The WWW as a Tool to Obtain Molecular Parameters

Igor V. Tetko*

Institute for Bioinformatics, GSF, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany and Biomedical Department, IBPC, Murmanskaya 1, Kyiv, 253660, Ukraine

Abstract: This article analyses molecular property calculation resources available on the Internet. The first section summarizes the on-line database resources that could be useful to search molecular and biological properties of chemicals, and indicates some principal databases with physicochemical, thermochemical, toxicity, cancer and HIV data. The second section overviews popular standalone programs for calculation of molecular descriptors. Some of these programs can be downloaded for free and used as standalone applications for calculation of molecular descriptors. The third section describes on-line tools for the prediction of molecular properties, activities and calculation of molecular descriptors. Analysis of emerging tools that can be useful to developing new on-line servers for the prediction of molecular parameters and properties is also given.

Keywords: on-line tools, Internet, physico-chemical parameters, toxicity, medicinal chemistry, drug design

1. INTRODUCTION

After a successful start 10 years ago, the Internet dramatically penetrated all aspects of life in modern society [1]. Internet activities were extended in both a qualitative and quantitative way. The qualitative extension is mainly connected with increasing the bandwidth speed of the Internet. While the first private Internet users could use mainly dial-up connections using low-qualitative telephone lines (the connection speed of 2 Kbit per second was considered to be very good), modern users can easily have access to two-three order faster lines using modern technologies. A similar qualitative change from MB to GB bandwidth lines is also now utilized by many corporate users. Such remarkable increases in the quality of Internet connections dramatically influenced the variety and quality of services provided over the Internet, and the development of secure technologies provided fast development of e-commerce. Indeed, the United States market for E-business services, which includes consulting, IT outsourcing, software development, and system integration had a $10.3 billion dollar revenue in 2000 and is expected to reach $59 billion by 2003 [2]. In addition to traditional services (such as selling plane tickets and goods), new services (such as interactive TV, casino, interactive virtual-reality games, etc.) have begun to appear. There are fast changes even in such traditional services as e-mail, where one can now send voice or video mail or have an interactive teleconference over the Internet.

The Internet has also developed significantly in the field of chemistry [3-7]. The impact of the World Wide Web (WWW) on society has dramatically increased, especially in the fields of education and scientific research, and a great deal of information is now available for the chemist in the form of chemical databases (such as ChemFinder [8], ChemExper Chemical Directory [5]), electronic conferences [9], etc. There are also growing numbers of chemical journals and pre-print servers available exclusively over the Internet [10].

Of particular interest for medicinal chemists is the development and availability of over-the-Internet data analysis methods and molecular properties calculation. In particular there are a number of molecular descriptors that have found useful applications in the field of medicinal chemistry because of their relevance to absorption, distribution, metabolism and excretion (ADME, see article by Lombardo et al. in this issue). As an example the lipophilicity of molecules represents one of the most frequently used parameters in the Quantitative Structure Activity Relationship (QSAR) models (see article by Caron et al. in this issue). A survey of all such models reported in the QSAR journal in 1988 showed that more than 40% of articles involved some hydrophobic descriptor, and this number increased to over 50% in 1998 [11]. The melting point represents another parameter that is frequently used to predict aqueous solubility of compounds, which is very important physicochemical property of a drug.

To predict molecular properties and/or activities, the user need readily available and operable data. Therefore the first section of the paper provides an overview of experimental data concerning physicochemical, toxicological and biological properties of the chemicals available on the WWW. The second section analyzes standalone resources available for calculation of molecular descriptors for drug design and QSAR. The third section describes on-line resources available for calculation of molecular indices. Some virtual library collections of the WWW resources are described in the fourth section. The last section provides an overview of some modern Internet technologies that can be useful to develop new on-line calculation servers. It also analyzes some technical details of client-server applications used in the Virtual Computational Chemistry laboratory site (VCCLAB, http://www.vcclab.org).

*Address correspondence to this author at the Institute for Bioinformatics, GSF - Forschungszentrum für Umwelt und, Gesundheit, GmbH, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany; Tel. +49-89-3187-3575; Fax. x.3585; E-mail: itetko@vcclab.org; Website: http://www.vcclab.org

1389-5575/03 $41.00+.00 © 2003 Bentham Science Publishers, Ltd.

Overall this paper provides an overview of the WWW resources to calculate molecular parameters available on the Internet and describes some new technologies used to provide such services.

2. DATA RESOURCES

Users, who are working on the development of property prediction methods, have to rely on reliable experimental data that can be used to develop such models and validate them. Twenty years ago in the 80s, the main source of information was published media, i.e., journals, reviews and monographs. Researchers who wanted to develop new models had to retype all data collections, which was a time consuming and unproductive procedure. The appearance of the Internet and development of computer technologies dramatically changed this situation. Many different databases were created and made available for research for free or for some limited fee that is incomparable with the time that the user would spend to collect such a database. The following subsections will overview some of the databases that can be useful to develop new property predictions. All these databases are available and searchable on-line (some of them require subscription and fee) or can be downloaded and used as standalone applications.

2.1 NIST Chemistry Web Book

A large database with different thermochemical, thermophysical, and ion energetics data with about 40,000 species compiled by the National Institute of Standard and Technology (NIST) is available on the Chemistry Web Book (http://webbook.nist.gov). This site is updated periodically and the most recent update was performed in March 2003. The thermochemical data include enthalpies of formation, entropies, constant pressure heat capacities, critical temperatures (melting and boiling points) and pressures, enthalpies of phase transitions and vapor pressures. Data provided for reactions include Gibbs free energy values, enthalpies and entropies. The total number of data points for some thermochemical properties is listed in Table 1.

Table 1. The Number of Different Entries in Chemistry Web Book (release of July 2001). The Major Data Collections have Teams of Compiler(s) who Support and Update Information

Data Type	Number of Species
Gas phase thermochemical data	6,500
Condensed phase thermodynamic data	5,600
Phase change thermodynamic data	13,800
Reaction thermodynamic data	9,800
Gas phase ion energetics data	15,600
Gas phase ion cluster data	1,100
IR spectrum	8,700
Mass spectrum	12,600
UV/Vis spectrum	400
Vibrational & electronic energy levels	4,100
Constants of diatomic molecules	600
Fluid property data	33

All these data can be searched on-line. The data in the Web Book mainly come from the peer-reviewed literature and include citations to the original literature. If errors are detected in a database, the user can submit error reports that are logged and reported to compiler(s), who curate and update this particular database. In addition, a larger database with physical and thermodynamic properties of 7,468 chemical compounds (497,000 property data points) is available for a license fee as WinTable database. This database was compiled by Thermodynamics Research Center for more than 50 years. A more detailed overview of NIST Web Book can be found in [12].

