharmacogenetic trials will be most important for drugs with a high risk for severe ADRs and a narrow therapeutic range, e.g., immunosuppressive or cytostatic drugs. However, as exemplified with the P450 enzymes, there are genetic polymorphisms that solely affect plasma levels of a drug. If this drug has a narrow therapeutic range, the consequence may be that drug monitoring rather than genetic tests prior to therapy will be performed. Pharmacogenetics will be also important for long-term treatments, whose outcome can be only predicted after years of treatment. In this case, it will be important to identify patients that will not respond to treatment early on to optimize the risk-benefit and the cost-benefit ratios of treatment. Therefore, therapeutic interventions that fall into one of these categories are predictably the first that will be assessed in pharmacogenetic trials.

Methodology:

For clinical analysis of pharmacogenetically relevant polymorphisms or mutations, several approaches can be taken. Pharmacologic approaches are usually not feasible in the clinical routine setting, because they require test doses of the drug or a compound supposed to have the same metabolism followed by serial determinations of blood levels of the drug and/or metabolites. There are few commonly used tests, e.g., the caffeine test to identify the NAT-2 phenotype, however, it may be expected that these tests are substituted by genetic tests or enzyme activity determinations. If enzyme activity is a critical factor and related to gene polymorphisms (e.g., TPMT), it may be preferable to measure this activity directly. Genetic analysis can be performed by different techniques, depending on the gene in question. While genetic tests for known point mutations are usually simple to perform, in the case of TPMT only approximately 75-80% of TPMT deficiencies are caused by the most common mutations. The remainder is caused by rare or still unknown mutations. Therefore, the simple genetic tests have a low sensitivity. The biochemical test, on the other hand, is far from trivial. In addition, it may yield false negative results in patients who received blood transfusions.

If genetic analysis is chosen, it usually requires the detection of known point mutations and rarely the analysis of whole genes for mutations. For most routine and research purposes, the analysis of SNPs will be the major requirement for a laboratory[26].

Single nucleotide polymorphism (SNP) analysis:
The analysis of SNPs can either focus on single genes or on the combination and pattern of many SNPs. While the former is a simple straightforward approach that is based on the knowledge about the association of an SNP with gene function or specific phenotypes, the latter is a complicated approach that requires extensive clinical studies on the clinical relevance of the combined genotypes. It is envisioned that upon identification of hundreds of thousands of sequence variations within the genome (human genome project) in the near future, simultaneous genetic analysis of the individual's DNA polymorphisms will be feasible. Technology is underway that will facilitate such massive parallel determinations in short time.

Currently, a large number of molecular strategies are being used for single nucleotide polymorphism (SNP) analysis[26]. In the majority of these methods, the target sequence is amplified and the polymorphism identified by various technologies, for some of which multiplexing has been demonstrated. In principle, there are two general approaches: mass spectrometry platforms and fluorescence-based platforms.

The distinction of the SNPs is either achieved by short hybridization probes or by restriction endonucleases, by discrimination of mismatched DNA substrates by polymerases or ligases or by observing the template-dependent choice of nucleotides incorporated by a polymerase (minisequencing). More recently, mass spectrometry (MALDI-TOF) has been demonstrated to be a good alternative. These various techniques have been adapted to assay formats that simplify scale-up in SNP analysis.

Mass spectrometry in SNP analysis:
The use of mass spectrometry for genotyping has been reviewed recently in detail[27,28]. Mass spectrometric methods are based on the determination of small mass changes of specific oligonucleotides that have been used to identify a SNP of interest. These oligonucleotides change their mass usually in a minisequencing reaction dependent on the genotype in the SNP analyzed. It is also possible to modify the oligonucleotide following the discriminatory reaction that way that the determination of the mass is much easier. Such oligonucleotides may be partially degraded or a specific mass-tag released.

Mass spectrometry is an alternative to fluorescence-based technologies, can be automated, enables high-throughput analysis and multiplexing. The result serves as quality control as well. The disadvantage may be that it is expensive and that the DNA-crystals needed for laser-mediated desorption, are not formed uniformly with proteins.

