An Integrated Web-based Grid-Computing Environment for Research and Education in Computational Science and Engineering

Thanh N. Truong
Henry Eyring Center for Theoretical Chemistry
Department of Chemistry
University of Utah
315 S 1400 E, Rm 2020
Salt Lake City, Utah 84112
Phone: (801) 581-4301
Email: Truong@chemistry.utah.edu

Abstract

We present the development of an integrated extendable web-based simulation environment called Computational Science and Engineering On-line (CSEO) that allows computational scientists to perform research using state-of-the-art tools, querying data from personal or public databases, discuss results with colleagues, and access resources beyond those available locally from a web browser. It is a user-driven web-based application tool. Currently, CSEO provides an integrated environment for multi-scale modeling of complex reacting systems. A unique feature of CSEO is in its framework that allows data to flow from one application to another in a transparent manner. A particular example is demonstrated to show how results from fundamental quantum chemistry simulations are used to calculate thermodynamic and kinetic properties of a chemical reaction, which subsequently are used in the simulation of a combustion reactor. Advantages, disadvantages, and future prospects of a web-based simulation approach are then discussed. CSEO can be accessed at http://cseo.net.

1. INTRODUCTION

The World Wide Web (often referred as the web) has been revolutionizing the way we communicate since its invention in the last decade. E-mail is becoming an alternative official form of written communication and chat is an alternative to a telephone conversation. Virtual universities allow students to obtain education without the restrictions of a course schedule and geophysical location. Scientists are also looking into web technology to help revolutionizing the way science is being conducted and taught. The last several years have seen tremendous efforts in the creation of 'collaboratories'. These are laboratories without walls, in which researchers can take advantage of web technology to expand their research capabilities 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 Research Collaboratory for Structural Bioinformatics is an excellent example of the data-oriented collaboratory that provides access to databases of biological structures and tools for determining and analyzing these structures. Most existing collaboratories are instrument driven and provide the capability for real-time data acquisition from remote research instruments through web-accessible servers in a seemingly transparent way 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 acquisitions from a remote site in Greenland. The Materials Microcharacterization Collaboratory provides remote access to facilities that perform electron-beam microcharacterization of materials. The Environmental Molecular Science Collaboratory allows remote instrument control of the NMR and FTICR spectrometers located at the Pacific Northwest National Laboratory. The focus of these collaboratories currently are 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 these collaboratories has undisputed

1080-241X/04 $20.00 © 2004 IEEE

potential for making significant impacts on science and technology in the 21st century.

The possibility of using web technology to provide a new framework for simulation has attracted a lot of interest in the mid-1990s. It is indicated by the growing size of the conference dedicated to web-based simulation from a whole session in the 1996 Winter Simulation Conference[5] to the first International Conference on Web-based Modeling and Simulation in 1998.[6,7] Numerous reports illustrate the potential of web-based framework for simulation.[5-9] However, realization of such potential has been rather limited and the interest in this area has significantly dropped since 2000. Kuljus and Paul[8] in 2001 gave a rather accurate appraisal of the field and concluded

> "We cannot but observe that most of the work is a playground for the idle for whom it is most important what tools are being used (e.g. Java and JavaBeans) than how the new medium can enhance our use of simulation as a modeling vehicle. This is partially due to the fact that there are no real applications and no real users who are pushing for a more adventurous approach. As it is, most applications are 'invented' and answer the question 'What can I put on the web?' rather than 'This simulation has to be available over the web – how can I design it to facilitate the needs of its users'. Applications that are not user driven rarely progress far or have an impact on society and this is the sad story of web-based simulation so far."