2.2 Syracuse Research Corporation

It has several on-line databases with various information about physicochemical and environmental data. The Environmental Fate Data Base (EFDB, http://esc.syrres.com/efdb.htm) was developed in order to allow rapid access to all the available fate data on a given chemical, to identify critical gaps in the available information, to facilitate planning of research, and to provide a data source for constructing structure-activity correlations for degradability and transport of chemicals in the environment. It includes four parts with bibliographic information (DATALOG), environmental fate and physical/chemical property information on commercially important chemical compounds (CHEMFATE and BIODEG), and microbial toxicity and biodegradation data (BIOLOG). The physical properties database (PHYSPROP, http://esc.syrres.com/interkow/PhysProp.htm) contains chemical structures, names, and physical properties of more than 25,000 compounds. The physical properties are collected from a wide variety of sources and include experimental, extrapolated, and estimated values for melting point, boiling point, water solubility, octanol-water partition coefficient, vapor pressure, pK_a, Henry's law constant, and OH rate constant in the atmosphere. The estimated values can be clearly distinguished from the experimental ones by their data type (EXP or EST), and corresponding calculation algorithms are published in peer-reviewed literature [13]. The on-line version of this database (http://esc.syrres.com/interkow/physdemo.htm) allows retrieval of these values using as input CAS RN of molecules. This database has been used by many researchers to develop different models for the prediction of physicochemical properties of compounds, see, e.g., [14-18]. The PHYSPROP database is also available as part of EPI suite reviewed in the next paragraph.

2.3 The United States Environmental Protection Agency (EPA)

It provides access to a large number of experimental databases that can be useful for the researches working on parameter modeling. This site contains several experimental

toxicology databases available as ECOTOX (http://www.epa.gov/ecotox). This resource includes more than 320,000 individual effect records abstracted from 17,195 peer-reviewed publications representing over 7,800 chemicals and 5,300 aquatic and terrestrial species [19]. The database can also be downloaded as an ASCII delimited file format [19]. Another interesting resource available at this site is Estimation Program Interface (EPI) Suite (http://www.epa.gov/oppt/exposure/docs/episuitedl.htm). The EPI Suite contains more than 150 quantitative structure-activity relationships (QSARs) developed by the EPA's Office of Pollution Prevention Toxics and SRC. This software can be downloaded and installed for free on Windows computers.

2.4 The AQUASOL

This database developed by Prof. S.H. Yalkowsky [20] is available for a license fee on-line at http://www.pharmacy.arizona.edu/outreach/aquasol. The database contains detailed information about temperature and solubility of more than 6,000 chemical compounds with corresponding literature references, and represents one of the largest collections of compounds with the aqueous solubility data. All the experimental evidences have references and, in addition, a five point score is provided to evaluate the quality of the reported data for each data point. The quality estimation is one of the most important features of this database, and can be useful for researches to assign confidence to each data point in model development. The on-line database is searchable using several criteria, such as formula name, CAS RN, and compound characteristic.

2.5 The National Cancer Institute (NCI)

It represents the largest public collection (more than 250,000 molecules) of chemical structures. It is developed using chemistry information toolkit CACTVS (http://cactus.nci.nih.gov). These structures include anticancer and anti-HIV data provided by NCI's Developmental Therapeutic Program (45,228 compounds), and experimental log P values for 3,576 compounds [21]. The 3D coordinates are available for each structure and the user can perform 3D pharmacophore searching using up to 25 conformations per molecule (average number is 10.5). By offering an access to the largest public database of compounds, the NCI site provides a very useful resource that, in addition to development of cancer prediction programs, can be used to conduct large-scale analysis, such as development and testing approaches for screening of virtual libraries. This site also contains a link to several other public databases available at U.S. Government web sites (http://cactus.nci.nih.gov/ncidb2/govt_dbs.html).

2.6 The Chemexper Chemical Directory (CCD, http://www.chemexper.com)

It is a database, which currently lists more than 100,000 chemicals from an international range of suppliers. In addition to chemical structures this database provides an access to various physicochemical parameters of molecules, such as boiling point, melting point, density, flash point and others. There are several possibilities to perform searches in this database. Simple searches can be made using molecular formula, or CAS RN. More complex searches can be performed by specifying the substructure of molecules, ranges of their physicochemical parameters, and combining such characteristics [5].

2.7 The ChemFinder (http://www.chemfinder.com) [8]

It is another popular database with 2D and 3D structures of molecules, melting and boiling points, density, evaporation rate, flash points, etc., and links to other sites with physicochemical characteristics of each compound. The product information of 450,000 chemicals from over 300 chemical suppliers is also provided.

2.8 The ChemWeb site (http://www.chemweb.com)

It provides access to a collection of databases with physicochemical properties of compounds. The ACD/labs physicochemical property database contains data on pK_a - over 8,900 structures with over 23,000 experimental values under different temperatures and ionic strengths in purely aqueous solutions; *log P* - 11,886 compounds with experimental log P values collected from different sources; *HNMR* - ca. 100,000 structures; *CNMR* - ca. 100,000 compounds; *FNMR* ca. 11,500 compounds; *PNMR* - ca. 18,500 compounds, (http://www.acdlabs.co.uk/clients/pr_chemweb.html). Properties of organic compounds are also available from Chapman & Hall/CRC Press chemical dictionary and the NIST Chemistry Web Book. This web server also supports the chemistry pre-print server, a freely available and permanent web archive and distribution medium for preprints [22].

2.9 Model Data Set Collections

The previous databases that mainly contained a large number of "raw" resources that can be used to develop new methods for the calculation of one or more molecular properties. However, when developing different software tools, a researcher could benefit from an access to datasets that were already used in some previous models. Such datasets can be very important to validate new ideas and approaches. The computer science and machine-learning fields contain one of the most widely developed free repository systems with such resources. The database at the UCI (http://www.ics.uci.edu/ ~mlearn/MLRepository.html) [23], DELVE (http://www.cs. toronto.edu/~delve/data/datasets.html) contains examples with various levels of complexity, number of cases (from tens to thousands of samples), dimensionality (from few to thousands of dimensions) and data structure (logical and/or missed variables), etc. These examples are annotated, and the people maintaining such databases from time-to-time update references with the analysis of such data sets. These repositories also contain a number of QSAR datasets.

2.10 QSAR

This society has a similar but, so far, smaller repository system with QSAR examples (http://www.qsar.org/resource/

Table 2. Descriptors Calculated by Dragon 3.1 Program

Constitutional descriptors	47	Geometrical descriptors	70
Topological descriptors	266	RDF descriptors	150
Molecular walk counts	21	3D-MoRSE descriptors	160
BCUT descriptors	64	WHIM descriptors	99
Galvez topological charge indices	21	GETAWAY descriptors	197
2D autocorrelations	96	Functional groups	121
charge descriptors	14	Atom-centred fragments	120
Aromaticity indices	4	Empirical descriptors	3
Randic molecular profiles	41	Properties	3

datasets.htm). At this site there are about 20 datasets representing some classical examples that were extensively used in the QSAR field. A number of references to articles that used each example is also provided. Thus this site provides very useful information for researches working in the QSAR and drug design field. Unfortunately, the number of such datasets is limited and there are no suggestions how one should perform analysis of each dataset and which training/validation approaches were used in the previous publications. Such information can be very useful to quickly select one or another dataset for validation of a new approach. It would also be useful to include molecules as SMILES or "sdf" files thus allowing different users to develop and to try different sets of molecular descriptors.

3. STANDALONE RESOURCES

A great number of software packages developed to calculate molecular descriptors are available as standalone applications. I will briefly overview several popular software packages developed by Academia and Industry.

3.1 Academic Resources: Non-Quantum Chemistry Descriptors

3.1.1 DRAGON ***(http://www.disat.unimib.it/chm/Dragon.htm)***

It was developed by Prof. Todeschini and his colleagues at the University of Milano-Bicocca. This software (Dragon v 3.1) calculates 1497 descriptors that are divided into 18 main categories listed in Table 2.

