



Review

Rational drug design

Soma Mandal^a, Mee'nal Moudgil^a, Sanat K. Mandal^{b,c,*}^a Manitoba Institute of Cell Biology, University of Manitoba, 675 McDermot Avenue, Winnipeg, Manitoba, Canada R3E 0V9^b Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada A1B 3V6^c Division of Science and Technology, College of the North Atlantic, Clarenville Campus, Clarenville, NL, Canada A5A 1V9

ARTICLE INFO

Article history:

Received 6 May 2009

Received in revised form 21 May 2009

Accepted 8 June 2009

Available online 14 October 2009

Keywords:

Rational drug design

Computer aided drug design

Drug targets

Gene drug toxicity

Bioinformatics tools

Docking

Lead drug like molecule

Ancient approach

ABSTRACT

In this article, current knowledge of drug design is reviewed and an approach of rational drug design is presented. The process of drug development is challenging, expensive, and time consuming, although this process has been accelerated due to the development of computational tools and methodologies. The current target based drug design approach is incomplete because most of the drugs developed by structure guided approaches have been shown to have serious toxic side effects. Otherwise these drugs would have been an ideal choice for the treatment of diseases. Hence, rational drug design would require a multidisciplinary approach. In this regard, incorporation of gene expression technology and bioinformatics tools would be indispensable in the structure based drug design. Global gene expression data and analysis of such data using bioinformatics tools will have numerous benefits such as efficiency, cost effectiveness, time saving, and will provide strategies for combination therapy in addition to overcoming toxic side effects. As a result of incorporation of gene expression data, partial benefit of the structure based drug design is slowly emerging and rapidly changing the approach of the drug development process. To achieve the full benefit of developing a successful drug, multidisciplinary approaches (approaches such as computational chemistry and gene expression analysis, as discussed in this article) would be necessary. In the future, there is adequate room for the development of more sophisticated methodologies.

© 2009 Elsevier B.V. All rights reserved.

Contents

1. Introduction	90
2. Rational drug design	91
2.1. Structure guided–computer aided drug design	91
2.2. Reactivity	93
2.3. Known targets for cancer therapy	93
2.4. State-of-the-art methods in drug design	95
2.5. Side effects of currently available structure based drugs	96
3. Ancient approach	96
4. Combining gene technology, bioinformatics tools in rational drug design	97
4.1. Global gene expression profiling	98
4.2. Global gene expression analysis—Bioinformatics tools	98
5. Conclusion	99
Acknowledgments	99
References	99

1. Introduction

Drugs are essential for the prevention and treatment of disease. Human life is constantly threatened by many diseases such as cancer. Therefore, ideal drugs are always in great demand. To meet the

challenges of ideal drugs, an efficient method of drug development is demanding. The process of drug development is challenging, time consuming, expensive, and requires consideration of many aspects. To fulfill these challenges, several multidisciplinary approaches are required for the process of drug development; collectively these approaches would form the basis of rational drug design. A drug target is a biomolecule which is involved in signaling or metabolic pathways that are specific to a disease process. As a prime example, a drug target would be a biomolecule (for example epidermal growth factor receptor)

* Corresponding author. Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada A1B 3V6. Tel.: +1 709 466 6914; fax: +1 709 466 2771.

E-mail addresses: sanat.mandal@cna.nl.ca, sanat_mandal@yahoo.com (S.K. Mandal).

that is frequently mutated or otherwise deregulated in the disease of cancer. Biomolecules play critical roles in disease progression by communicating through either protein–protein interactions or protein–nucleic acid interactions leading to the propagation of signaling events and/or alteration of metabolic processes. Therefore, modulation of biological functions performed by these biomolecules would be potentially beneficial and could be achieved either (i) by inhibiting their function with small molecules whose competitive binding affinity would be greater than their natural ligands that bind to the active sites (within the biomolecules), or (ii) by inhibiting the bimolecular interactions by small molecules (between the biomolecules, relatively less studied) (Fuller et al., 2009), to stop cross talks between biomolecules, or (iii) by activating biomolecules (for normal functions) that are functionally deregulated in some diseases such as cancer. Developing a lead molecule and an effective drug (small molecules with desired properties) is challenging even for known targets. Recently, drug discovery has significantly increased due to the availability of 3D X-ray or NMR structures of biomolecules, docking tools, and the development of computer aided methodologies (Greer et al., 1994; Müller, 2009; Henry, 2001). Currently, the Protein Data Bank (PDB) holds about 57,558 3D structures, but even this high number is insignificant. Some of the biomolecules have more than one structure bound to different molecules. 3D structures of many important targets are still unknown. Similarly, number of lead drug like molecules is also relatively less. Thus an improved approach of rational drug design is necessary to overcome problems associated with currently available drugs that are developed based on the sole approach of structure guided drug design.

2. Rational drug design

Rational drug design can be broadly divided into two categories:

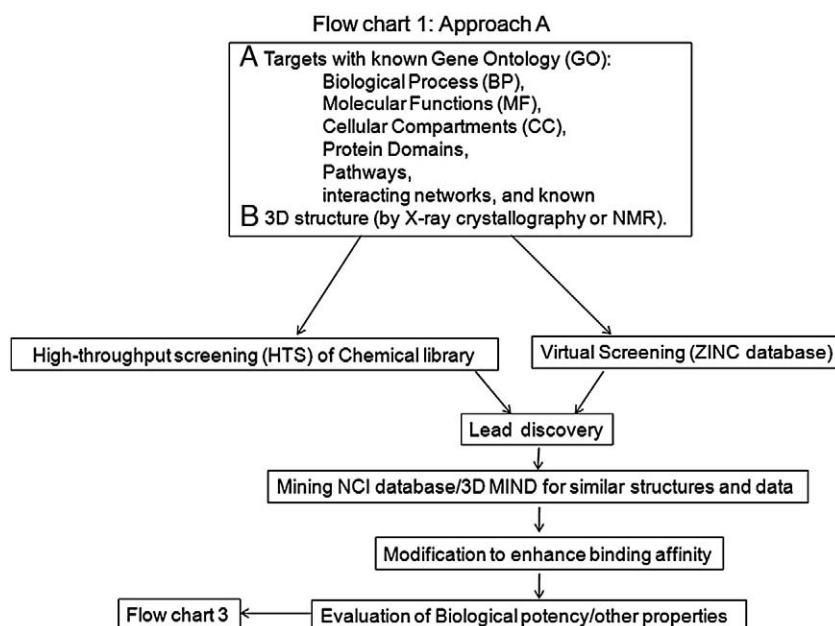
- (A) Development of small molecules with desired properties for targets, biomolecules (proteins or nucleic acids), whose functional roles in cellular processes and 3D structural information are known. This approach in drug design is well established and is being applied extensively by the pharmaceutical industries.
- (B) Development of small molecules with predefined properties for targets, whose cellular functions and their structural information may be known or unknown. Knowledge of unknown targets (genes and proteins) can be obtained by analyzing global gene

expression data of samples untreated and treated with a drug using advanced computational tools (refer to the following sections). Steps related to these two approaches and evaluation of other properties in rational drug design are presented in the following flow charts (Flow charts 1, 2, and 3).

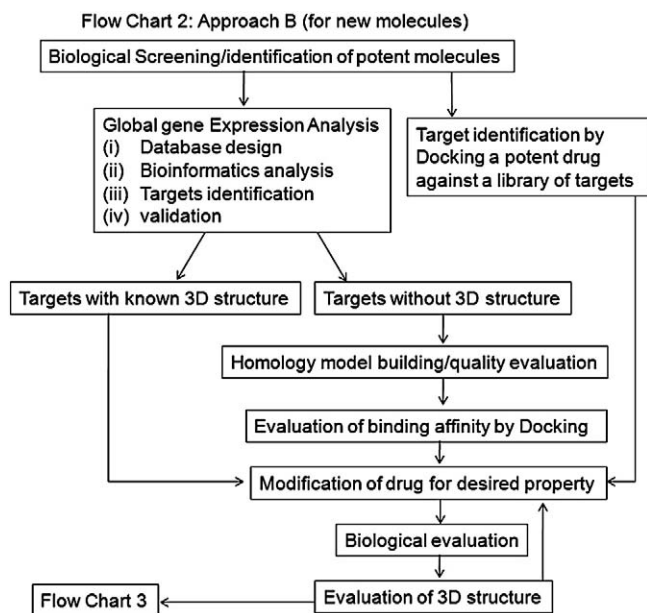
Once a target is identified, then both approaches (A) and (B) for development of small molecules require examination of several aspects (Flow chart 3). These aspects include, but are not limited to, the evaluation of binding scores (affinity/specificity), balance between hydrophilicity/lipophilicity, absorption, distribution, metabolism, and excretion (ADME), electrophilic, nucleophilic, and radical attack (biodegradation), toxicity of the parent small molecules, and products due to biotransformation in the different phases of metabolism, quantitative structure–activity relationship (QSAR), and quantitative structure–property relationship (QSPR) respectively. Most of these aspects including design of a small molecule could be performed initially using computational tools. After the initial evaluation and identification of lead molecules, gene expression profiling and bioinformatics analysis would be particularly important to gain insights in gene expression patterns. In turn, this knowledge can be utilized to improve drugs to accomplish desirable attributes such as disease free survival, eradication of disease, elimination or minimization of toxic side effects, reduction of undesirable biotransformation, improvement in distribution (bioavailability), overcoming of drug resistance, and improvement of immune responses. Therefore, rational drug design would be an integral approach to drug development and discovery.

