MOLECULAR PHYSICS, 20 FEBRUARY 2004, VOL. 102, NO. 4, 353–360

Taylor & Francis
Taylor & Francis Group

Computational Science and Engineering On-line: an integrated web-based environment for multi-scale modelling of complex reaction systems

THANH N. TRUONG*, TOM COOK, MANOHAR NAYAK, CHAIWOOT BOONYASIRIWAT, LE-THUY T. TRAN and SHAOWEN ZHANG

Henry Eyring Center for Theoretical Chemistry, Department of Chemistry, University of Utah, 315 S 1400 E, Rm 2020, Salt Lake City, Utah 84112, USA

(*Received 6 July 2003; accepted 19 November 2003*)

We present a development of an integrated extendable web-based environment called Computational Science and Engineering On-line (CSEO) to include different fields of computational science. Our initial efforts are focusing on an integrated environment for multi-scale modelling of complex reacting systems from fundamental quantum chemistry with different entry points. CSEO provides an information management system that allows data flow from one application to another in a transparent manner. In addition, it provides a set of web-based graphic-user interfaces (GUIs) to different scientific applications. Current available GUIs are for quantum chemistry, thermodynamics and kinetics. Work is in progress to allow CSEO accessing resources from the computing grids using the Globus technology. CSEO can be accessed at http://cseo.net. It can also be hosted at different mirror sites.

1. Introduction

Advances in computer technology and electronic communication in the last decade, especially the world-wide web (web), have changed the face of modern science by providing a ubiquitous environment for interacting with colleagues, for querying, retrieving, sharing and analysing data, for accessing forefront research tools and computational resources regardless of geographical location and personal schedule and thus has provided a mechanism for scientists to become more productive. The web has become the new medium of choice, when the opportunity presents itself, for remote data acquisition and analyses. The acquisition and analysis of data through web-accessible servers has provided access to unique remote research tools in a seemingly transparent way. Web-based 'transparent supercomputing' can provide a mechanism by which researchers gain access to computer systems that exceed their local computational resources.

In realization of such advances, the last several years have seen tremendous efforts in the creation of 'collaboratories', laboratories without walls, in which researchers can take advantage of the fusion between computer and information technology to expand their research capabilities, to perform research, and to collaborate in solving complex scientific problems [1–4]. These collaboratories can be classified into two types: data sharing oriented or remote access scientific instrument oriented. The Worm Community System is an example of data-driven collaboratories that provides a central repository for everything known about the soil-residing worm nematode *C. elegans* ranging from genome to behaviour level and is accessible to everyone in the community who contributes to the system. Another example is the Research Collaboratory for Structural Bioinformatics, which provides access to databases of biological structures and tools for determining and analysing these structures and the under-construction DOE Collaboratory for Multiscale Chemical Science, which focuses on a data-driven knowledge management system for chemical properties that facilitate multi-scale modelling. Most existing collaboratories are instrument driven to provide the capability for remote real-time data acquisition, and are often focused specifically on a particular complex scientific problem. The Space Physics and Aeronomy Research Collaboratory (SPARC) provides an internet-based collaborative environment for studies of space and upper atmospheric science facilitating real-time data acquisition from a remote site in Greenland. The Materials Microcharacterization Collaboratory provides remote access to facilities that

*Author for correspondence. e-mail: truong@chem.utah.edu

Molecular Physics ISSN 0026–8976 print/ISSN 1362–3028 online
http://www.tandf.co.uk/journals
DOI: 10.1080/00268970310001649392

perform electron-beam microcharacterization of materials. The Diesel Combustion Collaboratory provides a successful electronic collaboration between industries and government laboratories to design better diesel engines. The Environmental Molecular Science Collaboratory allows remote instrument control of the NMR and FTICR spectrometers located at the Pacific Northwest National Laboratory and access to scientific data and modelling software tools for studies of environmental problems at the molecular level. The focus of these collaboratories currently is on: (1) expanding research capabilities far beyond the limitation of the local facilities available to the researcher; (2) facilitating an environment for collaboration regardless of geographical location of the researchers; and (3) solving specific major complex scientific problems. The establishment of the above collaboratories, laboratories without walls, has undisputed potential for making significant impacts on science and technology in the 21st century. The benefits are undoubtedly substantial but are limited to small numbers of researchers who are either directly involved with the projects or are active participants in that focused scientific area.