The Dragon package became well known for a practical implementation of some algorithms after being described in a very successful book [24]. Ten out of 18 groups of descriptors do not require 3D structure of molecules, and can be calculated using molecules code, e.g., SMILES. Several sets of these descriptors, e.g., WHIM (weighted holistic invariant molecular) [25] and GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) [26], were proposed by Prof. Todeschini's group. The WHIM descriptors are based on principal component analysis of atomic coordinates and thus are invariant to rotation and translation (see article by Maiocchi in this issue). The GETAWAY descriptors are based on a leverage matrix called Molecular Influence Matrix (MIM). These descriptors are matching 3D-molecular geometry provided by the MIM and atom relatedness by molecular topology using different atomic weightings (e.g., atomic mass, polarizability, van der Waals volume, electronegativity). There are several versions of DRAGON software, including free Dragon Web version. The professional versions require a license fee.

3.1.2 Molconn-Z ***(http://www.eslc.vabiotech.com/molconn/molconnz.html)***

This software represents another popular academic package for calculation of molecular indices. This software developed by Profs. Kier and Hall includes several groups of topological indices (Table 3), such as E-state indices [27], which is one of the most popular sets of descriptors developed by these authors.

A detailed description of indices used in this software package was published in peer-reviewed literature, including several monographs. The total number of publications using

Table 3. Molecular Descriptors Calculated by Molconn-Z Program

Molecular Connectivity Chi Indices	Shannon Index
Kappa Shape Indices	Information Indices
Electrotopological State (E-State) Indices	Wiener Number
Molecular Connectivity Difference Chi Indices	Platt Number
Atom-type E-State Indices	Bonchev-Trinajstiç
Group-type E-State Indices	Total Topological Index
Topological Equivalence Classification of Atoms	Counts of Subgraphs: paths, rings, clusters, etc.
Other Topological Indices	Vertex Eccentricities

E-state descriptors was above 100 in 2001 (as summarized in http://www.eslc.vabiotech.com/molconn/mconpubs.html), and it continues to grow.

3.1.3 VolSurf ***(http://chemiome.chm.unipg.it)***

It was developed by Prof. Cruciani [28]. This method explores the physicochemical space of a molecule using 3D maps of interaction energies between the molecule and chemical probes. The VolSurf compresses the information present in 3D grid maps into 2D numerical descriptors representing size and shape of molecules, hydrophobic and hydrophilic regions, interaction energy moments, and some additional parameters as well. These descriptors are simple to interpret and have found many applications in the QSAR field, particularly in modeling the ADME properties of drugs (see article by Lombardo et al. in this issue).

3.2 Quantum Mechanics Descriptors

The quantum-chemical programs can be used both to provide optimization of molecular structure and also to calculate molecular descriptors to be used in QSAR studies. Traditionally such calculations were always time consuming, especially using *ab initio* methods. The Quantum Chemistry Program Exchange (QCPE, http://qcpe.chem.indiana.edu) holds about 770 computational chemistry systems, programs, and routines. Most programs are available in source code, and for some popular programs QCPE holds several versions. These programs can be ordered from QCPE for a nominal distribution fee. Many quantum mechanics programs are also freely distributed, e.g., GAMESS (http://www.msg.ameslab.gov), which uses different SCF wave functions, and can compute a variety of molecular properties ranging from simple dipole moments to frequency dependent hyperpolarizabilities. DALTON (http://www.kjemi.uio.no/software/dalton) and COLUMBUS (http://www.itc.univie.ac.at/~hans/Columbus/columbus.html) represent other freely available popular *ab initio* programs. Semi-empirical molecular orbital programs, such as Mopac are available as free resource (ftp://esca.atomki.hu/mopac7/LINUX) or can be acquired commercially (http://www.schrodinger.com). The semi-empirical Hamiltonians MNDO, MINDO/3, AM1, and PM3 are used in MOPAC to obtain molecular orbitals, the heat of formation and its derivative with respect to molecular geometry. Using these results, MOPAC calculates the vibrational spectra, thermodynamic quantities, isotopic substitution effects and force constants for molecules, radicals, ions, and polymers. A review of free software resources for analysis of small molecules that also includes a section about quantum-chemical programs has recently been published [29].

3.3 Commercial Software

A number of chemoinformatic companies provide software for calculation of molecular descriptors. This includes QuaSAR-Descriptor software (http://www.chemcomp.com/feature/descr.htm) from Chemical Computing group Inc. (300 descriptors) that proposes a number of physicochemical properties of molecules and 3D descriptors. C^2 descriptor+ (http://www.accelrys.com/cerius2/descriptor.html) calculates topological, information-content, fingerprint, charged partial surface areas, shadow and other indices. CODESSA (http://www.semichem.com/codessa/index.shtml) calculates over 600 descriptors (topological, geometric, constitutional, thermodynamic, electrostatic, quantum mechanical). SciQSAR (http://www.scivision.com/sciQSAR.html) calculates a number of 2D and 3D descriptors, including connectivity and shape descriptors, charge related descriptors, polarizability, log P, etc. The MDL QSAR (http://www.mdl.com/products/qsar.html) provides calculation of over 400 2D and 3D molecular descriptors. A number of companies license software packages, including "in-house" packages, and integrate all of them into a common easy-to use interface. For example, Tripos (http://www.tripos.com) provides different software packages, such as HQSAR, Distill (both fragment based methods), Moconn-Z, VolSurf, CLOGP that are all fully integrated with other proprietary Tripos software and can be used within SYBYL Molecular Spreadsheet for data analysis and visualization. This list of companies and software resources, of course, is not complete.

3.4 Open Source Software

The open source software projects represent a very attractive way of sharing and advancing ideas in computer programming. By making the source code of a useful and timing project available to other people, one can get a lot of recognition and avoid re-design of code by many users. The users' feedback can be very useful to improve such software and to speed up its development. There are about 100 chemistry-related projects (found using search on keyword chemistry/chemical, molecule/molecular) at the SourceForge (http://sourceforge.net). About half of these projects are releasing codes and thus can be considered as already available projects. Among these projects, there are a few that could be interesting for molecular descriptor software development. The cross-platform visualization software includes JChemPaint (http://sourceforge.net/projects/jchempaint) [30] and Jmol (http://sourceforge.net/projects/jmol), representing open source Java programs to draw 2D and 3D chemical structures, respectively. Both these projects are based on the Chemistry Development Kit (CDK) [31] that also provides a basis for several other projects working on NMR database of chemical shifts (http://nmrshiftdb.org) and computer-assisted structure elucidation. The OpenBabel (http://openbabel.sourceforge.net) project provides a further development of the Babel (http://www.eyesopen.com/babel.html) cross-platform program and library designed to interconvert between many file formats used in molecular modeling and computational chemistry. The interconversion of chemical structure can also be performed using JOELib, which is an open platform-independent computational chemistry package written in Java (http://www-ra.informatik.uni-tuebingen.de/software/joelib). This software package also includes calculation of various molecular indices, such as Kier shape, Zagreb, log P, molecular refractivity according to Wildman and Crippen, Gasteiger-Marsili atom partial charges, etc. A nice feature of this library is a possibility to perform searches of molecular substructures using SMART (http://www.daylight.com/dayhtml/doc/theory/theory.toc.html).

