Editorial

QSAR and Complex Networks in Pharmaceutical Design, Microbiology, Parasitology, Toxicology, Cancer and Neurosciences

Both, computer-aided Pharmaceutical Design and Drug Target Discovery using Bioinformatics are valuable tools in biomedical sciences. They may become useful in order to reduce costs in terms of material resources, personal, time and the use of animals of laboratory in the exploration of large databases. These techniques are not aimed to replace experimentation at all; we should understand these methods only as a guide to "seek the needle in the haystack". There are many computational techniques and mathematical models useful in this sense. In particular, Graph theory is of special interest due to its high flexibility to study many types of systems ranging from drug molecules to drug target proteins and beyond. In fact, many authors have used molecular graphs to represent the structure of drugs by means of vertices (represented by dots) that represent atoms and edges that represent chemical bonds. Consequently, molecular graphs express the structure of organic compounds in terms of atom connectivity. In addition, we can associate graphs with different classes of numeric matrices to carry out computational studies. The Boolean or Adjacency matrices are perhaps the more simple to explain. These matrices are square tables with elements $b_{ij} = 1$ for pair of connected nodes and 0 otherwise. At one higher structural level we can use essentially the same type of graphs to study complex networks used to represent the 3D structures of proteins (enzymes, molecular targets, channels, receptors). The construction of this type of graphs and matrices is straightforward to realize in an intuitive form taking into consideration the analogy between the previous situations. In these networks, aminoacids often play the role of nodes and links express spatial contact between two aminoacids (see also, contact maps or residue networks). In the same group with proteins we can find the graphs used to represent the secondary structure of RNAs. In this last class of networks, nucleotides often play the role of nodes and links express that a pair of bases are sequence neighbors or are involved in a hydrogen bond. In parallel, many authors have been used Graph and Complex Network theory to approach very large networks with low computational cost. These large networks are graphical representations of real bio-systems with essentially two components nodes and links in a broad sense. In the case of bio-systems of certain relevance for Current Pharmaceutical Design we can name drug-target networks, protein interaction networks (PINs) used to represent proteomes, drug-tissue action networks, and drug - disease/gen-disease networks for diseasome, to cite only some examples. These are the same type of above-mentioned graphs but nodes are not atoms or aminoacids but proteins, tissues, targets, patients, diseases, population groups, disease incidence regions, etc. Node-to-node links (edges or arcs) express different types of ties or relationships between two nodes as for instance: drug-target inhibition, gen-disease regulation. In all these cases, we can easily calculate different invariant parameters of the matrices associated to the graphs that may be used to describe the structure of these objects (drugs, proteins, or large bio-systems). As this numbers are based only on connectivity information they are often named as Connectivity measures or Topological Indices (TIs). To recommended readings connecting these topics are both the comprehensive handbook in graph and complex networks [1] and the handbook of molecular descriptors [2].

In fact, in our days, there is an explosion on the use of Topological Indices (TIs) of Graphs and Complex Networks on a broad spectrum of topics related to Drug metabolism and distribution research. Using TIs as inputs we can find Quantitative Structure-Property Relationships (QSPR) models for any kind of bio-systems in principle. We see QSPR model as a function that predict the properties of the system (drug, protein, RNA, diseasome) using parameters that numerically describe the structure of the system (like TIs). There are many QSPR-like terms that fit to more specific situations, for instance Quantitative Structure-Activity Relationships (QSAR), Quantitative Structure-Toxicity Relationships (QSPR), Quantitative Proteome-Property Relationships (QPPR), Quantitative Sequence-Action Model (QSAM), or Quantitative Structure-Reactivity Relationships (QSRR), to cite a few examples. In all this cases we can find models that use the TIs of the system as input to predict the properties of this system (output), see the recent book edited by González-Díaz and Munteanu in 2010 [3].