Fluorescence-based SNP analysis:
The methods that do not rely on mass spectrometry and in most cases use fluorescence for detection can be divided in six categories: i. Restriction digestion, ii. Homogeneous hybridization, iii. Mismatch distinction by polymerases and ligases, iv. Array hybridization assays, v. Minisequencing, and vi Rolling circle signal amplification. While all methods are useful for small sample series, real high-throughput analysis, as is needed for clinical trials, can be achieved only by some of the techniques. PCR can be performed on chips with on-board detection, assays on micro particles and DNA-micro arrays. For SNP analysis, specificity and discrimination of alleles is even of greater importance. Specificity of the amplification process applies equally to any application and methodology and is achieved by the use of two hybridizing events. Without amplification it is not possible to discriminate between alleles even if longer hybridization primers are used in combination with signal amplification procedures. Better discrimination of alleles is achieved with short primers.

Computational genomics:
Computational genomics can be defined as a discipline of computational biology, which deals with the analysis of entire genome sequences. But today, computational genomics is much more than mere sequence analysis. Although its roots lie in more traditional bioinformatics methods, there have been significant steps towards a more integral analysis of genome information, including

metabolic pathways[29], signalling networks[30], functional classes[31], phylogenetic patterns[32], protein fold types[33] and genome organisation[34]. This increasingly intensified computational approach to genome analysis has generated not only tools for experimental biologists but also interesting scientific results[35].

This section of review is divided into three parts. First, a brief description of the general methodological approaches, second, attempt to detect and describe molecular function by exploiting genome structure and third, future directions in the field of computational genomics.

General methodological approaches:
Computational genomics technique arises from the 'mainstream' bioinformatics activities and mainly focus on genome sequence analysis, for example gene finding, sequence diagnostics, database searching, sequence clustering and functional annotation.

Gene finding:
There are two approaches for gene finding first is extrinsic approach[36] (searching protein databases with the query DNA sequence for the identification of protein-coding genes) are not as effective as in the case of prokaryotes. Even for prokaryotic genomes, inconsistencies of open reading frame calling abound[37]. Second is intrinsic approaches[36] of gene detection (predicting genes from first principles such as exon/intron boundary detection) lack the appropriate amount of learning sets for the training of the algorithms[38]. Progress in this area includes the development of hidden Markov model (HMM)-based methods for gene structure that detect more accurately exon/intron boundaries, such as GeneMark[39] and Glimmer[40].

Sequence diagnostics:
Sequence analysis is fundamental for the characterisation of the query sequences, especially when no similarity to other sequences in the database is readily identifiable. In this category like the detection of coiled-coil[41], trans-membrane[42,43], cellular localization signal[44,45] and compositionally biased[46,47] regions.

Database searching:
The database search stage provides indications not only of family membership (when a set of sequences is identified as being homologous to the query sequence) but also of the possible function of the query sequence, when the homologues have been experimentally characterized and appropriately annotated[48-50]. New methods include profile vs. profile methods such as LAMA[51] and HMM-based methods such as Hmmer[52], MAST[53], SAM-T98[54] and the benchmarking of algorithms for their ability to detect weak sequence similarities[55,56].

Sequence clustering:
Further stage where quality control of the annotations can take place is the clustering of genes and proteins into families. Resources in this field include Pfam[57], COGS[58], WIT[59], Protomap[60], Emotif[61] and one exciting development is the detection of multi-domain proteins which may also provide clues to the function of their single-domain counterparts in complete genomes[62,63].

Functional annotations:
There are a number of problems which hamper accurate and, more importantly, consistent functional annotations for genome sequences. First, the transfer of function via homology is a subject of current research[64-67] and no clear-cut rules may immediately apply. Second, the transfer of this information, even when all other criteria are satisfied, crucially depends on the quality of transient database annotations[68] which may be far from satisfactory (no published material on the quality control of curated database annotations is available). Third, the reproducibility of sequence annotations is poor[69-73] and the result is a conundrum of descriptions for genome sequences without a clear consensus.

The best annotations currently available take the form of community-curated databases centred around model organisms, for example EcoCyc[74] for *Escherichia coli*, SGD[75] for *Saccharomyces cerevisiae* and FlyBase[76] for *Drosophila melanogaster*.

