feature

medications to treat Alzheimer's and cardiovascular diseases. The company aims to become a fully integrated pharmaceutical company with one-stop production and marketing capabilities. The company's work hasn't gone unnoticed by investors, as the company raised about US$65 million since 2004.

To achieve its other strategic goal of helping create and develop companies, the North Carolina Biotechnology Center provides low-interest loans to young companies long on promise and potential but short on cash. The funding comes at a critical time, when the companies struggle to obtain financing. To date, the Biotechnology Center has invested about US$14.5 million in ~90 companies, which have gone on to raise more than US$1 billion in other investments.

For example, Trimeris, a company spun out of Duke University Medical Center, was able to use a US$250,000 Biotechnology Center loan to attract venture capital, which later led to a US$33 million public stock offering. Today the company employs ~135 people and sells Fuzeon, a life-extending drug for the treatment of AIDS. The drug is the first in a new class of anti-HIV drugs known as entry inhibitors that work by blocking HIV from entering T cells.

Tony Laughrey of KBI BioPharma can relate to Trimeris' success story. KBI BioPharma used the Biotechnology Center's US$1 million loan to leverage US$32 million in private funding. That early success has affirmed the company's decision to locate in North Carolina.

'We knew we belonged in North Carolina', Laughrey said. 'And we are glad we are here.'

Boris Hartl
North Carolina Biotechnology Center,
15 T.W. Alexander Drive,
Post Office Box 13547,
Research Triangle Park,
NC 27709-3547, USA

feature

Computing chemistry on the web

Igor V. Tetko, itetko@vcclab.org

Despite the dramatic growth of the internet, the number of practical applications in drug design that are available online, particularly those that predict absorption, distribution, metabolism, excretion and toxicology (ADME–Tox) properties, remains limited. For example, the number of methodological publications about lipophilicity predictions has gradually increased over the past ten years and >25 of these articles are expected to be published in different journals in 2005 [1]. At the same time, the number of programs available for the online prediction of this important property is approximately ten applications [2]. Publicly available tools for predicting many other important physico-chemical and biological properties simply do not exist. Thus, there is a need to develop new web applications to boost drug design and chemoinformatic studies. This article explains why publishing programs on the internet is important and beneficial for authors and users alike, it describes several examples of such sites and, crucially, discusses why this field remains underdeveloped compared with its nearest rival, the field of bioinformatics.

Why develop web services?

The typical research activity of a computational chemist includes data preparation, optimizing molecular structures, calculating indices, selecting the most important indices and deriving property–activity correlations using statistical methods. Then, the calculated models are revised and poorly predicted compounds are analyzed to gain an insight into the new indices that are required to improve the overall model or to detect errors and inconsistency in data preparation. However, if a new method has been developed it might be interesting to compare it with previously existing approaches. This could be difficult if the algorithm is only referenced in a paper but even if the method is available as source or binary code a dedicated computer platform, particular operating system, configuration of system parameters, compiler, libraries, among others, might also be required to perform a study. Problems can arise because the distribution of a binary code can unintentionally propagate viruses or spy software or there could be a conflict of interests (e.g. if the authors plan to commercialize it and substantial efforts are required to support it).

Contrary to this, publishing on the web allows a program to be executed in the same environment where it was developed. This software is easy to maintain, update and every web user can access and use it in their workplace. The web applications also enable much better dissemination of information regarding the algorithm, because of powerful search engines (e.g. Google or Yahoo).

Some examples of web services

An increasing number of diverse tools for performing data analysis in chemistry on the internet is available for users (Table 1), this is also reviewed elsewhere [2–5]. The Virtual Computational Chemistry Laboratory

1359-6446/05/$ – see front matter PII: S1359-6446(05)03584-1

feature

TABLE 1

Some examples of free computational web resources in chemistry[a]

Number	Name	Information provided	Link
1	CORINA	2D to 3D conversion of molecular structures	http://www2.chemie.uni-erlangen.de/software/corina/
2	Osiris	logP, solubility, toxicity, drug likeness	http://www.organic-chemistry.org/prog/peo
3	Petra	Physico-chemical properties of compounds	http://www2.chemie.uni-erlangen.de/services/petra
4	Pre-ADME	Molecular descriptors and various ADME–Tox properties	http://preadme.bmdrc.org/preadme
5	VCCLAB	Molecular descriptors, physico-chemical properties and data analysis tools	http://www.vcclab.org

[a]Extended comprehensive lists of other resources can be found elsewhere, references [2–5].