4. ON-LINE RESOURCES

In addition to be the Largest Informational Database that already had more than 2 billion pages in 2000 [32], the Internet also provides on-line services and thus can actively interact with the user. Instead of downloading and installing a program at his local computer, the user can perform data analysis on-line. Recently, several free and for fee services appeared on the Internet and their numbers are continuously increasing.

4.1 Actelion

It has developed a nice property explore applet (http://www.actelion.com/page/property_explorer) as a part of its in-house substance registration system. This applet draws chemical structures and simultaneously calculates on-the-fly various drug-relevant properties, using a fragment-based approach, whenever a structure is valid. The predicted values are shown both as numbers and are coded in colors. The applet predicts log P, solubility and drug-likeness, estimated as proportion of fragments in drugs and various commercially available chemicals. The program also predicts toxicity risks, such as mutagenic, irritant, tumorigenic and reproductive effects. The toxicity alerts are an indication that the drawn structure may be harmful concerning the risk category specified (Fig. (**1**)). An overall drug score that takes into account drug-likeness, cLogP, log S, molecular weight and toxicity risks is also provided. This applet is easy and convenient in use.

4.2 Advanced Chemistry Development (http://www2.acdlabs.com/ilab)

It provides free calculation of molar refractivity, molar volume, parachor, polarizability and a number of other molecular properties. The same company provides commercial services, including calculation of different physicochemical properties, such as log P, log D, aqueous solubility, pK_a, etc. The analysis is limited to non-charged, non-organometallic structures with up to 255 heavy atoms. The Web interface is provided as an ACD Structure Drawing Applet (ACD/SDA http://www.acdlabs.com/products/java/sda). The first-time users can request two-weeks evaluation access to all ACDlab services free of charge. A detailed overview of on-line resources of ACDlabs was published recently [33].

4.3 Computer Chemie Centrum, the University of Erlangen-Nuremberg (http://www2.ccc.uni-erlangen.de/services/petra)

It provides on-line access to a large number of molecular, atomic and bond properties calculated by the Parameter Estimation for the Treatment of Reactivity Applications (PETRA) program (see article by Gasteiger in this issue). This program can be used to quantify heats of formation, bond dissociation energies, sigma/pi charge distribution, inductive, polarizability and resonance effects and delocalization energies. The input data can be provided as SMILES, created using JME editor of Dr. Peter Ertl

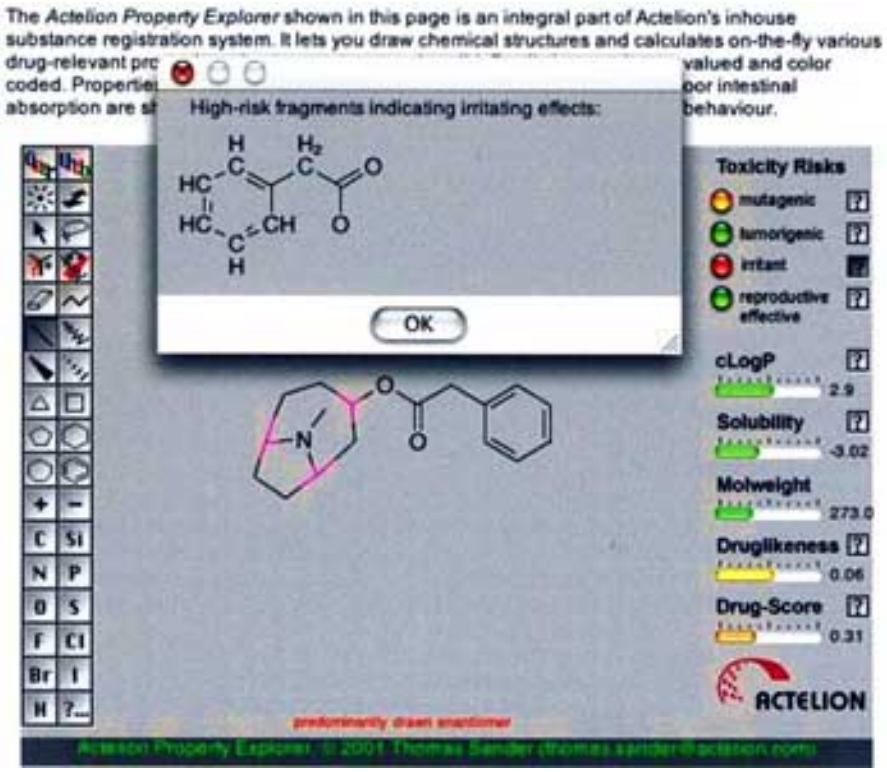

Fig. (1). Actelion Property Explorer applet developed by Dr. Thomas Sander (http://www.actelion.com/page/property_explorer). A medium level mutagenic toxicity risk (yellow) and high level irritant risk (red) is predicted for the analyzed compound. A click on question mark displays fragments that could be relevant for this toxicity risk.

(Novartis) [34] or uploaded in one of 13 available data formats. The batch analysis is limited to 100 molecules. The PETRA manual includes detailed literature references and a description of methods implemented in this software. Another popular software developed by this group, CORINA (http://www2.chemie.uni-erlangen.de/software/corina), provides 2D => 3D conversion of molecules. This program is available on-line at (http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html). The user can convert 1,000 molecules for free.

4.4 Daylight Chemical Information Systems (http://www.daylight.com/daycgi/clogp)

It provides on-line calculation of log P values using CLOGP. A molecule is depicted on the WWW and fragment contributions are provided. If experimental values for the analyzed compound are available, they are also displayed. The experimental values are selected from the BIOBYTE database that was curated by Dr. Leo. At the present moment this database contains about 11,000 compounds with experimental log P values (http://www.biobyte.com/bb/prod/cqsar.html). The CLOGP software and the data used to develop it are published in numerous references [35-38]. The same site also has on-line calculations (http://www.daylight.com/daycgi/cmr) of molecular refractivity, which is an important descriptor of steric properties of molecules.

4.5 Interactive Analysis (www.logp.com)

It provides on-line access to prediction of lipophilicity (IA_LogP) and aqueous solubility (IA_LogS) of chemicals. This software was developed using electron-state indices [39] and neural networks. The site contains a number of slides with results of software validation using cross-validation and test sets. A new version of software developed with methodological approaches by this company is available from ChemSilico.

4.6 ChemSilico (http://www.chemsilico.com)

It provides on-line calculation of different physicochemical (aqueous solubility, log D, pK_a) and biological parameters (blood-brain barrier partition, plasma protein binding, mutagenicity prediction). The site contains information with detailed statistics about performance of different prediction modules. All these parameters were developed using E-state indices and neural network. A registered user (registration is sent to his e-mail address) can analyze 50 molecules for free. The compounds can be downloaded as "sdf" or "mol" file or can be created using JME editor. The calculated results are sent to the user's e-mail address.

4.7 Molinspiration Cheminformatics (www.molinspiration.com/services)

It provides calculation of molecular physicochemical properties relevant to drug design, including log P, molecular polar surface area (PSA) [40], drug likeness and others. The molecules can be introduced as SMILES code or prepared using JME.