In a recent, preliminary review in the field published in Proteomics in 2008 González-Diaz *et al.* discussed the use of these methods but only from the point of view of proteins [4]. Next we extended the discussion to a collective of authors edited a special issue on TIs but ever restricted to the field of protein and proteomics; published in Current Proteomics in December 2009 [5-11]. In other recent issue, we guest-edited [12] a series of papers devoted to QSPR techniques but only from the point of view of low-molecular-weight drugs without discussion of metabolism or distribution; this issue was published in Current Topics in Medicinal Chemistry in 2008 [12-21]. Last, we guest-edited [22] an issue focused on graph TIs approach to Drug ADMET processes and Metabolomics, see the papers published in the issue of may 2010 for the journal Current Drug Metabolism [23-30]. In any case, we believe that there is necessity of a collection of manuscripts or issue more focused on QSAR, TIs and networks applied to pharmaceutical design at all structural levels. Based on all these reasons we edited the present issue including QSPR/QSAR studies with applications to Pharmaceutical Design and related areas like Microbiology, Parasitology, Pharmacology, Chemoterapy, Epidemiology, Toxicology, and others.

In this sense, the present issue provides state of the art reviews of some of these new computational approaches in this rapidly expanding area. Taking these aspects into consideration, in the first work of this issue Marero-Ponce *et al.* [31] reviewed the uses of QSAR methods for *in silico* identification of new families of tyrosinase inhibitors. Assembling, validation of models through prediction series, and virtual screening of external data sets are also shown to prove the accuracy of the QSAR models. The authors put special emphasis on QSAR models obtained with the TOMOCOMD-CARDD (TOpological MOlecular COMputational Design-Computer-Aided Rational Drug Design) software and Linear Discriminant Analysis (LDA) as statistical technique. Finally, a translation to real world applications is shown by the use of QSAR models in the identification and posterior *in vitro* evaluation of different families of compounds like tetraketones, cycloartanes, ethylsteroids, lignans, dicoumarins and vanilloid derivatives.

In the second work, Roy and Ghosh [32] explored other of the above-mentioned directions. This communication reviews published reports of QSARs/QSPRs with Extended Topochemical Atom (ETA) indices for modeling chemical and drug induced toxicities and some physicochemical properties relevant to such toxicities. In each study, ETA models have been compared to those developed using various non-ETA models and it was found that the quality of the QSARs involving ETA parameters were comparable to those involving non-ETA parameters. ETA descriptors were also found to increase statistical quality of the models involving non-ETA parameters when used in combination. On the basis of the reported studies, the authors concluded that the ETA descriptors are sufficiently rich in chemical information to encode the structural features contributing to the toxicities and these indices may be used in combination with other TIs and physicochemical descriptors for development of predictive QSAR models. Such models may be used for virtual screening and *in silico* prediction of toxicities, and if appropriately used, these may be proved helpful for regulatory decision support and decision making processes.

In the third paper, Munteanu *et al.* [33] started recognizing the need for a study of the complex diseases, like Cancer and Neurodegenerative disorders and others, due to their important impact on our society. The authors suggest that one of the solutions involves

the theoretical methods which are fast and efficient tools that can lead to the discovery of new active drugs specially designed for these diseases, including QSAR and the complex network theory. Both become important solutions for screening and designing efficient pharmaceuticals by coding the chemical information of the molecules into molecular descriptors. In this review the authors presents the most recent studies on drug discovery and design using QSAR of several complex diseases in the fields of Neurology, Cardiology and Oncology.

In the next work of this issue, Speck-Planché *et al.* [34] discussed the increasing resistance of *Mycobacterium tuberculosis* to the existing drugs that has alarmed the worldwide scientific community dedicated to Pharmaceutical Design and Microbiology. In an attempt to overcome this problem the authors recommend to apply computer-aided drug design, which provide an extraordinary support to the different strategies in drug discovery. The authors stated that there are around 250 biological receptors like enzymes that can be used in principle, for the design of anti-tuberculosis compounds that act by a specific mechanism of action. Also, there more than 5000 compound available in the literature, and that constitute important information in order to search new molecular patterns for the design of new anti-tuberculosis agents. In this paper the authors explored the current state of drug discovery of anti-tuberculosis agents with QSAR based on TIs and other indices.