In this study, we describe our current efforts in developing an integrated web-based research environment for multi-scale computational modelling of complex reacting systems. The goal is to bring the collaboratory technology to the greater scientific community of computational science and engineering while gradually introducing the psycho-social changes to facilitate the paradigm shift in scientific research. The new environment is called Computational Science and Engineering On-line (CSEO).

2. Development of CSEO

CSEO provides a Web-based grid-computing environment in which a researcher from a university, an industry, or a national lab can perform research using a variety of state-of-the-art scientific application tools, access and analyse information from public databases and personal electronic notebook, share and discuss results with colleagues, access computational resources on the computing grids that are far beyond those available locally, and master subjects in other areas of computational science and engineering without regard to geographical location and schedule (see figure 1).

CSEO provides an extensible web-computing environment that allows a seamless interface and data flow between different areas of computational science and engineering and thus allows collaborations on multi-scale modelling of complex scientific problems. There are no limitations on what aspects of computational science and engineering can be incorporated into CSEO.

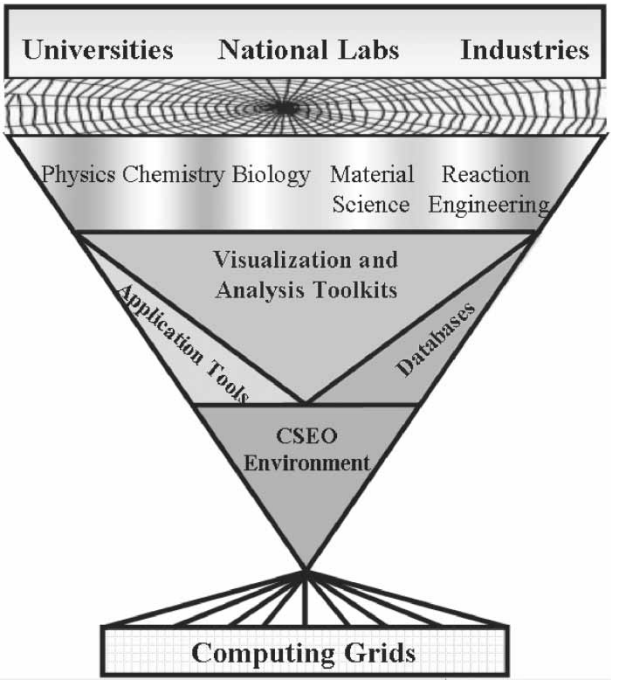

Figure 1. Schematic illustration of the Computational Science and Engineering On-line (CSEO).

Our initial effort is on providing a workflow environment to bridge fundamental chemistry and reaction engineering.

CSEO is not meant to be a central web portal but rather a world-wide extensive network of many mirror sites, hosted by different universities, computer centres, national laboratories, even industries, that share their public databases via a secure network. This will maximize resource utilization, data generation and sharing. We will use the case scenario below to illustrate how CSEO can be used to model complex reacting systems and how data flow among scientific applications that have different time and length scales.

Case scenario: A chemical engineering researcher from a chemical company needs to develop a mechanism for an important chemical process that includes several specific toxic by-products that he would like to monitor. He can use existing kinetic models but needs to add specific reactions that allow for destruction and creation of these toxic species or use an automated mechanism generator. To complete the mechanism, he needs to supply both thermodynamic and kinetic properties of the newly added reactions. He can look into available public databases or his proprietary company databases for such information. In case such information does not exist, he needs to either estimate them from an empirical method or from first-principles quantum mechanical methods. He can collaborate with a com-