(VCCLAB, www.vcclab.org) is one of the most comprehensive resources for data analysis in chemistry on the web [6]. The front-end of the laboratory, visible to the users, is represented by applets (Figure 1). They are used to start, control and display the results of different tasks ranging from the generation of descriptors to the development of predictive models. The Java-based system seamlessly integrates programs running on computers and various operating systems in five countries in Europe. For example, the unsupervised forward selection algorithm [7] is implemented on Silicon Graphics® in Portsmouth (UK), CORINA [8] is executed on a Linux machine in Erlangen (Germany), ALOGPS [9] runs on a MacOsX system, DRAGON [10], E-state and fragment-based descriptors are performed on Windows in Milan (Italy), Kyiv (Ukraine) and Moscow (Russia), respectively. A simple click on a mouse starts a sequence of tasks that will be executed on computers, located 1000s of miles from one another. Currently, the VCCLAB site calculates 100s of tasks every day and has >600 registered users, including 46% with PhDs. The largest number of users came from the USA (115) and the second largest number of users (81) came from India. There were 207 users from the European Union in total, the majority of these came from the UK (31), followed by Germany (27). 76% of all users came from academia, 17% came from industry and the remaining 7% came from governmental organizations.

Similar projects that provide comprehensive resources on the web also exist in industry [11] or in collaborative research programmes between academia and industry [12]. The LINK3D project [12] developed tools and 3D software for synchronous collaboration in the field of drug design, in particular the virtual meeting software. The Novartis system supports >1000 users with molecular modelling and molecular processing tasks. These include the calculation of molecular and substituent properties, property-based virtual screening, visualization of molecules and bioisosteric design, among others [11]. However, these systems are usually not available publicly. The success of the VCCLAB site clearly indicates that there is a demand for the development of public versions of such systems.

The chemical community is also actively involved in the development of new protocols for the internet. The Chemical Markup Language (CML™, www.xml-cml.org) was one of the first XML-based standards for the scientific exchange of information [13]. The group that developed CML™ is actively involved in the popularization of the Semantic Web and in new ways of scientific publishing on the web [14,15], which is, without doubt, very important work. However, it looks like a large part of the chemical community is not yet actively involved in this development and the number of web services in the field remains limited. With this in mind it is interesting to compare chemo- with bioinformatics; the latter is definitely leading the way in web developments.

Why are chemoinformatics and bioinformatics different with respect to publishing on the web?

An explanation for the disparity between chemoinformatics and bioinformatics is a dramatic difference in the amount of data available and the computer resources that are, consequently, required for this data storage. For example, human genome data are stored as 2.7 GB of zipped files at Ensembl (www.ensembl.org) and the methods used to analyze these data are frequently changed (so far in 2005, four updates of the human genome have already been released). The analysis of bioinformatics data is very time consuming, for example, InterPro domain [16] calculations take several days to be completed on one CPU, even for a single human chromosome. The clustering of sequences is also computationally expensive and has stimulated the development of specialized methods [17]. The human genome is just one of >400 genomes that are annotated and publicly available (e.g. on MIPS [18]) and their number continues to grow. These apparent computational difficulties and the huge amounts of data strongly encourage cooperation between different research centres, thus boosting the development of web technologies.

The situation in drug design and chemoinformatics is different. The largest commercial database of chemical compounds, the iResearch library (www.chemnavigator.com), comprises no more than 15,000,000 unique SMILES. All these structures can be stored in a 50 MB compressed file. The public database, ZINC [19], offers a smaller set of just a few million unique compounds and these data can be processed in a few hours by, for example, the ALOGPS program [9]. These general databases only have limited applications in the field. Unlike bioinformatics, where the genome sequence and its position on the chromosome provides a lot of information that can be used in many studies, in chemoinformatics the structure of a molecule should be accompanied by measured physical or biological activities. Such measurements are extremely expensive; therefore, the databases for the development of new methods (e.g. physico-chemical properties) are orders of magnitude smaller. One of the largest databases in the field, Physical

