# Human metabolic network reconstruction and its impact on drug discovery and development

**Hongwu Ma and Igor Goryanin**

School of Informatics, the University of Edinburgh, Mayfield Road, Edinburgh EH9 3JB, United Kingdom

**This paper describes the process for the reconstruction of a high quality human metabolic network from the genome information, the existing problems in the reconstruction and why a time consuming literature based consolidation process is needed. The reconstructed metabolic network provides a unified platform to integrate all the biological and medical information on genes, proteins, metabolites, disease, drugs and drug targets for a system level study of the relationship between metabolism and disease. System analysis of metabolic networks will help us, not only in identifying new drug targets but also in developing a system-oriented drug design strategy.**

The recently developed high throughput technologies allow us to measure all cellular components with just a few experiments [1]. The challenge is to understand how the cellular components are organized into a complex network based on the measurement of individual components, and, particularly for humans, to understand the causes of diseases and find effective drugs for their treatment. The reconstruction and analysis of complex biological networks, such as metabolic networks, gene regulatory networks and protein–protein interaction networks (PPN), are among the central topics of Systems Biology [2,3]. Mapping the experimental data to a network reconstructed based on previous knowledge can help to consolidate the 'noisy' high throughput data and discover new interactions, thereby complementing and improving the existing network. On the other hand, analyzing the data from a new network perspective can provide new knowledge about the system and lead to the discovery of new therapy strategies or new drug targets. This review will focus on the reconstruction of the human metabolic network and its impact on the drug discovery process.

The metabolic network is of special interest among the different types of biological networks because firstly it is the most complete and reliable network thanks to decades of biochemical research. This is in contrast to other networks such as the PPN that has not been studied until recently and is not very reliable because it is mainly based on two-hybrid experiments [4,5]. Secondly, the metabolic network is the only network that can integrate the experimental data for different types of molecules (transcriptomics for genes, proteomics for enzymes and metabolomics for metabolites). In comparison, the gene regulatory network and signal transduction network can only make use of the transcriptomics or proteomics data. Thirdly, compared to genes and proteins, metabolites are more closely related to the phenotype of an organism. Specifically, the health and disease states of the human can be described more meaningfully by the metabolic state of human cells, tissues, organs and the organism as a whole.

The cause or effect of many human diseases is an abnormal metabolic state, such as a high glucose concentration in the blood of diabetes patients and high urine amino acid level resulting from liver or renal disorders. For a long time, metabolites have been used as biomarkers for diagnosing diseases from blood or urine samples [6–8]. Several hundred diseases are classified as metabolic diseases which are directly caused by a deficiency of metabolic enzymes and the subsequent accumulation of toxic substances or the lack of essential metabolites. Besides these metabolic disorders, many other common diseases such as cardiovascular disease, cancer and diabetes are also related to the human metabolism. Metabolic Syndrome is a complex disorder characterized by extensive metabolic changes in the patients such as the levels of glucose, cholesterol, uric

☆ This paper discusses the processes for the reconstruction of a high quality human metabolic network and how it can help in better understanding human diseases and in facilitating the drug research and development process.

*Corresponding author:* Goryanin, I. (goryanin@inf.ed.ac.uk)



1359-6446/06/$ - see front matter  doi:10.1016/j.drudis.2008.02.002

acid, and so on. The prevalence in the USA is 25% of the population, increased to 43% for people aged 60–69 years [9]. The patients have an increased risk for cardiovascular disease, obesity and diabetes. It is also known that the metabolic state in tumor cells is different from that in normal cells [10]. Even diseases seemingly unrelated to human metabolism may actually be closely related. For instance, in patients with Parkinson's disease, the activities of the two enzymes for dopamine synthesis (tyrosine 3-monooxygenase [1.14.16.2] and dopa decarboxylase [4.1.1.28], as shown in Figure 1) in the brain are very low. The subsequent lack of dopamine (a neurotransmitter) leads to the reduced inhibitory effect of dopamine receptors and this causes excessive muscle contraction which is the primary symptom of Parkinson's disease.