putational chemist who will perform first-principles calculations to predict such properties. If he elects to use the first-principles approach, he needs to build the 3D molecular structures of the species as well as estimates of transition state structures for the reactions, then submit them to an available quantum chemistry package for calculating potential energy surface information. Specific potential information is then collected and submitted to different software packages for determination of thermodynamic and kinetic properties. At this point, he can elect to deposit the molecular and reaction properties just calculated to his personal electronic notebook, or his company databases for use within his company, or in the public database to share with others. Once the kinetic model is completed, he needs to analyse the mechanism and reduce it to a more manageable size for reaction engineering simulations. Subsequently, he can perform a simulation of a plug-flow reactor model using, for example, the Chemkin package and then analyse the concentration profiles of the toxic species at a particular reaction condition.

The above scenario describes a complex scientific problem that requires an integrated multi-disciplinary approach. The flowchart (figure 2) illustrates how data flows from one scientific application or one area of science to the others, starting from first-principles quantum chemistry to reactor modelling. The information management system provides the backbone for extracting valuable data from one application to be stored or to be input parameters for another application. Below are the descriptions of individual components.

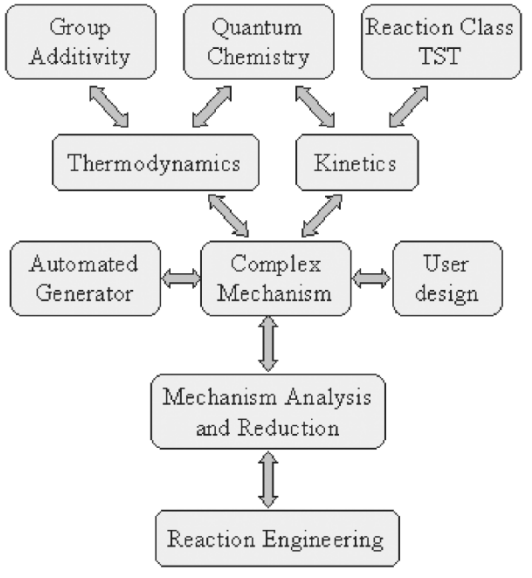

Figure 2. Schematic flowchart to show how data can flow from different scientific applications.

2.1. *Quantum chemistry*

First-principles (*ab initio*) quantum mechanics at the molecular level offers enormously powerful techniques for predicting molecular properties such as structures, dipole moments, polarizabilities, ionization potentials, electron affinities, etc., following the reaction paths of chemical transformations, predicting electrostatic effects in a variety of environments, estimating pK_a shifts, assisting interpretations of spectroscopic probes of molecular environments, and provides potential energy surface information that can be used to predict thermodynamic properties of molecular species or kinetics of chemical reactions. There are a number of available quantum chemistry packages, each with different license restrictions. Each of these packages has its unique strength. CSEO provides a universal web-based graphic-user interface (GUI) (e.g. see figure 3) that allows building molecular species from atoms and fragments, provides a means to generate or convert input files for these packages, and allows job submission to any package that is available at the host site.

The results will then be parsed into a common data format for storage, visualization and analysis. By combining the capabilities of the available packages, CSEO will provide users with a unique and comprehensive research and educational environment for quantum chemistry. Such an environment not only increases the productivity level but also improves the quality of research and education. Currently, CSEO has an interface with the Gaussian program since it is widely used and is available at most research institutions. Incorporation of other programs such as GAMESS, NWChem and Molpro will follow.

2.2. *Thermodynamics*

CSEO inherited the capability for calculating thermodynamic properties, namely enthalpy, entropy and heat capacity of molecular species from VKLab. The main tool is for direct calculation based on standard statistical thermodynamic formalism by using the potential energy information calculated from first-principles quantum chemistry discussed above (e.g. see figure 4).

A much quicker but less accurate method is to use the group additivity approach [5]. However, use of the group additivity approach requires the recognition of chemical groups in a given molecular species. Such a feature will be made available to CSEO in the future.

