Journal of Computer-Aided Molecular Design (2005) 19: 453–463
DOI 10.1007/s10822-005-8694-y

Virtual computational chemistry laboratory – design and description

Igor V. Tetko[a,b,*], Johann Gasteiger[c], Roberto Todeschini[d], Andrea Mauri[d], David Livingstone[e], Peter Ertl[f], Vladimir A. Palyulin[g], Eugene V. Radchenko[g], Nikolay S. Zefirov[g], Alexander S. Makarenko[h], Vsevolod Yu. Tanchuk[a] & Volodymyr V. Prokopenko[a]

[a]*Institute of Bioorganic & Petroleum Chemistry, Kyiv, Ukraine;* [b]*Institute for Bioinformatics, GSF-Forschungszentrum für Umwelt und Gesundheit, GmbH, Ingolstädter Landstraße 1, D-85764, Neuherberg, Germany;* [c]*University of Erlangen-Nürnberg, Computer-Chemie-Centrum, D-91052, Erlangen, Germany;* [d]*Milano Chemometrics and QSAR Research Group, Department of Environmental Sciences, Università di Milano – Bicocca, Milano, Italy;* [e]*Centre for Molecular Design, Portsmouth and ChemQuest, UK;* [f]*Novartis Institute for BioMedical Research, Cheminformatics, Basel, Switzerland;* [g]*Department of Chemistry, Moscow State University, Moscow, Russia;* [h]*Institute of Applied System Analysis, Kyiv, Ukraine*

Received 30 March 2005; accepted 13 June 2005

Key words: on-line analysis, physico-chemical property predictions, indices calculation, model generation and validation, drug design

Abstract

Internet technology offers an excellent opportunity for the development of tools by the cooperative effort of various groups and institutions. We have developed a multi-platform software system, Virtual Computational Chemistry Laboratory, http://www.vcclab.org, allowing the computational chemist to perform a comprehensive series of molecular indices/properties calculations and data analysis. The implemented software is based on a three-tier architecture that is one of the standard technologies to provide client-server services on the Internet. The developed software includes several popular programs, including the indices generation program, DRAGON, a 3D structure generator, CORINA, a program to predict lipophilicity and aqueous solubility of chemicals, ALOGPS and others. All these programs are running at the host institutes located in five countries over Europe. In this article we review the main features and statistics of the developed system that can be used as a prototype for academic and industry models.

Introduction

Internet activities have become in a few years a major investment in information, business, communication, teaching technologies and chemistry [1, 2]. The WWW (World Wide Web) impact on society dramatically increased especially in the field of education and scientific research. The Internet is becoming a major system for knowledge extraction and education in the new century. There is an increasing number of universities offering on-line courses or even complete education over the Internet. Teleconferences and e-conferences are becoming more and more a standard means of communication that substitute for the usual meetings. The Web services [3] have recently appeared as an industrial standard for calculation over the Web and this field is developing quickly. A great deal of information is available for chemists in the form of chemical databases with physicochemical, thermochemical, toxicity, bioactivity, etc. data,

*To whom correspondence should be addressed. Phone: +49-89-3187-3575, Fax: +49-89-3187-3585
E-mail: itetko@vcclab.org

on-line resources for prediction of molecular properties, activities and calculation of molecular descriptors that were recently reviewed elsewhere [4].

The academic scientific research can have a specific place in this system by providing access to scientific programs developed here. Such programs developed by professionals can become available to a worldwide audience thus providing applications across several disciplines of science and industry. A large number of available scientific programs have been developed in FORTRAN and C/C++ programming languages. The question is how to make these software products publicly available through the Internet. The main idea is to make such integration as general as possible, flexible for extension of programs and incorporation of new modules with minimal changes in the existing software. This can be important to share scientific programs and methods of data analysis over the Internet.

This article presents an example of an academic software site, Virtual Computational Chemistry Laboratory, developed by several academic groups in Europe. We describe the underlying technology and the services available at our site, provide user statistics and share the experiences gained during the work. This article can be interesting to users of our services as well as to the scientists working on similar projects.

Methods and VCCLAB structure

The VCCLAB site was inspired following our development of on-line tools for analysis of neuroscience data [5] and an early version of a program for the calculation of logP and logS (ALOGPS) [6] at the University of Lausanne, Switzerland.

The VCCLAB is a Java based server

The Java language is increasingly used by the Internet community and successively covers newer and newer areas, from 3D graphics and on-line games to e-business. Since the beginning, Java creators envisioned the same Java program running on different types of computer chips and in many different operating environments. Thus, the Java compiler does not convert a program to a machine language specific code but to a byte code of so-called Java Virtual Machine (JVM). The JVM is implemented in software and represents the Java interpreter. The JVM is available on most computer platforms and computer systems allowing the same Java program to run on different machines.