4.8 Pre-ADME (http://camd.ssu.ac.kr/adme)

It was developed in Computer Aided Molecular Design Research Center, Soong Sil University, Seoul. This software calculates different molecular descriptors (15 structural, 350 topological), lipophilicity, aqueous solubility, molecular refractivity, polarizability, water solvation free energy, drug-likeness and drug absorption data, such as Caco-2, MDCK, BBB, HIA. The user can draw input structures using ACD Structure Drawing Applet (ACD/SDA, http://www.acdlabs.com/products/java/sda) or upload input file in "sdf" or "mol" file format. The use of this software requires membership.

4.9 SPARC server (http://ibmlc2.chem.uga.edu/sparc)

It was developed at the Department of Chemistry, University of Georgia. The main purpose of this server was to facilitate chemical fate modeling. The input data are introduced as SMILES or created using ACD/SDA applet. The server calculates pK_a, boiling point, solubility, refractive index, polarizability, vapor pressure and aqueous solubility (two last parameters require melting point) and a number of other parameters. The user can also perform kinetics and hydrolysis analysis. The server provides database search of pK_a and physical property values.

4.10 Syracuse Research Corporation (http://esc.syrres.com/interkow/kowdemo.htm)

It provides on-line interface to the KOWWIN program [13] to estimate lipophilicity of chemical compounds. The compounds can be submitted as SMILES or CAS RN. In addition to calculated values, this site also provides experimental log P values of compounds from the PHYSPROP database.

4.11 VCCLAB (http://www.vcclab.org)

It provides calculation of lipophilicity and aqueous solubility of chemicals using ALOGPS program. This service is provided on a permanent basis since 2000, and it was initially available from the University of Lausanne (http://www.lnh.unil.ch/~itetko/software.html), but this site is no longer supported. In addition to values calculated with ALOGPS program [16,41], the applet displays CLOGP, KOWWIN, IA_LogP, IA_LogS and XLOGP [42] values for single molecule analysis. All these values are retrieved from the corresponding WWW sites of the programs available on-line (except XLOGP that is executed locally) and are integrated on the VCCLAB site. The ALOGPS program analyzes molecules in SMILES format. However, if the user has his/her molecules in different formats, such data are first converted to SMILES using the Open Babel program (http://openbabel. sourceforge.net), which is installed locally. In addition, starting March 2003 this site also performs on-line calculation of molecular descriptors using e-DRAGON (http://www.vcclab.org/lab/edragon) using client-server interface to the DRAGON software. Another program, PCLIENT (http://www.vcclab.org/lab/pclient), that

will integrate DRAGON and E-state indices under the same interface is being developed now.

4.12 Prediction of Activity Spectra for Substances PASS (http://www.ibmh.msk.su/PASS)

It is able to predict more than 700 pharmacological effects, mechanisms of action, mutagenicity, carcinogenicity, teratogenicity and embryotoxicity. The PASS system was developed using 45,466 biologically active compounds that were collected from articles and electronic databases since 1972 [43]. The user can upload "mol" file or create a molecule using MarvinSketch (http://www.chemaxon.com/marvin). The PASS predictions of biological activities of molecules were recently done for molecules from the NCI database, and the users can easily access them at the NCI site. While the PASS system cannot be considered as program to generate molecular indices, this is a very interesting and nice example of an on-line system for molecule analysis.

5. ON-LINE ARCHIVES

An increasing amount of virtual library sites contain comprehensive information about chemistry resources on WWW. The QSAR site (http://www.qsar.org) contains a list with a number of different resources grouped in databases, software, web resources, companies, etc. The Computational Chemistry (http://www.ccl.net) list in addition to providing a collection of WWW links also contains archives with software packages for various computer platforms and languages. The Australian Computational Chemistry via the Internet Project (ACCVIP, http://www.chem.swin.edu.au) also contains a large collection of WWW teaching modules, including Introduction to the use of the Internet for chemists. LINUX users will be interested in Linux4Chemistry (http://zeus.polsl.gliwice.pl/~nikodem/linux4chemistry.html) which contains an extended list of free and commercial software tools available for the LINUX operating system. The Network Science (http://www.netsci.org/Resources) provides a collection of information on different topics in Science, including standalone programs and web resources as well as featured articles. Unfortunately, it looks that this site is not updated for several years. The WWW Virtual Library (http://www.vlib.org) contains a rich collection of chemistry links (http://www.liv.ac.uk/Chemistry/Links) joined according to different topics.

6. JAVA, XML AS EMERGING TOOLS FOR THE DEVELOPMENT OF ON-LINE SOFTWARE

This section will overview some new trends in the development of Internet resources, and will provide analysis of some technological resources that can be useful to develop new sites for the calculation of molecular parameters.

6.1 Java and the VCCLAB Site Development Experience

The VCCLAB site is based on its prototype, Data Analysis for Neuroscience (DAN), that was initially developed in the Laboratory of Neuro-heuristique, University of Lausanne (http://www.lnh.unil.ch) using Java 1.1 language. The current version includes more than 100 Java classes and contains about 500 Kb of source code. Actually, this is the third version of the code that was completely rewritten to correspond to the increasing requirements of the virtual laboratory. In this sub-section I will overview some major changes in our site from the time of our previous publication and share some experiences about the use of Java for integration of native code for scientific calculation. A more detailed description of the VCCLAB site will soon be published elsewhere.

6.1.1 Why Java?

The initial choice of Java platform for the development of the virtual laboratory was inspired by several nice features of Java, such as its portability, security, and convenience in development [44]. Indeed, growing number of Java-based applications developed for web-based calculation indicates an increasing popularity of this language. For example, all on-line software tools described in this article, except Syracuse Research Corporation (but in fact, it also uses Java applets in their chemistry substructure search system, http://esc.syrres.com/Chems3), use one or another kind of Java applets to facilitate input of chemical structures or/and display calculated results. Let me remind that the possibility of applets to be executed on different platforms is provided by the specific features of this language. The Java compilers convert a program to a machine independent code, so–called byte code, that can be executed by Java Virtual Machine (JVM), i.e., the Java interpreter. JVMs are developed for major computer operating systems, such as OC-Linux, Windows, Macintosh, etc., thus allowing the same Java program to run on different machines. This makes the Java software easily portable for different platforms. The Internet packages of Java make it possible to run the programs across the Internet. The main WWW browsers, such as Netscape or Internet Explorer contain plugins or built-in support of Java language and can display Java applets.

6.1.2 Distributed Computing

The calculation at VCCLAB site is organized using so-called three-tier architecture (Fig. (**2**)).

The main server (SuperServer) is running on a LINUX computer and provides Web interface support and redistribution of tasks submitted by the users. The calculation servers that actually perform data analysis are located on 11 computers, including a server (Corina) at the University of Erlangen and a server (DRAGON) at the University of Milano-Bicocca. The Super Server provides a wide flexibility of the software and easily permits extending the number of calculation servers and services without a need to restart. The Super Server recognizes the applications according to an identification TASK keyword, i.e., "logP" in case of log P/log S program or "asnn" in case of neural networks. The tasks submitted by a user and/or subtasks provided by the calculation servers are stored on the SuperServer. The CalculationServer sends a request to the SuperServer to verify if there are any tasks available for it and if so (this is verified by matching the TASK keyword of server and available task), the server uploads the corresponding task and calculates it. The SuperServer is also used to upload data files using Java servlets.