2.3. *Kinetics*

A unique feature of CSEO is the on-line availability and accessibility of a rather complete range of computational methods for predicting rate constants of elementary reactions from first principles (e.g. see figure 5). In particular, these methods include the simple transition

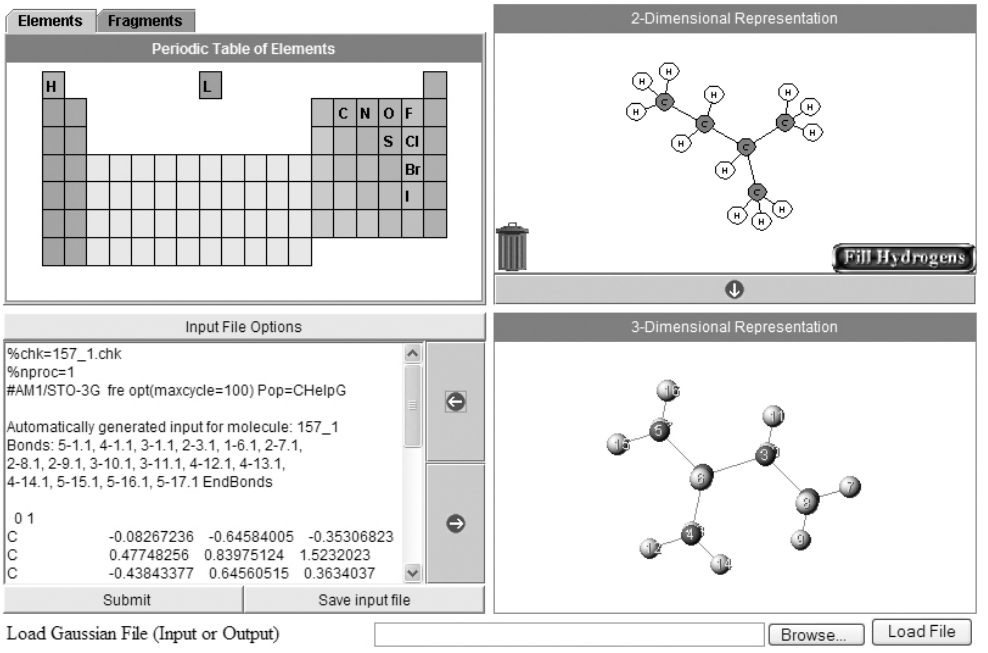

Figure 3. A snapshot of the GUI for the quantum chemistry programs.

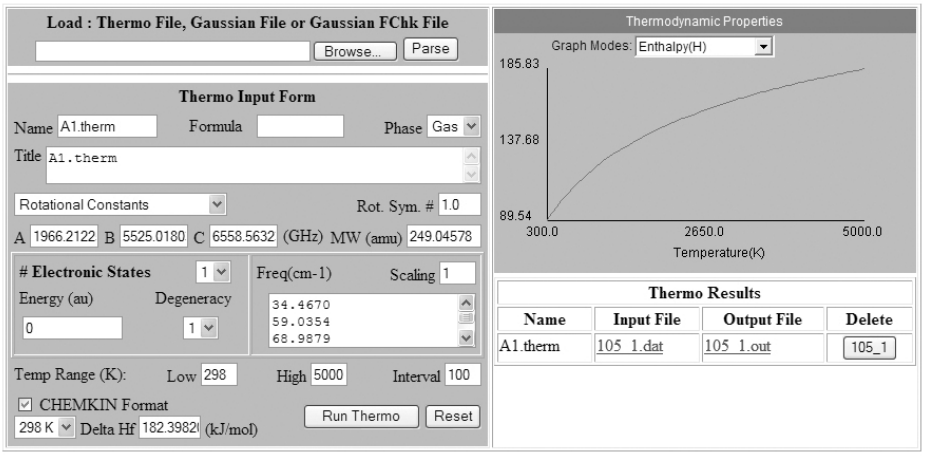

Figure 4. A snapshot of the GUI for the thermo program for calculating thermodynamic properties of molecular species.

state theory (TST) [6], canonical variational TST (CVT) [7, 8], microcanonical TST (also known as QRRK or RRKM) [9–11], microcanonical variational TST (mVT) [12–14] for gas-phase elementary reactions. These methods can be augmented with various methodologies for including quantum mechanical tunnelling effects, namely, the one-dimensional Wigner [15] and Eckart [16] models within the TST framework, and the multi-dimensional semiclassical zero-curvature and small-curvature tunnelling methods [17] within the CVT framework.