## Genome based human metabolic network reconstruction

From the above discussion, we can appreciate that metabolism is involved in various aspects of human disease and that it would be valuable to have a complete and reliable human metabolic network for a better understanding of the relationship between human metabolism and diseases. As a positive result of the human genome project and the powerful bioinformatics analysis methods developed, it is now straightforward to reconstruct the human metabolic network at a system level based on gene annotation information [3,11,12]. The first step in the reconstruction is to identify all genes encoding metabolic enzymes in the human genome and to assign correct Enzyme Commission classification (EC) numbers to them. Information about the enzyme coding genes can be extracted from many human genome databases and metabolic enzyme databases (see Table 1 for the most popular databases) [13–16]. However, mainly because of the different gene recognition algorithms and gene annotation methods used in different databases, the enzyme genes obtained from different databases are often not the same. Therefore to obtain a comprehensive enzyme–gene list for the network reconstruction one needs to integrate the gene annotation information from different databases.

It should also be noted that not all databases use EC numbers for enzyme–gene annotation. Some databases may use EC numbers but not assign EC numbers for all the enzyme genes. For example, in HGNC [16] only 988 genes are annotated with EC numbers, while in KEGG [17] the number of genes with EC numbers is nearly 3000. Another useful database is Uniprot (Universal Protein Resource) [18] which includes extensively curated protein function information. Among the 67668 human proteins in Uniprot (release 11.3, July 2007), 5517 are annotated with at least one clear or unclear EC numbers (such as 1.-.-.-). One can link a human gene to EC numbers through the database cross-reference section in Uniprot which includes links to genes in Entrez [19], Ensembl [15], and so on.

An alternative way to finding enzyme genes is through GO (Gene Ontology) [20]. The Gene Ontology project aims to provide a controlled vocabulary to describe gene and gene product attributes in any organism. Each term in the vocabulary is assigned a unique GO ID and the terms are organized in a hierarchical way. The terms related to cellular metabolism are listed under 'metabolic process' (GO:0008152) and 'catalytic activity' (GO:0003824). Most of the human gene databases including Entrez gene and