One can push this observation one step further to provide the main reason for the current status of web-based simulation. That is the lack of domain knowledge for those who have the expertise in web technology to create an actual web-based application simulation tool for real users while those who have the domain knowledge often do not have the necessary skills to do so. Furthermore, web-based simulation approach would drastically change the way research is being conducted and thus would face resistance due to the psycho-social changes it entails. In this study, we describe our current efforts in developing an integrated web-based simulation environment for multi-scale computational modeling of chemical and biological systems. The new environment is called Computational Science and Engineering Online (CSEO). The goal is to create a web-based simulation environment that would benefit the greater simulation community of computational science and engineering while gradually introduce the psycho-social changes to facilitate the paradigm shift in scientific research. To do so, we would focus our attention on how research is currently being done and then design the environment that would enhance current research capabilities while not significantly altering the research culture.

2. DEVELOPMENT OF CSEO

The main goal of CSEO is to provide a Web-based grid-computing environment in which a researcher can perform the following functions:

1. research using a variety of state-of-the-art scientific application tools
2. access and analyze information from databases and electronic notebooks
3. share and discuss results with colleagues or teach a class
4. access computational resources from the computing grids that are far beyond those available locally
5. master subjects in other areas of computational science and engineering

without regard to geographical location and schedule. In designing CSEO, our vision is for it not to be a central web portal but rather a world-wide extensive network of many mirror sites, hosted by different universities, computer centers, national laboratories, even industries, that share their public databases via a secure network. This will maximize resource utilization, data generation and sharing.

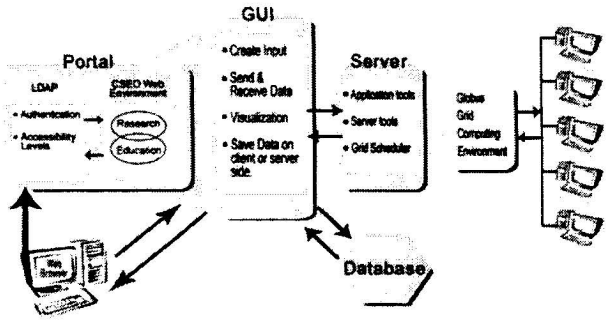


Figure 1. Flowchart showing different components of the CSEO enviroment and their functionalities

Figure 1 illustrates how a user from a web browser can login to the CSEO web portal and use its functionalities. The login initiates the authentication process, which also determines the accessibility level of the user since CSEO supports commercial application tools that have different license restrictions. Each application tool has a graphic-user-interface (GUI) to create input, submit a job to the server or the computing grid, receive and visualize data, and save data either to the database on the server or to the hard-disk on the client computer. Application tools consist of both commercial and open-source legacy codes residing on the server. A unique feature of CSEO is the seamless interface that allows data flow from one application to another and thus facilitate multi-disciplinary research. We will demonstrate this feature by the illustrative case scenario below to show how results

from quantum chemistry simulations are used in simulations of a combustion reactor. It is interesting to point out that CSEO is the first application environment that allows multi-scale simulations of complex reaction systems which require bridging length scale from 10^{-8} to 10^{0} meter and time scale from 10^{-15} to 10^{2} second.

Illustrative case scenario:

A chemical engineering researcher needs to develop a mechanism to monitor several specific toxic by-products from an important chemical process. He realizes that he needs to add to an existing kinetic model for the process specific reactions that allow for destruction and creation of these toxic species and then to supply both thermodynamic and kinetic properties of the newly added reactions. To provide understanding on the importance of these reactions, he needs to perform a sensitivity analysis. Finally, before using the model in simulations of an actual reactor, he validates it by performing premixed reactor simulations to compare with experimental data.

The above scenario describes a complex scientific problem that requires an integrated multi-disciplinary approach. Currently, there has not been any tool to date that can model the above scenario. In fact in order to carry this out, the researcher has to perform the following tasks.

1. Perform quantum chemistry simulations using a commercial legacy program for individual molecular species, and the transition states for the considered reactions.
2. Manually parse the results of quantum chemistry runs and create input files that have different formats for calculating thermodynamic properties of molecular species and rate constants of selected reactions using two different application tools.
3. Manually edit the mechanism files to insert new information in a specific fixed format then run the interpreter, another legacy program that is part of the commercial Chemkin package, to ensure that the modified mechanism have the correct format.
4. Manually create input files for the Senkin and Premix codes for sensitivity analysis and simulations of premixed flames, respectively, then execute these programs.

