Bio- and chemo-informatics beyond data management: crucial challenges and future opportunities

Florence L. Stahura and Jürgen Bajorath

Bio- and chemo-informatics are now thought to be crucial to the success and integration of biotechnology and drug discovery. Research in this area has expanded to go beyond data- and information-management. Here, we review exemplary areas, such as target identification and validation, virtual screening, and prediction of downstream characteristics of leads, where further research will play a key role in progressing the field.

Florence L. Stahura
Albany Molecular Research
Bothell Research Center
(AMRI-BRC)
18804 North Creek
Parkway
Bothell
WA 98011, USA
***Jürgen Bajorath**
AMRI-BRC and Dept of
Biological Structure
University of Washington
Seattle
WA 98195, USA
tel: +1 425 424 7297
fax: +1 425 424 7299
*e-mail: jurgen.bajorath@
albmolecular.com

▼ The current status and future perspective of bio- and chemo-informatics research is complicated by the significantly expanding scope of these disciplines. Clearly, the development and implementation of computational infrastructures to process, analyze, and organize the rapidly growing volume of primary biological and chemical data, continues to be a major issue for these informatics disciplines. However, at the same time, research and development in bio- and chemo-informatics has gone far beyond data analysis and management. Novel concepts and methods are being developed to support many stages of biotechnology and pharmaceutical research from target identification to lead optimization, and the prediction of pharmacological compound characteristics. We focus here on these aspects outside data and information management. As gene-to-lead paradigms become increasingly popular, researchers need to bear in mind some of the major challenges that informatics research is currently facing in biotechnology and drug discovery. We review in detail, target identification and validation strategies, virtual screening (VS), and ADME (absorption, distribution, metabolism, elimination) predictions as some of the crucial topics in bio- and chemo-informatics research. These areas could provide either major bottlenecks or significant opportunities for growth in the future.

Biotechnology, drug discovery and informatics

When considering the vast amounts of genomic data generated to date, it is not surprising that biotechnology will have increasingly a 'front-end' role to play in pharmaceutical research [1], in particular, the identification and characterization of molecular targets for drug discovery. However, the introduction of new technologies and paradigms in biology and drug discovery during the 1990s has, at least to date, generally had relatively little impact on the number of newly introduced drugs [1]. For example, preclinical and clinical attrition rates of drug candidates continue to be extremely high [2]. At best, 1% of drug leads produced by early-phase discovery efforts pass preclinical development [3] and only about 4% of candidate compounds that reach investigational new drug (IND) status and enter clinical trials will eventually become drugs [2].

Among the new technologies thought to help establish well-defined target-to-drug pathways were bio- and chemo-informatics. These technologies initially evolved driven by the need to analyze and manage exponentially growing amounts of primary data in biology and chemistry, and expanded to become largely

1359-6446/02/$ – see front matter PII: S1359-6446(02)02271-7

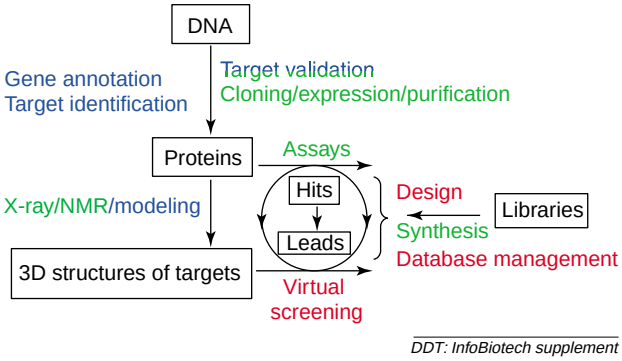

Figure 1. Merging informatics components: the interplay between bioinformatics (blue), chemoinformatics (red) and experimental (green) efforts in drug discovery. In this scheme, target validation is considered a hybrid component dependent on both computational and experimental analysis.

independent areas of research [4]. As biotechnology and pharmaceutical research become more complementary, bio- and chemo-informatics efforts are also beginning to merge [4,5]. Figure 1 shows a combination of different informatics components that provides a coherent infrastructure for research and development at the macromolecular (target) and small molecular (drug candidate) level. There is little doubt that the medium- to long-term impact of informatics disciplines on biotechnology and drug discovery will depend crucially on the ability to interface them effectively with experimental programs [4]. We therefore also discuss what we believe are some of the major challenges for bio- and chemo-informatics concepts.