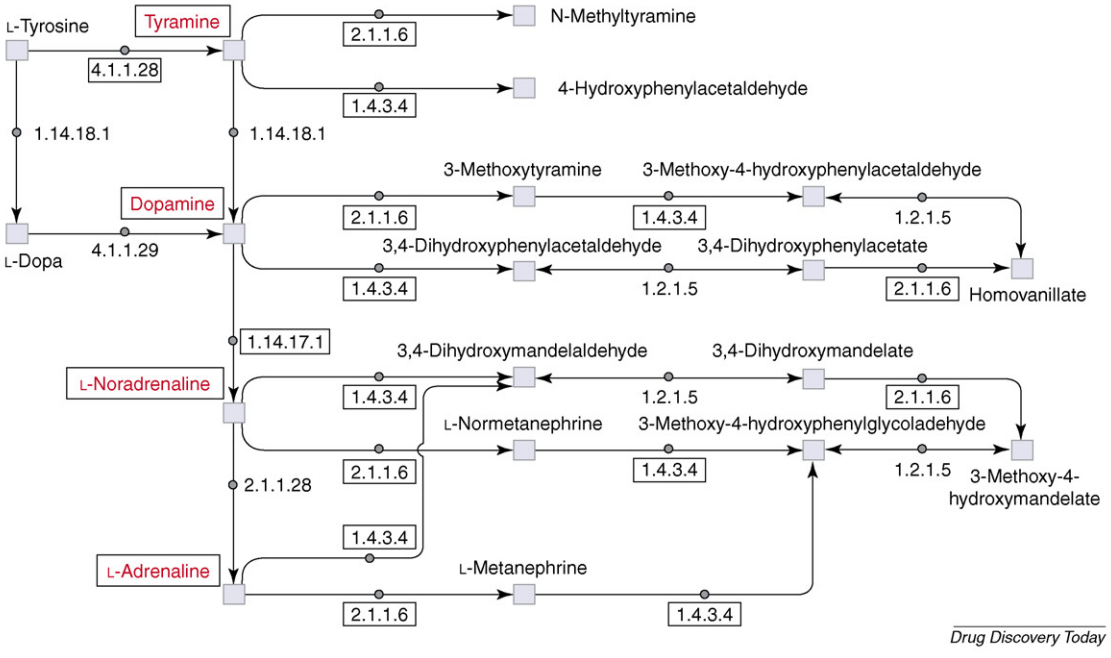


**FIGURE 1**

Metabolic pathways for the tyrosine derived monoamine neurotransmitters. The four compounds in the square are neurotransmitters. The four enzymes: DOPA decarboxylase (4.1.1.28), Monoamine oxidase (1.4.3.4), Catechol-*O*-methyl transferase (2.1.1.6) and dopamine beta-monooxygenase (1.14.17.1) are drug targets. Metabolites (e.g. Normetanephrine, Homovanillate) resulting from the degradation of these neurotransmitters are excreted in the urine and used as diagnostic markers.

**TABLE 1**

**Useful resource for the human metabolic network reconstruction**

| *Resource* | Web address | Description |
|---|---|---|
| **Genes and proteins** | | |
| Uniprot | http://www.ebi.uniprot.org | The most comprehensive resource for information on proteins |
| Entrez Gene | http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene | One of the most extensive gene database including automatically generated and human curated information on genes for the fully sequenced organisms |
| Ensembl | http://www.ensembl.org/ | Automatic annotation of human (and other eukaryotes) genome |
| HGNC | http://www.genenames.org/ | Manually curated and approved human gene names and symbols |
| **Metabolic pathways and reactions** | | |
| KEGG (LIGAND) | http://www.genome.ad.jp/kegg/ | One of the most useful databases for metabolic networks. Organization of metabolic genes, enzymes, reactions and metabolites into hundreds of metabolic pathways |
| Biocyc | http://biocyc.org/ | Includes literature derived metabolic pathways (Metacyc and Ecocyc) and hundreds of computationally derived genome based metabolic networks (including Humancyc) |
| Reactome | http://www.reactome.org | A curated resource of core metabolic and signal transduction pathways and reactions in human biology |
| **Metabolites and compounds** | | |
| HMDB | http://www.hmdb.ca/ | Small molecule metabolites found in the human body with experimental measurement information |
| ChEBI | http://www.ebi.ac.uk/chebi/ | Manual annotated and non-redundant information on small compounds of biological interest |
| PubChem | http://pubchem.ncbi.nlm.nih.gov/ | Extensive information for millions of small molecule compounds with emphasis on biological properties |

Ensembl provide GO annotation for the genes. This allows the identification of all metabolic genes through the GO terms assigned to a gene. The GO team also maintains a file (Ec2go) including the mapping relationships between EC numbers and GO terms, making it possible to assign EC numbers to genes through GO.

## From genes to metabolites and the high quality network reconstruction

Identification of the metabolic genes and the corresponding enzymes is just the first step of the network reconstruction. To link the genes and proteins with metabolites, one needs to find out all the metabolic reactions. To this end, a standard and consistent reaction database and an associated compound database that describes all the compounds involved in the reactions are necessary. Although many enzyme databases use reactions to describe the enzyme function, the reaction equations are often not standardized to be used in the network reconstruction. For example, in IUBMB enzyme nomenclature, the reaction for EC 1.1.1.1 is 'an alcohol + NAD+ = an aldehyde or ketone + NADH + H+'. The general compound name 'alcohol', 'aldehyde' can match to many different compounds in a network. Using 'an', 'or' in the reaction equation makes it difficult to process such equations using computer programs.

One of the most useful reaction and compound database is the KEGG LIGAND database [21]. The current version (July 2007) includes more than 7000 reactions and about 15,000 compounds. The information for compounds in the database includes ID, name and synonyms, formula, structure and links to other compound databases such as PubChem [22] and ChEBI [23]. The reaction information in the KEGG LIGAND database includes the reaction equations written in both compound names and IDs, and the EC numbers corresponding to that reaction. Therefore, together with the EC numbers obtained from the annotated human genome, one can get nearly 2000 reactions from KEGG which form the skeleton of the human metabolic network. Some other useful reaction and compound databases are listed in Table 1. These databases can be used to add more reactions and compounds to the human metabolic network. For example, in the Edinburgh human metabolic network (EHMN) there are nearly 1000 compounds and reactions that are missing in KEGG [24].