Association with functional roles:
This particular aspect of genome analysis is where the whole activity transcends the boundaries of 'classical' sequence analysis and necessitates technology that has yet to be developed. The idea is that the appropriately structured (and potentially formal) function descriptions of gene products can be integrated into systems that represent a general network of cellular processes, including metabolic pathways, transcription activation mechanisms and intracellular control cascades[77]. Metabolic databases[78] have formed a basis upon which other complex categorization schemes have been developed. Some of the most successful attempts here include various approaches to metabolic reconstruction, defined as the prediction of the metabolic complement for a species based on the analysis of genome sequence[79-84].

Molecular function:

Even a bird's eye view of the recent advances in genomics, is sufficient to establish bioinformatics as an essential utility for the experimental biologists[85,86]. Genome subtraction can only pick out unique genes if the genome sequence is completely known. Whole-genome alignment requires entire genomes, almost by definition. Finally, the precision performance of the last two methods crucially relies on completeness[63]. Thus, all the approaches described next can be defined as computational genomics methods according to our original definition.

Genome subtraction:

Entire genomes allow the detection of unique sequences, genes that are not present anywhere in the database or in the close relatives of the species under investigation[87]. These elements are sometimes components of cellular pathways that remain to be discovered and can be interesting drug targets in pathogenic organisms. To identify unique sequences, however, one has to detect equivalent (or orthologous) genes, which are not always easy to define[88]. Despite this shortcoming, this method will be most valuable for the comparison of bacterial strains or other, closely related species. Two interesting studies using this virtual subtraction method have appeared for *Haemophilus influenzae*[87] and *Helicobacter pylori*[89].

Whole-genome alignment:

Another area where technical advances resulted in some deeper understanding of the genome structure and thus function of certain species is whole-genome alignment[34]. Previous systems could not cope with hundreds of kilobases of raw DNA sequence. This advance will facilitate detailed comparisons of genome organisation (revealing single nucleotide polymorphisms, translocations or inserts, repeats and syntenic regions in chromosomes). Another application is strain comparison, reminiscent of genome subtraction. This method has been applied to the comparison of two *Mycobacterium tuberculosis* strains, *Mycoplasma genitalium* and *Mycoplasma pneumoniae* and regions from mouse chromosome 6 against human chromosome 12[34].

Functional coupling of gene cluster:

Another method that exploits genome structure and organisation is the prediction of functional association of neighbouring genes. It has been observed that certain conserved gene clusters (which may be operons) contain functionally related genes[90-92]. Thus, even for genes of unknown function, there is a possibility to predict their cellular roles[90] or a more specific functional property[92] on the basis of their neighboring genes. Applications include the comparative analysis of two bacterial genomes[90], the comparison of nine bacterial and archaeal genomes to propose physical interactions of gene products[91] and the use of gene clusters from 31 complete genomes to infer functional coupling and reconstruction of metabolic networks[92].

Fusion analysis:

Finally, based on the observation that the homologues of certain genes appear to fuse during the course of biological evolution, this approach attempts to predict functional association and protein interactions on the basis of gene fusion. The methods rely on the assumption that individual component proteins whose homologues are involved in a fused, multi-domain protein must be involved with each other in a protein complex, biochemical pathway or another cellular process[62,63]. Detection of false positive predictions by this approach is difficult, mainly due to the lack of extensive experimental information about protein interactions.

Towards a scientific discipline:

The explosion in computational analysis methods for complete genomes brought out not only technologies but also some key scientific results. Some very interesting developments that have appeared in the recent literature in the areas of metabolic reconstruction and comparative genomics, using computation alone have been listed here.

For metabolic reconstruction, examples include the reconstruction of the metabolic networks of *Methanococcus jannaschii*[93], the analysis of the tricarboxylic acid cycle across a number of species[94], the characterisation of the known metabolic complement of *E. coli*[95], the distribution of functional classes across the domains of life[96] and the prediction of functional networks in yeast[97].

For comparative genomics, examples include the detection of an archaeal genomic signature[98], the compilation of universal protein families[99], the comparison of three entire eukaryotic genomes[100], the detection of eukaryotic signalling domains in archaea and bacteria[101], the distribution of individual protein families across species[102], the patterns of protein fold usage in microbial genomes[103], the derivation of the universal tree based on enzyme families[104] and the derivation of species relationships based on gene content[105].

Taking a closer look at the properties of entire genome sequences, two things become apparent: first, comparative analysis greatly enhances our abilities to 'predict' and detect molecular function using sequence information and second, the current bottleneck in genomics appears to be the turnover of experimentally obtained novel properties for molecular families of unknown function. Once the function universe is covered, it may be that computation will acquire a truly central role in biological science.