Three-tier organisation of the VCCLAB site

The VCCLAB includes three main parts: Applet Clients, Super Server and Calculation Servers (Figure 1). The Applet Clients represent a front-end part of the site and allow the users to provide data, specify parameters, execute tasks and collect calculated results (Table 1). The Calculation Servers execute the tasks submitted by the clients (Table 2). The Super Server provides a link between the Applet Clients and Calculation Servers. Below, we describe each component of the VCCLAB site in details.

Super Server is the central processing unit, or heart, of the VCCLAB site. It organises a queue of tasks submitted by the Applet Clients, dispatches them to the corresponding Calculation Servers, collects and returns the calculated results to the Client. The Super Server recognizes the applications according to an identification TASK keyword, i.e. 'logp' in the case of the ALOGPS program or 'asnn' in the case of neural networks. The tasks submitted by a user and/or subtasks provided by the Calculation Servers are stored on the Super Server. The Calculation Servers send a request to the Super Server to verify if there are any tasks available for them. If some tasks are available (this is determined by matching the TASK keyword of server and available task) the server uploads the corresponding task and calculates it. The Super Server is also used to upload data files using Java servlets. The analysed data and calculated results are stored internally in zip format. This allows one to decrease the network traffic and to increase the speed of network connections. This is particularly important for users from the developing and third-world countries, where Internet connections are still mainly done through dial-up and telephone lines. The Super Server also logs statistics of submitted tasks. The tasks analysed by the users are not normally stored unless they cause a crash of the corresponding Calculation Server. In this case the task is stored on disk and it can be re-run to debug the

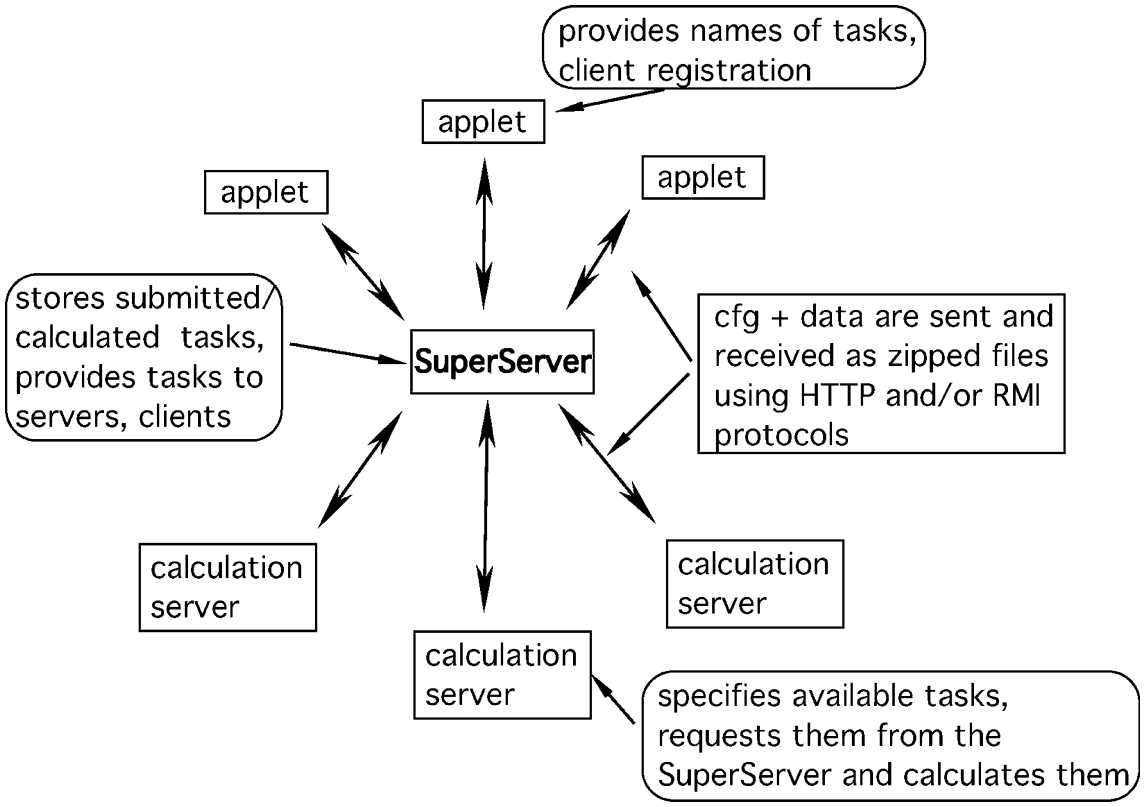

Figure 1. Data-flow in the Virtual Computational Chemistry Laboratory.

Table 1. Applet Clients available at the VCCLAB site.