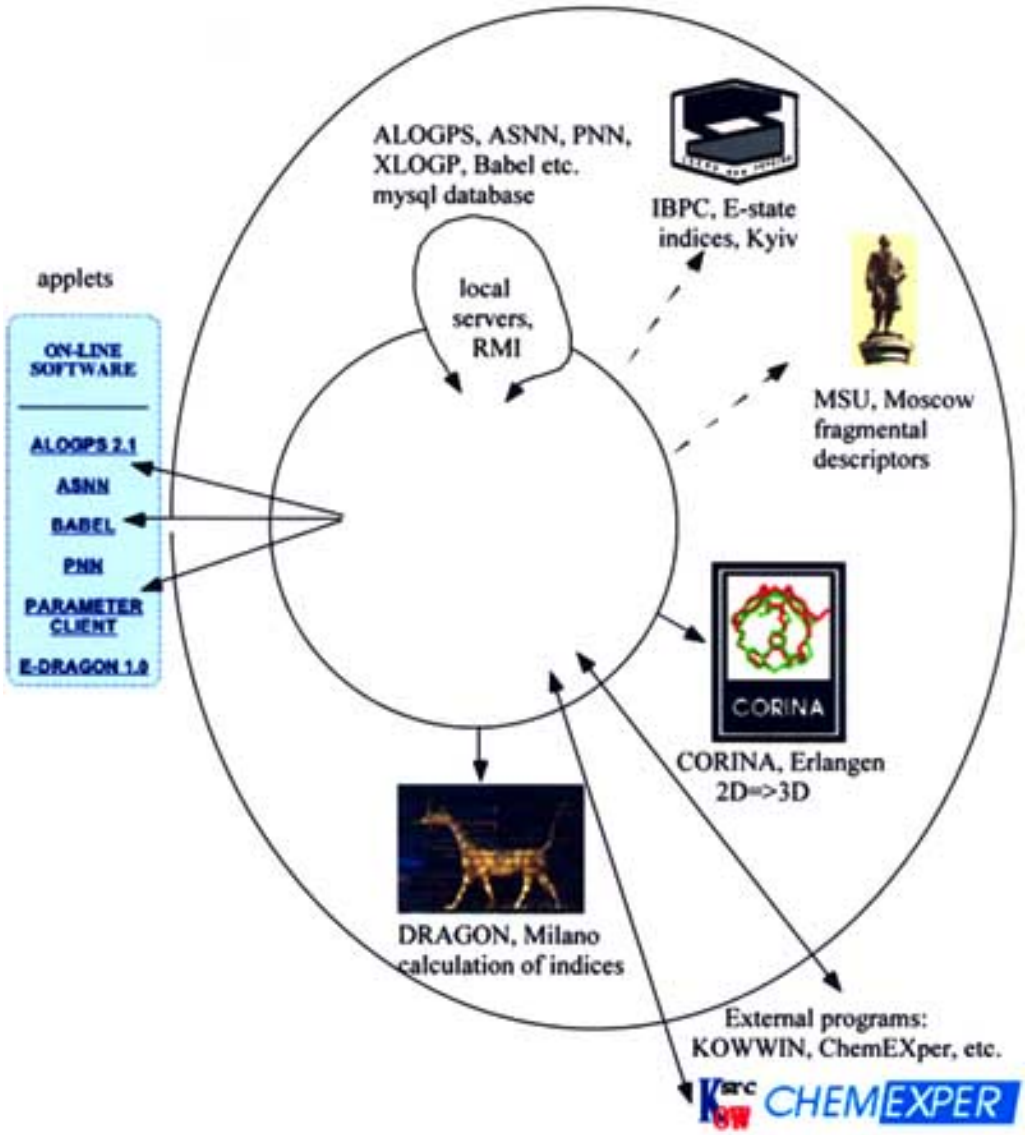

Fig. (2). Flow-chart of data analysis at the VCCLAB (http://www.vcclab.org). The heart of laboratory, the SuperServer, provides integration and distribution of tasks. The tasks are calculated by CalculationServers installed locally in Munich and two servers, DRAGON and CORINA, located in the University of Milano-Bicocca and in the University of Erlangen, respectively. The Remote Method Invocation, RMI, is used between the SuperServer and local servers. Two servers in Kyiv, Institute of Bioorganic & Petroleum Chemistry, IBPC, and Moscow State University are under development now (dashed lines). Some external programs, such as KOWWIN or ChemExper are accessed using HTPP protocol. The results calculated by all servers are displayed in client applets using Java servlets.

6.1.3 Java Native Interface

A nice feature of Java programs is a possibility of their integration with C/C++ programs using Java Native Interface (JNI) [45,46]. This makes possible an easy integration of different computer programs running on various computer systems. The JNI is used to run ALOGPS program, programs to calculate E-state indices, and calculation methods, such as neural networks and polynomial neural networks. We have found that use of JNI makes it possible to easily and smoothly re-compile the code under different platforms. The ALOGPS program was running on several different platforms, such as Sun Solaris, Windows, Mac Os 8-9, Mac OS X and Linux platforms. The native code is compiled and stored as dynamic libraries that are loaded in memory during execution of the Java programs (Java calculation servers). However, such use of Java and native source requires some (considerable) work to integrate Java and C/C++ code smoothly. One of the most serious and unpleasant discoveries was the inconsistent behavior of dynamic libraries under different platforms. We found that because different mechanisms were used to load such libraries on different platforms, some global variables could not be re-initialized if Java applications did not quit. This problem was solved by redesign of some native code applications and/or developing of a recycling mechanism that quits the Java calculation servers after the calculation of each task.

In addition to JNI, we also developed an ExecServer that made possible the execution of standalone applications without the need to recompile them. The ExecServer was found to be a flexible and convenient mechanism to integrate standalone programs that would require considerable effort to be implemented using JNI. At the present moment, the ExecServer is being used to execute CORINA and DRAGON software.

6.1.4 Internet protocols

The integration of SuperServer and Calculation Servers was done using Remote Method Invocation (RMI). Unfortunately, the increased number of attacks on private and public servers forced the network administrators to implement stronger computer security and to close all ports except a few standard ones used for Web navigation. This left the HTPP protocol as the only one available for communication between remote computers located on the Internet. Thus, we have rewritten the communication software to use HTPP protocols both for client (applets) and calculation servers.