Future directions:

What is to be expected from computational genomics in the near future? As illustrated in the previous sections, our battery of tools is becoming increasingly sophisticated and our ability to detect protein function using computation is generally improving. However, to resolve the issue of function description and detection, we need to progress from methods mostly derived from traditional sequence analysis that examine genome sequences individually to algorithms and databases that exploit the inherent properties of entire genomes. We are in the process of discovering the constraints that apply to entire genomes so that genomic context can be reflected in our future methods, enhancing the quality of function descriptions.

We argue here that all our approaches towards the elusive goal of predicting function from sequence have to take into account the genomic context and describe molecular function in terms of actions and interactions within the cell. In other words, our procedures from sequence to function require the development of models that describe cells as systems, using their genetic blueprint, i.e., genome sequence.

Querying biological databases:

It is indisputable that publicly available databanks play a fundamental role in disseminating sequence data to the biological community. However, one of the most important problems of biological data repositories is their archive-like nature. Public databases are designed to store information in an unstructured way, largely in free-text flat-file format without defined object relations. This may help end-users that occasionally browse to retrieve individual entries, but it is very far from making the database amenable to large-scale computation. In this sense, these repositories are not genuine database systems, designed for flexible querying and large-scale data mining.

Classifications of biological function:

To accurately describe function, biologists have to agree on a common vocabulary and classification of molecular roles for all genes and proteins. This is an immense task, made much more difficult by its nature as a community project. The amount of data is significant, but finite. It is the complexity of the information that makes this task daunting.

The issue of data complexity can be tackled by portable ontology designs, exact specifications of various conceptualizations for a given domain. This strategy is one way of dealing with highly complex, qualitatively rich features of specific domains of discourse. Such systems attempt to classify and further process various aspects of molecular function in terms of general hierarchies for genome sequence and biochemical pathways, ribosome structure and function, cellular processes and function categories and generalized functional classes. Although the latter are simple, general and also automatically derived[106], they have yet to be widely accepted. One reason may be the clash of opinions on the definition of functional classes, and the relatively restricted utility of a high-precision but low-coverage classification of protein functions[107].

Structural genomics:

The function of all proteins will be determined by the sheer knowledge of their structure, a well-known motto in structural biology[108-110]. Significant progress has already been made in terms of assigning structural homologues to proteins of known function for a number of completely sequenced species[111]. It should be noted, however, that some recent claims for structural genomics might be slightly overstated[112,113].

By way of rising knowledge about interactions among genes and drug treatment, there will be an equally increasing demand for speedy and consistent diagnostic tests prior to the institution of therapy. One can say that analyzing genomes constitutes much more than mere sequence analysis, but it also includes essential characteristics of reaction detection and reconstruction of biological metabolism and therapeutics. Genome analysis is a highly consistent and dependable integration of functional information. However, one must clearly understand the fact that biological databases should exactly mimic the actual biological reality, as closely as possible, for this information to be useful for computation.