No.	Applet Client	Required Calculation Servers
1	Associative Neural Network (ASNN)	asnn, asnnp
2	Polynomial Neural Networks (PNN)	pnn
3	Open Babel	babel, mol2sm
4	Electronic Dragon (E-Dragon)	dragon, corina, alogps
5	Parameter Client (PCLIENT)	mainserv, dragon, corina, e-state, alogps, density, cfrag, cfrag-l, jme
6	Artificial neural network logP and logS calculation program (ALOGPS)	alogps, mol2sm, babel, jme, xlogp
7	Unsupervised Forward Selection (UFS)	ufs
8	Partial Least Squares (PLS)	pls

corresponding Server. The First In First Out (FIFO) queue is used for all submitted tasks. The users can also register at our site to request a download of some programs and/or to submit tasks requiring long calculation times. The registration and storage of information about registered users in a relational database is also handled by the Super Server.

Applet Clients or simply applets represent an important feature of the Java language since they are accessible over the Internet using WWW browsers. Sometimes browsers do not include support of Java applets as a default option. This can be easily changed by downloading and installing a recent version of the Java runtime software from http://www.java.sun.com. The applets were

Table 2. Description of Principal Calculation Servers used at the VCCLAB server.

No.	Module	Developed by	Programming language and operation system	Integration with Java	Short description	Number of servers
1	alogps	IBPC, UNIL	C++/MacOsX[a]	JNI	Calculation of logP (lipophilicity) and logS (aqueous solubility)	1
2	asnn	IBPC, UNIL	C++/Linux[a]	Standalone	Neural network data analysis	14
3	asnnp	IBI	Java	Java code	Parallelisation of the ASNN calculations	2
4	babel	Open Babel project[b]	C++/MacOsX[a]	Standalone	Interconversion of molecules in different formats	2
5	cfrag	MSU	C/Windows	Standalone	Calculation of fragment-based indices	2
6	cfrag-l	MSU	C/Windows	Standalone	Calculation of fragment-based indices	2
7	corina	UNIE	C/Linux	Standalone	Conversion of molecules to 3D	2
8	density	MSU	C++/Windows	Standalone	Prediction of density of chemicals	2
9	dragon	UNIMIB	Basic/Windows	Standalone	Calculation of 1600 indices of molecules	1
10	e-state	IBPC, UNIL	C++/Windows	JNI	Calculation of E-state indices	1
11	jme	Novartis	Java	Java code	Drawing of Molecules	–[c]
12	mainserv	IBPC	Java	Java code	Management of Parameter Client tasks	2
13	mol2sm	IBPC, UNIL	C++/MacOsX[1]	JNI	Conversion of sdf and mol2 files to SMILES and vice versa	1
14	pnn	IASA, IBI	C++/MacOsX	JNI	Non-linear modelling of data	1
15	ufs	UNIP, IBI	C++/SGI, Linux	Standalone	Dimensionality reduction	2
16	xlogp[d]	UNIL, Beijing University	C++/Linux	Standalone	XLOGP calculation	1
17	plsr	MSU	C++/Windows	Standalone	PLS regression	2

IASA – Institute of Applied System Analysis, IBI – Institute for Bioinformatics, IBPC – Institute of Bioorganic Chemistry & Petrochemistry, MSU – Moscow State University, UNIE – University of Erlangen-Nuremberg, UNIL – University of Lausanne, UNIMIB – University of Milano-Bicocca, UNIP – University of Portsmouth.
[a]The running system is indicated. The program was also tested on Linux, Windows, MacOsX and Sun systems.
[b]Open Babel is open source software available at http://openbabel.sf.net.
[c]JME is available as part of the Applet Clients and thus runs on the computers of users.
[d]XLOGP[15] program is freely distributed by Beijing University [16].

developed as front-ends for all programs accessible at the VCCLAB site. They are used to upload data, to configure the calculated tasks, to check the consistency of the parameters, to submit tasks and to visualize calculated results. The parameters of the applets include short names (keywords) that are highlighted and appear as links in the applet. A click on such a link opens an HTML window with an explanation of the keyword. This allows the users to become familiar with the site and to quickly learn how it can be used. Most of the applets contain pre-loaded data and a pre-configured task. The first-time user can simply press a 'submit task' button and launch his first analysis without entering any data or specifying any configuration parameters. The applet links contain descriptions of the input data format and of the calculated results. All these features make it possible for a non-experienced user to quickly become familiar with the software. Some applets contain a task manager panel. It displays the status of submitted and calculated tasks, statistics of connected users and available servers and allows one to save or upload completed tasks. The Registered Users can submit a task requiring a long calculation time and to close the browser with the applet. The user will be notified about the completion of the task by e-mail.

Calculation Servers. The academic software tools integrated into the VCCLAB site include more than ten programs and modules that were developed in Fortran, C/C++ and Basic programming languages (Figure 2). Two strategies were used.

The ALOGPS, PNN and program for the calculation of E-state indices (see Table 2) were integrated using the Java Native Interface (JNI) [7]. An interface between Java and native code was programmed by declaring in a Java program native methods implemented using C/C++ code. Calls of such methods allowed a fast execution of time-critical code as well as re-use of the C/C++ libraries. The results of such calls were directly sent to the Java program.