6.2 Chemical Markup Language (CML http://www.xml-cml.org)

It represents an implementation of Extensible Markup Language (XML, http://www.xml.org) technology for representing chemical structure developed by Profs. Murray-Rust and Rzepa in the UK [47-49]. The appearance of XML was started thanks to the development of Internet technologies that require some standard interfaces to provide an easy data transfer over the Internet. XML is a successor of HTML language that was used to describe how the information (text, images, applets, etc.) should be displayed within a browser. XML is designed allowing the description of any structured data by the use of a user defined set of tags. It also meets a number of needs, such as flexibility, extensibility, and simplicity of structure. For example, one can simply describe molecule as

```
<molecule name = "Ethane" >
    <formula>C2H6</formula>
    <weight>30</weight>
</molecule>
```

The simplicity is the result of the syntax and descriptive tags. The tags are usually selected in such a way that another user can easily understand meanings of the values from their names. The flexibility and extensibility result from the fact that one can easily add a new field. For example, one can add SMILES to the molecule:

```
<molecule name = "Ethane" >
    <formula>C2H6</formula>
    <SMILES>CC</SMILES>
    <weight>30</weight>
</molecule>
```

Notice, that the above presentations are not comprehensive or compatible with the CML and are used only for demonstration purposes. XML also provides mechanisms to verify that XML files comply ("are valid") with a general set of rules using Document Type Definition (DTD) or XML Schemas associated with each particular document. One can specify in the XML Schema that molecular weight should be positive float and SMILES should be string containing some specific characters. Such checks can be automatically performed by the XML software and identify documents that do not contain the proper syntax. Since XML provides only rules for "how elements should be represented", one can develop different alternative languages that can be used to describe the same entity. For example, it is also possible to describe the same molecule using abbreviations as

```
<mol name = "Ethane" >
    <formula>C2H6</formula>
    <SMILES>CC</SMILES>
    <mw>30</mw>
</mol>
```

A nice feature of XML allows its representations to be handled in many different ways. XML files are displayed using specified stylesheets defined by extensive stylesheet language transformation (XSLT) [50]. The conversion of an XML file can be performed using the server (the browser will receive the HTML page) or using an XML-compatible browser, such as Internet Explorer 5+ or Netscape 6+ (the browser will receive both XML and XSLT files and will perform the conversion itself). Different XSLT schemes can be integrated with Java applets or plugins to display 2D or 3D structures of molecules.

The CML language was extensively elaborated to comprehensively describe molecules. It has support for atom- and bond-based stereochemistry; can hold both 2D & 3D coordinates simultaneously; and manages hierarchy of molecules. This makes it possible to use the same file for different representations (e.g., as 2D or 3D structures) as demonstrated at the ChiMeral site (http://www.ch.ic.ac.uk/rzepa/chimeral). Unfortunately, due to incompatibility of XML browsers, the results can be observed only for IE 5 for PC and could not be seen from Mac. Nevertheless, in general this feature of XML, i.e., separation of data and their representation, is very useful and will definitely find a wide application in chemistry.

It is also quite possible that in addition to the CML, several XML specifications for molecules, similar to a number of different file formats used in chemistry, will appear. Such languages would be simpler than the CML and will be developed to meet some specific requirements (i.e., support of only SMILES, etc.). An appearance of such languages would speed up development of the CML as well, since it is easier and more straightforward to write a conversion program between two different XML files than to export ASCII files. The particular challenge of CML is that it was the first comprehensively developed XML language for chemistry. Thus one can consider CML as a language for external publication of data on the Internet or in journals, provided that sufficient resources to do this will be developed by the CML community. The former will also depend on the acceptance of CML as standard language for

the description of chemical data by major chemical publishers, such as ACS, etc. [51].

The extensibility of XML makes it possible to easily incorporate physicochemical parameters of data into molecules. The user has several possibilities to choose from. First, new element, like <SMILES> in the example above can be defined and described in a corresponding XML Schema file. Another possibility can be to use some standard elements, such as <float>. For example, CML language supports several attributes, such as @title, @convention, @units and @builtin. Using these attributes one can describe the water solubility of a molecule as

<float title="water solubility" units="mg/100 ml at 25 degC">30</float>

The advantage of the second way is higher flexibility. New parameters can be easily added without a need to change the XML description. However, when using the second way there is a danger that different users will give different names to describe the same property. Indeed, different titles, such as title="Water Solubility" or "aqueous solubility" will require some human intervention to recognize them as the same property. Thus, some convention for naming of the same parameters should be developed.

Prof. Kehiaian in Paris was sponsored by IUPAC, ICSTI, and CODATA to develop a Standard Electronic Data File (SELF) format for physicochemical properties. This language is used in ELDATA, The International Electronic Journal of Physico-Chemical Data. In collaboration with CML-developers, SELF was extended to the XML and is known as SELFML. Some examples of on-line systems to search physicochemical parameters are available at http://www.xml-cml.org. The SELFML is used in a preliminary version of DataExplorer (http://www.fiz-karlsruhe.de/dataexplorer, id: everyone; password: sesame [52]), that provides access to numerical physicochemical property data sources from 16 databases from 8 contributing organizations. The search can be performed using 4155 chemical components, 998 original data sources, 41 property types, and 3805 SELF files [52]. The SELFML can be a possible solution for incorporation and transferring of physicochemical parameters over the Internet.

7. CONCLUSIONS

The development of Internet technologies provides new challenges in the field of medicinal chemistry. The data resources available on the WWW allow the scientists to avoid the time consuming steps in data collection and focus their work on the development of models. At the same time the WWW repository also provides significant resources to test new approaches and to validate them by comparison with previous models.

The Internet contains a growing number of free and commercial standalone programs that can be downloaded or purchased on-line. It also provides the possibility to calculate various descriptors and physicochemical parameters of molecules on-line. The use of Internet technologies to boost ADME studies was recently discussed [53]. The leading pharmaceutical firms, such as Novartis [34,54] or Actelion (see above) have developed integrated systems for data analysis on the WWW that also include calculation of various physicochemical parameters. Apparently, similar WWW based systems are also used in other companies, but these results are not published in peer-reviewed literature. Commercial systems, such as JChem (http://www.chemaxon.com) are also available [55]. Such systems can facilitate creation of web applications that access chemical structures and corporate data in a database over the Internet or in Intranets. The Academic users can also gain profit from open source developments, such as projects available at the Source Forge site (http://sourceforge.net). The increasing speed of Internet also makes it possible to calculate and transfer bulky molecular parameters, such as 3D interaction maps used in CoMFA or GRID approaches. LINK3D (http://www.tecn.upf.es/prj/link3d/) was recently started by a consortium of Academia and Industry to bring new technologies into this field.

The use of XML is providing new ways of data representation and transfer over Internet. An adoption of CML as a standard language for molecule description can boost up chemical communication. A use of this language for the calculation of chemical descriptors can also facilitate integration of data programs on the Internet. A possibility to use CML/SELFML for data analysis at the VCCLAB site is also considered by us now. The CML is easily integrated with Java language and the combination of XML and Java can be a convenient tool to develop a client-server for the calculation of new systems for molecular properties on the WWW.

The development of on-line tools requires a sufficient level of informatics skills by the developer. This includes knowledge in Java language, particularly servlets and JDBC (database connectivity interface to access databases) as well as understanding XML-XSLT technologies. For the development of methods, a knowledge of mathematical and computer science approaches, such as neural networks and support vector machines, is also required. Of course, in addition to this, one should also have knowhow of chemistry! There is a great demand for such specialists in Academia and Industry [56]. A number of Universities world-wide started to provide chemoinformatics courses to teach students these diverse skills (http://www.indiana.edu/~cheminfo/informatics/cinformacad.html). An availability of such specialist is a requisite for the advancement of the field.

Concerning software for calculation of molecular parameters and descriptors, the Virtual Computational Laboratory (http://www.vcclab.org) will keep a list of sites providing such on-line services.

ACKNOWLEDGEMENTS

This study was partially supported by Virtual Computational Laboratory INTAS-INFO 00-0363 grant. I am grateful to Prof. J. Gasteiger, Prof. R. Todeschini, Dr. V. Yu. Tanchuk, Dr. P. Ertl, and their colleagues for participation in developing software modules used at the Virtual Computational Laboratory site. I would like also to thank Dr. G. Caron for her useful remarks, and Dr. M. Wormke for his help with English.