feature

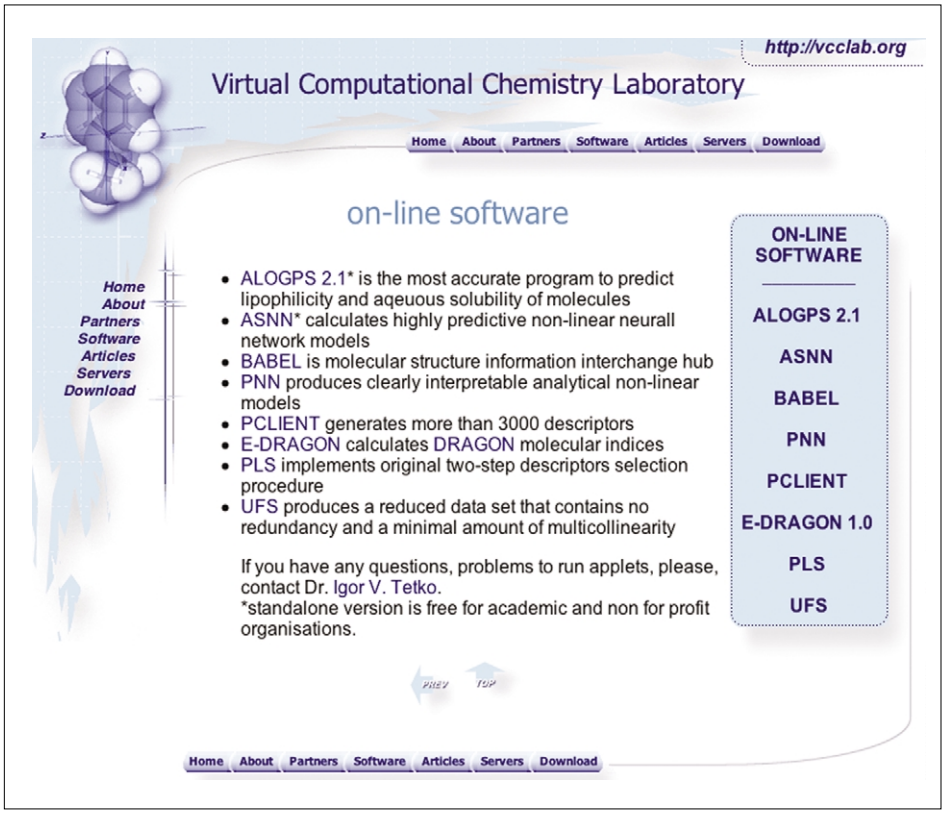

FIGURE 1

Online software tools available at the Virtual Computational Chemistry Laboratory site. ALOGPS applet calculates and compares lipophilicity with aqueous solubility of molecules using six methods including the ALOGPS 2.1 [9,25], CLOGP (www.daylight.com/daycgi/clogp), KOWWIN (www.syrres.com/esc/est_kowdemo.htm), MiLogP (www.molinspiration.com), IA_logP (www.logp.com) and XLOGP [26]. The user can draw the molecule using the JME applet (from Peter Ertl, Novartis) or submit it in a format supported by OpenBabel (http://openbabel.sourceforge.net). Associative neural network method (ASNN) [27,28] calculates models with high prediction ability, by correcting the bias of the neural network ensemble. Polynomial Neural Networks (PNN) [29] calculates analytical nonlinear models between descriptors of molecules and the target activity and it provides a clear interpretation of the detected relations. e-DRAGON and its extension Parameter Client (PCLIENT) calculate more than 1600 and 3000 descriptors per molecule, respectively. The user can either provide optimized molecules or seamlessly convert molecules from 2D to 3D representations using the integrated CORINA program [8]. The unsupervised forward selection (UFS) [7] decreases the number of descriptors and produces a reduced dataset that contains no redundancy and a minimal amount of multicollinearity.

Properties Database (PHYSPROP) (www.syrres.com), comprises just over 25,000 compounds and this is a commercial, not a public, database. The largest datasets of biological properties such as the blood–brain barrier [20] or intestinal absorption [21] contain just a few hundred compounds. Larger collections of these and other ADME–Tox properties exist in industry but they are not available publicly because of privacy issues. Thus, an inadequate amount of data and limited data availability significantly slow down the development of chemoinformatics compared with the development of bioinformatics. Of course, some other differences such as legacy and data complexity, among others, also contribute to the problem [15].