It should be mentioned that the enzyme-reaction relationships are often not simple one-to-one relationships. One enzyme may catalyze several metabolic reactions and one reaction can also be catalyzed by different enzymes. Because of these complex relationships, EC numbers themselves are often not regarded as good guides to what reactions are in a network. In fact, orthologous proteins from two different organisms may have the same EC number but with slightly different catalysis function. For example, the GBA3 gene in human codes for cytosolic beta-glucosidase that hydrolyses certain flavonoid glucosides [25]. The EC number for this enzyme function is 3.2.1.21. In certain organisms proteins with this EC number also function as catalysts for the degradation of cellulose [26]. However, this degradation reaction apparently does not occur in humans. Unfortunately there is still no automatic way to obtain the human specific EC number-reaction (human specific function) relationships. The reactions often need to be manually checked against the literature to exclude non-human reactions.

There are also other reasons why we need to consolidate and improve the reconstructed network using literature. Firstly, there may be inconsistent gene annotation in different databases. For example, about 600 gene-EC relationships are different between KEGG and Uniprot. Secondly, many genes are annotated with an incomplete EC number like 1.2.-.- and for those genes it is difficult to find the proper reaction by automatic database search. Thirdly, many enzymes that have been experimentally proved in the human body are not found encoded in the genome because the function of a large part of the human genome is still unknown [27,28]. Such enzymes and reactions should be added from the

literature to avoid gaps in the reconstructed network. In addition, certain bioinformatics methods such as those based on genome-context can also be used to find the candidate genes for filling gaps in the pathways [29].

The reconstruction of a high quality metabolic network with extensive literature support may take several years. However, this labour intensive and time-consuming process is necessary if a complete and reliable network is required for further functional analysis. So far, two high quality human metabolic networks have been developed independently: EHMN and Human Recon 1 reconstructed by Palsson's group [24,30]. A brief comparison of the two networks can be seen in Table 2. It should be noted that Recon 1 is a compartmented network including more than 1000 transport/exchange reactions and the same reactions occurring in different compartments are also counted as different reactions. This is the reason why Recon 1 has more reactions but less genes and metabolites. Only 1069 genes are common to both networks. This implies that both networks have some missing parts and a combination and cross-validation of the two networks can generate a more complete and reliable network. However, because of the different compound names used in the two networks, it is not easy to merge the two networks. EHMN mainly uses the KEGG compound ID, but there are about 1000 compounds obtained from literature without KEGG ID. About a half of the metabolites in Recon 1 do not have a link to KEGG compound ID. A compound database which provides cross-links to the existing compound databases and includes more synonyms would be useful for comparison and combination of the reconstructed networks from different research groups.

## Human metabolic network, diseases and drug targets

As stated in the introduction, human metabolism is closely related with human diseases. To quantitatively evaluate the relationship, we obtained the disorder-gene association information from the Morbid map of the OMIM (Online Mendelian Inheritance in Man) database [31]. The current (July 2007) Morbid map contains 2980 genes associated with 4542 disorders. The percentage of the disease related genes in the human genome is 8.2% (2980/36547). Among the disease related genes, 528 are in EHMN, accounting for 22.9% of the total metabolic genes (2308). Because an enzyme reaction is often catalyzed by proteins encoded by different genes, the percentage of the reactions affected by disease related genes is even higher at 48.0%. This fact clearly indicates the deep involvement of human metabolism in disease processes.