The other programs were integrated using a dedicated Calculation Server, the Exec Server, which executed the standalone programs and collected the calculated results using a standardised command-line interface. Both the name of the programs to be executed and the result files to be collected by the Exec Server are specified in a configuration file. The list of programs to be configured for each computer is determined by the server administrator and it depends on the resources available on each site. This made possible a flexible distribution of programs amid different servers.

Results

The software available at the VCCLAB site can be divided into two major categories: indices and property calculation software and data analysis tools. However, we start its description from two additional important programs, a hub to convert molecules between different formats, Open Babel, and a molecular editor to draw molecules, JME.

Molecule conversion and preparation programs

Open Babel. The molecules used for analysis by the VCCLAB server are internally represented either as SMILES codes or sdf files. These are the two main formats supported by the Server. Since the user can have his/her data in different molecular formats, the molecular structure information interchange hub, BABEL, originally started by Pat Walters and Matt Stahl at the University of Arizona and later-on developed as the open source project, Open Babel (http://openbabel.sf.net) was used to provide a conversion between 48 chemical data formats. This software is accessed as a separate Applet Client and it is also integrated into the property calculation programs to allow 'on-the-fly' interconversion of the molecules in different data formats.

JME Molecular Editor. Some programs, ALOGPS and PCLIENT, include the possibility to prepare SMILES codes using the JME Molecular Editor [8]. This editor was developed in Java and allows easy creation and editing of molecules. The editor generates a SMILES code that is passed for further analysis to the server. JME is a very convenient tool especially for people who are not familiar with the SMILES notation.

Indices and property calculation programs

The ALOGPS 2.1 program predicts lipophilicity and aqueous solubility of non-charged chemical compounds [9, 10]. This program is one of the first

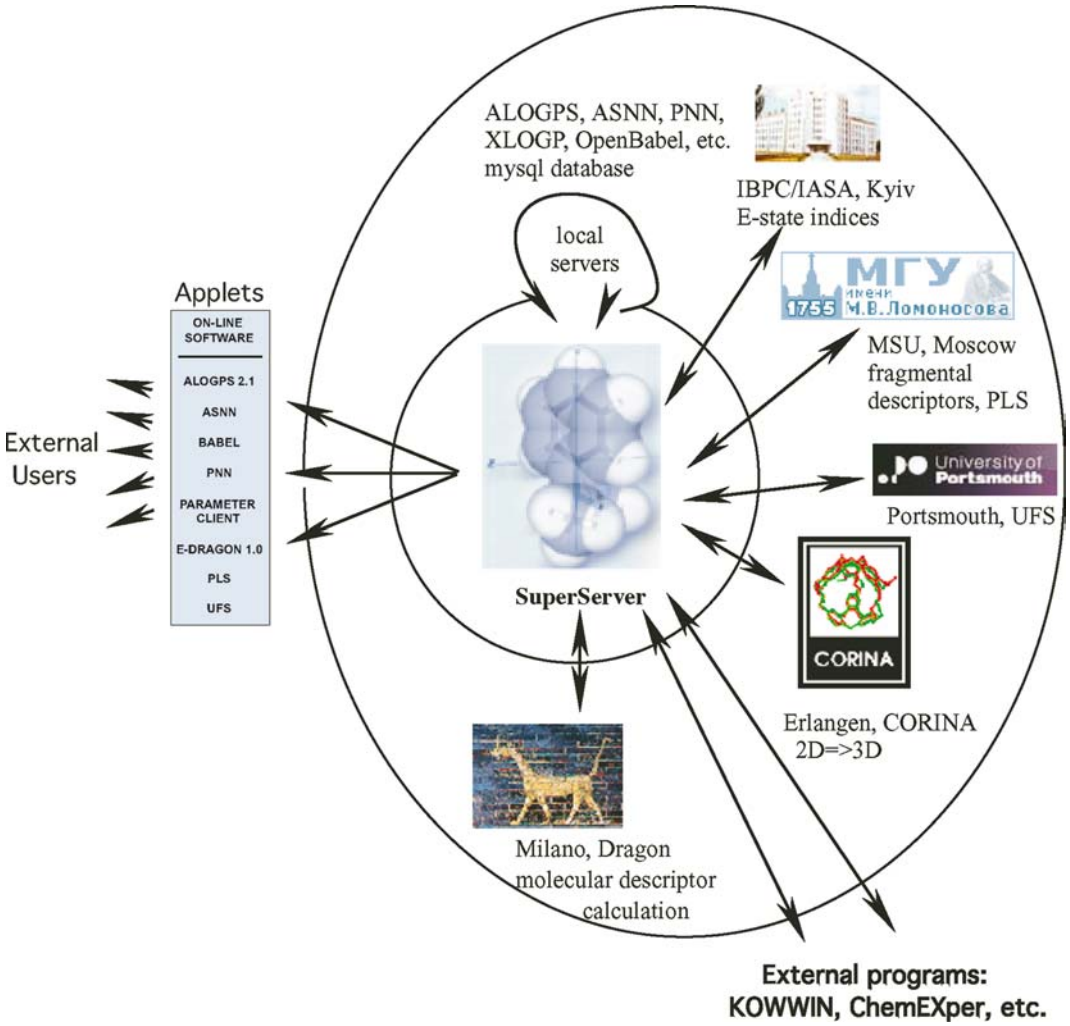

Figure 2. The main Calculation Servers running at the organisations of the grant participants at the VCCLAB site.