Revising gene annotations

Annotation of genes and identification of therapeutic targets in human genome sequences [6,7] are major focal points of bioinformatics. Initial estimates of the total number of human genes differed greatly, ranging from approximately 30,000 to 120,000 genes [8]. More-detailed annotation of the human genome sequence led to an adjusted number of approximately 40,000 genes [6,7,9]. However, recent comparisons suggest that there are considerable differences between human genome annotations, and that the total number of human genes could be underestimated at present [10]. Even for well-studied genomes, annotations are probably incomplete, as indicated by the recent identification of more than 100 previously unknown genes in yeast [11]. Taken together, these findings suggest that much work remains to be done before gene annotations reach a high level of confidence. Therefore, research in the post-genomic era might initially need to focus on further analysis of primary genomic data, rather than large-scale evaluation of gene products.

Paradigm shift in target identification

Drugs available at present act on a relatively small number of targets, probably only 200–500 [1]. To date, target identification has typically proceeded from an assay observation to the identification of a key protein, and the subsequent cloning of its gene [12]. Although complicated by the uncertainties of annotation and functional assignments, this situation is rapidly changing, as many more potential targets are being considered at the gene level. Thus, the more conventional 'function-to-(single) gene' effort is shifting to a '(multiple) gene-to-function' paradigm [12]. These trends can be rationalized as a change from a 'deductive' to an 'inductive' approach in the analysis of molecular and cellular functions of gene products (Fig. 2). Given this scenario, how will it be possible to effectively evaluate thousands of new genes as potential targets? For bioinformatics, this is a considerable challenge, and the field would benefit from the availability of more-precise bioinformatics tools to assign functions to putative targets. Some promising trends can already be observed, including advances in microarray analysis [13,14] and protein–protein interaction [15,16] analysis, modeling of metabolic or regulatory pathways to predict cellular functions of genes [17,18], and analysis of 3D protein structures to obtain functional insight and predict promising drug targets [19,20]. A meaningful combination of some of these efforts is expected to make a substantial impact, not only on target identification, but also on validation.

Target validation

Finding homologs of a newly identified gene, making use of expert knowledge, and mining the scientific literature, might provide some important clues about the potential function(s) of a gene product and its putative involvement in human disease. However, in drug discovery, where target validation strategies continue to be a major bottleneck [1], theoretical analysis is, in general, not sufficient [4].

Theory and experiment

Without doubt, identification of molecular and cellular functions and confirmation of their role in disease states are important initial steps in therapeutic target validation. However, such targets might not be 'drugable' (amenable to small-molecule-based discovery approaches), and if there are no ligands and assays, no leads can be identified. Thus, without generating an appropriate experimental infrastructure, effective target validation is extremely difficult. Therefore, informatics approaches should, in principle, not aim to replace experimental strategies but rather to complement them [4].

Therapeutic target proteins and their interactions

Substantial progress is being made in understanding protein–protein interactions that are relevant to therapy, either by

comprehensive mapping of these interactions using yeast two-hybrid screens [15] or emerging proteomics technologies such as affinity purification combined with mass spectrometry [16]. Microarray analysis at the DNA level [13,21] can yield functional and pathway information and, in addition, array analysis at the protein level [14,22] can aid in the identification of specific ligands and functions. For array-based experiments, data representation and analysis have become major challenges for informatics efforts. Moreover, the design and implementation of databases that contain detailed protein–protein-interaction maps [15,23], functional information, and predictions about the drugability of protein families based on structural considerations [19,20], should help to prioritize putative targets.