On the other hand GSK-3 (glycogen synthase kinase 3) inhibitors are interesting candidates to develop Anti-Alzheimer compounds, which is also a complex disease. GSK-3β are also interesting in Pharmaceutical Design and Parasitology as Anti-parasitic compounds active against *Plasmodium falciparum*, *Trypanosoma brucei*, and *Leishmania donovani*; the causative agents for Malaria, African Trypanosomiasis and Leishmaniosis. In a very interesting work of this issue García, Fall, and Gómez [35] noted that the high number of possible candidates creates the necessity of QSAR models in order to guide the GSK3 inhibitor synthesis. In this work, the authors revised different computational studies for a very large and heterogeneous series of GSK-3 inhibitors. First, they revised QSAR studies with conceptual parameters such as flexibility of rotation, probability of availability, etc. Next, they used the method of regression analysis and QSAR studies in order to understand the essential structural requirement for binding with receptor. Last, they reviewed 3D-QSAR, CoMFA and CoMSIA with different compounds to find out the structural requirements for GSK-3 inhibitory activity.

In another review of this special issue Estrada *et al.* [36] discussed new applications of the TOPological Substructural MOlecular DEsign approach (TOPS-MODE), which has been used to formulate structural rules for binding of substrates of P-glycoprotein (P-gp). The authors first review some of the QSAR models developed in the recent literature for predicting binding to P-gp. Then, they develop their own QSAR model using TOPS-MODE, which is able to identify 88.4% of substrates and 84.2% of non-substrates. When the model is presented to an external prediction set of 100 substrates and 77 non-substrates it identifies correctly 81.8% of all cases. Using TOPS-MODE strategy they found structural contributions for binding to P-gp, which identifies 24 structural fragments responsible for such binding. This group carried out a chemico-biological analysis of some of the structural fragments found as contributing to P-gp binding of substrates and show that in general the model developed so far can be used as a virtual screening method for identifying substrates of P-gp from large libraries of compounds.

In the next paper of this issue Concu *et al.* [37] discussed applications of QSAR and Complex Networks to a higher structural level in the manuscript entitled: Review of QSAR models for Enzyme Classes of Drug targets: Theoretical background and Applications in Parasites, Hosts, and other organisms. In fcat, the number of protein 3D structures without function annotation in Protein Data Bank (PDB) has been steadily increased. Many of these proteins are relevant for Pharmaceutical Design because they may be enzymes of different classes that could become drug targets. This fact has led in turn to an increment of demand for theoretical models to give a quick characterization of these proteins. Consequently, in this work the authors present a review and discussion of Alignment-Free Methods (AFMs) for fast prediction of the Enzyme Classification (EC) number from structural patterns. The review refers to both type of methods based on linear techniques such as Linear Discriminant Analysis (LDA) and/or non-linear models like Artificial Neural Networks (ANN) or Support Vector Machine (SVM) in order to compare linear vs. non-linear classifiers. We also detected which of these models have been implemented as Web Servers free to the public and compiled a list of some of these web sites. For instance, we reviewed the servers implemented at portal Bio-AIMS (http://miaja.tic.udc.es/Bio-AIMS/EnzClassPred.php) and the server EzyPred (http://www.csbio.sjtu.edu.cn/bioinf/EzyPred/).

After that, Vazquez-Naya *et al.* [38] reviewed the application complex diseases in the field of Neurology, Cardiology and Oncology computational ontology methods used to unravel items that are linked with the molecule metabolism and the treatment of these diseases giving us the possibility to correlate information from different levels and to discover new relationships between complex diseases such as common drug targets and disease patterns. This review presents the ontologies used for drug discovery and Pharmaceutical Design in the most common complex diseases.