REFERENCES

1. Lazarou, J., Pomeranz, B.H and Corey, P.N., **J. Amer. Med. Assoc.**, 1998, 279, 1200.

2. Evans, W.E. and Trease, M.V. **Science,** 1999, 286, 487.
3. Roses, A.D., **Nature**, 2000, 405, 857.
4. Sainty, D., Liso, V., Cantù-Rajnoldi, A. and Head D., **Blood,** 2000, 96, 1287.
5. Bertz, R.J. and Grannemann, G.R., **Clin. Pharmacokinet.,** 1997, 32, 210.
6. Ingelman-Sundberg, M., Oscarson, M., and McLellan, R.A., **Trends Pharmacol. Sci.,** 1999, 20, 342.
7. Mason, D.A., Moore, J.D., Green, S.A. and Liggett, S.B., **J. Biol. Chem.,** 1999, 274, 12670.
8. Green, S.A., Cole, G., Jacinto, M., Innis, M. and Liggett, S.B., **J. Biol. Chem.,** 1993, 268, 23116.
9. Beinborn, M., Lee, Y.M., McBridge, E.W., Quinn, S.M. and Kopin, A.S., **Nature,** 1993, 362, 348.
10. Kopin, A.S., McBridge, E.W., Gordon, M.C., Quinn, S.M. and Beinborn, M., **Proc. Natl. Acad. Sci. USA,** 1997, 94, 11043.
11. Bond, C., LaForge, K.S. and Tian M., **Proc. Natl. Acad. Sci. USA,** 1998, 95, 9608.
12. Kopin, A.S., McBride, E.W., Schaffer K. and Beinborn, M., **Trends Pharmacol. Sci.,** 2000, 21, 346.
13. Parma, J., Duprez, L. and Van Sande, J., **Nature,** 1993, 365, 649.
14. Shenker, A., Laue, L. and Kosugi S., **Nature,** 1993, 365, 652.
15. Schipani, E., Kruse, K., and Juppner, H., **Science,** 1995, 268, 98.
16. Issa, A.M., **Trends Pharmacol. Sci.,** 2000, 21, 247.
17. Weinshilboum R.M. and Sladek, S.L., **Amer. J. Hum. Genet.,** 1980, 32, 651.
18. Lennard, L., Van Loon, J.A. and Weinshilboum, R.M., **Clin. Pharmacol. Ther.,** 1989, 46, 149.
19. Ford, G.A., Wood S.M. and Daly, A.K., **Brit. J. Clin. Pharmacol.,** 2000, 50, 77.
20. Poolsup, N., Po, A.L.W. and Knight, T.L., **J. Clin. Pharm. Ther.,** 2000, 25, 197.
21. Linder, M.W., Prough R.A. and Valdes, **R., J. Clin. Chem.,** 1997, 43, 254.
22. Kuivenhoven, J.A., Jukema, J.W. and Zwinderman, A.H., **N. Engl. J. Med.,** 1998, 338, 86.
23. Poirier, J., Delisle, M.C. and Quirion, R., **Proc. Natl. Acad. Sci. USA,** 1995, 92, 12260.
24. Editorial, Freely associating, **Nat. Genet.** 1999, 22, 1.
25. McCarthy J.J. and Hilfiker, R., **Nat. Biotechnol.,** 2000, 18, 505.
26. Shi, M.M., Bleavins, M.R and De La Iglesia, F.A., **Mol. Diagn.,** 1999, 4, 343.
27. Jackson, P.E., Scholl P.F. and Groopman, J.D., **Mol. Med. Today,** 2000, 6, 271.
28. Griffin T.J. and Smith, L.M., **Trends Biotechnol.,** 20
29. Karp, P.D., Krummenacker, M., Paley, S. and Wagg. J., **Trends Biotechnol.,** 1999, 17, 275.
30. Takai-Igarashi, T., Nadaoka, Y., and Kaminuma. T., **J. Comput. Biol.,** 1998, 5, 747.
31. Riley, M., **Curr. Opin. Struct. Biol.,** 1998, 8, 388.
32. Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D. and Yeates, T.O., **Proc. Natl. Acad. Sci. USA,** 1999, 96, 4285.
33. Gerstein., M., **J. Mol. Biol.,** 1997, 274, 562.
34. Delcher, A.L., Kasif, S., Fleischmann, R.D., Peterson, J., White, O. and Salzberg, S.L., **Nucleic Acids Res.,** 1999, 27, 2369.
35. Andrade, M.A., and Sander, C., **Curr. Opin. Biotechnol.**, 1997, 8, 675.
36. Borodovsky, M., Rudd, K.E., and Koonin, E.V., **Nucleic Acids Res.,** 1994, 22, 4756.
37. Raghavan, S. and Ouzounis, C.A., **Nucleic Acids Res.,** 1999, 27, 4405.
38. Burset, M., and Guigo, R., **Genomics**, 1996, 34, 353.
39. Lukashin, A.V. and Borodovsky, M., **Nucleic Acids Res.,** 1998, 26, 1107.
40. Salzberg, S.L., Delcher, A.L., Kasif, S. and White, O., **Nucleic Acids Res.,** 1998, 26, 544.
41. Lupas, A., Van Dyke, M. and Stock, J., **Science**, 1991, 252, 1162.
42. Kihara, D., Shimizu, T., and Kanehisa, M., **Protein Eng.,** 1998, 11, 961.