programs that was available for on-line calculation at our site since May 2002. Because a user can be interested in comparing results calculated using several different methods, the applet displays the lipophilicity results calculated by our own model, ALOGPS, and five other programs, namely CLOGP [11], KOWWIN [12], IA_LOGP [13], miLogP [14] and XLOGP [15]. Demo versions of the first four programs are available on-line (the links are provided on the ALOGPS page) and the applet collects values calculated by these programs using the HTTP protocol. The XLOGP program [15], which is freely distributed by the Institute of Physical Chemistry, Beijing University [16], was implemented as an independent Calculation Server. The aqueous solubility data include ALOGPS and IA_LOGS [13] program values. The applet also displays logP and logS experimental values that were used in references [17, 18] and constitutes our internal database. If there are no values in this database, they are retrieved from the on-line version of PHYSPROP [19]. The ALOGPS 2.1 program includes an innovative feature, the so-called LIBRARY mode. The LIBRARY allows users to include molecules in the training set without retraining the neural network weights. This increases the prediction ability of the method for the user's molecules. Moreover, when using a library with logD values the program can also predict logD, i.e. the lipophilicity of charged compounds [20, 21]. The LIBRARY mode is available on-line and more than 50 users have

already used it to analyse their data using the WWW interface.

E-Dragon is an electronic remote version of the popular DRAGON indices calculation program, which is an application for the calculation of molecular descriptors developed by the Milano Chemometrics and QSAR Research Group. DRAGON [22] provides more than 1600 molecular descriptors that are divided into 20 logical blocks. The user can calculate not only the simplest atom type, functional group and fragment counts, but also topological and geometrical descriptors, autocorrelation and information indices, 3D molecular descriptors, molecular properties, etc. (see http://www.vcclab.org/lab/indexhlp/). Besides the standard parameters of the Dragon program the molecular properties also include logP and logS values calculated using the ALOGPS program. If the 3D atomic coordinates are not available for the provided molecules, the user can calculate them using CORINA [23].

Parameter Client (PCLIENT) [24] is an extension of E-Dragon and one of the main developments of the VCCLAB project. The PCLIENT receives tasks from the user, plans tasks for the Calculation Servers, receives intermediate data and results and, if required, prepares new tasks, and forms final results for the clients. PCLIENT is scalable and extensible. Any number of different descriptor calculation modules can be incorporated into it.

If some molecular parameters are dependent on the others, the PCLIENT determines this situation and includes the required parameters into its calculation plan. This feature allows us to incorporate into the VCCLAB not only a number of molecular descriptors, but also a number of QSAR and QSPR models, based on them. The system is easily extensible and can include any number of calculation servers to calculate molecular descriptors and properties according to the QSAR/QSPR models.

The flexibility and extensibility of PCLIENT is achieved due to its internal representation of data. It is organized as a relational database and stored in a number of tables in a MySQL database (http://www.mysql.com). In general, PCLIENT receives arguments and calculates parameters. Both arguments and parameters do not have direct relation to molecules or other chemical notions. They could be represented by anything (molecules, different parameters of the calculation methods, etc.). There is also a table to provide a 'multiple-to-multiple' relationship between parameters and arguments. This table stores arguments needed to calculate parameters. Parameters are divided into groups that could be calculated by different Calculation Servers. The servers are provided by the authors (at the present time INTAS project participants, though other contributors are highly encouraged). The parameters can also be used as the arguments. Such parameters should be calculated first and later used as arguments to calculate the other parameters dependent on them.

Both parameters and arguments are typed values. Each type is a structure that consists of the fields of simple types. Each field may contain also an array of fixed or dynamic length. At the present time there are five such basic types – integer, float (4-byte floating point value), double (8-byte floating point value), String and MOLEC String (a string with a molecule in any format). Such flexible representation of data types allows the definition and use of complex structures of any nature as parameters and arguments for the Calculation servers. The data of any complexity is represented by a set of data fields. Each data field is uniquely identified by 4 integers and may be accompanied by the value itself in the string form.

The PCLIENT consists of three parts – Java applet, the so-called 'Main Server' and Calculation Servers. The Java applet provides the user interface. It is accessible at the VCCLAB site, though it may be also used as a standalone application by the registered users. It receives input from users, creates tasks and sends them to the 'Main Server'. 'Main Server' is the server that manages the whole process. It creates the tasks for the calculation servers, collects the results, prepares the final result and sends it back to the PCLIENT.