The master equation method [18–20] for pressure-dependent rate constants and dynamics of multi-channel reactions will also be made available in the near future. These methods require potential energy surface information, which can be extracted from *ab initio* quantum chemistry calculations and stored in the information management system. It is important to point out that the accuracy of the thermodynamic and kinetic properties depends strongly on the quality of the input potential energy surface information. There exists a variety of quantum chemistry methods providing different levels

Figure 5. A snapshot of the GUI for the kinetic program for calculating rate constants.

of accuracy and computational costs. Thus, care must be taken in selecting the appropriate level quantum chemistry method to ensure accurate thermodynamic and kinetic properties. In addition to those well-established methodologies above, the new reaction class TST (RC-TST) method coupled with the linear energy relationship [21–25] provides a cost-effective approach for estimating thermal rate constants from first principles using only the reaction energy information.

2.4. *Complex reaction mechanism*

The key bridge that connects first-principles quantum chemistry and reaction engineering is the computational tool for automatically generating detailed mechanisms of complex reaction systems such as combustion of hydrocarbon fuels or atmospheric chemistry which can consist of the order of hundreds of thousands of reactions. We have recently developed a new automatic complex mechanism generator based on the chemical graph theory (called COMGEN). Currently, COMGEN can generate a complete mechanism based on all reaction types reported in the literature for combustion of hydrocarbons [26]. Molecular species are represented externally by 1D SMILE string notation [27]. There are several challenges remaining to be addressed. One is the combinatorial explosion problem, i.e. the number of the generated reactions and intermediates can increase rapidly. We can avoid such problems by using a rate-based approach similar to that proposed by Broadbelt and co-workers [28] to eliminate unimportant reactions and intermediate species. Doing so, however, can run the risk of making the mechanism depend on the initial conditions of the reacting systems. Alternatively, we propose to develop a generic algorithm approach to generate all possible reactions with physical rules for making offspring (new species). Another challenge in bridging this stage with first-principle quantum chemistry is the ability to convert 1D SMILE notation to 3D structure and vice versa. Progress in these directions is being made.

2.5. *Mechanism analysis and reduction*

A reaction mechanism generator as described above can construct a rather ‘complete’ mechanism, which may consist of a rather large number of reactions and stable intermediates. Embedding such mechanisms into computational fluid dynamics simulations of real combustion devices in the reactive flow conditions presents many computational challenges. To circumvent such challenges, some approximations must be made to reduce the size of the mechanism to the computationally feasible range. Mechanism reduction is an area of intensive research and much progress has already been made [29–33]. As the first step, we provide an interface to the SENKIN program for sensitivity analysis [34].

2.6. *Reaction engineering*

To add the reaction-engineering component to CSEO, we will first develop an interface to the CHEMKIN package for fluid dynamics simulations of complex reaction systems. In particular, CHEMKIN provides a rather comprehensive tool for modelling different types of reactors such as steady or transient reactor networks; laminar boundary-layer flow; opposed-flow flames; well and partially stirred reactors; plug-flow reactors; steady, laminar, one-dimensional premixed flames; one-dimensional rotating-disc reactors for chemical vapour deposition, etc. This will allow direct validation of computational approaches by comparisons with macroscopic observables. More importantly, CSEO allows researchers in chemical engineering to take advantage of the integrated environment to incorporate fundamental chemistry into their reaction engineering modelling.

3. CSEO software engineering infrastructure

CSEO infrastructure consists of three main components: (1) visualization tools and GUI to scientific applications; (2) information management system; (3) access, security and submission.

