

Features

2. Molecular Modeling, Principles and Applications. Ed., A. R. Leach, Addison Wesley Longman Limited, 1996
3. Guidebook on Molecular Modeling in Drug design, Ed., N. C. Cohen, Academic Press, Inc., 1996.
4. Practical Application of Computer-Aided Drug Design. Ed., P. S. Charifson, Marcel Dekker, Inc., 1997.
5. M. Davis and S. J. Teague, *Angew. Chem. Int. Ed.*, 1999, 38, 736-749.
6. G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J. Mol. Biol.*, 1997, 267, 727-748.

Genomics and Novel Drug Targets

Bhupesh Taneja, Radha Chauhan and Shekhar C. Mande

Institute of Microbial Technology, Sector 39A, Chandigarh-160036.

Introduction

Diseases and pathogenic organisms have always been a great threat to human life. Medicinal scientists have attempted to fight the battle against diseases by integrating expertise from as distinct fields as biology, chemistry and computer science to generate new and powerful therapies. The biotechnological revolution of 1990s and its role in understanding the molecular basis of disease holds the promise of changing the face of medical science in the coming years with the likely discovery of more specific and efficient novel drugs.

Approaches for drug discovery

Discovery of a new drug follows a long and circuitous path. There are essentially two approaches which may be followed for the discovery of new drugs. The first is the empirical or classical approach which is based on the non-target-specific principle. In this approach, potent drug molecules are screened for their biological activity against a specific disease without the need of characterizing their target receptors. The second approach is a target-specific or rational approach which identifies a biological target followed by screening of a specific inhibitor for that target. Both these approaches have been followed extensively in the past for novel drug discovery.

Inspite of generating many important therapeutics, the empirical approach has its own limitations. It is an expensive and time consuming process, results of which are mostly serendipity

based in the absence of the knowledge of the target. In addition, it identifies lead compounds with ill-defined molecular mechanisms of action and toxicities. Hence, there is a need to circumvent these limitations by a more rational target-based approach.

Rational drug design

Rational drug design is a target-based drug discovery approach in which potent drug molecules are identified against a single specific target of known mode of action. Key biochemical pathways, virulence factors or the multi drug resistance genes of an infectious organism constitute potential drug targets. The drug target is then used to locate specific ligands or inhibitors to modulate their function.

Various biochemical pathways essential to the survival of the pathogen have been explored to identify new drug targets. The potential drugs may target the proteins involved in pathogenicity. For instance, virulence factors for survival in the changing environment in the host cell, or bacterial proteases involved in degradation of host defense and signaling. These potential drug targets may also be proteins or nucleic acids involved in cell wall synthesis, lipid biosynthesis, cell division, DNA and RNA metabolism etc. Moreover, the differences in the physiology and the structure of the diseased cell and the normal cell may also be exploited. Unfortunately, the complexity of the human genome as well as genetic polymorphisms may lead to the ineffectiveness of these potential

Features

drug targets. Multi-drug resistance, poor bioavailability, strong immunogenicity or high toxicities of known drugs to the previously defined targets throws open the challenge for the identification of more new drug targets. Some of the current approaches in new target identification and selection are discussed below.

Genomes and new targets

The complete sequencing of the genomes of various organisms including pathogenic microorganisms has provided a large number of potential drug receptor genes to treat various diseases. As many as 25 genomes of various microorganisms have been completely sequenced in the last five years and several others are underway [12].

Genomics information has been used to predict the potential selectivity of drug target genes. Sequence homology is among the first steps in the identification of new genes from the genome sequences. Using sequence information, genes of pathogens can be identified that have no obvious mammalian counterpart. In general, amino acid identities of over 30% are generally considered sufficient to imply functional relatedness. Whether sequence homology can be used effectively to predict and assume selectivity remains to be determined. Moreover, approximately half of the entire genome sequences of various organisms are composed of putative ORFs with no known homologs. Putative functions may, however still be attributed to these ORFs by analogy. Recently, an inositol-1-synthase gene has been identified in *M. tuberculosis* [1]. This gene, absent in other eubacteria, is postulated to be present in *M. tuberculosis* through complex horizontal transfers from eukaryotes and archaea [2].

In another study, a new drug target was established in *Plasmodium falciparum* from the genome sequence. The identification of genes of the non-mevalonate pathway and the absence of HMG CoA reductase, an important enzyme of the mevalonate pathway in *Plasmodium*, led to the reconstitution of the non-mevalonate pathway of isoprenoid synthesis in *Plasmodium falciparum* [8]. The non-mevalonate pathway is found in

prokaryotes and plants but is absent in animals. The *Plasmodium* parasite, however, seems to rely solely on the non-mevalonate pathway for isoprenoid synthesis. The enzymes of this pathway are, hence, valuable targets for malarial drug development. Two compounds, fosmidomycin and FR900098 have already been identified which block this pathway and cure malaria in a mouse model of the disease [8].