Many small metabolites function as signalling molecules such as hormones and neurotransmitters that can bind with protein receptors, and through complex signal transduction pathways regulate the concentrations and activities of many other proteins. Therefore, malfunction of the enzymes involved in synthesis and degradation of such metabolites results in their abnormal concentrations and subsequently causes various diseases because of the broad downstream effects. For example, lack of dopamine is the main cause of Parkinson's disease and a shortage of acetylcholine in the brain has been associated with Alzheimer's disease. On the other hand, drugs targeted at those enzymes to recover the concentrations of the signal metabolites could be developed for disease treatment. Recently several papers have studied the distribution of drug targets in the human genome [32–34]. Not surprisingly, the most-targeted proteins are membrane receptors such as the histamine H1 receptor and the dopamine receptor $D_2$. Drugs targeted at these proteins are often the agonists or antagonists of the receptors which can restore their activities. An alternative way to restore the receptor activity is to bring the concentrations of the endogenous signal molecules to normal by targeting the enzymes. However, it is often very difficult to change the concentration of one metabolite while keeping others unchanged (to avoid side effects) because the metabolic network is a complex system. That is why enzymes are not widely targeted. Nevertheless, many human metabolic enzymes have been chosen as drug targets [33,34]. An analysis of the drug targets in the DrugBank database [35] revealed that 351 proteins in the EHMN are drug targets. These proteins are involved in more than 1000 reactions in the network. The pathways with a large number of drug targets include purine metabolism, tyrosine metabolism and tryptophan metabolism, glutamate-derived amino acid metabolism, bile acid synthesis, androgen and estrogen biosynthesis, glycerophospholipid metabolism, and so on. The purine metabolism pathway contains the highest number of drug targets (25 targets). However, most of them are for maintaining the levels of cAMP and cGMP, recovering the nucleotide through the salvage pathway, Na+, H+ coupled ATP dephosphorylation but not in the de novo purine synthesis pathway. The drug targets in the tyrosine metabolism pathway are mainly for the synthesis of monoamine neurotransmitters and hormones. As shown in Figure 1 for part of the pathway, four enzymes that are important for controlling the levels of the neurotransmitters (mainly dopamine) have been selected as drug targets. Compounds that can affect the activities of these enzymes have been developed as drugs for the treatment of Parkinson's disease and other neurological disorders by restoring the dopamine level in the brain cell.

**TABLE 2**

**Comparison of EHMN and Homo sapiens Recon 1**

| | # *Genes* | # Reactions | # Metabolites | # Pathways |
|---|---|---|---|---|
| **EHMN** | 2322 | 2823 | 2671 | 70 |
| **Recon 1** | 1496 | 3748 | 1469 | 88 |

## How the reconstructed network helps in drug discovery

As mentioned above, many enzymes have already been selected as drug targets. The high quality human metabolic network will help identify more enzyme targets through systematic analysis of the metabolic network. Methods for theoretical analysis of metabolic networks such as metabolic control analysis, flux balance analysis, elementary flux mode analysis and kinetic modeling have been developed to find the optimal pathways for producing specific metabolites, minimal enzyme sets for blocking a metabolic conversion process and the best enzymes for changing the concentration of a metabolite [2,36,37]. These methods can also be used in studying the mechanism of diseases and finding the most effective drug targets. For example, Cakir et al. [38] have shown how enzyme deficiencies in the red blood cell could affect the metabolism using elementary flux mode analysis. Cascante *et al.* [39] have shown several examples of using metabolic control analysis

to find enzyme targets in a recent review paper. A common problem of those methods is that they cannot be directly applied to very large-scale networks. Therefore the whole network needs to be decomposed to obtain small-scale structurally and functionally independent modules or pathways [40]. By combining theoretical analysis methods with existing knowledge of drug targets, we would be able to check how many enzymes in a related pathway have been targeted and if we can approach the same or better results by targeting other enzymes in the pathway.

Metabolic network analysis can also be used to predict drug side effects in the drug discovery process. Computational screening of drugs to find potential harmful side effects will allow pharmaceutical companies to eliminate drug candidates before they undergo expensive animal testing and clinical trials. Recently, Xie *et al.* [41] reported a computational approach to predict the side effects of drugs by finding the off-target proteins through protein structure analysis. Besides this, drug side effects are also related to the unexpected propagation of drug effects in the human body. The reconstructed human metabolic network can be used to predict how an enzyme modification would affect the downstream proteins and metabolites. Graph theory based pathway analysis methods could be very useful for such analysis. First, the metabolic network needs to be converted to a metabolite graph (metabolites as nodes) or a reaction graph (reactions as nodes). Then all the nodes (called output domain in graph theory) that can be reached by the modified nodes can be found and then we can check which enzymes or metabolites in the output domain have important physiological functions and thus could lead to serious side effects. It is important to exclude the currency metabolites such as $H_2O$, $CO_2$ in graph analysis to avoid the biologically meaningless pathways through those metabolites [3].