2.1. Structure guided–computer aided drug design

Structure guided methods are an integral part of drug development for known 3D structure of potential drug binding sites, which are the active sites. In structure guided drug design, a known 3D structure of a target bound to its natural ligand or a drug is determined either by X-ray crystallography or by NMR to identify its binding site, the so called active site. For a lead discovery, this is the starting point of structure guided drug design for a known target. Once the ligand bound 3D structure is known, a virtual screening of large collections of chemical compounds, such as ZINC (Irwin and Shoichet, 2005), can be performed. Such screening enables the identification of potential new drugs by



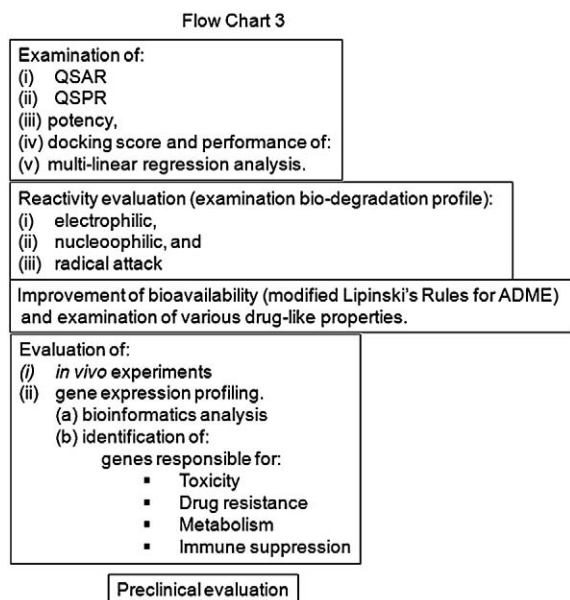
Flow chart 1. Shows the possible steps in drug design for known targets.



Flow chart 2. Shows the possible steps in drug design for unknown targets.

performing docking experiment of this collection of molecules. To enhance binding and hence to improve binding affinity/specificity, a group of molecules with similar docking scores is generally used for potency determination; this is High-Throughput Screening (HTS) (Flow chart 1). After the determination of biological potency, several properties such as relationships (QSAR, QSPR, between potency and docking scores) including statistical analysis can be performed to ascertain the potential molecule(s) for lead drug discovery. Before optimization, the lead molecules could be examined further to understand the ADME and reactivity. Investigation of reactivity (examination of electrophilic, nucleophilic or radicals attack) and spectra such as UV-Visible of large molecules can be performed applying Gaussian, a powerful quantum mechanical procedure.

Instead of virtual screening of a collection of small molecules, a virtual screening of a collection of targets against a single potent drug whose target is unknown could be performed. Such screening would



Flow chart 3. Shows examination of additional properties towards the improvement of drug like properties.

help in the identification of a potential target for the potent drug. At the end a target identified by this docking technique must be verified experimentally. To our knowledge, this approach of identification of target for a potent drug has not been applied. This approach of target identification for a potent drug with unknown target offers a unique opportunity for lead discovery. In our ongoing studies, we are applying this approach in combination with gene technology for target identification.

Examination of reactivity will allow further modification (functionalization) of the lead molecule for preventing its possible undesirable metabolic fate. Reactivity of a molecule is mainly of three types; these are attacked either by electrophiles (*electron-lover*), nucleophiles (*nucleus-lover*), or by radicals (atoms, molecules or ions with unpaired electrons) respectively. As an example, the water molecule can act as a nucleophile. Therefore, the modulation of degradation pathways could be achieved through functionalization/modification of the lead, without compromising its potency.

Besides the evaluation of potency, binding affinity/specificity as well as other properties including drug like properties (pharmacokinetics) such as log *P*, molar refractivity, number of hydrogen donor and hydrogen bond acceptor and molecular weight are also determined (Flow chart 3). These parameters are important molecular properties as formulated by Lipinski et al (Lipinski et al., 1997) and later developed by Ghose et al. (1999). Toxicity predictions of the drug itself and its metabolic products can also be examined initially by computational methods; however these properties should be verified by experimental methods. A small molecule must possess several properties to be a drug like molecule to exert the desired pharmacological effect and thus, rational drug design should focus in the evaluation of all such properties.

Several successful drugs have been developed applying the drug design approach and some of them are already in use in the market. In this regard, the development of imatinib (compound 3, Chart 1) is worth mentioning. This drug has been used to treat certain types of cancers including chronic myelogenous (or myeloid) leukemia (CML). A high-throughput screening of chemical libraries was performed to identify the starting molecule, 2-phenylaminopyrimidine. 2-phenylaminopyrimidine served as a lead compound, which was then tested and modified to develop imatinib (Druker and Lydon, 2000). Imatinib has enhanced binding properties. However, imatinib is much less effective in patients with mutation, and these patients represent with high number of cases with mastocytosis. Similarly, structure based drug design is also applied to develop inhibitors for different diseases including HIV/AIDS, (ritonavir is used as a protease inhibitor), hepatitis C (infectious disease), malaria, and for other diseases. Protease inhibitors are an important class of drugs which are used as antiretroviral agents. These inhibitors are used to treat or to prevent viral infection including hepatitis C and HIV/AIDS. The drug ribavirin is worth mentioning; ribavirin is used in treating hepatitis C in combination of interferon α . But ribavirin produces serious adverse side effects. Originally ritonavir was developed as an HIV protease inhibitor, but currently it is used as a booster for other protease inhibitors. Chart 1 shows the chemical structures of few anti-cancer drugs (1 to 5) that were developed applying target based knowledge, some of which are in clinical use.

These drugs are the only currently available choice of treatment despite their toxic side effects. If the undesirable side effects of these drugs could be removed, then the same drugs could be more beneficial as mentioned earlier. The drug nilotinib (compound 4, Chart 1) is the second generation of the drug imatinib and has been used to treat CML patients, who are BCR-ABL positive. BCR-ABL is a good target for the treatment of some forms of leukemia and investigation is ongoing for further drug development. Nilotinib has an improved profile than imatinib but not free from toxic side effects. Fig. 1 shows the binding site of the drug nilotinib (Weisberg et al., 2005; Cowan-Jacob et al., 2007) to its target, the BCR-ABL. The binding affinity and specificity of nilotinib has improved significantly by modifying based on the active site of BCR-ABL. Fig. 2 shows an approach (adjacent surface) of a target guided