Another important aspect is how much motivation there is to develop such resources. One of the main outcomes of academic activity is the publication of articles but the leading chemical journals do not yet have experience of accepting and publishing articles describing web resources and the number of these publications is limited. At the same time, the ratio of regular articles to application notes (usually describing web applications) published in *Bioinformatics* in the first half of 2005 is approximately 5:3. In this journal, it is common to publish a methodological article that will be followed, a few months later, by an application note. The apparent success of this strategy is illustrated by the impact factor of *Bioinformatics*, it jumped from 3.4 to 6.7 in just two years, according to the Institute for Scientific Information (www.isinet.com).

Is there a light at the end of the tunnel?

There is hope that the situation in the chemoinformatics field will change. The recent PubChem initiative of the National Institutes of Health [22] will house compound information from scientific literature as well as screening and probe data from the Molecular Libraries Screening Centre. This could provide large amounts of high-quality data, including details on the biological and ADME–Tox activity of chemical compounds used for drug design studies. The availability of these data could dramatically change the chemoinformatic field and boost the development of web resources until they rival those available in the field of bioinformatics. Considering the expense of drug failures, in the final stages of development, as a result of unsatisfactory ADME–Tox properties [23] there is an increasing motivation for large pharmaceutical companies to release some of their data to promote the development of new chemoinformatic methods. Because protecting the privacy of molecular structures is of paramount importance for the success of the drug industry, a set of procedures designed to release data, but not the underlying molecule structures, is being actively explored in the field [24]. The attitude towards publishing web services is also changing and one of the leading publications in the field, *Journal of Chemical Information and Modeling*, plans to publish a dedicated issue on web services in 2006 (W. Warr, personal communication).

Conclusions

Given the benefits that have been brought to bioinformatics by web applications, it should be beneficial to encourage the development of these technologies in the chemoinformatics field. Publishing of data and/or methods on the web allows other researchers to avoid duplication, to reuse and to validate the results of previous studies in a new development. The web servers increase awareness about

feature

existing software and can increase citations of articles. The appearance of new protocols and standards for data sharing on the web makes developing new applications easier and more straightforward. The VCCLAB can be used as a prototype to develop such projects. The developed technology allows integration of new third-party applications, which could be made available to the worldwide community.

Acknowledgements

Thanks to Johann Gasteiger, Roberto Todeschini, Peter Ertl, David Livingstone, Vladimir Palyulin, Vsevolod Tanchuk, Alexander Makarenko and members of their teams for their contributions and development of the VCCLAB site and to Louise Riley for her valuable remarks. This work was partially supported by the INTAS 00–0363 project.