High throughput technologies such as microarray profiling and metabolomics are increasingly used in the drug discovery process [42,43]. A huge amount of data is generated from patient samples, tissues and cells treated by drug candidates, and so on. It is a major challenge to find the disease related genes (as possible targets) and evaluate the effects of drugs from the data. A very promising application of the reconstructed network is to map the experimental data onto it and analyze the data from a network perspective [44]. Actually, several recent studies have shown the usefulness of the human protein–protein interaction networks and signal transduction networks in drug discovery [32,45,46]. By contrast, research initiatives on the use of the human metabolic network in drug discovery are very limited [47]. The main reasons for this are: (1) the high quality human network was not available until recently; (2) the graph representation of a metabolic network is not quite as straightforward as for a PPN because a metabolic network is naturally a hypergraph where one edge (a protein or reaction) often connects more than two nodes (metabolites). However, in comparison with PPN, a metabolic network has several features that are important in drug discovery. First, the literature based metabolic network is more reliable than the PPN derived from high throughput experiments. Second, a metabolic network is a directed network but PPN is not. We can only predict the downstream effect of a perturbation (by drugs) in a directed network. Last but not least, we do not really know the biological function of an interaction in PPN and therefore cannot derive a dynamic model to predict the state change of the proteins in the network. Actually, PPN is mainly used to find previously unknown functional links between genes and thus find new candidate genes associated with a disease or a drug. Further experimental studies are required to determine the exact function of the new genes. By contrast, in a metabolic network, all the functions of the proteins are known. The main objective is to understand how the whole system is related with a disease and to find good targets by systems analysis of the network. From this point of view, more systems analysis methods should be developed for making use of the human metabolic network in drug discovery processes, but it will also generate more in-depth understanding of the mechanism of diseases and thus provide better guidance for drug discovery.

The reconstructed human metabolic networks provide an unprecedented large system for the study of diseases and development of drugs at a new systems level. The systems analysis will help us not only in discovery of novel drug targets but also in developing new systems-based therapy strategies. Structure analyses of metabolic networks and other biological networks have revealed certain important network organization principles such as scale free, hierarchical modular organization and the bow-tie connectivity structure [48]. These structure features contribute to the robustness and flexibility of the complex biosystems and may explain, in general, the fact that many drug candidates are ineffective (drug effect is compensated by other paths in the network) or show unexpected severe side effects (drug effect propagates globally) [49–51]. Prompted by these findings, many scientists are proposing a systems-oriented drug design strategy to replace the current one-drug-one-target-one disease approach [51–53]. A disease is regarded as a wrong state of the system and thus the therapy should aim at changing the state of the system. Metabolic control analysis already indicated that to change the flux in a pathway it is often better to manipulate two or more enzymes than one key enzyme [36]. For a complex system, like the human metabolic network, it is also reasonable that multiple target modifications can more effectively convert the system from a disease state to a normal state than a single target modification. Actually successful applications of multi-component therapies have been reported and multi-component drugs are already on the market [54,55]. For example, lovastatin is an inhibitor for 3-hydroxy-3-methylglutaryl-coenzyme in the cholesterol synthesis pathway. Niacin can reduce the blood cholesterol level by reducing the transportation of cholesterol. Lovastatin and Niacin have been combined together as a new drug 'Advicor' by Kos Pharmaceutical to achieve improved results [55].

The main challenge for applying a systems-oriented strategy in drug discovery processes is the increasing complexity and cost. For example, for 100 drug candidates for target A and 100 candidates for target B it will require only 200 experiments to screen the best one-target drug but will need 10000 experiments for a two-component drug, in addition to which the relative proportions of the two components may also vary. Therefore, a prior analysis of the system to identify a small subset of proper combinations for further drug screening experiments is of particular importance in a systems-oriented drug design strategy. The reconstruction of the human metabolic networks as well as other biological networks provides a basis for such an analysis.

## Future directions

To make good use of the reconstructed human metabolic network, efforts should be made to integrate disease and drug information into the networks which are available as regularly updated databases. Thanks to recent work on developing databases for drugs and disease [56], it is now comparatively simple to collect and integrate that information into the metabolic network. Such an integrated database will make it easy to study the diseases in a network framework and to study the complex relationship between diseases, which may lead to a new systems-based disease classification [57]. As shown in recent studies from various aspects [57–59], different diseases may link to the same or closely related genes and thus often co-occur in patients.