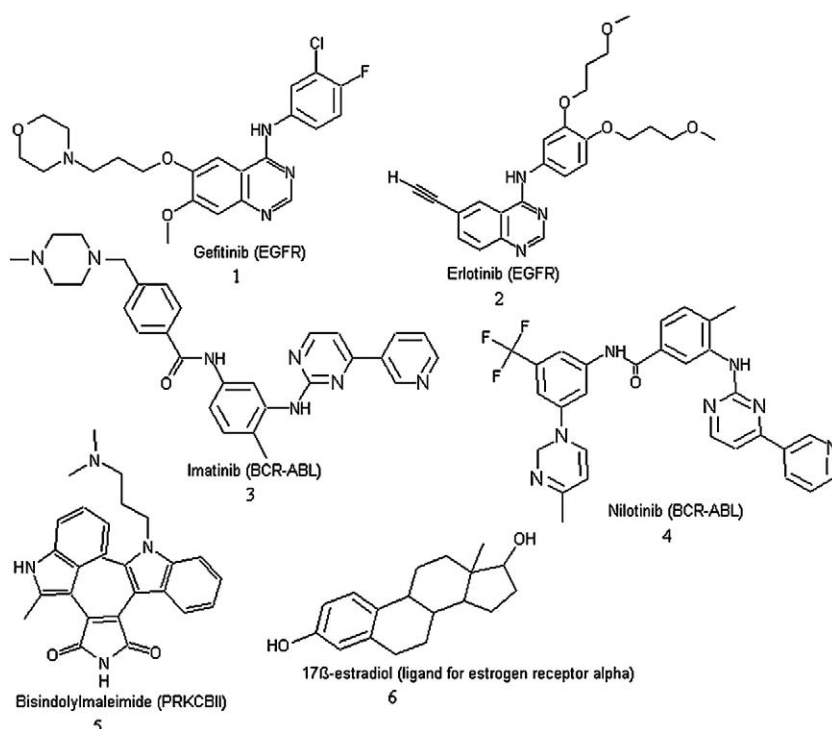


Chart 1. Structure of few drugs (1 to 5) developed based on targets: name of the drug followed by its target inside the parenthesis is presented as drug name (target); compound 6, 17β-estradiol, is a ligand for estrogen receptor alpha.

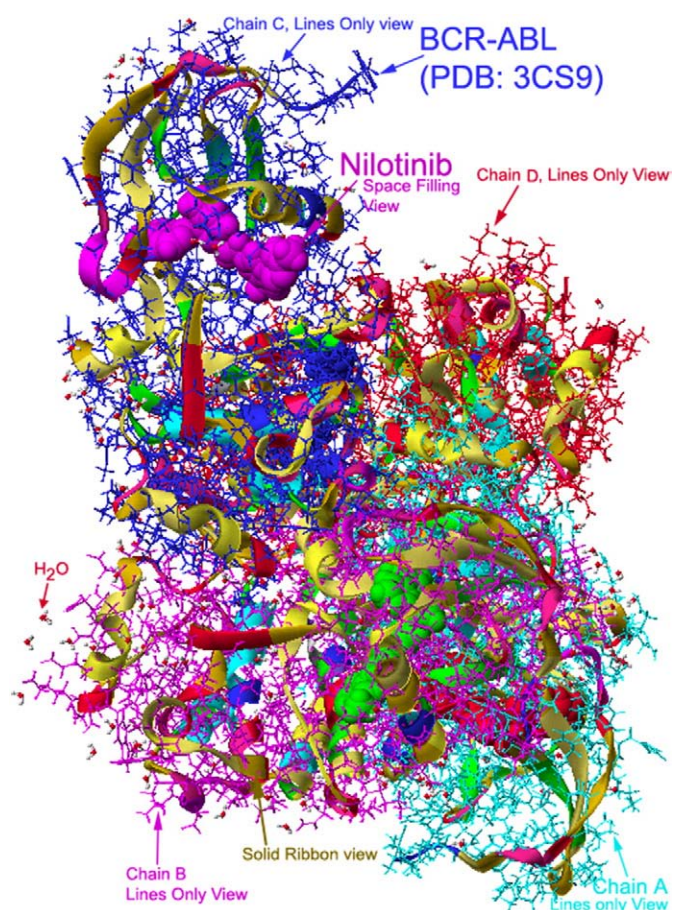


Fig. 1. BCR-ABL kinase domain: showing the binding pocket of nilotinib (purple) bound to the active site of the target BCR-ABL in Chain C. In this figure, BCR-ABL is the cluster of four chains (chain A [green], chain B [purple], chain C [blue], and chain D [red]).

computer aided drug design. This figure shows nilotinib within 3 Å of the active site of the target, the BCR-ABL kinase domain. The adjacent surface provides information necessary for modification of bound drug. Red marks a surface where the protein needs H-acceptors (e.g.—C O), blue marks a surface where the protein needs H-donors and cream marks the hydrophobic surface (refer to the web version for color interpretation). The pocket surface is colored so that it is easy to design ligands. Ligands that bind well should have H-bond acceptors (e.g.—C O) groups touching the red surface and N—H groups with the N—H bond poking through the surface.

2.2. Reactivity

To understand the reactivity of the nilotinib as an example, the susceptibility to an electrophilic, or a nucleophilic, or a radical attack is generated by a MO-G/AM1 wavefunction for nilotinib, at a geometry determined by performing an optimized gradient calculation in MO-G using AM1 parameters. The nucleophilic or the electrophilic or the radical frontier density (Fukui et al., 1954) measures the susceptibility of the substrate to attack by a nucleophilic or an electrophilic or a radical respectively. It reveals reactive sites based on the electron distribution of a set of active orbitals near the HOMO and LUMO. It is especially useful for large molecules where several orbitals may have energies nearly equal to the HOMO and LUMO. Fig. 3 shows the possible reactive site for nilotinib. To improve accuracy, these susceptibility attacks could be calculated by performing a DGauss/DFT (density function theory) or by Gaussian procedures in the presence of solvents. Calculations using Gaussian procedure in the presence of solvent consume significant amount of time.

2.3. Known targets for cancer therapy

Besides structure based drug design for the target BCR-ABL (Padmanabhan et al., 2008), this structure based drug design approach has been applied for several other targets. A few such targets are P-glycoprotein (Aller et al., 2009), vascular endothelial growth factor

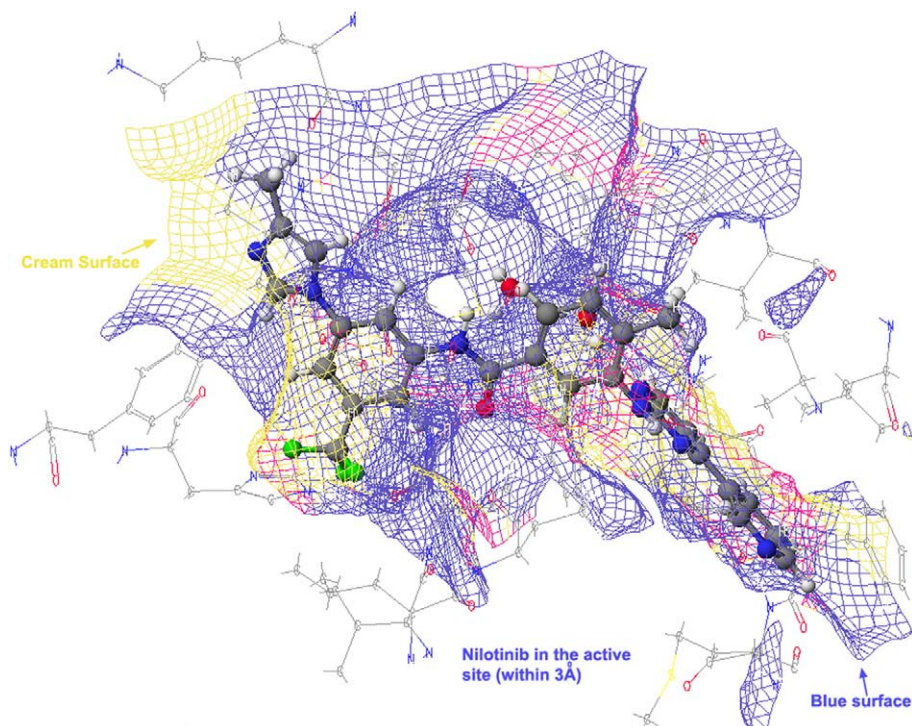


Fig. 2. Adjacent surface for nilotinib: shows the adjacent surface pocket that is the surface within 3 Å of the drug to the active site of the target (BCR-ABL). This surface provides guidance in the process of drug development.

receptor 2 (La et al., 2008), protein kinase C beta II (Grodsky et al., 2006), BCL2 B-cell CLL/lymphoma 2 (Nickells et al., 2008), tumor protein p53 (Tp53) (Mandal et al., 2007a,b), estrogen receptor (Bazer et al., 2009; Gupta et al., 2008; Provencher-Mandeville et al., 2008; Gagnon et al., 2004; Desco^{teaux} et al., 2003), epidermal growth factor receptor (Mandal et al., 2002; Sharma et al., 2009; Nautiyal et al., in press; Ricciardi et al., 2009). These targets are considered as potential targets and provide opportunity for drug development directed towards cancer or any other disease treatment. These targets and other targets that are unknown at present could serve as potential target candidates for drug discovery be-

cause they are involved in signaling pathways operating to accomplish different cellular events such as cell growth, differentiation and proliferation. Fig. 4 (recreated using Microsoft Office PowerPoint 2007) (Harvey, 2003) demonstrates the involvement of signaling pathways currently known for the above mentioned targets. These targets are involved in a number of pathways as tabulated in Tables 1 and 2.