At the present moment the Calculation Servers run Corina, Dragon, E-state indices calculation module, ALOGPS, density and two fragment-based indices calculation programs. These programs are running in four countries (Germany, Italy, Russia and Ukraine) on three different computer platforms (MacOsX, Linux and Windows). The PCLIENT provides a seamless integration of all these modules and makes it possible to generate more than 3000 indices listed at http://www.vcclab.org/lab/indexhlp. The indices calculated by

these programs can be analysed by the data analysis programs that are also available at the VCCLAB site.

Data analysis programs

Unsupervised Forward Selection (UFS) [25] is a data reduction algorithm that selects from a data matrix a maximal linearly independent set of columns with a minimal amount of multiple correlation. UFS was designed for use in the development of Quantitative Structure-Activity Relationship (QSAR) models, where the m by n data matrix contains the values of n variables (typically molecular properties) for m objects (typically compounds). QSAR data sets often contain redundancy (exact linear dependencies between subsets of the variables), and multicollinearity (high multiple correlations between subsets of the variables). Both of these features prevent the development of predictive QSAR models, i.e. models with the ability to generalise successfully to new objects. UFS produces a reduced data set that contains no redundancy and a minimal amount of multicollinearity.

Associative Neural Network (ASNN) [26, 27] represents an innovative method to calculate nonlinear models between indices and molecular properties. The method represents a combination of an ensemble of feed-forward neural networks and the k-nearest neighbour technique. If new data become available, the network further improves its predictive ability and provides a reasonable approximation of the unknown function without a need to retrain the neural network ensemble. We also programmed a Java proxy-server, which makes possible execution of one task on several computers simultaneously.

Polynomial Neural Network (PNN) [28, 29] correlates input and target variables using (non) linear regression. In this software the user can define the desired properties of the solution such as the number of terms and the maximum degree of polynomials. The PNN calculates analytical models that can be easily interpreted. This is a substantial advantage of this method over other neural network approaches. Both approaches were recently compared to several other neural network methods using several standard QSAR datasets [30].

Partial Least Squares (PLS) uses a two-step descriptor selection procedure. The first step eliminates low-variable descriptors and the second step optimises the descriptor subset using a Q^2-guided descriptor selection by means of a genetic algorithm. The computational experiments demonstrate the stability and good prediction accuracy of models [31].

The use of Java-based technology provided a seamless integration of these software products. The developed software programs are executed on four different computer platforms (MacOsX, Linux, SGI and Windows) in five countries over Europe (Table 2). A possibility of the Calculation Servers to execute numerous tasks makes possible a flexible configuration of them. Some computer-demanding tasks, such as ASNN, are running in parallel on up to 14 computers. This enables a fast execution of even sophisticated analyses.

User statistics

The VCCLAB has been available for public use since 2002. A number of programs were developed, tested and made accessible at the site during this time. In June 2004 a final version of the site was released and the external users were also given an opportunity to register and request standalone versions of ALOGPS and ASNN programs. Since that time the number of registered users reached 384 scientists (as of March, 2005). The geographic distribution of registered users indicates a wide interest in the project of scientists from all over the world (Figure 3). Currently, we have Registered Users from more than 59 countries and from all continents with the exception of Antarctica. It is interesting, that the number of Registered users from India is the second largest after the USA. This indicates an increasing involvement of this country in the development of drug design and chemoinformatics. Figure 4 illustrates the dynamics of the site statistics as a function of time. The ALOGPS program for MS Windows accounted for 3/4 of all downloads and about 25% of users requested the ASNN program. Only 8 and 2 users requested the ALOGPS program for Linux and MacOsX, respectively. The ALOGPS program was the most frequently used at the VCCLAB site, too. Out of more than 51,324 tasks that were executed on our site during 2004 the ALOGPS and its subtasks accounted for about 80% of them.

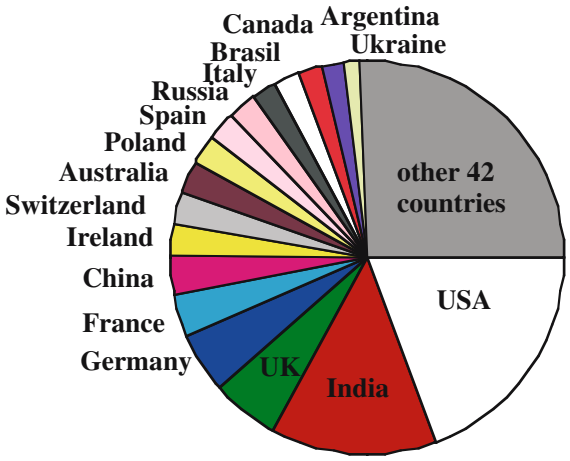

Figure 3. The distribution of 384 users registered on our site (June 2004 to March 2005) from different countries of the world. The 'other' category comprises users from countries with four or less registered users.