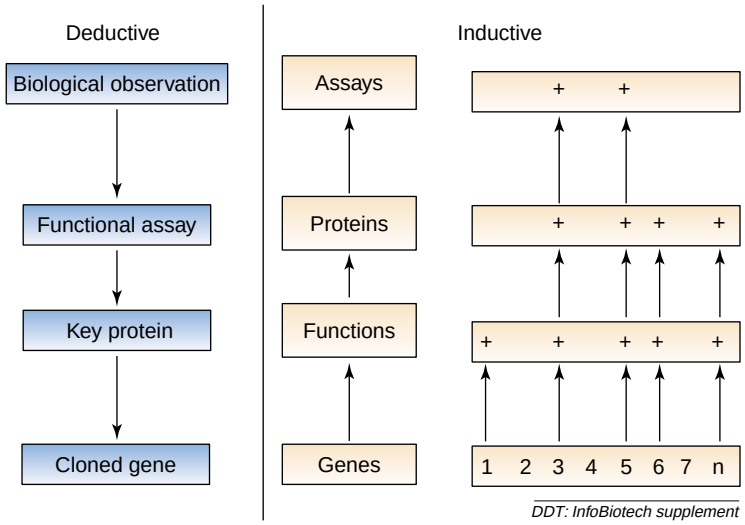

Figure 2. Paradigm shift in the analysis of therapeutic targets. The more conventional 'deductive' approach essentially aims to identify the gene responsible for a therapeutically relevant effect or observation. By contrast, the availability of many more potential targets at the gene level necessitates an 'inductive' approach to functional analysis. Not all potential targets are expected to be validated.

Chemical genetics

In the context of target validation, the chemical genetics [24,25] discipline is also intriguing. In part, this concept is equivalent to the (reverse genetics) approach of introducing genetic knockouts. The principle of the technique is the identification of small molecules that specifically affect each potential cellular target, thereby providing chemical probes for large-scale target validation. Such efforts crucially depend on the availability of well-designed chemical libraries [26] and high-throughput assay systems or protein arrays [22]. As a potential drawback, the technique is based on the premise that all gene products are potential small-molecule targets, which is questionable, at best [20]. Nevertheless, the concept of chemical genetics offers significant potential in target validation as it can be applied in various ways. For example, when combined with microarray analysis, target-specific modulation of cellular expression profiles can be detected, or drug sensitivity of mutated genes can be studied [25].

Focus on target structure

In the context of the structural genomics initiative [27], significant progress is being made with high-throughput structure determination [28], which is, in part, due to technical advances in miniaturization and automation of many steps in protein expression, purification and crystallization [27,28]. Consequently, the 'gene-to-structure-to-drug' paradigm has become an attractive theme in biotechnology. The structural genomics initiative aims to determine representative structures for all protein families [27] so that reasonable models of related proteins can be obtained by application of comparative structure prediction methods [29]. By contrast, protein models generated by *de novo* prediction methods [29], and not by comparative extrapolation from known structural templates, are, at present, usually not sufficiently accurate for drug discovery applications.

From target structures to leads

A basic idea of the 'gene-to-structure-to-drug' theme is to use the growing number of experimentally determined or modeled 3D target structures for rapid discovery of hits or leads. In this approach, high-throughput docking of large compound databases into binding sites of target structures [30] plays an important role in the identification of hits, and in closing the gap between structural biology and medicinal chemistry. Similar to the situation in small-molecule-based VS [31], where different techniques are rapidly being developed but commonly accepted computational standards are yet to be established [31], a diverse array of methods is currently available for docking simulations [30,32].

However, although docking algorithms have significantly matured over time, scoring functions to evaluate predicted complexes, calculate binding energies, and rank putative hits, continue to be of limited accuracy [30]. An example of these shortcomings is shown in Fig. 3. The limitations of current scoring functions present a major problem for effective structure-based VS, especially considering the rapidly growing number of compounds available for evaluation. Therefore, the 'gene-to-structure' part of the structural-biology-driven paradigm mentioned previously might, in fact, be easier to realize than its 'structure-to-drug' component.