Among these targets, the epidermal growth factor (EGF) receptor has been studied extensively due to its aberrant expression in certain types of cancer such as non-small cell lung cancer and breast cancer. EGF receptor is known to interact directly or indirectly with other 151

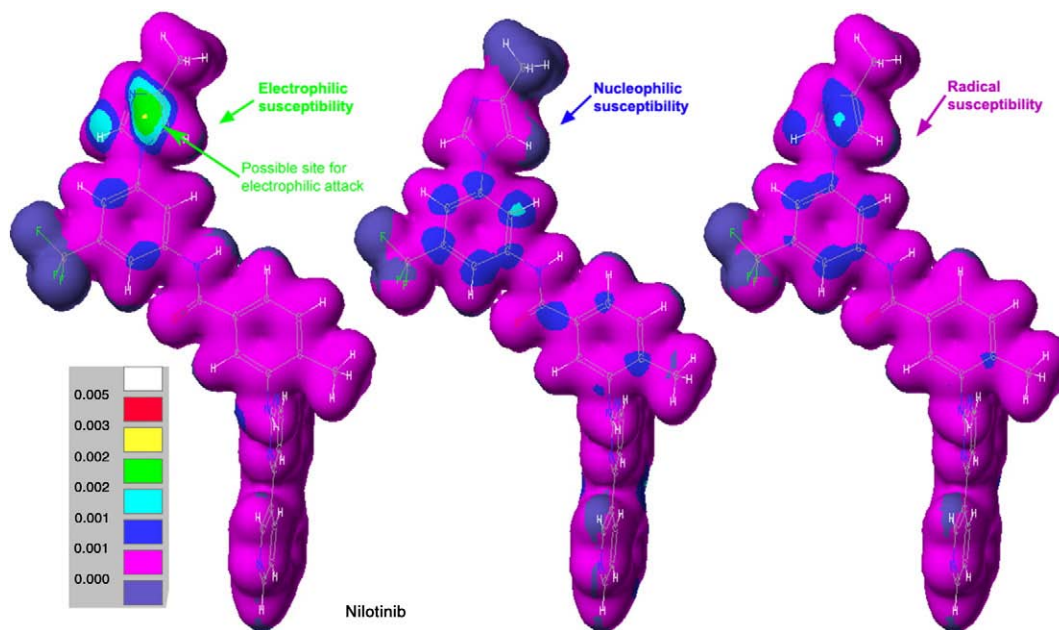


Fig. 3. Reactivity for nilotinib: shows the susceptibility towards the electrophilic, nucleophilic and radical attack for the drug nilotinib. The susceptibility to an electrophilic, or a nucleophilic, or a radical attack is generated by a MO-G/AM1 wavefunction for nilotinib, at a geometry determined by performing an optimized gradient calculation in MO-G using AM1 parameters.

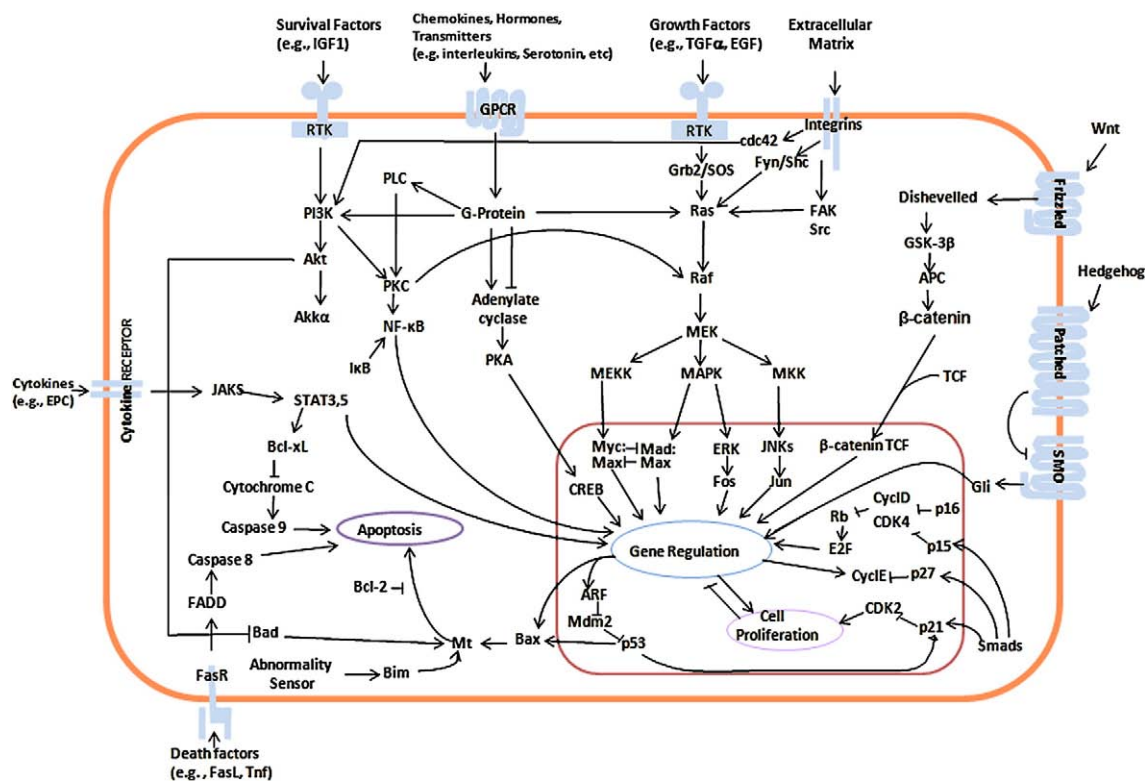


Fig. 4. Signal transduction pathway: shows the known signaling events performed by various target molecules.

proteins, but it also represents one of the most common genetic aberrations in various malignancies. Some of the most common genetic aberrations of EGF receptor comprise of somatic mutation or gene amplification which lead to the abnormal expression of the receptor protein. Somatic mutations of EGF receptor occur in a very high frequency in lung cancer. Fig. 5 shows the different types of EGF receptor mutations.

As shown in Tables 1 and 2, EGF receptor is involved in a number (19) of different pathways representing one of the most important targets. Fig. 6 shows the KEGG pathway for the MAPK signaling pathway in which EGF is involved as one of the signaling molecules (EGF is the natural ligand for EGF receptor) and mediates the signaling events by binding to the EGF receptor.

A representative interacting network for EGF receptor and other proteins is also shown in Fig. 7. This network suggests that an ideal drug or a combination of drugs would be specific to disease processes and will not disturb the intricate balances maintained by normal cellular processes. In the interacting network, each terminal gene/protein further interacts with other genes/proteins and these interactions continue as a cascade. The status (up/down/unchanged) of interacting

genes/proteins can be computed by performing analysis of gene expression data of treated versus untreated and can be visualized using bioinformatics tools (Ingenuity or BioGRID) to examine their status.

Targeted therapies for the management of non-small cell lung cancer are mostly directed at the EGF receptor. These include monoclonal antibodies against the EGF receptor protein (cetuximab) and EGF receptor tyrosine kinase inhibitors (gefitinib, erlotinib). Gefitinib (compound 1, Chart 1) or erlotinib are generally well tolerated with relatively less severe systemic side effects usually seen with cytotoxic drugs (Ricciardi et al., 2009). Currently gefitinib is an important drug and is a choice for treatment for lung cancer, besides radiation. Fig. 8 shows the tyrosine kinase domain of the EGF receptor in complex with gefitinib. Interestingly, gefitinib was developed based on structure guided design.