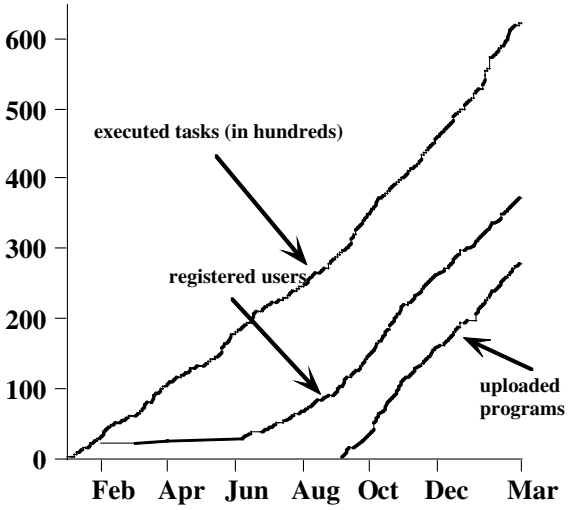

Figure 4. The VCCLAB server statistics for January 2004 to March 2005. The downloaded programs are counted only starting September 2004 when an automatic licensing system was made available at the server.

Discussion

The VCCLAB project created an inter-platform software that can be used for comprehensive computer-aided molecular design on the WWW. The increasing interest of users from different countries all over the world indicates that this server has received the wide attention of many groups of scientists. The user statistics indicate a growing interest of the Internet community in the software tools offered by the VCCLAB project. The ALOGPS program so far received the highest attention.

The VCCLAB site was developed in Java language. We believe that it was a correct strategic choice that allowed us to achieve the goals of the project, i.e. development of truly platform independent software that can be used by many scientists all over the world. Thus this technology can be recommended for the development of similar complex software servers in different fields of science. However, the use of Java also raised some technical issues that should be taken into account when developing similar software suites. Firstly, the use of JNI appears to be more complex than was foreseen. The difficulties appeared due to principal differences in library management on various computer systems. Thus, if speed is not the main issue of the server, one should better consider interaction on the level of applications. Secondly, even if software is developed in C or C++ codes, it is sometimes not simple to make it portable to another computer platform. The problems could arise due to different initialisation of variables, memory management, differences in compiler brands, etc. Thus, it is better to run the program using the environment in which it was developed and debugged. Thirdly, each developed server should be duplicated and should run on at least two different computers. This will help to avoid dead-lock situations, when one of such servers is down, e.g. due to network connections or operation system failures.

The VCCLAB project limits the maximum number of molecules that can be calculated in one task using E-Dragon and PCLIENT to 150. Also maximum of 150 atoms per molecule is allowed. These limits appear due to Corina and Dragon programs that are distributed commercially. If a user is interested in running really large sets of molecules in a batch mode on a daily basis, he/she should contact Molecular Networks GmbH (http://www.mol-net.com) and Talete Srl (http://www.talete.mi.it) to obtain commercial versions of the programs without such restrictions. The memory limitations are also important since, e.g., an analysis of 150 molecules in PCLIENT generates on average a file with >4 MB of indices. The display of such a file in the PCLIENT applet requires even larger memory, ca. 40 MB, of the Applet. We do not think that the current

limitation on the maximum number of molecules per task is a strictly restrictive one, since the user can subdivide the molecules into several subtasks and analyse them one by one. The other programs do not have limitations on the number of simultaneously analysed molecules.

The VCCLAB site uses a customised HTTP-based protocol for communication between Applet Clients, Calculation Server and Super Server. This technology requires some knowledge and understanding of the principles of Java classes used in the project. However, the VCCLAB can easily be integrated with other software tools on the WWW. The Web Services deserve a particular mention [3]. This technology was introduced in June 2000 and is becoming a standard for business development on the Internet. The Web Services uses SOAP (Simple Object Access Protocol) that is an extensible XML messaging protocol to describe each service. It makes possible an easy integration of programs written in different computer languages and running on different computer platforms. The VCCLAB site can be converted to the Web Services by exporting some of its Client Applets. This can provide in the future an easy integration of the VCCLAB site with this technology, providing that Web Services will find a wide application in the chemoinformatics and drug design field.

There are more and more new developments for science and research on the Internet including a number of Open Source initiatives [32]. For example, the LINK3D project [33] developed tools and software for synchronous collaboration in the field of drug design. The Virtual Institute for Chemometrics and Industrial Metrology (VICIM, http://www.vicim.urv.es) was organized to advance e-school and to provide a chemometric channel for knowledge transfer and pre-competitive research to industry and academia.

In summary, we have developed Internet software for calculation of molecular indices and properties of chemical compounds and data analysis tools and described the main components of this software. Our package so far includes software tools developed in laboratories of the project participants but it can also include new programs developed by other scientists who can contact us in order to receive sample code and instructions how to implement this software. This can provide a worldwide dissemination of academic results and will have a positive impact on research in chemistry and the drug design field.

Acknowledgements

This study was partially supported by 'Virtual Computational Chemistry Laboratory' INTAS grant 00-0363. We thank members of our teams, particularly Alexander Sashko, Anna Ivanova, David Whitley, Giorgi Lekishivili, Igor Baskin, Kristina Danilova, Maria Skvortsova, Markus Sitzmann, Natalia Halbershtam, Tetyana Aksenova, Vasily Kovalishyn, Vladimir Sykora who contributed to the testing, support and development of this site.