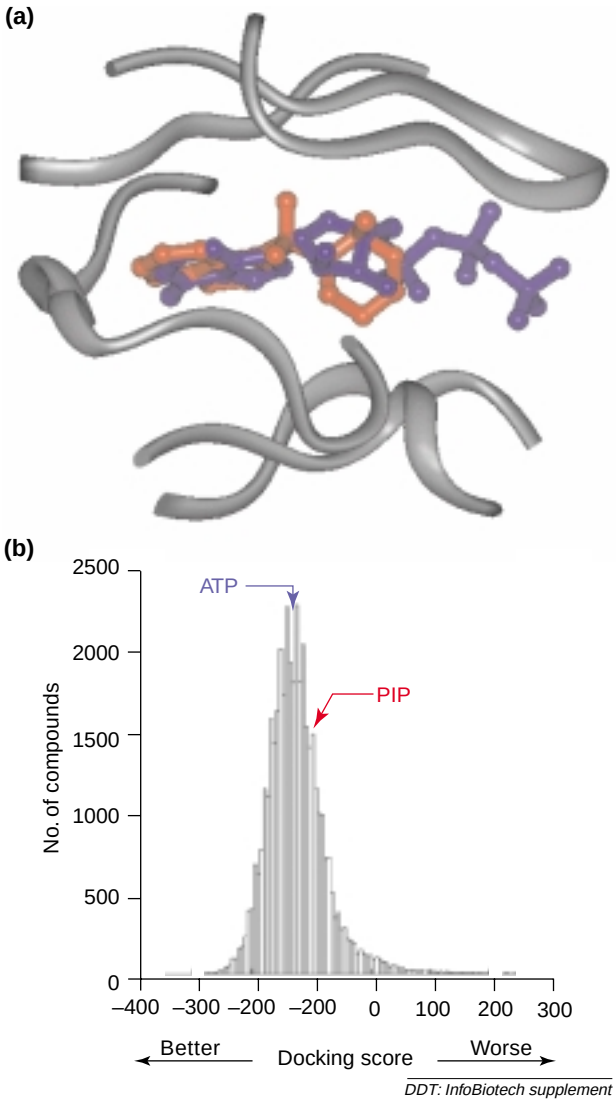

Figure 3. Docking and scoring. The calculations summarized in this figure were performed in our laboratory and highlight the problems of compound selection based on docking scores. (a) The ATP binding site of cAMP-dependent protein kinase (gray ribbon) is shown with ATP (blue) and a known inhibitor of this kinase, 1-(5-isoquinoline-sulfonyl)-2-methyl-piperazine (PIP; red), superimposed in their crystallographic binding modes (Protein Data Bank entries '1ATP' and '1YDR', respectively). A collection of ~60,000 commercially available compounds was docked into the ATP binding site and scores were calculated using a function that evaluates the quality of the fit of ligands into receptor sites. (b) The distribution of obtained scores is shown, yielding a distinct peak for the majority of database compounds. In essence, the peak area represents scores for random binding of many compounds. By contrast, the best scoring compounds (that would usually be selected for testing) are found in the scoring range to the left of the peak. Equivalent scores were calculated for the two experimentally determined protein–ligand complex structures shown in (a) and mapped to the distribution of scores obtained by database docking. As can be seen, both 'experimental' scores fall into the peak area. Thus, even if these structures would have been correctly predicted in docking simulations, the ligands could not have been identified or selected based on their computed scores.

Some progress in improving scoring schemes and hit rates from VS is being undertaken. These advances include systematic comparison and further refinement of energy functions [32], design of protein–ligand databases for 'parameterization' of new functions [33], improvement of the quality of compound libraries for docking [34,35], and development of protocols for comparative docking against target families [35]. Nevertheless, research in bio- and chemo-informatics is challenged with the task of developing more-precise and robust functions for the evaluation of high-throughput virtual screens.