An important target for the treatment and prevention of cancer are the protein kinases (PK), which play important roles in many of the malignancy promoting processes. Aberrant PK activity is associated with many forms of cancer and thus these proteins are attractive targets for anti-cancer drug development. As shown in Table 1, PKs are also involved in 19 different pathways. Bisindolylmaleimide (Compound 5, Chart 1) is one of the lead molecules for the PK domain. Another important target is BCL2. BCL2 encodes for an integral outer mitochondrial membrane protein that blocks the apoptotic death of some cells such as lymphocytes. Overexpression of this gene is observed in many forms of cancer. Hence, this is an attractive target for drug development for the treatment or prevention of cancer (Frenzel et al., 2009).

2.4. State-of-the-art methods in drug design

In rational drug design, for targets with known 3D structure, one must be familiar with the resources that are useful for this purpose. These resources including the PDB, the NCI (National Cancer Institute, USA) database, DTP (Development Therapeutics Program), the 3D MIND, and the ZINC database of molecules, are required for virtual screening respectively. Users should also be equipped with docking

Table 1

A tabulation of the selected targets and number of known pathways for each target.

Target	Pathways	Example
Epidermal growth factor receptor (EGF receptor)	Involved in 19 pathways	Non-small cell lung cancer
Protein kinase C beta II (PRKC β II)	Involved in 19 pathways	Non-small cell lung cancer
Vascular endothelial growth factor receptor 2 (VEGFR2)	Involved in Three pathways	VEGF signaling pathway
BCL2 B-cell CLL/lymphoma 2 (BCL2)	Involved in eight pathways	Small cell lung cancer
P-glycoprotein (P-gp)	Transporters	ABC transporters
Estrogen receptor alpha (ER α)	Involved in five pathways	Role of ERBB2 in signal transduction and oncology

Table 2
Known pathways for the selected targets.

EGF receptor	PRKCB	VEGFR2	BCI-2
Non-small cell lung cancer	Non-small cell lung cancer		
Bladder cancer			
Melanoma			
Prostate cancer			Prostate cancer
Glioma	Glioma		
Endometrial cancer			
Pancreatic cancer			
Colorectal cancer			Colorectal cancer
Cytokine–cytokine receptor interaction		Cytokine–cytokine receptor interaction	
ErbB signaling pathway	ErbB signaling pathway		
MAPK signaling pathway	MAPK signaling pathway		
GnRH signaling pathway	GnRH signaling pathway		
Calcium signaling pathway	Calcium signaling pathway		
Dorso-ventral axis formation			
Epithelial cell signaling in <i>Helicobacter pylori</i> infection			
Regulation of actin cytoskeleton			
Gap junction	Gap junction		
Adherens junction			
Focal adhesion	Focal adhesion	Focal adhesion	Focal adhesion
	B-cell receptor signaling pathway		
	VEGF signaling pathway	VEGF signaling pathway	
	Tight junction		
	Wnt signaling pathway		
	Phosphatidylinositol signaling system		
	Melanogenesis		
	Long-term depression		
	Long-term potentiation		
	Leukocyte transendothelial migration		
	Fc epsilon RI signaling pathway		
	Natural killer cell mediated cytotoxicity		
			Amyotrophic lateral sclerosis (ALS)
			Apoptosis
			Neurodegenerative disorders
			Prion disease
			Small cell lung cancer

tools such as AutoDock, DOCK6, and Molecular Modeling tools (computational chemistry software packages, such as Gaussian), to solve complex chemical problems. All these valuable resources may provide an excellent advantage for rational drug design.

The PDB contains information about experimentally determined structures of proteins, nucleic acids, drug–protein, drug–nucleic acid, and protein–nucleic complex assemblies. Drug-bound target molecules can be searched in a number of ways. The DTP is a valuable database for dose response bulk data for over 100,000 small molecules which can be downloaded for any molecule using a NSC number (number assigned by the NCI for a compound). Bulk data is a valuable source for the examination of QSAR for lead discovery.

Two other very useful tools used for drug design are the 3D MIND tool and the OSIRIS Property Explorer respectively. The 3D MIND is a database searchable tool providing information of the cytotoxic potency of over 100,000 small molecules for 60 human cancer cell lines (NCI, USA), cellular gene expression data for these cell lines, and protein–ligand information of possible targets for these small molecules. The 3D MIND tool is user friendly; it can accept an NSC number, the structure of a small molecule that can be sketched as a searchable query or in the form of smile. This tool provides valuable information (structure of small molecules and cytotoxic potency of most of the compounds unless protected by a secrecy agreement between the NCI and the inventors), and can significantly facilitate the lead discovery process towards drug discovery. The OSIRIS Property Explorer can be used to draw chemical structures, to predict various drug like properties. This tool is also capable of predicting properties associated with high risk of undesired effects like mutagenicity or poor intestinal absorption.

Tools mentioned above can be used in the early developmental stage to evaluate lead-like properties of compounds. Evaluation of these properties will save time and would be cost effective and perhaps be a

guide to a better starting point. The web links of some of the important resources and tools are given below the references (web links).

2.5. Side effects of currently available structure based drugs

Many drugs developed using structure based criteria have been discontinued due to variety of reasons. These reasons include safety problems, adverse toxic side effects, cardiac toxicity, and development of drug resistance and so on. Hence, it is worth mentioning that a relatively high docking score or binding affinity does not necessarily mean that a substance is going to be a potent inhibitor for that target or it would be free from undesirable side effects. However, these criteria provide valuable information for drug design.

As mentioned earlier, gene expression profiling/protein profiling and advanced computational tools can be used to gain insights to overcome adverse side effects of drugs. This can be achieved within the existing classes of structure based drugs through modification of parent drugs or by the application of combination therapy; this in turn, is based on genes that are expressed due to drug treatment and the expression of such genes is undesirable.

3. Ancient approach

Ancient Asian (China and India) herbal medicine used the concept of combination. Medicinal preparations had a combination of herbs for the purpose of their recommended usage and for achieving high potency for cure and well being of people (Kong et al., 2009; Samy et al., 2008; Patwardhan and Bodeker, 2008; Garodia et al., 2007). An example of such a recommended combination in Ayurveda (system of traditional Indian medicine) is as follows: *Azadirachta indica* (bark) 20%, *Bauhinia variegata* (bark) 15%, *Crataeva nurvala* (bark) 15%, *Terminalia chebula* (fruits) 15%,

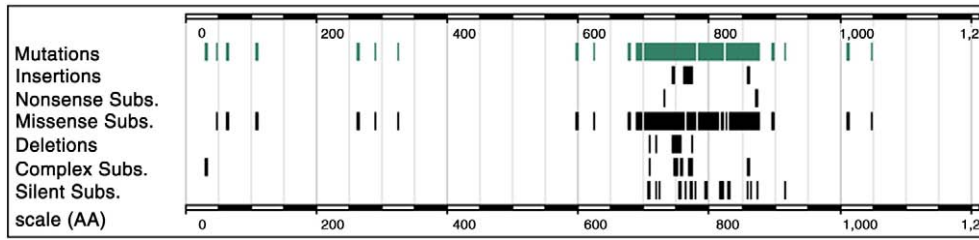


Fig. 5. Types of mutation of EGF receptor: shows the different types of mutation and their position in the EGF receptor.

Terminalia bellerica (fruits) 10%, *Holarrhena antidysenterica* (bark) 10%, and *Tinospora cordifolia* (stems) 15%. Four grams of a mixed powder (combination) made up of these above herbs should be given to the patient two times a day (morning and night) with lukewarm honey for the treatment of cancer (Samy et al., 2008). The principle is that besides treatment of a specific disease (as in this example), a preparation of herbs or other medicinal components in combination would also ward off the side effects of any particular individual component. A similar approach of combination can also be applied in modern medicine. A combination of drugs could be developed for disease prevention, treatment, and for the improvement of the quality of life. With the advent of high profile

genomic and proteomic approaches and tools, a detailed analysis of gene expression can be obtained. This knowledge serves as a valuable resource for gaining insights to combination therapy.