A key part of our technology is to enhance the user's experience by using Java Applets connected to Java Servlets. As with most on-line applications, interfaces to databases are implemented by connecting a hierarchy of web pages in which the user navigates to do research. For CSEO, the philosophy is that the user should stay on one web page while relevant information and resources should come to her. In order to do this, a combination of Java Applets on the client's machine connected to Java Servlets on the CSEO server has been created that make very flexible user interfaces. This allows not only a more simple interface since there is no navigating, but it also allows a more interactive experience than traditional HTML (since Java is a full-fledged programming language, unlike HTML and DHTML). Thus, visualizations, high-level mathematics, highly complicated user interfaces, etc. all become available. In this way, the user's experience is less like navigating the web, and more like operating a normal desktop application.

3.1. *Visualization tools and GUI to scientific applications*

Visualization is perhaps the most effective means to communicate scientific data to people. Thus, visualization tools and GUIs to scientific applications are a crucial component of CSEO. Due to the diverse data types and the vast amount of data that will be available for mining in CSEO, the demand for visualization and analysis tools is far greater than a simple graphic user interface (GUI) for a single application tool. In addition, the requirement that such tools provide real-time interactive performance over the internet adds a new challenge to the development. So despite the existence of GUIs for some of the application tools that we plan to incorporate into CSEO, and besides potential license agreement restrictions, these GUIs were typically *not* designed for use in a web browser as they often lack platform independence and therefore cannot be easily ported to CSEO. We are developing a new class of general visualization and analysis tools written mostly in Java to support such a multi-disciplinary, broad-based, web-computing environment as CSEO (e.g. see figure 6).

In developing such tools, we have to balance between the resolution of the graphic and the real-time interactive performance over the network. Our philosophy is to focus on the user's psychological experience by presenting data efficiently with minimal processor time.

3.2. *Information management system*

It provides the key avenue for data flow within the whole CSEO environment, i.e. data can be stored, retrieved, shared and transferred in a transparent manner. To achieve this, we are developing a multi-level database system. The multi-level database system allows storage, retrieval and transfer of data at the personal, group and public levels. At the present time, database technology is dominated by relational databases and SQL. Although easy to use in their simplest phases, as their size and complexity grow, relational databases become too complicated for non-professionals to use. Since most scientists have highly sophisticated and specialized database needs, the use of simple, pre-programmed relational databases is limited. In lieu of this, we are developing a new type of database that is designed to have the power and efficiency of a relational database, but with the ease of construction of XML [35]. In the initial version of this database, the ability to do sophisticated queries and to ship large quantities of heterogeneous data over the web quickly was excellent. We are improving the speed of these queries and also interconnecting databases to allow a larger range of queries. With this system, any non-professional can make his own databases and still have a great deal of query ability (perhaps equivalent to SQL) but require less training.

3.3. *Access, security and submission*

Due to limiting resources, we are able to address this component only at a very simple level at the present time. In particular, we have implemented the Linux htaccess technology to limit access to the CSEO site. However, since CSEO will have access to many resources or information that are protected by different levels of proprietary and/or copyright concerns, we plan