References

1 Tetko, I.V. and Livingstone, D.J. Rule-based systems to predict lipophilicity. In *Comprehensive Medicinal Chemistry II:* In silico *Tools in ADMET* (Vol. 5) (Testa, B. and van de Waterbeemd, H) Elsevier (in press)
2 Tetko, I.V. (2003) The WWW as a tool to obtain molecular parameters. *Mini Rev. Med. Chem.* 3, 809–820
3 Van de Waterbeemd, H. and De Groot, M. (2002) Can the Internet help to meet the challenges in ADME and e-ADME? *SAR QSAR Environ. Res.* 13, 391–401
4 Marchand-Geneste, N. and Carpy, A.J. (2004) e-Quantum chemistry free resources. *SAR QSAR Environ. Res.* 15, 43–54
5 Jonsdottir, S.O. *et al.* (2005) Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates. *Bioinformatics* 21, 2145–2160
6 Tetko, I.V. *et al.* (2005) Virtual Computational Chemistry Laboratory design and description. *J. Comput. Aided Mol. Des.* 19, 453–463
7 Whitley, D.C. *et al.* (2000) Unsupervised forward selection: a method for eliminating redundant variables. *J. Chem. Inf. Comput. Sci.* 40, 1160–1168
8 Sadowski, J. *et al.* (1994) Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J. Chem. Inf. Comput. Sci.* 34, 1000–1008
9 Tetko, I.V. and Tanchuk, V.Y. (2002) Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *J. Chem. Inf. Comput. Sci.* 42, 1136–1145
10 Todeschini, R. and Consonni, V. (2000) In *Handbook of Molecular Descriptors*, 667, WILEY-VCH
11 Ertl, P. *et al.* (2003) Web-based cheminformatics and molecular property prediction tools supporting drug design and development at Novartis. *SAR QSAR Environ. Res.* 14, 321–328
12 Pastor, M. *et al.* (2002) Distant collaboration in drug discovery: the LINK3D project. *J. Comput. Aided Mol. Des.* 16, 809–818
13 Murray-Rust, P. and Rzepa, H.S. (1999) Chemical markup language and XML Part I. Basic principles. *J. Chem. Inf. Comp. Sci.* 39, 928–942
14 Murray-Rust, P. *et al.* (2004) Representation and use of chemistry in the global electronic age. *Org. Biomol. Chem.* 2, 3192–3203
15 Curcin, V. *et al.* (2005) Web services in the life sciences. *Drug Discov. Today* 10, 865–871
16 Mulder, N.J. *et al.* (2005) InterPro, progress and status in 2005. *Nucleic Acids Res.* 33, 201–205
17 Tetko, I.V. *et al.* (2005) Super paramagnetic clustering of protein sequences. *BMC Bioinformatics* 6, 82
18 Riley, M.L. *et al.* (2005) The PEDANT genome database in 2005. *Nucleic Acids Res.* 33, 308–310
19 Irwin, J.J. and Shoichet, B.K. (2005) ZINC–a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182
20 Platts, J.A. *et al.* (2001) Correlation and prediction of a large blood-brain distribution data set–an LFER study. *Eur. J. Med. Chem.* 36, 719–730
21 Zhao, Y.H. *et al.* (2001) Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J. Pharm. Sci.* 90, 749–784
22 Morrissey, S.R. (2005) NIH initiatives target chemistry. *Chem. Eng. News* 83, 23–24
23 Landers, P. (2003) Cost of developing a new drug increases to about $1.7 billion. In *The Wall Street Journal*
24 Wilson, E.K. (2005) Is safe exchange of data possible? *Chem. Eng. News* 83, 24–29
25 Tetko, I.V. *et al.* (2001) Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. *J. Chem. Inf. Comput. Sci.* 41, 1407–1421
26 Wang, R. *et al.* (2000) Calculating partition coefficient by atom-additive method. *Perspect. Drug Discov. Des.* 19, 47–66
27 Tetko, I.V. (2002) Neural network studies. 4. Introduction to associative neural networks. *J. Chem. Inf. Comput. Sci.* 42, 717–728
28 Tetko, I.V. (2002) Associative neural network. *Neur. Proc. Lett.* 16, 187–199
29 Aksyonova, T.I. *et al.* (2003) Robust polynomial neural networks in quantitative-structure activity relationship studies. *SAMS* 43, 1331–1339

Igor V. Tetko
Institute for Bioinformatics,
GSF, Forschungszentrum fuer Umwelt und Gesundheit,
Ingolstaedter Landstrasse 1,
D-85764 Neuherberg,
Germany

Institute of Bioorganic and Petrochemistry,
Murmanskaya 1,
UA-02094,
Kyiv,
Ukraine

Novel treatment options for infectious exacerbations

Fred Del Vecchio, Lamine Mbow and Robert T. Sarisky*, *rsarisky@cntus.jnj.com

In acute exacerbations and the onset of asthma, considerable data exist that implicate the role of infectious agents, particularly atypical bacteria, *Chlamydia pneumoniae* and *Mycoplasma pneumonia*, and the common cold viruses, respiratory syncytial virus and rhinovirus. These agents have been identified in the airways of stable asthmatics and are presumed to contribute to chronic lower-respiratory inflammatory disease. It is current practice for treating acute exacerbations of asthma, although perhaps inappropriate albeit typical, to prescribe antibiotics. This occurs despite the fact that respiratory viruses represent the primary trigger of an exacerbation. However, consistent with this approach, macrolides have been shown to confer immune modulatory effects beneficial to those suffering from chronic pulmonary inflammatory syndromes. Using macrolide concentrations below the minimal inhibitory concentration can modulate expression of virulence factors, which could prevent establishment or expansion of an infection. Herein, we detail characteristics shared between two common therapeutic approaches, macrolide antibiotic therapy and systemic corticosteroids, in an attempt to propose an alternative treatment paradigm.

Macrolides: empirical therapy

Weinberger [1] recently reviewed treatment strategies for respiratory infections and asthma, highlighting the typical practice of prescribing antibiotics for acute exacerbations, although such events are often elicited by respiratory viruses. In addition to macrolide-directed inhibition of bacterial protein synthesis via binding to the 50S ribosomal subunit [2], documented reports of improved

1359-6446/05/$ – see front matter PII: S1359-6446(05)03597-X