4. Combining gene technology, bioinformatics tools in rational drug design

This process of combined approaches of drug development would be particularly important for a number of reasons. To list a few, benefits would range from improvement in disease free survival, containment/eradication of disease, elimination/minimization of toxic side effects,

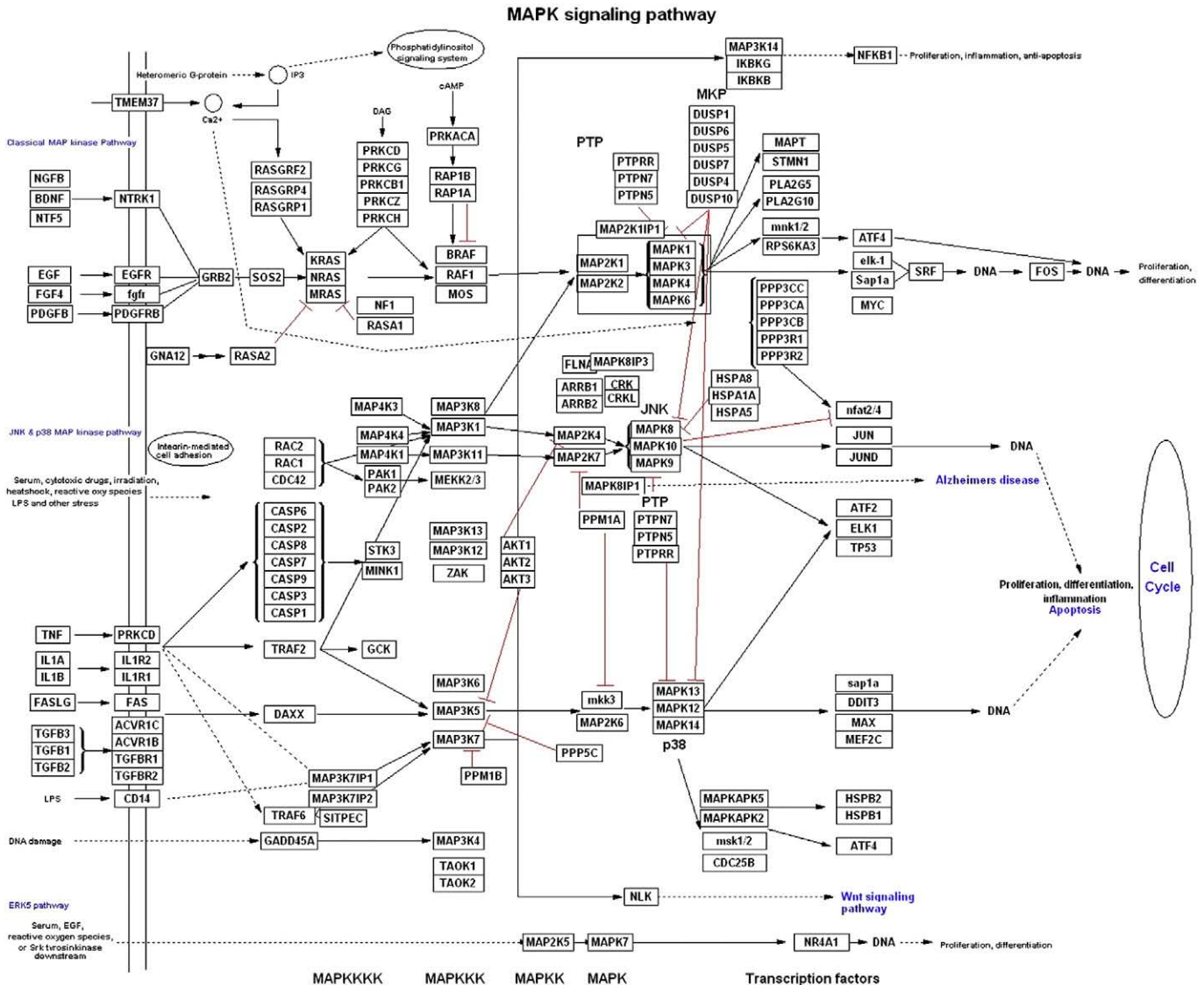


Fig. 6. MAPK signaling pathway (KEGG): shows one of the pathways in which the EGF receptor plays a critical role.

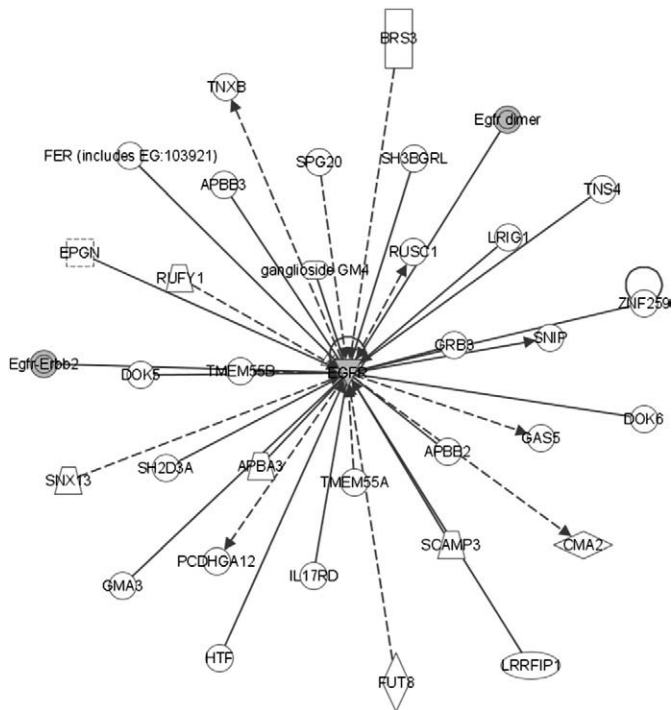


Fig. 7. Protein–protein interaction network for EGF receptor: shows the interacting network between EGF receptor and other proteins.

reduction in biotransformation, improvement in distribution (bioavailability), overcoming drug resistance, and improvement of immune responses respectively.

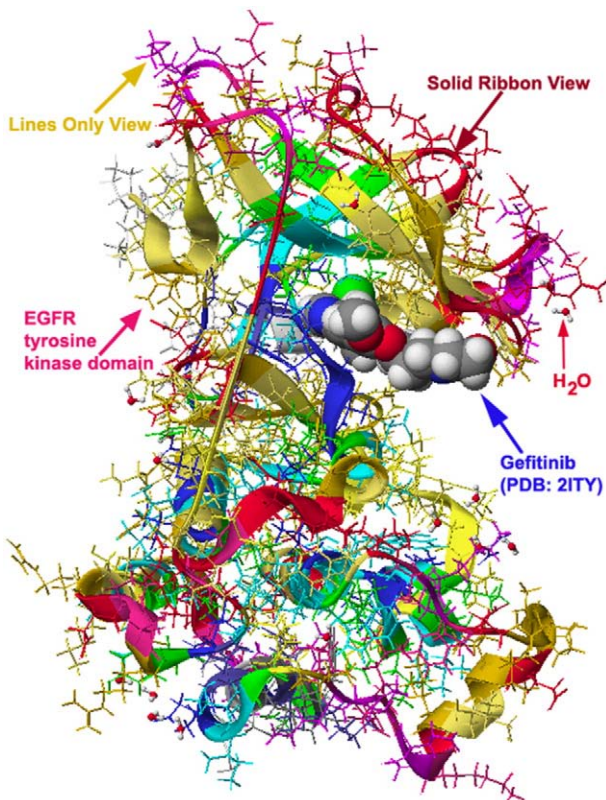


Fig. 8. The tyrosine kinase domain of the EGF receptor: showing the binding of gefitinib.

4.1. Global gene expression profiling

Global gene expression profiling is an invaluable technology which reveals novel insights into the pathogenesis of diseases including cancers. This is accomplished through the identification of distinct molecular subtypes of the disease in groups that were classified as homogenous diagnostic categories based on existing classical clinicopathological parameters. Many such platforms have been developed over the past few years. The most widely used global gene expression platforms include SAGE (serial analysis of gene expression) and Microarray (MA) respectively. Both these platforms have their own pros and cons, nevertheless, these platforms have provided empowerment to the scientific community to elucidate several aspects of the biology in question from a global perspective. SAGE produces a comprehensive gene expression profile without an *a priori* gene sequence information, which results in the identification of novel transcripts (Velculescu et al., 1995; Yamashita et al., 2008). The MA platform also analyses genome-wide gene expression patterns and is a very promising technology used to discern the classification of minute details present within a diversity of various tumor populations of a particular organ. Nevertheless, a combination of both platforms can be successfully used to harness the intricacies of gene expression data (Mandal et al., 2007c; Mandal and Davie, 2007). The end result is rewarding as minute differences even in seemingly similar cellular types can be accurately determined (Mandal and Davie, 2007; Yang et al., 2009). Over the last few years, a variety of MA platforms have become commercially available. The MA platform is gaining popularity due to the cost effectiveness, reproducibility, and speed of data acquisition. These include high-density chips such as cDNA arrays, Affymetrix arrays, and oligo nucleotide/cDNA chips respectively. These high-density MA platforms produce a global view of the gene expression patterns present in the tissues under scrutiny (Orr and Scherf, 2002). However, to harness the fullest potential of such massive data produced by SAGE or MA, relational databases management systems, and software interfaces are now available to incorporate and analyze the biological information in a meaningful way. Collectively, these tools are termed as bioinformatics tools. Bioinformatics tools allow the orderly arrangement of the global gene expression data, such that genes can be grouped into various pathways and interaction networks.