Structure-based library design

In addition to increasing hit rates, accurate docking of ligands to proteins, and combinatorial docking [36], are also important starting points for structure-based (and target-focussed) design of compound libraries [36]. However, ligand docking to single proteins or families [35] presents only one of several possible routes to support library design by use of structural information. In addition, the development of active-site-directed 3D pharmacophores for protein families [37] provides another computational approach to ligand and library design that is less dependent on the accurate prediction of ligands in binding sites. However, provided at least some weak initial hits can be obtained, series of experimental structures of protein–ligand complexes [28] and specialized protein–ligand databases [33] provide a more accurate basis for the combinatorial exploration of ligand binding to target protein families.

Predicting drugs and *in vivo* effects

The tremendous attrition rates of preclinical and clinical candidates (as mentioned previously), represent, without doubt, the major obstacle in drug discovery. It is therefore not surprising that chemoinformatics research is increasingly focussed on the analysis and prediction of compound features beyond potency, such as ADME characteristics [38,39]. Effects such as solubility or toxicity of compounds and pharmacokinetic parameters are often considered part of ADME analysis [39]. Implicitly, many ADME parameters contribute to the potential drug-likeness of molecules [40], in addition to selected structural or pharmacophore features shared by many drugs [41]. To better understand and predict what exactly renders a compound drug-like or not, is currently another intensely studied area in chemoinformatics [40,42]. The basic idea behind theoretical analysis and prediction of ADME and drug parameters is to apply this information as early as possible in the discovery process; for example, when profiling [38] or designing [42,43] chemical libraries. This is thought to increase the probability of identifying hits or leads having favorable development characteristics and

therefore a greater chance of surviving preclinical and clinical evaluation.

At present, ADME and drug prediction models are typically applied as various filter functions at some point in the search for active compounds [38,43]. However, an open question is when such filters should be best applied; for example, early on during library design (or a pre-experimental analysis of screening libraries) or later in the evaluation of hits or leads. ADME analysis is, in general, challenged by the fact that many of the *in vivo* effects that decide the fate of a clinical candidate are not well understood, at present. The derivation of predictive models relies mainly on the quality of learning sets that comprise both 'successful' and 'unsuccessful' molecules [38], which is a rather limited (and probably not very reliable) knowledge base. Similarly, we do not yet understand what drugs should, in general, 'look like' (if this is indeed possible). In fact, many of the models developed to distinguish drugs from non-drugs are based on fairly complex neural network simulations [40,43], the results of which are very difficult to interpret in physical or chemical terms. Therefore, the quality of molecular learning sets available to develop such models also plays a pivotal role here, similar to the situation in ADME analysis. It is currently unclear what the success rate of predictive drug and ADME models could ultimately be if they were more generally applied. However, any (computational or experimental) test that measurably reduces attrition rates of preclinical and clinical candidates will make a significant impact. It will therefore be exciting to monitor further developments in this area of chemoinformatics research.