To understand disease mechanisms better, it is also required to integrate the metabolic network with signal transduction and gene regulatory networks. As discussed previously, many diseases are related to abnormal concentrations of signal metabolites. The signal metabolites can bind to protein receptors and coordinate their activities which then regulate a set of proteins and genes through the signal transduction networks. For example, there are two classes of dopamine receptors, D1 and D2, which activate adenylyl cyclase (increasing cAMP concentration) and the cyclic nucleotide phosphodiesterases (decreasing cAMP concentration), respectively. Through the cAMP-dependent protein kinase pathway they affect the activity of many other proteins [60]. Therefore only by carefully integrating the metabolic pathways and the signal transduction pathways can we obtain a comprehensive picture of disease mechanisms.

## References


1 Thomas, C.E. and Ganji, G. (2006) Integration of genomic and metabonomic data in systems biology – are we 'there' yet? *Curr. Opin. Drug. Discov. Dev.* 9, 92–100
2 Palsson, B. (2006) *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press
3 Ma, H.W. and Zeng, A.P. (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 19, 270–277
4 Stelzl, U. and Wanker, E.E. (2006) The value of high quality protein–protein interaction networks for systems biology. *Curr. Opin. Chem. Biol.* 10, 551–558
5 Futschik, M.E. *et al.* (2007) Comparison of human protein–protein interaction maps. *Bioinformatics* 23, 605–611
6 Riley, R.D. *et al.* (2004) A systematic review of molecular and biological tumor markers in neuroblastoma. *Clin. Cancer Res.* 10 (Pt 1), 4–12
7 Schwedhelm, E. and Boger, R.H. (2003) Application of gas chromatography-mass spectrometry for analysis of isoprostanes: their role in cardiovascular disease. *Clin. Chem. Lab. Med.* 41, 1552–1561
8 Schlotterbeck, G. *et al.* (2006) Metabolic profiling technologies for biomarker discovery in biomedicine and drug development. *Pharmacogenomics* 7, 1055–1075
9 Ford, E.S. *et al.* (2002) Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *JAMA* 287, 356–359
10 Menendez, J.A. and Lupu, R. (2007) Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* 7, 763–777
11 Romero, P. *et al.* (2005) Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* 6, R2
12 Reed, J.L. *et al.* (2006) Towards multidimensional genome annotation. *Nat. Rev. Genet.* 7, 130–141
13 Schomburg, I. *et al.* (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* 32, D431–D433 Database issue
14 Bairoch, A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.* 28, 304–305
15 Hubbard, T.J. *et al.* (2007) Ensembl 2007. *Nucleic Acids Res.* 35, D610–D617 Database issue
16 Eyre, T.A. *et al.* (2006) The HUGO Gene Nomenclature Database, 2006 updates. *Nucleic Acids Res.* 34, D319–D321 Database issue
17 Kanehisa, M. *et al.* (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280
18 Wu, C.H. *et al.* (2006) The universal protein resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* 34, D187–D191 Database issue
19 Maglott, D. *et al.* (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* 35, D26–D31 Database issue
20 Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29
21 Goto, S. *et al.* (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* 30, 402–404
22 Wheeler, D.L. *et al.* (2006) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 34, D173–D180
23 Brooksbank, C. *et al.* (2005) The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res.* 33, D46–D53 Database issue
24 Ma, H. *et al.* (2007) The Edinburgh Human Metabolic Network reconstruction and its functional analysis. *Mol. Syst. Biol.* 3, 135
25 Berrin, J.G. *et al.* (2003) Substrate (aglycone) specificity of human cytosolic beta-glucosidase. *Biochem. J.* 373 (Pt 1), 41–48
26 Mukherjee, S. *et al.* (2006) Cellobiase from Termitomyces clypeatus: activity and secretion in presence of glycosylation inhibitors. *Biotechnol. Lett.* 28, 1773–1778
27 Karp, P. (2004) Call for an enzyme genomics initiative. *Genome Biol.* 5, 401
28 Lespinet, O. and Labedan, B. (2006) ORENZA: a web resource for studying orphan enzyme activities. *BMC Bioinform.* 7, 436
29 Green, M.L. and Karp, P.D. (2007) Using genome-context data to identify specific types of functional associations in pathway/genome databases. *Bioinformatics* 23, i205–i211
30 Duarte, N.C. *et al.* (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1777–1782
31 Hamosh, A. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 33, D514–D517 Database issue
32 Yildirim, M.A. *et al.* (2007) Drug-target network. *Nat. Biotechnol.* 25, 1119–1126
33 Imming, P. *et al.* (2006) Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Discov.* 5, 821–834
34 Overington, J.P. *et al.* (2006) How many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993–996
35 Wishart, D.S. *et al.* (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 34, D668–D672 Database issue
36 Fell, D.A. (1998) Increasing the flux in metabolic pathways: A metabolic control analysis perspective. *Biotechnol. Bioeng.* 58, 121–124
37 Klamt, S. and Gilles, E.D. (2004) Minimal cut sets in biochemical reaction networks. *Bioinformatics* 20, 226–234
38 Cakir, T. *et al.* (2004) Metabolic pathway analysis of enzyme-deficient human red blood cells. *Biosystems* 78, 49–67
39 Cascante, M. *et al.* (2002) Metabolic control analysis in drug discovery and disease. *Nat. Biotechnol.* 20, 243–249
40 Ma, H.W. *et al.* (2004) Decomposition of metabolic network based on the global connectivity structure of reaction graph. *Bioinformatics* 20, 1870–1876
41 Xie, L. *et al.* (2007) In silico elucidation of the molecular mechanism defining the adverse effect of selective estrogen receptor modulators. *PLoS Comput. Biol.* 3, e217
42 Dohrmann, C.E. (2004) Target discovery in metabolic disease. *Drug Discov. Today* 9, 785–794
43 Fan, T.W. *et al.* (2006) Integrating metabolomics and transcriptomics for probing SE anticancer mechanisms. *Drug Metab. Rev.* 38, 707–732
44 Nikolsky, Y. *et al.* (2005) Biological networks and analysis of experimental data in drug discovery. *Drug Discov. Today* 10, 653–662
45 Ruffner, H. *et al.* (2007) Human protein–protein interaction networks and the value for drug discovery. *Drug Discov. Today* 12, 709–716
46 Araujo, R.P. *et al.* (2007) Proteins, drug targets and the mechanisms they control: the simple truth about complex networks. *Nat. Rev. Drug Discov.* 6, 871–880
47 Sridhar, P. *et al.* (2007) An iterative algorithm for metabolic network-based drug target identification. *Pac. Symp. Biocomput.* 12, 88–99
48 Ma, H.W. and Zeng, A.P. (2003) The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19, 1423–1430
49 Kitano, H. (2004) Biological robustness. *Nat. Rev. Genet.* 5, 826–837
50 Csete, M. and Doyle, J. (2004) Bow ties, metabolism and disease. *Trends Biotechnol.* 22, 446–450