43. Pasquier, C., Promponas, V.J., Palaios, G.A., Hamodrakas, J.S. and Hamodrakas, S.J., **Protein Eng.,** 1999, 12, 381.
44. Nakai, K. and Horton, P., **Trends Biochem. Sci.**, 1999, 24, 34.
45. Nielsen, H., Brunak, S. and Von Heijne, G., **Protein Eng.**, 1999, 12, 3.
46. Wootton, J.C., **Comput. Chem.**, 1994, 18, 269.
47. Wootton, J.C. and Federhen, S., **Methods Enzymol.**, 1996, 266, 554.
48. Andrade, M.A., **Bioinformatics,** 1999, 15, 391.
49. Karp, P.D., **Bioinformatics,** 1998, 14, 753.
50. Altschul, S.F., Boguski, M.S., Gish, W. and Wootton, J.C., **Nat. Genet.,** 1994, 6, 119.
51. Henikoff, S., Henikoff, J.G. and Pietrokovski, S., **Bioinformatics**, 1999, 15, 471.
52. Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Finn, R.D., and Sonnhammer, E.L., **Nucleic Acids Res.**, 1999, 27, 260.
53. Bailey, T.L. and Gribskov, M., **J. Comput. Biol.**, 1998, 5, 211.
54. Karplus, K., Barrett, C. and Hughey, R., **Bioinformatics,** 1998, 14, 846.
55. Muller, A., MacCallum, R.M., and Sternberg, M.J., **J. Mol. Biol.,** 1999, 293, 1257.
56. Henikoff, S., Pietrokovski, S. and Henikoff, S., **Nucleic Acids Res.**, 1998, 26, 309.
57. Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe K.L. and Sonnhammer, E.L., **Nucleic Acids Res.,** 2000, 28, 263.
58. Tatusov, R.L., Koonin, E.V., and Lipman, D.J., **Science,** 1997, 278, 631.
59. Overbeek, R., **Nucleic Acids Res.,** 2000, 28, 123.
60. Yona, G., Linial, N. and Linial, M., **Proteins,** 1999, 37, 360.
61. Nevill-Manning, C.G., Wu, T.D. and Brutlag, D.L., **Proc. Natl. Acad. Sci. USA,** 1998, 95, 5865.
62. Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O. and Eisenberg, D., **Science**, 1999, 285, 751.
63. Enright, A.J., Iliopoulos, I., Kyrpides, N.C. and Ouzounis, C.A., **Nature**, 1999, 402, 86.
64. Shah, I. and Hunter, L., **ISMB,** 1997, 5, 276.
65. Des Jardins, M., Karp, P.D., Krummenacker, M., Lee, T.J. and Ouzounis, C.A., **ISMB,** 1997, 5, 92.
66. Wilson, C.A., Kreychman, J., and Gerstein, M., **J. Mol. Biol.**, 2000, 297, 233.
67. Hegyi, H., and Gerstein, M., **J. Mol. Biol.,** 1999, 288, 147.
68. Wheelan, S.J., and Boguski, M.S., **Genome Res.**, 1998, 8, 168.
69. Brenner, S.E., **Trends Genet.,** 1999, 15, 132.
70. Andrade, M., Casari, G., De Daruvar, A., Sander, C., Schneider, R., Tamames, J., Valencia, A. and Ouzounis, C., **Comput. Appl. Biosci.,** 1997, 13, 481.
71. Tsoka, S., Promponas, V. and Ouzounis, C.A., **FEBS Lett.,** 1999, 451, 354.
72. Pallen, M., Wren, B. and Parkhill, J., **Mol. Microbiol.**, 1999, 34, 195.
73. Kyrpides, N.C. and Ouzounis, C.A., **Mol. Microbiol.**, 1999, 32, 886.
74. Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Paley, S.M. and Pellegrini-Toole, A., **Nucleic Acids Res.**, 2000, 28, 56.
75. Cherry, J.M., **Nucleic Acids Res.,** 1998, 6, 73.
76. Gelbart, W.M., **Nucleic Acids Res.**, 1997, 25, 63.
77. Karp, P.D and Paley, S., **J. Comput. Biol.,** 1996, 3, 191.
78. Karp, P.D., **Trends Biochem. Sci.**, 1998, 23, 114.
79. Gaasterland, T. and Selkov, E., **ISMB,** 1995, 3, 127.
80. Bono, H., Ogata, H., Goto, S., and Kanehisa, M., **Genome Res.,** 1998, 8, 203.
81. Karp, P.D., Ouzounis, C. and Paley, S., **ISMB,** 1996, 4, 116.
82. Kanehisa, M., and Goto, S. **Nucleic Acids Res.**, 2000, 28, 27.
83. Bailey, L.C., Fischer, Jr., S., Schug, J., Crabtree, J., Gibson, M. and Overton, G.C., **Genome Res.**, 1998, 8, 234.
84. Ashburner, M., **Nat. Genet.,** 2000, 25, 25.
85. Benton, D., **Trends Biotechnol.,** 1996, 14, 261.
86. Huynen, M.A., Diaz-Lazcoz, Y. and Bork, P., **Trends Genet.,** 1997, 13, 389.