Expressed sequence tags (ESTs) are at the forefront of technological change to provide a high throughput means for identifying gene transcripts and monitoring complex gene expression patterns. ESTs have increased the rate of discovery of new, biologically relevant target molecules significantly. The power of EST approach is derived from the fact that only a subset of the genome is actively transcribed in a given cell. By producing a cDNA library of the cell, clones can be isolated and partially sequenced. These partially sequenced 'tags' are then used to obtain a digital image of gene expression levels for a given tissue. These tags, in some instances, reveal the functional or disease-related importance of newly identified genes. Efforts to obtain ESTs for each gene in the human genome are in progress in various laboratories of the world and are generating massive databases for drug discovery purpose [23].

Functional prediction of the gene products is the bottleneck to the identification of new drug targets. Gene disruption, in chromosomal genes, and their characterization is the traditional approach to functional assignment of genes. Such an approach, however, is tedious, time-consuming and lacks efficiency. However, targeted gene disruption mediated by transposons is now being used on a high-throughput basis. Although attempts have been made to mutate each gene of *Saccharomyces cerevisiae* chromosome V by the traditional approach [18], the absence of phenotypic variations may make the identification of mutants difficult.

Another genetic method of identifying genes essential for viability is the use of conditionally lethal mutants. A conditional lethal mutation causes the gene product to function normally under

Features

one set of environmental conditions, but fail to function under another set. Gene products that can mutate to conditional lethality are generally considered essential for viability. Temperature sensitivity is the most studied example in the identification of essential genes in the bacterial system. One of the pitfalls of temperature sensitive mutant strategies is the differential ability of some proteins to achieve thermolabile mutant forms. In addition, a few genes have been identified that are required for viability at high temperatures, which will lead to false assignment of a few genes as essential under all growth conditions.

Computational approaches have also been used to identify the 'minimal genome set' that is necessary for viability in bacteria [13]. Comparison of the completed genomes of *Haemophilus influenzae* and *M. genitalium* led to the estimation that approximately 250 genes were necessary to support the minimal bacterial cell. However, with additional genomes sequenced, the computationally estimated minimal set continues to shrink [11], suggesting that different species may accomplish the necessary tasks using different gene set, and perform certain essential functions using structurally different proteins.

Current approaches to the assignment of function to gene products requires unraveling of their molecular properties viz. enzymatic activity, interacting partners etc. The major tools currently used to assign molecular properties to these gene products and their identification and selection as potential drug targets can broadly be classified into three main categories:

- (i) Functional genomics
- (ii) Proteomics
- (iii) Structural genomics

Functional genomics

Study of the expression behavior of gene products and assessment of their biological function is termed as functional genomics. Among the major tools for the comprehensive analysis of transcriptional expression of genes are DNA microarray methods, antisense technology and immunological methods. In the micro-array

method, cDNAs or short oligonucleotides arrayed on chips are used to probe cellular mRNA populations for gene transcripts on a high throughput basis. These tools help us "to predict a large number of novel drug target molecules. Recently, a DNA microarray containing 97% of the predicted ORFs of the *M. tuberculosis* genome was used to monitor changes in gene expression in response to the anti-tuberculosis drug isoniazid [21]. It was found that isoniazid induced several genes encoding proteins that were physiologically relevant to the drug's mode of action viz. fatty acid synthase enzymes as well as others that were not within directly affected biosynthetic pathways.

DNA antisense technology is another important tool to select gene products as potent drug targets. In this method, designed antisense molecules deplete the cell of the target mRNA. Ribozyme, a catalytic RNA molecule, is also exploited to deplete a cell of the target mRNA. Depletion of the proto-oncogene HER-2/neu mRNA exemplifies the contemporary use of ribozymes in functional genomics [8]. Controlled expression of ribozymes with a tetracycline regulated promoter diminished the cellular content of HER-2/neu mRNA and protein by >90%. This modulated expression of HER-2/neu mRNA helped in understanding the susceptibility of transformed cells to tumorigenicity. Novel immunochemical methods based on antibody-epitope interaction is another useful technique to reveal the functional role of the targeted macromolecule. In this technique, monoclonal or polyclonal antibodies are directed against a macromolecular target which may be intracellular or on the surface of cells.

Intracellular targets are now accessible to immunochemical neutralization through the use of intrabodies (intracellular antibodies). Intrabodies are single chain antibodies consisting of variable region of heavy chain and variable region of light chain [15]. cDNA constructs of intrabodies are transfected into targeted cells which on expression, neutralize the targeted molecules inside the cell. Piche and colleagues [14] used anti-Bcl-2 intrabodies to demonstrate their importance in

Features

inducing apoptosis in breast cancer cell lines.