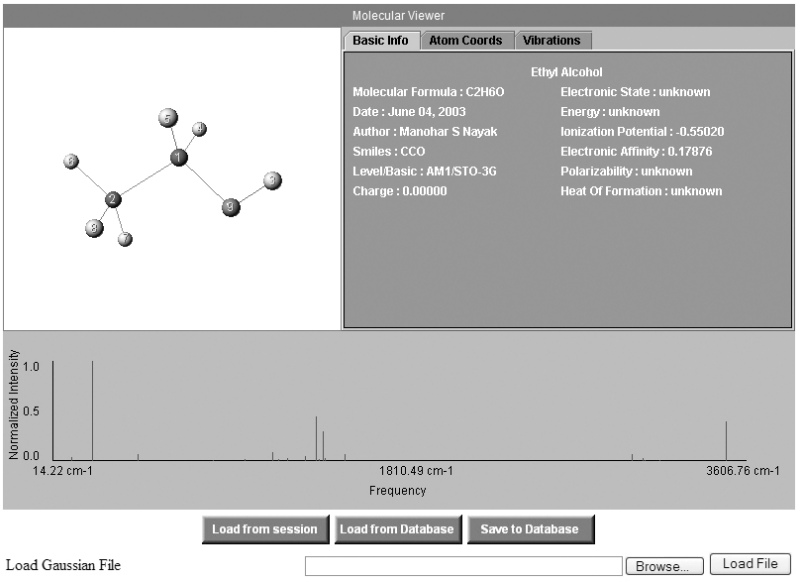

Figure 6. A snapshot of the GUI for accessing the molecular property databases and for viewing results from quantum chemistry calculations.

to implement components from the NSF middleware initiative (http://www.nsf-middleware.org/) and the DOE SciDAC collaboratory pilot projects, for the authentication, authorization and resource control mechanisms necessary to make CSEO a truly production-quality environment.

The second important issue in this component is the allocation of computer resources like storage and CPU. While many of the functions that will be supported at the CSEO require a modest amount of computer resources, calculations using realistic models typically use more than trivial resources. We do not expect that all CSEO mirror sites will be able to provide these resources to its potential users. CSEO users can secure computational resources at a remote site needed for large simulations. We are implementing in CSEO tools using Globus GRID technologies [36] to allow execution of complex simulations in a remote system.

4. Distribution

The first version of CSEO with interfaces to quantum chemistry, thermodynamics and kinetics was released on 10 June 2003. There are two modes for accessing the tools available in CSEO. One is to log into the CSEO website at http://cseo.net from a web-browser and the other is to host a mirror site. Note that for interfacing with quantum chemistry only the semi-empirical method is allowed in MolBuild at the CSEO central website for evaluation purposes due to the limiting computing resource of the server and to the license restrictions of the Gaussian98 program. To access full features of MolBuild one should host a CSEO mirror site. Interested parties should contact Professor Truong for more information. The next major release is planned for 1 January 2004 with new tools and interfaces to reaction engineering applications.

5. Conclusion

We present our efforts in developing an integrated web-based environment called CSEO for multi-scale modelling and simulations of complex reacting systems. This environment allows data flow from the most basic fundamental level of quantum chemistry to be used in flame simulations of combustion systems. The development utilizes advances in web programming such as Java Applets and Servlets with new development in databases to create a user-friendly environment for both research and education. There are a lot of tasks remaining to be done. However, its current release already provides useful tools for interfacing with quantum chemistry, thermodynamics and fundamental kinetics. More importantly, we would like to have feedback from the chemistry and chemical engineering communities to shape CSEO's future development.

This work is supported in parts by the National Science Foundation through the Information Technology Research Initiative and by a generous donation from the Dow Chemical Company. We are also grateful to the University of Utah Center for High Performance Computing for their help and technical support.

References

[1] North Carolina Board of Science and Technology, and National Research Council, 1999, *Collaboratories: Improving Research Capabilities in Chemical and Biomedical Sciences: Proceedings of a Multi-site Electronic Workshop* (Washington, DC: National Academy Press).

[2] Kouzes, R. T., Myers, J. D., and Wulf, W. A., 1996, *IEEE Comput.*, **29**.

[3] Kiernan, V., March 1999, Internet-Based 'Collaboratories' Help Scientists Work Together, *The Chronicle of Higher Education.*

[4] Agarwal, D. A., Sachs, S. R., and Johnston, W. E., 1998, *Comput. Phys. Commun.*, **110,** 134.

[5] Cohen, N., and Benson, S. W., 1993, *Chem. Rev.*, **93,** 2419.

[6] Glasstone, S., Laidler, K. J., and Eyring, H., 1941, *The Theory of Rate Processes* (New York: McGraw-Hill).

[7] Truhlar, D. G., and Garrett, B. C., 1984, *Annu. Rev. phys. Chem.*, **35,** 159.