To complete these tasks, the researcher is required to be quite familiarized with the specific input format for at least five different legacy codes and would have spent a significant amount of time to just parse data and prepare input files. CSEO provides an integrated environment where application tools are naturally connected according to their input and output information.

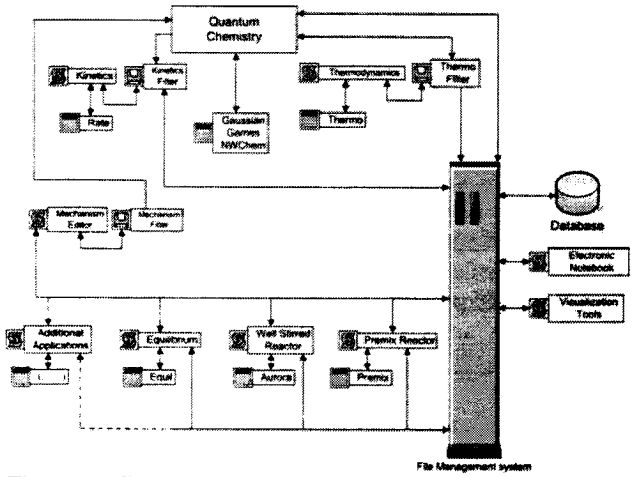
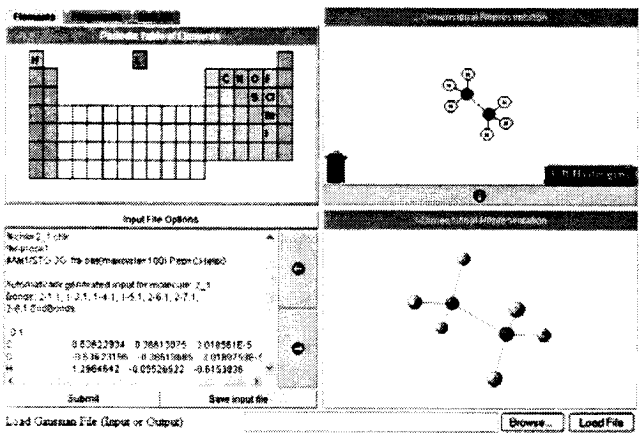


Figure 2. Flowchart to show how applications are connected.

The flowchart above illustrates how data flows from one scientific application or one area of science to the others, starting from first-principles quantum chemistry to reactor modeling. The file management system provides the backbone for extracting valuable data from one application to be stored or to be input parameters for another application using different filters. Below are the descriptions of individual components.

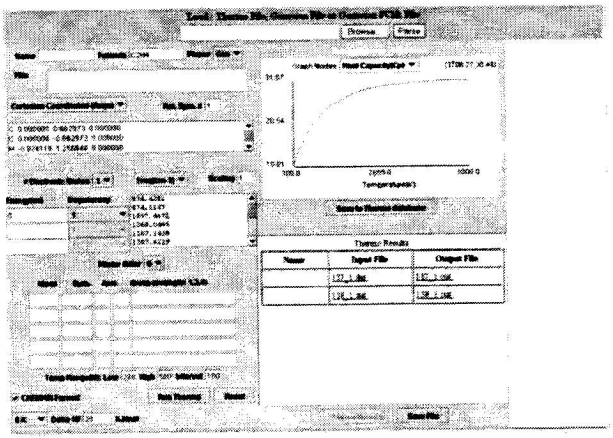
2.1. Quantum Chemistry



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., for following the reaction paths of chemical transformations, and for providing 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