References

1. Wiggins, G., J. Chem. Inf. Comput. Sci., 38 (1998) 956.
2. Murray-Rust, P., Rzepa, H.S., Tyrrell, S.M. and Zhang, Y., Org. Biomol. Chem., 2 (2004) 3192.
3. McGovern, J., Java Web Services Architecture. Morgan Kaufmann, 2003.
4. Tetko, I.V., Mini Rev. Med. Chem., 3 (2003) 809.
5. Villa, A.E.P., Tetko, I.V. and Iglesias, J., Neurocomputing, 38–40 (2001) 1025.
6. Tetko, I.V., Tanchuk, V.Y., Kasheva, T.N. and Villa, A.E., J. Chem. Inf. Comput. Sci., 41 (2001) 246.
7. Gordon, R., Essential JNI: Java Native Interface, 1st edn. Prentice Hall, Colorado, 1998.
8. JME Homepage is at http://www.molinspiration.com/jme/, 06/06/2005.
9. Tetko, I.V., Tanchuk, V.Y. and Villa, A.E., J. Chem. Inf. Comput. Sci., 41 (2001) 1407.
10. Tetko, I.V. and Tanchuk, V.Y., J. Chem. Inf. Comput. Sci., 42 (2002) 1136.
11. Leo, A.J. and Hoekman, D., Persp. Drug Discov., Design (2000) 18–19.
12. Meylan, W.M. and Howard, P.H., J. Pharm. Sci., 84 (1995) 83.
13. Interactive Analysis Bedford MA, http://www.logp.com, 06/06/2005.
14. Molinspiration Cheminformatics Bratislava Slovakia, http://www.molinspiration.com, 06/06/2005.
15. Wang, R., Gao, Y. and Lai, L., Persp. Drug Discov. Design, 19 (2000) 47.
16. XLOGP program is available by anonymous ftp to ftp2.ipc.pku.edu.cn, directory 'pub/software/xlogp'.
17. Huuskonen, J.J., Livingstone, D.J. and Tetko, I.V., J. Chem. Inf. Comput. Sci., 40 (2000) 947.
18. Tetko, I.V., Tanchuk, V.Y., Kasheva, T.N. and Villa, A.E., J. Chem. Inf. Comput. Sci., 41 (2001) 1488.
19. PHYSPROP database is available from Syracuse, Inc. http://www.syrres.com, 06/06/2005.
20. Tetko, I.V. and Bruneau, P., J. Pharm. Sci., 93 (2004) 3103.
21. Tetko, I.V. and Poda, G.I., J. Med. Chem., 47 (2004) 5601.

22. Todeschini, R. and Consonni, V., Handbook of Molecular Descriptors. Wiley–VCH, Weinheim, 2000.
23. Sadowski, J., Gasteiger, J. and Klebe, G., J. Chem. Inf. Comput. Sci., 34 (1994) 1000.
24. Tanchuk, V.Y., Tetko, I.V., Sasko, A. and Prokopenko, V.V., Parameter Client: Calculation of molecular descriptors in the Internet as a part of the Virtual Computational Chemistry Laboratory. QSAR2004, Istanbul, 2005.
25. Whitley, D.C., Ford, M.G. and Livingstone, D.J., J. Chem. Inf. Comput. Sci., 40 (2000) 1160.
26. Tetko, I.V., J. Chem. Inf. Comput. Sci., 42 (2002) 717.
27. Tetko, I.V., Neural Proc. Lett., 16 (2002) 187.
28. Aksyonova, T.I., Volkovich, V.V. and Tetko, I.V., SAMS, 43 (2003) 1331.
29. Tetko, I.V., Aksenova, T.I., Volkovich, V.V., Kasheva, T.N., Filipov, D.V., Welsh, W.J., Livingstone, D.J. and Villa, A.E.P., SAR QSAR Environ. Res., 11 (2000) 263.
30. Ajmani, S., Tetko, I.V., Livingstone, D.J. and Salt, D., A comparative study of neural network architectures for QSAR. QSAR2004, Istanbul, 2005.
31. Palyulin, V.A., Radchenko, E.V., Baranova, O.D., Oliferenko, A.A. and Zefirov, N.S. MFTA: Recent Extensions of Molecular Field Topology Analysis. EuroQSAR2002. Blackwell Publishing, Bournemouth, UK, 2003, pp. 188–190.
32. Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., Luttmann, E. and Willighagen, E., J. Chem. Inf. Comput. Sci., 43 (2003) 493.
33. Pastor, M., Benedetti, P., Carotti, A., Carrieri, A., Diaz, C., Herraiz, C., Holtje, H.D., Loza, M.I., Oprea, T., Padin, F., Pubill, F., Sanz, F. and Stoll, F., J. Comput. Aided. Mol. Des., 16 (2002) 809.