Approaches involving ribozymes and intrabodies for molecular target validation have their own limitations. For instance, synthesis of metabolically stable ribozymes and antisense molecules is a rate limiting step. Similarly, intrabody and antibody engineering and their expression is still in infancy. In addition, transcriptional expression and regulation of a gene is very complex and not fully understood yet. Its expression may also be regulated at the level of translation or the protein product may be subjected to further control by post-translational modifications. Moreover, the presence of introns, organellar localizations and pre- and post-transcriptional modifications, such as RNA editing and splicing has further necessitated the direct identification of their gene products.

Proteomics in target selection

Proteins are the workhorses of the cell. The study of proteins as gene products of the genome and their properties viz. expression levels, post-translational modifications, interactions etc. on a large scale to obtain an integrated view of cellular processes is known as proteomics. The classical proteomics' approach aims at identification of the protein complement of the genome. Briefly, the total protein complement of a cell or a tissue is applied to a two-dimensional polyacrylamide gel electrophoresis to separate them on the basis of their charge and mass. These proteins are then identified through a matrix-assisted laser-desorption-ionization-time-of-flight (MALDI-TOF) mass spectrometry of tryptic digests of the protein spots[3]. Differences between the reference or normal and the diseased or altered states of a cell or microorganism may be measured by quantifying the ratios of spot intensities of the protein fingerprints obtained from two separately run two-dimensional gels. The proteome analysis of virulent strains of *M. tuberculosis* have recently been compared with the non-virulent *M. bovis* BCG to identify proteins for the development of vaccines, diagnosis and therapeutics [9].

A more accurate 'differential display proteomics' approach makes use of mass

spectrometric-isotope labeling. In this approach, a stable isotope is introduced into the protein sample of one of the states to be identified permitting more accurate quantifications [12]. Other experimental methods for example: co-immuno-precipitation and cross-linking for proteins are used to study the interaction of various protein molecules. Molecular biology tools viz. two-hybrid system and phage display has been introduced recently to assist proteomics. In phage display technique, a peptide or protein is expressed as fusion with a coat protein of bacteriophage, hence fused protein displays on the surface of phage while coat protein partner of fusion remain within the bacteriophage. The display of protein domains and proteins on the surface of bacteriophage represents a powerful new methodology which provides a direct physical link between phenotype and genotype [20]. Once a new drug target is identified by way of its unique function and/or structure, various screening tools are synergized to identify ligands that modulate the biological functions of these proteins.

Structure-based drug design: Structural genomics

Structure-based drug design (SBDD), an advance phenomenon of RDD, is visualization of a lead molecule, bound to specific target receptor which helps to direct structural modification of the drug candidate. The available three-dimensional structural information will help to develop targeted combinatorial libraries, based on pharmacophores that are known to be active for the target class. Recent advances in parallel combinatorial chemistry integrated with SBDD provides an empirical understanding of drug binding affinities. Knowledge of the three-dimensional structure of a protein can also throw much light on its probable function and help in the identification of new drug targets.

Structural genomics aims at the assignment of three-dimensional structures to the proteome of a cell on a high throughput basis either by computational (homology modeling) or experimental (X-ray crystallography, NMR) methods [16, 19]. One of the major goals of this high throughput structure determination is to obtain

Features

enough structures so that every family of fold is represented. This dataset of folds would help in classification of novel proteins into their respective families and hence to predict its function since structural homologs are predicted to have molecular properties similar to the homologs.

Assignment of biochemical functions has been possible to hypothetical proteins of *Methanococcus jannaschii* [6,22] of hitherto unknown function by this approach. In one such structure-based assignment of molecular function, a hypothetical protein MJO226 of *M. jannaschii*, was classified into a new fold [6]. A probable nucleotide triphosphatase function for the hydrolysis of non-standard nucleotides was proposed for this protein from its structure, which was later confirmed by biochemical analysis. Large scale predictions of structures for proteins of various organisms by homology modeling has also been attempted [4, 17]. For instance, three-dimensional structures were proposed for as many as 17% of the sequences of the yeast genome with high confidence values by homology modeling [17]. An interesting result emerged in another such study in which a large number of protein sequences across the major groups of organisms were compared for prediction and classification into different protein folds [5]. It was found that different groups of organisms have characteristically distinct distribution of folds. This difference in fold usage is directly related to biological function. Hence, the need for more structures so that proteins with unique folds and probably unique functions may be used as new and novel drug targets.

Conclusions

The completion of numerous microbial genome sequences has provided a vast number of new potential drug targets. Science has taken up the challenge to study the entire genomes by developing tools to assess the anticipated spectrum of the targets as well as the selectivity of inhibitors from the genome database. In the near future, however, we can predict a tremendous influx of truly novel targets that will require rapid and convincing validation before undertaking drug discovery efforts. Functional as well as structural

genomics along with proteomics are the emerging technologies involving DNA microarray, ribozymes and intrabodies are attractive for filtering molecular drug target options. Furthermore, structure based drug design can play a key role in the development of new drugs whose molecular targets are being discovered as a result of genome sequencing and analysis efforts.