REFERENCES

[1] Tetko, I. V.; Tanchuk, V. Y.; Kasheva, T. N.; Villa, A. E. P. *J. Chem. Inf. Comput. Sci.* **2001***, 41*, 246.
[2] Friedman, B. B.; Sunseri, A. *Stud. Health. Technol. Inform.* **2002***, 80*, 175.
[3] Wiggins, G. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 956.
[4] Heller, S. R. *J. Chem. Inf. Comput. Sci.* **1995**, 36, 205.
[5] Patiny, L. *Internet J. Chem.* **2000***, 3*, 2.
[6] Tonge, A. P.; Rzepa, H. S.; Yoshida, H. *J. Chem. Inf. Comput. Sci.* **1999***, 39*, 483.
[7] Sadowski, J.; Kubinyi, H. *J. Med. Chem.* **1998***, 41*, 3325.
[8] Brecher, J. S. *Chimia* **1998***, 52*, 658.
[9] Tissue, B. M.; Van Bramer, S. E.; Rosenthal, D. *J. Chem. Inf. Comput. Sci.* **2002***, 42*, 23.
[10] Warr, W. A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 362.
[11] Livingstone, D. J. *J. Chem. Inf. Comput. Sci.* **2000***, 40*, 195.
[12] Linstrom, P. J.; Mallard, W. G. *J. Chem. Eng. Data* **2001***, 46*, 1059.
[13] Meylan, W. M.; Howard, P. H. *J. Pharm. Sci.* **1995***, 84*, 83.
[14] Tetko, I. V.; Tanchuk, V. Y.; Kasheva, T. N.; Villa, A. E. P. *J. Chem. Inf. Comput. Sci.* **2001***, 41*, 1488.
[15] Huuskonen, *J. Comb. Chem. High. Throughput Screen.* **2001**, 4, 311.
[16] Tetko, I. V.; Tanchuk, V. Y.; Villa, A. E. P. *J. Chem. Inf. Comput. Sci.* **2001***, 41*, 1407.
[17] Stahura, F. L.; Godden, J. W.; Bajorath, J. *J. Chem. Inf. Comput. Sci.* **2002***, 42*, 550.
[18] Engkvist, O.; Wrede, P. *J. Chem. Inf. Comput. Sci.* **2002***, 42*, 1247.
[19] Russom, C. L. *Toxicology* **2002***, 173*, 75.
[20] Ran, Y.; He, Y.; Yang, G.; Johnson, J. L.; Yalkowsky, S. H. *Chemosphere* **2002***, 48*, 487.
[21] Ihlenfeldt, W. D.; Voigt, J. H.; Bienfait, B.; Oellien, F.; Nicklaus, M. C. *J. Chem. Inf. Comput. Sci.* **2002***, 42*, 46.
[22] Weeks, J. R.; Kuras, J.; Town, W. G.; Vickery, B. A. *J. Chem. Inf. Comput. Sci.* **2002***, 42*, 765.
[23] Blake, E. K.; Merz, C. UCI repository of machine learning databases, http://www.ics.uci.edu/~mlearn/MLRepository.html, **1998**.
[24] Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; WILEY-VCH: Weinheim, **2000**.
[25] Todeschini, R.; Vighi, M.; Finizio, A.; Gramatica, P. *SAR QSAR Environ. Res.* **1997***, 7*, 173.
[26] Consonni, V.; Todeschini, R.; Pavan, M. *J. Chem. Inf. Comput. Sci.* **2002***, 42*, 682.
[27] Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Academic Press: London, **1999**.
[28] Cruciani, G.; Pastor, M.; Guba, W. *Eur. J. Pharm. Sci.* **2000***, 11*, S29.
[29] Carpy, A. J. *SAR QSAR Environ. Res.* **2002***, 13*, 403.
[30] Steinbeck, C.; Krause, S.; Willighagen, E. *Molecules* **2000***, 5*, 93.
[31] Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. *J. Chem. Inf. Comput. Sci.* **2003***, 43*, 493.
[32] Murray, B. H.; Moore, A. *Cyveillance* **2000**, http://www.cyveillance.com/web/newsroom/releases/2000/2000.
[33] Masunov, A. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1093.
[34] Ertl, P.; Jacob, O. *J. Mol. Model.* **1997***, 419*, 113.
[35] Hansch, C.; Leo, A.; Hoekman, D. H. *Hydrophobic, Electronic, and Steric Constants*; ACS: Washington D.C., **1995**.
[36] Hansch, C.; Leo, A. J. Subsistent constants for correlation analysis in chemistry and biology; Wiley: New York, **1979**.
[37] Leo, A. J. *Chem. Rev.* **1993***, 93*, 1281.
[38] Leo, A. J.; Hoekman, D. *Persp. Drug Discov. Design* **2000***, 18*, 19.
[39] Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **1995***, 35*, 1039.
[40] Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000***, 43*, 3714.
[41] Tetko, I. V.; Tanchuk, V. Y. *J. Chem. Inf. Comput. Sci.* **2002***, 42*, 1136.
[42] Wang, R.; Gao, Y.; Lai, L. *Persp. Drug Discov. Design* **2000***, 19*, 47.
[43] Poroikov, V. V.; Filimonov, D. A.; Ihlenfeldt, W. D.; Gloriozova, T. A.; Lagunin, A. A.; Borodina, Y. V.; Stepanchikova, A. V.; Nicklaus, M. C. *J. Chem. Inf. Comput. Sci.* **2003***, 43*, 228.
[44] Simkin, S.; Bartlet, N.; Leslie, A. *Java Programming Explorer*; The Coriolis Group, Inc.: Scottsdale, Arizona, **1996**.
[45] Gordon, R. *Essential JNI: Java Native Interface*; 1 ed.; Prentice Hall: Colorado, **1998**.
[46] Bloch, J. *Effective Java Programming Language Guide*; Addison Wesley Professional: N.Y., **2001**.
[47] Murray-Rust, P.; Rzepa, H. S. *J. Chem. Inf. Comp. Sci.* **1999***, 39*, 928.
[48] Murray-Rust, P.; Rzepa, H. S. *J. Chem. Inf. Comput. Sci.* **2001***, 41*, 1113.
[49] Gkoutos, G. V.; Murray-Rust, P.; Rzepa, H. S.; Wright, M. *J. Chem. Inf. Comput. Sci.* **2001***, 41*, 1124.
[50] Kay, M. *XSLT Programmer's Reference*; Wrox: Birmingham, UK, **2000**.
[51] Liao, Y. M.; Ghanadan, H. *Anal. Chem.* **2002***, 74*, 389A.
[52] Kehiaian, H. Publication, Retrieval and Exchange of Data: an Emerging Web-based Global Solution. *CODATA 2002*: Montréal, Canada, **2002**.
[53] Van de Waterbeemd, H.; De Groot, M. *SAR QSAR Environ. Res.* **2002***, 13*, 391.
[54] Ertl, P. *Chimia* **1998***, 52*, 673.
[55] Csizmadia, F. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 323.
[56] Russo, E. *Nature* **2002***, 419*, 4.

Copyright of Mini Reviews in Medicinal Chemistry is the property of Bentham Science Publishers Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.