51 Kitano, H. (2007) A robustness-based approach to systems-oriented drug design. *Nat. Rev. Drug Discov.* 6, 202–210
52 Zimmermann, G.R. *et al.* (2007) Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov. Today* 12, 34–42
53 Morphy, R. and Rankovic, Z. (2007) Fragments, network biology and designing multiple ligands. *Drug Discov. Today* 12, 156–160
54 Nelson, H.S. (2001) Advair: combination treatment with fluticasone propionate/salmeterol in the treatment of asthma. *J. Allergy Clin. Immunol.* 107, 398–416
55 Gupta, E.K. and Ito, M.K. (2002) Lovastatin and extended-release niacin combination product: the first drug combination for the management of hyperlipidemia. *Heart Dis.* 4, 124–137
56 Wishart, D.S. *et al.* (2007) HMDB: the Human Metabolome Database. *Nucleic Acids Res.* 35, D521–D526 Database issue
57 Oti, M. and Brunner, H.G. (2007) The modular nature of genetic diseases. *Clin. Genet.* 71, 1–11
58 Goh, K.I. *et al.* (2007) The human disease network. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8685–8690
59 Rzhetsky, A. *et al.* (2007) Probing genetic overlap among complex human phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11694–11699
60 Girault, J.A. and Greengard, P. (2004) The neurobiology of dopamine signaling. *Arch. Neurol.* 61, 641–644