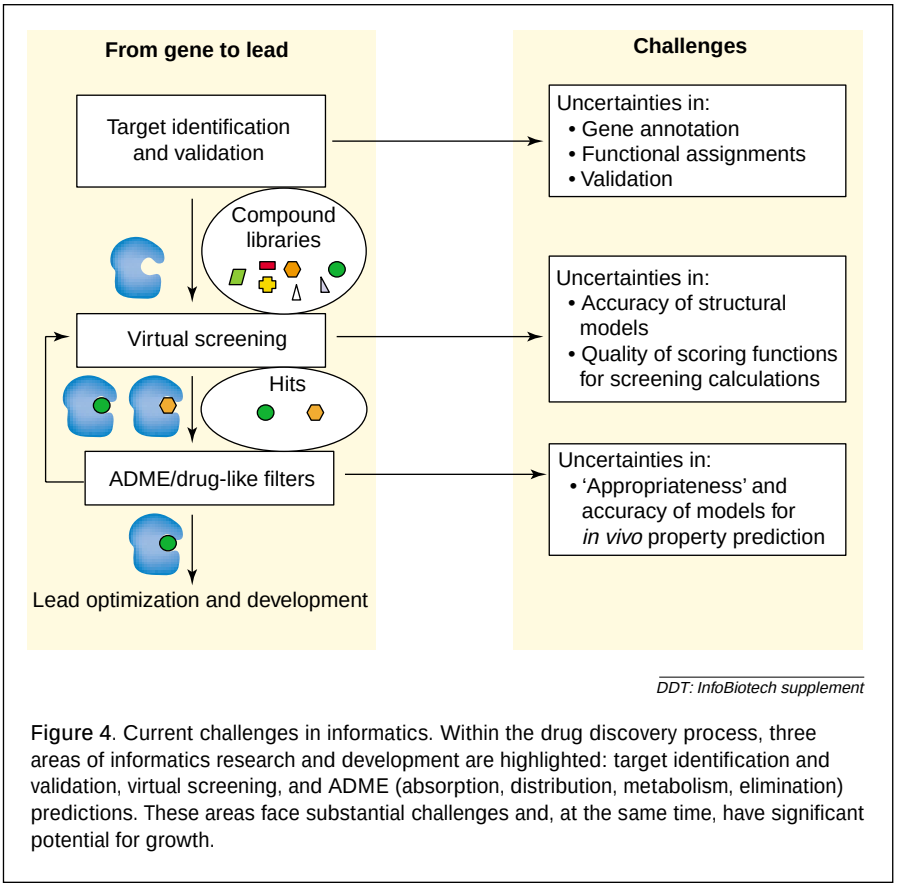

Figure 4. Current challenges in informatics. Within the drug discovery process, three areas of informatics research and development are highlighted: target identification and validation, virtual screening, and ADME (absorption, distribution, metabolism, elimination) predictions. These areas face substantial challenges and, at the same time, have significant potential for growth.

Summary and perspective

Research in bio- and chemo-informatics has rapidly progressed over the past few years. A wealth of new computational methods for the analysis and prediction of chemical and biological features has been, and continues to be, developed. A positive trend can be observed: the original distinction between bio- and chemo-informatics research is beginning to disappear through the realization that successful target validation and drug discovery programs require a concerted effort of many theoretical and experimental components. The fact that we can pinpoint several informatics topics that are equally prone to becoming either obstacles or catalysts of future progress, indicates that these computational methods are beginning to make a real impact on biotechnology and pharmaceutical research. Areas where informatics concepts and tools play a crucial role, span the entire spectrum of 'target-to-drug' efforts, from target discovery to lead evaluation and optimization (Fig. 4). Table 1 gives examples of links to institutions that focus on some of the key aspects involved.

Although a discussion of specific information technology tasks is outside the scope of this review, a few key issues should at least be mentioned. As an example, the design of relational databases to link diverse sets of chemical, biological and, ultimately, clinical data, and enable even a non-expert user to efficiently access, assemble and communicate this information, will be of significant importance for the future impact of informatics disciplines on biotechnology and drug discovery. Presently, these efforts are complicated by the fact that commonly accepted database standards and architectures are yet to be established [44]. Consequently, transfer and sharing of information continues to be difficult between different sites or institutions. Here, another interesting trend can be observed: major computer companies are making significant efforts to expand into the life-science area and offer comprehensive hardware and/or software solutions to informatics problems [44].

Table 1. Examples of commercial and academic organizations that focus on informatics-related R&D topics in biotechnology and drug discovery[a]