This special issue close with a review work after González-Díaz *et al.* [39]. The authors review QSAR and/or complex network models have been used in Pharmaceutical design and Parasitology for the discovery of anti-parasite drugs. They focus on QSAR models that predict biological activity using as input TIs. The first topic reviewed was: TIs and QSAR for anti-parasitic drugs. This topic included: Theoretical Background, QSAR for anti-malaria drugs, QSAR for anti-toxoplasma drugs. The second topic was: TOMO-COMD approach to QSAR of anti-parasitic drugs, including TOMO-COMD theoretical background and TOMO-COMD models for anthelmintic activity, trichomonacidals, anti-malarials, anti-trypanosome compounds. The third section was introduced to discuss TIs in the context of Complex Networks. The last section is devoted to the MARCH-INSIDE approach to QSAR of anti-parasitic drugs and targets. This begins with a theoretical background for drugs and parameters for proteins. Next, we reviewed MARCH-INSIDE models for Pharmaceutical Design of anti-parasitic drugs including: flukicidal drugs and anti-coccidial drugs. We close MARCH-NSIDE topic with a review of multi-target QSAR of antiparasitic drugs, MARCH-INSIDE assemble of Complex networks of anti-parasitic drugs. We closed the MARCH-INSIDE section discussing the prediction of proteins in parasites and MARCH-INSIDE web-servers for Protein-Protein interactions in parasites: Plasmod-PPI and Trypano-PPI web-servers. We closed this revision with an important section devoted to review some on legal issues related to QSAR models.

REFERENCES

[1] Bornholdt S, Schuster HG. Handbook of graphs and complex networks: from the genome to the internet. Wheinheim: WILEY-VCH GmbH & CO. KGa. 2003.
[2] Todeschini R, Consonni V. Handbook of molecular descriptors. USA: Wiley-VCH 2002.
[3] González-Díaz H, Munteanu CR. Topological indices for medicinal chemistry, biology, parasitology, neurological and social networks. Kerala: Transworld Research Network 2010.
[4] González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. Proteomics 2008; 8: 750-78.
[5] Chen J, Shen B. Computational analysis of amino acid mutation: a proteome wide perspective. Curr Proteomics 2009; 6: 228-34.
[6] Chou KC. Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology. Curr Proteomics 2009; 6: 262-74.