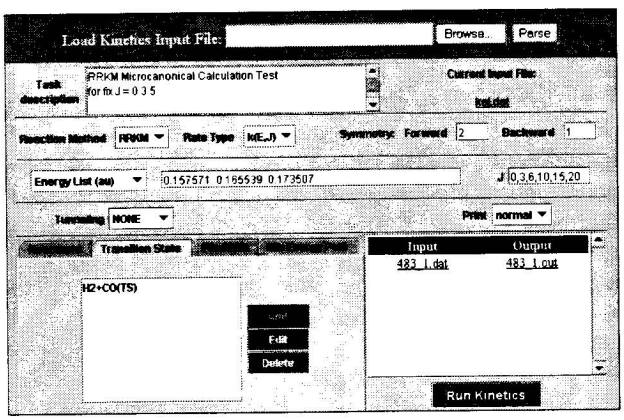
its unique strength. CSEO provides a universal web-based GUI that allows building molecular species from atoms and fragments, a means to generate or convert input files for these packages, and allows job submission to any package that is accessible to the user.
The results will then be parsed into a common data format for storage, visualization and analysis. By combining the capabilities of the available packages, CSEO provide users a unique and comprehensive research and educational environment for quantum chemistry.

2.2. Thermodynamics



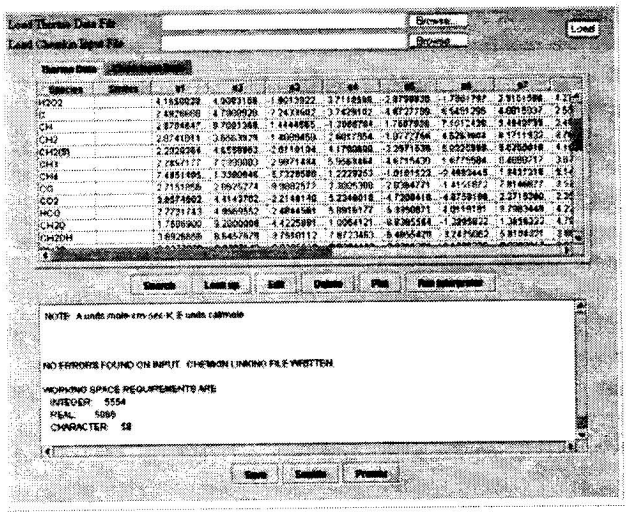
CSEO provides an interface to an application tool for calculating thermodynamic properties namely enthalpy, entropy and heat capacity of molecular species using the potential energy information calculated from first-principles of quantum chemistry discussed above. So selected information from the output of the quantum chemistry simulations will be parsed and used as input for thermodynamic calculations. The results are then stored in the format that can be used in premixed reactor simulations.

2.3. Kinetics



CSEO also provides an interface to a rather complete range of computational methods for predicting rate constants of elementary reactions. These methods require potential energy surface information, which can be extracted from *ab initio* quantum chemistry calculations. Similarly the results can be stored in the format that can be used for reactor simulations later.

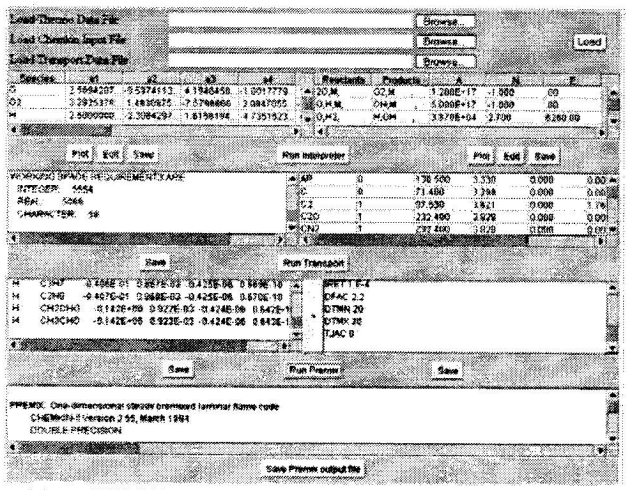
2.4. Mechanism Editor



Having calculated necessary thermodynamic properties of new molecular species and rate constants of new reactions, the researcher is required to update the kinetic model that has thermodynamic properties of all species and the associated reactions in the reacting system. The Mechanism Editor application tool provides a GUI with functionalities to edit and analyze the kinetic model. Necessary data can be uploaded from the database or parsed from a file. A visualization tool is also available for data analysis and comparisons. From here, the user can directly interface with reaction engineering application tools such as Senkin and Premixed codes.