4.2. Global gene expression analysis—Bioinformatics tools

When the interpretation of global gene expression data through manual in-depth inspection and literature research is coupled with the use of data analysis tools, the whole process can be easier and efficient. Data analysis tools would allow researchers to visualize gene expression data by various ways such as by gene grouping, by pathways, or by protein–protein interactions by functional categories. In addition, certain databases also provide researchers with repositories of global gene expression data that are publicly available. Examples of such databases include the Gene Expression Omnibus (GEO), SAGE data repository (Absolute Level Lister, SAGEMap), Stanford microarray database, and many other similar such databases.

An essential component of arrangement of gene expression data is the grouping of genes. This enables the identification of gene signatures within the annotated genes. Annotation tools such as DAVID (Data Annotation Validation and Integrated Discovery) are very useful to annotate a number of genes for Gene Ontology, protein domain, and pathways respectively (Dennis et al., 2003). This can be logically followed by the identification of gene signatures within annotated genes, which can be performed using cluster analysis (Cluster/TreeView tool, European Bioinformatics site, EBI) (Eisen et al., 1998). This gene cluster tool is an important tool and has been used extensively. Another useful computational tool is the Gene Set Enrichment Analysis (GSEA) tool. GSEA determines whether an *a priori* defined set of genes shows

statistically significant and concordant differences between two biological states (for example, treated vs. untreated or normal vs. abnormal states). Annotated genes can be visualized in a pathway comprised of the protein products of the annotated genes, using the pathway visualization tool, GenMAPP (Dahlquist et al., 2002). Yet another way to visualize genes is by network analysis. This can be done in a very pictorial and informative way using the Ingenuity Pathway Analysis (IPA) tool, which is an excellent data mining tool. However to use this tool, the institution needs a site license. All other tools mentioned in this section are publicly available with academic license and are being upgraded on a regular basis for the benefit of the scientific community. These tools are usually user friendly. However, bioinformatics is an ever expanding area and both publicly and commercially available software are now within the reach of investigators, interested in gene expression data mining. The use of appropriate tools will be determined by the end user according to the data that needs to be harnessed. In the context of rational drug design, a thorough knowledge of genes up/down-regulated or genes with unchanged expression is important especially to find biological targets for a lead molecule. Gene expression data for normal, treated and untreated states can be compared to identify those genes that need further attention. Information on gene expression is helpful either to modify the experimental drug or to develop strategies for combination therapy. This process of evaluation may have to be repeated a number of times or until the desired result can be achieved based on the outcome of the combination treatment. In addition, knowledge of database management systems (DBMS) such as Microsoft Access or other such DBMS is desirable because each bioinformatics tool requires different input file format. DBMS are important software for manipulation of different databases. To prepare input file for each bioinformatics tool requires the use of a combination of Microsoft Access and Excel. Excel is excellent software; however use of Excel for gene expression data manipulation requires attention because error in gene symbols can happen due to the automatic conversion features of Excel. One such example is the gene name septin 9 and its symbol SEPT9; Excel automatically tabulates it as Sep-09. Similar mistakes generally happen for other genes as well lending a chance for misinterpretation of gene data. Thus successful data mining sometimes requires careful manual inspection, especially when using these DBMS.

5. Conclusion

Cancer has common characteristics such as proliferation, negative regulation of proliferation checking cellular mechanisms such as cell death (apoptosis, necrosis etc.) or growth arrest and enhancement of proliferation promoting cellular mechanisms such as invasion, angiogenesis and metastasis. Cancer is one of the most common causes of death which affects people at all ages. Hence the development of new drugs for this disease needs continuous effort on the part of researchers and pharmaceutical companies. Rational drug design would be a multidisciplinary approach in developing drugs that will help in combating disease to improve the quality of life, and give better scope for the prospect of a disease free survival. This approach of drug design can be applied to develop drugs to treat a wide variety of diseases and can also be used for designing drugs for disease prevention. Except for the toxic side effects, drugs currently in clinical use are quite tolerable. These drugs in combination with other drugs could be beneficial if one performs a comparative analysis of gene expression profiles generated from drug treated and untreated individuals. Such analysis of gene expression profiles will help not only to identify the expression of undesirable genes resulting from drug treatment, but will also help to identify the suppression of disease promoting genes, and also will provide guidance for combination therapy. Gene expression profile could be examined *in vitro*, *in vivo*, and in clinical settings. An initial assessment of differences in gene expression profiles could be made from *in vitro* and *in vivo* study models. However, these models do not represent the exact nature of

the human cellular system. Therefore, clinical gene expression would be ideal for identifying gene expression profiles resulting from drug treatment.

Acknowledgments

We thank the division of BioMedical Sciences, Memorial University, and the College of the North Atlantic for their support. We are grateful to Mr. Nick Donovan for critically reading the manuscript and giving his valuable suggestions.

References

- Aller, S.G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., Harrell, P.M., Trinh, Y.T., Zhang, Q., Urbatsch, I.L., Chang, G., 2009. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323, 1718–1722.
- Bazer, F.W., Spencer, T.E., Johnson, G.A., 2009. Interferons and uterine receptivity. *Semin. Reprod. Med.* 27, 90–102.
- Cowan-Jacob, S.W., Fendrich, G., Floersheimer, A., Furet, P., Liebetanz, J., Rummel, G., Rheinberger, P., Centeleghe, M., Fabbro, D., Manley, P.W., 2007. Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr. Sect. D* 63, 80–93.
- Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C., Konkin, B.R., 2002. GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.* 31, 19–20.
- Dennis Jr., G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID. Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4, 3.
- Desco^{teaux}, C., Provencher-Mandeville, J., Mathieu, I., Perron, V., Mandal, S.K., Asselina, E., Be^{rube}, G., 2003. Synthesis of 17-estradiol platinum(II) complexes: biological evaluation on breast cancer cell lines. *Bioorg. Med. Chem. Lett.* 13, 3927–3931.
- Druker, B.J., Lydon, N.B., 2000. Lessons learned from the development of an Abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J. Clin. Invest.* 105, 3–7.
- Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868.
- Frenzel, A., Grespi, F., Chmielewski, W., Villunger, A., 1998. BCL2 family proteins in carcinogenesis and the treatment of cancer. *Apoptosis* 14, 584–596.
- Fukui, K., Yonezawa, T., Nagata, C., Shingu, H., 1954. Molecular orbital theory of orientation in aromatic, heteroaromatic, and other conjugated molecules. *J. Chem. Phys.* 22, 1433–1442.
- Fuller, J.C., Burgoyne, N.J., Jackson, R.M., 2009. Predicting druggable binding sites at the protein–protein interface. *Drug Discov. Today* 14, 155–161.
- Gagnon, V., St-Germain, M., Desco^{teaux}, C., Provencher-Mandeville, J., Parent, S., Mandal, S.K., Asselin, E., Be^{rube}, G., 2004. Biological evaluation of novel estrogen–platinum(II) hybrid molecules on uterine and ovarian cancers—molecular modeling studies. *Bioorg. Med. Chem. Lett.* 14, 5919–5924.
- Garodia, P., Ichikawa, H., Malani, N., Sethi, G., Aggarwal, B.B., 2007. From ancient medicine to modern medicine: ayurvedic concepts of health and their role in inflammation and cancer. *J. Soc. Integr. Oncol.* 5, 25–37.
- Ghose, A.K., Viswanadhan, V.N., Wendoloski, J.J., 1999. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. *J. Combin. Chem.* 1, 55–68.
- Greer, J., Erickson, W.J., Baldwin, J.J., Varney, M.D., 1994. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* 37, 1035–1054.
- Grodsky, N., Li, Y., Bouzida, D., Love, R., Jensen, J., Nodes, B., Nonomiya, J., Grant, S., 2006. Structure of the catalytic domain of human protein kinase C beta II complexed with a bisindolylmaleimide inhibitor. *Biochemistry* 45, 13970–13981.
- Gupta, A., Mandal, S.K., Leblanc, V., Descôteaux, C., Asselin, E., Bérubé, G., 2008. Synthesis and cytotoxic activity of benzopyran-based platinum(II) complexes. *Bioorg. Med. Chem. Lett.* 18, 3982–3987.
- Harvey, L., 2003. *Molecular Cell Biology*. 5 ed. W. H. Freeman and Co, New York. 973 s. b ill. ISBN: 0-7167-4366-3. Libris: 8926100.
- Henry, C.M., 2001. Structure-based drug design. *C & EN.* 79, 69–74.
- Irwin, J.J., Shoichet, B.K., 2005. ZINC is a free database of commercially-available compounds for virtual screening. ZINC contains over 8 million purchasable compounds in ready-to-dock, 3D formats. *J. Chem. Inf. Model* 45, 177–182.
- Kong, D.X., Li, X.J., Zhang, H.Y., 2009. Where is the hope for drug discovery? Let history tell the future. *Drug Discov. Today* 14, 115–119.
- La, D.S., Belzile, J., Bready, J.V., Coxon, A., Demelfi, T., Doerr, N., Estrada, J., Flynn, J.C., Flynn, S.R., Graceffa, R.F., Harriman, S.P., Larrow, J.F., Long, A.M., Martin, M.W., Morrison, M.J., Patel, V.F., Roveto, P.M., Wang, L., Weiss, M.M., Whittington, D.A., Teffera, Y., Zhao, Z., Polverino, A.J., Harmange, J.C., 2008. Novel 2, 3-dihydro-1, 4-benzoxazines as potent and orally bioavailable inhibitors of tumor-driven angiogenesis. *J. Med. Chem.* 51, 1695–1705.
- Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* 23, 3–25.
- Mandal, S., Davie, J.R., 2007. An integrated analysis of genes and pathways exhibiting metabolic differences between estrogen receptor positive breast cancer cells. *BMC Cancer* 7 (181), 1–17.