References

1. Bachhawat N., Mande S. C. 1999. Identification of the Ino-1 gene of *Mycobacterium tuberculosis* H37Rv reveals a novel class of Inositol-1-synthase enzyme. *J. Mol. Biol.* 291, p. 531-536.
2. Bachhawat N., Mande S. C. 2000. Complex evolution of the Inositol-1-synthase gene among archaea and eubacteria. *Trends Genet.* 16, p.111-113
3. Blackstock W. P., Weir M. P. 1999. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17, p.121-127
4. Fischer D., Eisenberg D. 1997. Assigning folds to the proteins encoded by the genome of *Mycoplasma genitalium*. *Proc. Natl. Acad. Sci. USA* 94, p.11929-11934.
5. Gerstein M., Levitt M. 1997. A structural census of the current population of protein sequences. *Proc. Natl. Acad. Sci. USA* 94, p. 11911-11916.
6. Hwang K. Y., Chung J. H., Kim S.H., Han Y.S., Cho Y. 1999. Structure-based identification of a novel NTPase from *Mycoplasma jannaschii*. *Nature Struct. Biol.* 6, p. 691-696.
7. Jomaa H., Wiesner J., Sanderbrand S., Altincicek B., Weidemeyer C., Hintz M., Turbachova I., Eberl M., Zeidler J., Lichtenthaler H. K., Soldati D., Beck, E. 1999. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science*, 285, p.1573-1576.
8. Juhl H., Downing S. G., Wellstein A., Czubayko F. 1997. HER-2/neu is rate limiting for ovarian cancer growth. Conditional depletion of HER-2/neu by ribozyme targeting. *J. Biol. Chem.* 272, p.29482-29486.

Features

9. Jungblut P. R., Schaible U. E., Mollenkopf H. J., Zimny-Arndt U., Raupach B., Mattow J., Halada P., Lamer S., Hagens K., Kaufmann S. H. E. 1999. Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol. Microbiol.* 33, p.1103-1117.
10. Kerlavage A. R. 1999. TIGR Microbial genome database. <http://www.tigr.org/>
11. Koonin E.V. 1997. Genome sequences: genome sequence of a model prokaryote. *Curr. Biol.* 7, p.R656-R659.
12. Mann M. 1999. Quantitative proteomics? *Nature Biotechnol.* 17, p. 954-955.
13. Mushegian A., Koonin E. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA* 93, p.10268-10273.
14. Piche A., Grim J., Rancourt C., Gomez-Navarro J., Reed J. C., Curiel D. T. 1998. Modulation of Bcl-2 protein levels by an intracellular anti-Bcl-2 single-chain antibody increases drug-induced cytotoxicity in the breast cancer cell line MCF-7. *Cancer Res.* 58, p. 2134-2140.
15. Rondon U., Marcaso W. A. 1997. Intracellular antibodies (intrabodies) for gene therapy of infectious diseases. *Ann. Rev. Microbiol.* 51, p. 257-283.
16. Sali A. 1998. 100,000 protein structures for the biologist. *Nat. Struct. Biol.* 5, p. 1029-1032.
17. Sanchez R., Sali A. 1998. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc. Natl. Acad. Sci. USA* 95, p. 13597-13602.
18. Smith V., Chou K., Lashkari D., Botstein D., Brown P. O. 1996. Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* 274, p. 2069-2074.
19. Teichmann S. A., Chothia C., Gerstein. M. 1999. Advances in structural genomics. *Curr. Opin. Struct. Biol.* 9, p. 390-399.
20. Widnersten M., and Mannervik B., 1997. *Glutathione transferases* with novel active sites isolated by phage display from library of random mutants. *J. Mol. Biol.* 250, p.115-122.
21. Wilson M., DeRisi J., Kristensen H.-H., Imboden P., Rane S., Brown P. O., Schoolnik P. K. 1999. Exploring drug-induced alterations in gene-expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc. Natl. Acad. Sci. USA* 96, p. 12833-12838.
22. Zarembinski T. I., Hung L.-W., Mueller-Dieckmann H.-J., Kim. K.-K., Yokota H., Kim R., Kim S.-H. 1998. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc. Natl. Acad. Sci. USA* 95, p. 15189-15193.
23. Zweiger G., Scott R. W., 1997. From expressed sequence tags to 'epigenomis': an understanding of disease process. *Curr. Opin. Biotech.* 8, p684-687.

Delay Regretted

We sincerely regret the delay in bringing out this issue of the journal, due to technical reasons. We hope that our readers will bear with us and continue to encourage this venture.

Editor