R&D topic	Company/institution	URL
Gene annotation	Celera Genomics	http://www.celera.com
	Incyte Genomics	http://www.incyte.com
	EMBL-EBI/The Sanger Center	http://www.ensembl.org
	Genomics Institute of the Novartis Research Foundation	http://www.gnf.org
Target identification	Athersys	http://www.athersys.com
	Cellular Genomics	http://www.cellulargenomics.com
	Galapagos Genomics	http://www.galapagosgenomics.com
Target validation	Exelixis	http://www.exelixis.com
	Lexicon Genetics	http://www.lexgen.com
Structural genomics	Syrrx	http://www.syrrx.com
	Structural GenomiX	http://www.stromix.com
	Inpharmatica	http://www.inpharmatica.co.uk
	Joint Center for Structural Genomics	http://www.jcsg.org
Proteomics technologies	Serenex	http://www.serenex.com
	Cellzome	http://www.cellzome.com
Chemical genomics	Stuart Schreiber lab at Harvard	http://www-schreiber.chem.harvard.edu
	Infinity Pharmaceuticals	http://www.infinitypharm.com
	Vertex Pharmaceuticals	http://www.vpharm.com
Protein–protein interactions	CuraGen	http://www.curagen.com
	Axcell Biosciences	http://www.axcellbio.com
	Stan Fields lab at University of WA	http://depts.washington.edu/sfields
Microarrays	Affymetrix	http://www.affymetrix.com
	Agilent Technologies	http://www.agilent.com
	Rosetta Inpharmatics	http://www.rii.com
Virtual screening	Irwin Kuntz lab at UCSF	http://www.cmpharm.ucsf.edu/kuntz/dock.html
	Chemical Computing Group	http://www.chemcomp.com
ADME predictions	Camitro	http://www.camitro.com
	Lion Bioscience	http://www.lionbioscience.com

[a]Abbreviations: EMBL-EBI, European Molecular Biology Laboratory – European Bioinformatics Institute; UCSF, University of California at San Francisco.

What are the crucial tasks for R&D in bio- and chemo-informatics in the near future? Considering the current drug discovery landscape, establishing more effective target identification and validation strategies appears to be as important as making progress with VS calculations and predictions of *in vitro* and *in vivo* compound characteristics. Any of these areas has significant growth potential for computational approaches and the opportunity to streamline the discovery process. As the amount of experimental data grows and is available as a knowledge base, it will also become easier to develop and tune computational models for the prediction of biological activity and compound characteristics. Of the areas discussed in this review, the derivation of reliable predictive models of *in vivo* profiles of candidate compounds, could perhaps be the most challenging but also the most promising topic for future research. This is because a reduction in the number of compounds that fail during development stages will be most crucial for drug discovery in the future, regardless of the number of therapeutic targets or leads that might become available.

References

1 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964

2 Caldwell, G.W. *et al*. (2001) The new pre-clinical paradigm: compound optimization in early and late phase drug discovery. *Curr. Topics Med. Chem.* 1, 353–366

3 Lakings, D.B. (2000) Non-clinical drug development: pharmacology, drug metabolism, and toxicology. *New Drug Approv.* 100, 17–54

4 Bajorath, J. (2001) Rational drug discovery revisited: interfacing experimental programs with bio- and chemo-informatics. *Drug Discov. Today* 6, 989–995

5 Tropsha, A. (2000) Recent trends in computer-aided drug discovery. *Curr. Opin. Drug Discov. Develop.* 3, 310–313

6 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921

7 Venter, J.C. *et al*. (2001) The sequence of the human genome. *Science* 291, 1304–1351

8 Aparicio, S.A. (2000) How to count…human genes. *Nat. Genet.* 25, 129–130

9 Gaasterland, T. and Oprea, M. (2001) Whole-genome analysis: annotations and updates. *Curr. Opin. Struct. Biol.* 11, 377–381

10 Hogenesch, J.B. *et al.* (2001) A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell* 106, 413–415

11 Kumar, A. *et al.* (2002) An integrated approach for finding overlooked genes in yeast. *Nat. Biotechnol.* 20, 58–63

12 Harrington, J. and Brunden, K.R. (2002) Drug screening in the genomics era. *Current Drug Discov.* 2, 17–20

13 Shoemaker, D.D. *et al.* (2001) Experimental annotation of the human genome using microarray technology. *Nature* 409, 922–927

14 Emili, A.Q. and Cagney, G. (2000) Large-scale functional analysis using peptide or protein arrays. *Nat. Biotechnol.* 18, 393–397