- Mandal, S., Richardson, V.J., Banoub, J., Church, J.G., 2002. Identification of myo-inositol 1, 2-cyclic monophosphate by electrospray tandem mass spectrometry, a major constituent of EGF-stimulated phosphoinositide turnover in MDA 468 cells. *Exp. Cell Res.* 277, 131–138.
- Mandal, S., Be'rubé, B., Asselin, E., Mohammad, I., Richardson, V.J., Gupta, A., Pramanik, S.K., Williams, A.L., Mandal, S.K., 2007a. A novel series of potent cytotoxic agents targeting G2/M phase of the cell cycle and demonstrating cell killing by apoptosis in human breast cancer cells. *Bioorg. Med. Chem. Lett.* 17, 4955–4960.
- Mandal, S., Be'rubé, G., Asselin, E., Richardson, V.J., Church, J.G., Bridson, J., Pham, T.N.Q., Pramanik, S.K., Mandal, S.K., 2007b. A new platinum complex of triazine demonstrates G1 arrest with novel biological profile in human breast cancer cell line, MDA-MB-468. *Bioorg. Med. Chem. Lett.* 17, 2139–2145.
- Mandal, S., Curtis, L., Pind, M., Murphy, L.C., Watson, P.H., 2007c. S100A7 (psoriasin) influences immune response genes in human breast cancer. *Exp. Cell Res.* 313, 3016–3025.
- Müller, B.A., 2009. Imatinib and its successors—how modern chemistry has changed drug Development. *Curr. Pharm. Des.* 15, 120–133.
- Nautiyal, J., Majumder, P., Patel, B.B., Lee, F.Y., Majumdar, A.P.N., 2009. Src inhibitor dasatinib inhibits growth of breast cancer cells by modulating EGFR signaling. *Cancer Letts.* 283, 143–151.
- Nickells, R.W., Semaan, S.J., Schlamp, C.L., 2008. Involvement of the Bcl2 gene family in the signaling and control of retinal ganglion cell death. *Prog. Brain Res.* 173, 423–435.
- Orr, M.S., Scherf, U., 2002. Large-scale gene expression analysis in molecular target discovery. *Leukemia* 16, 473–477.
- Padmanabhan, S., Ravella, S., Curiel, T., Giles, F., 2008. Current status of therapy for chronic myeloid leukemia: a review of drug development. *Future Oncol.* 4, 359–377.
- Patwardhan, B., Bodeker, G., 2008. Ayurvedic genomics: establishing a genetic basis for mind–body typologies. *J. Altern. Complement. Med.* 14, 571–576.
- Provencher-Mandeville, J., Desco'teaux, C., Mandal, S.K., Leblanc, V., Asselina, E., Be'rubé, G., 2008. Synthesis of 17 β -estradiol-platinum(II) hybrid molecules showing cytotoxic activity on breast cancer cell lines. *Bioorg. Med. Chem. Lett.* 18, 2282–2287.
- Ricciardi, S., Tomao, S., de Marinis, F., 2009. Toxicity of targeted therapy in non-small-cell lung cancer management. *Clin. Lung Cancer* 10, 28–35.
- Samy, R.P., Pushparaj, P.N., Gopalakrishnakone, P., 2008. A compilation of bioactive compounds from ayurveda. *Bioinformation* 3, 100–110.
- Sharma, P.S., Sharma, R., Tyagi, T., 2009. Receptor tyrosine kinase inhibitors as potent weapons in war against cancers. *Curr. Pharm. Des.* 15, 758–776.
- Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. *Science* 270, 484–487.
- Weisberg, E., Manley, P.W., Breitenstein, W., Brueggen, J., Cowan-Jacob, S.W., Ray, A., Huntly, B., Fabbro, D., Fendrich, G., Hall-Meyers, E., Kung, A.L., Mestan, J., Daley, G.Q., Callahan, L., Catley, L., Cavazza, C., Azam, M., Neuberger, D., Wright, R.D., Gilliland, D.G., Griffin, J.D., 2005. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell* 7, 129–141.
- Yamashita, T., Honda, M., Kaneko, S., 2008. Application of serial analysis of gene expression in cancer research. *Curr. Pharm. Biotechnol.* 9, 375–382.
- Yang, Y., Adelstein, S.J., Kassis, A.I., 2009. Target discovery from data mining approaches. *Drug Discov. Today* 14, 147–154.

Web links

The web links of some of the important resources and tools are given here.

<http://www.rcsb.org/pdb/home/home.do>.
<http://salilab.org/modeller/modeller.html>.
http://swissmodel.expasy.org/workspace/index.php?func=modelling_project1.
<http://www.expasy.ch/spdbv/text/modeling.htm>.
<http://swift.cmbi.ru.nl/servers/html/index.html>.
<http://autodock.scripps.edu/wiki/AutoDock4/>.
<http://www.organic-chemistry.org>.
http://www.dtp.nci.nih.gov/docs/dtp_search.html.
<http://spheroid.ncifcrf.gov/spheroid/>.
<http://www.ncbi.nlm.nih.gov/geo/>.
<http://cgap.nci.nih.gov/SAGE>.
<http://smd.stanford.edu/>.
<https://genome.unc.edu/cgi-bin/SMD/umad.pl>.
<http://cgap-stage.nci.nih.gov/Pathways>.
<http://www.broad.mit.edu/gsea/>.
<http://niaid.abcc.ncifcrf.gov/>.
<http://www.genmapp.org/>.
http://www.ingenuity.com/products/pathways_analysis.html.
<http://rana.lbl.gov/EisenSoftware.htm>.
<http://www.bioconductor.org/>.
<http://www.ebi.ac.uk/services/>.
<http://www.thebiogrid.org/SearchResults/summary/107068>.
<http://www.thebiogrid.org/SearchResults/summary/108276>.

Computational tools/software

<http://www.gaussian.com/>.
<http://www.fujitsu.com/us/services/solutions/lifesci/>.
<http://accelrys.com/products/additional-products.html%>.
<http://ambermd.org/>.
<http://www.chemcomp.com/>.
<http://serenasoft.com/>.
<http://classic.chem.msu.su/gran/gamess/>.