[8] Truhlar, D. G., Isaacson, A. D., and Garrett, B. C., 1985, *Generalized Transition State Theory. In Theory of Chemical Reaction Dynamics*, edited by M. Baer (Boca Raton, FL: CRC Press), p. 65.

[9] Rice, O. K., and Ramsperger, H. C., 1927, *J. Am. chem. Soc.*, **49,** 1617.

[10] Kassel, L. S., 1932, *Kinetics of Homogeneous Gas Reactions* (New York: van Nostrand Rheinhold).

[11] Marcus, R. A., 1952, *J. chem. Phys.*, **20,** 359.

[12] Truhlar, D. G., Garrett, B. C., and Klippenstein, S. J., 1996, *J. phys. Chem.*, **100,** 12771.

[13] Hase, W., 1998, *Acc. Chem. Res.*, **31,** 659.

[14] Garett, B. C., 2000, *Theor. Chem. Acc.*, **103,** 200.

[15] Wigner, E., 1937, *J. chem. Phys.*, **5,** 720.

[16] Truong, T. N., and Truhlar, D. G., 1990, *J. chem. Phys.*, **93,** 1761.

[17] Truhlar, D. G., Isaacson, A. D., Skodje, R. T., and Garrett, B. C., 1982, *J. phys. Chem.*, **86,** 2252.

[18] Troe, J., 1977, *J. chem. Phys.*, **66,** 4745.

[19] Gilbert, R. G., Luther, K., and Troe, J., 1983, *Ber. Bunsenges. Phys. Chem.*, **87,** 169.

[20] Robertson, S. H., Pilling, M. J., Gates, K. E., and Smith, S. C., 1997, *J. comput. Chem.*, **18,** 1004.

[21] Truong, T. N., Duncan, W. T., and Tirtowidjojo, M., 1999, *Phys. Chem. chem. Phys.*, **1,** 1061.

[22] Truong, T. N., and Truong, T.-T. T., 1999, *Chem. Phys. Lett.*, **314,** 529.

[23] Truong, T. N., 2000, *J. chem. Phys.*, **113,** 4957.

[24] Truong, T. N., Maity, D. K., and Truong, T.-T. T., 2000, *J. chem. Phys.*, **112,** 24.

[25] Zhang, S., and Truong, T. N., 2003, *J. phys. Chem.* A, **107,** 1138.

[26] Ratkiewicz, A., and Truong, T. N., 2003, *J. chem. Inf. comput. Sci.*, **43,** 36.

[27] Weinenger, D., 1988, *J. chem. Inf. comput. Sci.*, **28,** 31.

[28] Susnow, R. G., Dean, A. M., Green, W. H., Peczak, P., and Broadbelt, L. J., 1997, *J. phys. Chem.* A, **101,** 3731.

[29] Tomlin, A. S., Turanyi, T., and Pilling, M. J., 1997, *Comprehensive Chemical Kinetics: Low-Temperature Combustion and Autoignition*, edited by M. J. Pilling (Amsterdam: Elsevier), pp. 293–424.

[30] Maas, U., and Pope, S. B., 1992, *Combust. Flame*, **88,** 239.

[31] Lam, S. H., and Goussis, D. A., 1994, *Int. J. chem. Kinet.*, **26,** 461.

[32] Keck, J. C., 1990, *Prog. energy combust. Sci.*, **16,** 125.

[33] Fournet, R., Warth, V., Glaude, P. A., Battin-Leclerc, F., Scacchi, G., and Come, G. M., 2000, *Int. J. chem. Kinet.*, **32,** 36.

[34] Koda, M., McRae, G. J., and Seinfeld, J. H., 1979, *Int. J. chem. Kinet.*, **11,** 427.

[35] Fallside, D. C., 2001, *XML Schema Part 0: Primer, W3C Recommendation* (http://www.w3.org/TR/xmlschema-O).

[36] Foster, I., and Kesselman, C., 1997, *Int. J. supercomput. Appl.*, **11,** 115.