87. Ouzounis, C., **Trends Genet.,** 1999, 15, 445.
88. Huynen, M., Dandekar, T. and Bork, P., **FEBS Lett.,** 1998, 426, 1.
89. Tamames, J., Casari, G., Ouzounis, C., and Valencia, A., **J. Mol. Evol.,** 1997, 44, 66.
90. Dandekar, T., Snel, B., Huynen, M., and Bork, P., **Trends Biochem. Sci.**, 1998, 23, 324.
91. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D. and Maltsev, N., **Proc. Natl. Acad. Sci. USA,** 1999, 96, 2896.
92. Selkov, E., Maltsev, N., Olsen, G.J., Overbeek, R. and Whitman, W.B., **Gene,** 1997,197, GC11.
93. Huynen, M.A., Dandekar, T. and Bork, P., **Trends Microbiol.**, 1999, 7, 281.
94. Ouzounis, C.A. and Karp, P.D., **Genome Res.,** 2000, 10, 568.
95. Andrade, M.A., Ouzounis, C., Sander, C., Tamames, J. and Valencia, A., **J. Mol. Evol.**, 1999, 49, 551.
96. Marcotte, E.M., Pellegrini, M., Thompson, M.J., Yeates, T.O. and Eisenberg, D., **Nature**, 1999, 402, 83.
97. Graham, D.E., Overbeek, R., Olsen, G.J. and Woese, C.R., **Proc. Natl. Acad. Sci. USA**, 2000, 97, 3304.
98. Kyrpides, N., Overbeek, R., and Ouzounis, C., **J. Mol. Evol.** 1999, 49, 413.
99. Rubin, G.M., **Science,** 2000, 287, 2204.
100. Ponting, C.P., Aravind, L., Schultz, J., Bork, P. and Koonin, E.V., **J. Mol. Biol.,** 1999, 289, 729.
101. Tomii, K. and Kanehisa, M., **Genome Res.**, 1998, 8, 1048.
102. Gerstein, M., **Proteins**, 1998, 33, 518.
103. Doolittle, W.F., **Science,** 1999, 284, 2124.
104. Snel, B., Bork, P. and Huynen, M.A., **Nat. Genet.,** 1999, 21, 108.
105. Abernethy, N.F., Wu, J.J., Hewett, M. and Altman, R.B., **IEEE Intell. Syst.,** 1999, 14, 79.
106. Ouzounis, C., Casari, G., Sander, C., Tamames, J. and Valencia, A., **Trends Biotechnol.**, 1996, 14, 280.
107. Tamames, J., Ouzounis, C., Casari, G., Sander, C. and Valencia, A., **Bioinformatics,** 1998, 14, 542.
108. Burley, S.K., **Nat. Genet.**, 1999, 23, 151.
109. Brenner, S.E., and Levitt, M., **Protein Sci.**, 2000, 9, 197.
110. Orengo, C.A., Todd, A.E. and Thornton, J.M., **Curr. Opin. Struct. Biol.**, 1999, 9, 374.
111. Teichmann, S.A., Chothia, C. and Gerstein, M., **Curr. Opin. Struct. Biol.**, 1999, 9, 390.
112. Eisenstein, E., **Curr. Opin. Biotechnol.**, 2000, 11, 25.
113. Shapiro, L., and Harris, T., **Curr. Opin. Biotechnol.**, 2000, 11, 31.

Accepted 5 January 2007
Revised 2 August 2006
Received 23 December 2004
Indian J. Pharm. Sci., 2007, 69 (1): 1-9