15 Uetz, P. *et al.* (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627

16 Gavin, A-C. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147

17 Hasty, J. *et al.* (2001) Computational studies of gene regulatory networks: *in numero* molecular biology. *Nat. Rev. Genet.* 2, 268–279

18 Wiechert, W. (2002) Modeling and simulation: tools for metabolic engineering. *J. Biotechnol.* 94, 37–63

19 Thornton, J.M. (2001) From genome to function. *Science* 292, 2095–2097

20 Weir, M. *et al.* (2001) Insights into protein function through large-scale computational analysis of sequence and structure. *Trends Biotechnol.* 19 (Suppl.), S61–S66

21 Lockhart, D.J. and Winzeler, E. A. (2000) Genomics, gene expression and DNA arrays. *Nature* 405, 827–836

22 MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–1763

23 Boulton, S.J. *et al.* (2001) Use of protein interaction maps to formulate biological questions. *Curr. Opin. Chem. Biol.* 5, 57–62

24 Caron, P.R. *et al.* (2001) Chemogenomic approaches to drug discovery. *Curr. Opin. Chem. Biol.* 5, 464–470

25 Zheng, X.F.S. and Chan, T-F. (2002) Chemical genomics in the global study of protein functions. *Drug Discov. Today* 7, 197–205

26 Schreiber, S.L. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* 287, 1964–1969

27 Stevens, R.C. *et al.* (2001) Global efforts in structural genomics. *Science* 294, 89–92

28 Jhoti, H. (2001) High-throughput structural proteomics using x-rays. *Trends Biotechnol.* 19 (Suppl.), S67–S71

29 Baker, D. and Sali, A. (2001) Protein structure prediction and structural genomics. *Science* 294, 93–96

30 Schneider, G. and Böhm, H-J. (2002) Virtual screening and fast automated docking methods. *Drug Discov. Today* 7, 64–70

31 Bajorath, J. (2001) Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Comput. Sci.* 41, 233–245

32 Perez, C. and Ortiz, A.R. (2001) Evaluation of docking functions for protein-ligand docking. *J. Med. Chem.* 44, 3768–3785

33 Roche, O. *et al.* (2001) Ligand–protein database: linking protein–ligand complex structures to binding data. *J. Med. Chem.* 44, 3592–3598

34 Su, A.I. *et al*. (2001) Docking molecules by families to increase the diversity of hits in database screens: computational strategy and experimental evaluation. *Proteins: Struct., Funct., and Genet.* 42, 279–293

35 Lamb, M.L. *et al.* (2001) Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins: Struct., Funct., and Genet.* 42, 296–318

36 Böhm, H-J. and Stahl, M. (2000) Structure-based library design: molecular modeling merges with combinatorial chemistry. *Curr. Opin. Chem. Biol.* 4, 283–286

37 Mason, J.S. *et al.* (2001) 3D pharmacophores in drug discovery. *Curr. Pharm. Des.* 7, 567–597

38 Beresford, A.P. *et al.* (2002) The emerging importance of predictive ADME simulations in drug discovery. *Drug Discov. Today* 7, 109–116

39 Podlogar, B.L. *et al.* (2001) Computational methods to estimate drug development parameters. *Curr. Opin. Drug Discov. Develop.* 4, 102–109

40 Clark, D.E. and Pickett, S.D. (2000) Computational methods for the prediction of drug-likeness. *Drug Discov. Today* 5, 49–58

41 Muegge, I. *et al*. (2001) Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* 44, 1841–1846

42 Mitchell, T. and Showell, G.A. (2001) Design strategies for building drug-like chemical libraries. *Curr. Opin. Drug Discov. Develop.* 4, 314–318

43 Matter, H. *et al.* (2001) Computational approaches towards the rational design of drug-like compound libraries. *Combin. Chem. High Throughput Screen.* 4, 453–475

44 Augen, J. (2002) The evolving role of information technology in the drug discovery process. *Drug Discov. Today* 7, 315–323