[7] Giuliani A, Di Paola L, Setola R. Proteins as networks: a mesoscopic approach using haemoglobin molecule as case study. Curr Proteomics 2009; 6: 235-45.
[8] González-Díaz H, Prado-Prado F, Pérez-Montoto LG, Duardo-Sánchez A, López-Díaz A. QSAR models for proteins of parasitic organisms, plants and human guests: theory, applications, legal protection, taxes, and regulatory issues. Curr Proteomics 2009; 6: 214-27.
[9] Pérez-Montoto LG, Prado-Prado F, Ubeira FM,González-Díaz H. Study of parasitic infections, cancer, and other diseases with mass-spectrometry and quantitative proteome-disease relationships. Curr Proteomics 2009; 6: 246-61.
[10] Torrens F, Castellano G. Topological charge-transfer indices: from small molecules to proteins. Curr Proteomics 2009: 204-13.
[11] Vázquez JM, Aguiar V, Seoane JA, *et al*. Star graphs of protein sequences and proteome mass spectra in cancer prediction. Curr Proteomics 2009; 6: 275-88.
[12] Gonzalez-Diaz H. Quantitative studies on structure-activity and structure-property relationships (QSAR/QSPR). Curr Top Med Chem 2008; 8: 1554.
[13] Ivanciuc O. Weka machine learning for predicting the phospholipidosis inducing potential. Curr Top Med Chem 2008; 8: 1691-709.
[14] Gonzalez-Diaz H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. Curr Top Med Chem 2008; 8: 1676-90.
[15] Duardo-Sanchez A, Patlewicz G, Lopez-Diaz A. Current topics on software use in medicinal chemistry: intellectual property, taxes, and regulatory issues. Curr Top Med Chem 2008; 8: 1666-75.
[16] Wang JF, Wei DQ, Chou KC. Drug candidates from traditional chinese medicines. Curr Top Med Chem 2008; 8: 1656-65.
[17] Helguera AM, Combes RD, Gonzalez MP, Cordeiro MN. Applications of 2D descriptors in drug design: a DRAGON tale. Curr Top Med Chem 2008; 8: 1628-55.
[18] Gonzalez MP, Teran C, Saiz-Urra L, Teijeira M. Variable selection methods in QSAR: an overview. Curr Top Med Chem 2008; 8: 1606-27.
[19] Caballero J, Fernandez M. Artificial neural networks from MATLAB in medicinal chemistry. Bayesian-regularized genetic neural networks (BRGNN): application to the prediction of the antagonistic activity against human platelet thrombin receptor (PAR-1). Curr Top Med Chem 2008; 8: 1580-605.
[20] Wang JF, Wei DQ, Chou KC. Pharmacogenomics and personalized use of drugs. Curr Top Med Chem 2008; 8: 1573-9.
[21] Vilar S, Cozza G, Moro S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. Curr Top Med Chem 2008; 8: 1555-72.
[22] Gonzalez-Diaz H. Network topological indices, drug metabolism, and distribution. Curr Drug Metab 2010; 11: 283-4.
[23] Khan MT. Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. Curr Drug Metab 2010; 11: 285-95.
[24] Mrabet Y, Semmar N. Mathematical methods to analysis of topology, functional variability and evolution of metabolic systems based on different decomposition concepts. Curr Drug Metab 2010; 11: 315-41.
[25] Martinez-Romero M, Vazquez-Naya JM, Rabunal JR, *et al*. Artificial intelligence techniques for colorectal cancer drug metabolism: ontology and complex network. Curr Drug Metab 2010; 11: 347-68.
[26] Zhong WZ, Zhan J, Kang P, Yamazaki S. Gender specific drug metabolism of PF-02341066 in rats--role of sulfoconjugation. Curr Drug Metab 2010; 11: 296-306.
[27] Wang JF, Chou KC. Molecular modeling of cytochrome P450 and drug metabolism. Curr Drug Metab 2010; 11: 342-6.
[28] Gonzalez-Diaz H, Duardo-Sanchez A, Ubeira FM, *et al*. Review of MARCH-INSIDE & complex networks prediction of drugs: ADMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. Curr Drug Metab 2010; 11: 379-406.
[29] Garcia I, Diop YF, Gomez G. QSAR & complex network study of the HMGR inhibitors structural diversity. Curr Drug Metab 2010; 11: 307-14.
[30] Chou KC. Graphic rule for drug metabolism systems. Curr Drug Metab 2010; 11: 369-78.
[31] Marrero-Ponce Y, Casañola-Martín GM, Khan MTH, Torrens F, Rescigno A, Abad C. Ligand-based computer-aided discovery of tyrosinase inhibitors. Applications of the TOMOCOMD-CARDD method to the elucidation of new compounds. Curr Pharm Des 2010; 16(24): 2601-24.
[32] Roy K, Ghosh G. Exploring QSARs with extended topochemical atom (ETA) indices for modeling chemical and drug toxicity. Curr Pharm Des 2010; 16(24): 2625-39.
[33] Munteanu CR, Fernández-Blanco E, Seoane JA, *et al*. Drug discovery and design for complex diseases through QSAR computational methods. Curr Pharm Des 2010; 16(24): 2640-55.
[34] Speck-Planche A, Scotti MT, de Paulo-Emerenciano V. Current pharmaceutical design of antituberculosis drugs: Future perspectives. Curr Pharm Des 2010; 16(24): 2656-65.
[35] García I, Fall Y, Gómez G. QSAR, Docking, and CoMFA studies of GSK3 inhibitors. Curr Pharm Des 2010; 16(24): 2666-75.
[36] Estrada E, Molina E, Nodarse D, Uriarte E. Structural contributions of substrates to their binding to P-glycoprotein. A TOPS-MODE Approach. Curr Pharm Des 2010; 16(24): 2676-709.
[37] Concu R, Podda G, Ubeira FM, González-Díaz H. Review of QSAR models for enzyme classes of drug targets: theoretical background and applications in parasites, hosts and other organisms. Curr Pharm Des 2010; 16(24): 2710-23.
[38] Vázquez-Naya JM, Martínez-Romero1 M, Porto-Pazos AB, *et al*. Ontologies of drug discovery and design for neurology, cardiology and oncology. Curr Pharm Des 2010; 16(24): 2724-36.
[39] González-Díaz H, Romaris F, Duardo-Sanchez A, *et al*. Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, aplications and legal issues. Curr Pharm Des 2010; 16(24): 2737-64.

Humberto González-Díaz

Guest Editor
Applied Bioinformatics,
Department of Microbiology and Parasitology,
Faculty of Pharmacy,
University of Santiago de Compostela, 15782,
Spain