2.5. Mechanism Analysis and Reaction Engineering

The modified kinetic model can be further analyzed using a sensitivity analysis to determine a set of most important reactions for a given experimental condition. CSEO provides interfaces to the CHEMKIN package for fluid dynamics simulations of complex reaction systems. In particular, CHEMKIN provides a rather comprehensive tool for modeling 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-disk reactors for chemical vapor



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 modeling.

3. SOFTWARE ENGINEERING INFRASTRUCTURE

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

A key part of our technology is using Java Applets on the client's computer which are connected to Java Servlets on the CSEO server to create very flexible user interfaces. This not only allows a simpler interface since there is no navigating, but it also allows a more interactive experience than the 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. Our current effort includes migration of these GUI's to Java Webstart and thus all application tools can be accessed from one window and the entired environment behaves much more like a 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 GUI's 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 are far greater than a simple graphic user interface (GUI) for any given 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 GUI's for some of the application tools that we plan to incorporate into CSEO, and besides potential license agreement restrictions, these GUI's 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 and Java3D to support such a multi-disciplinary, broad-based, web-computing environment as CSEO.

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 file management system similar to a file directory tree system. This allows the user to access all of the data that exists on the server anywhere in the environment. Transfering this data to a more permanent multi-level database system is also being constructed. The multi-level database system allows storage, retrieval, and transfer of data at the personal, group, and public levels.

3.3. Access, security and submission

Since CSEO will have access to many resources or information that are protected by different levels of proprietary and/or copyright concerns, we are implementing 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 making progress in installing Globus 3.0 GRID technologies[10] on our experimental grid Linux clusters

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 June 10, 2003. There are two modes for accessing the tools available in CSEO. One is to login to 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. The next major release is planned for July 1st, 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 modeling and simulations of complex reacting systems. This environment allows data to flow from the most basic fundamental level of quantum chemistry to be used in flame simulations of combustion systems. A demonstration is given to illustrate the versatility of the environment. The development utilizes advances in web technologies to create a user-friendly environment for both research and education. There is much more work remaining to be done. However, its current release already provides useful tools for interfacing with quantum chemistry, thermodynamics, and fundamental kinetics and is already being used worldwide by many research groups.

Acknowledgement: This work is supported in part by the National Science Foundation through the Information Technology Research Initiative. 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 Technology and National Research Council, *Collaboratories: Improving Research Capabilities in Chemical and Biomedical Sciences: Proceedings of a Multi-site Electronic Workshop*. 1999, Washington, D.C.: National Academy Press.

[2] Agarwal, D.A., Sachs, S.R., and Johnston, W.E., *The Realities of Collaboratories*. Comp. Phys. Comm., 1998. **110**: p. 134.

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

[4] Kouzes, R.T., Myers, J.D., and Wulf, W.A., *Collaboratories: Doing Science on the Internet*. IEEE Computer, 1996. **29**(8).

[5] Charnes, J.M., Morrice, D.J., Brunner, D.T., and Swain, J.J., eds. *Proceedings of the 1996 Winter Simulation Conference*. 1996: Coronado, CA.

[6] Fishwick, P.A., Hill, D.R.C., and Smith, R., eds. *Proceedings of the 1988 International Conference on Web-Based Modeling and Simulation*. 1998, The Society for Computer Simulation International: San Diego.

[7] Bruzzone, A.G., Uhrmacher, A., and Page, E.H., eds. *Proceedings of the 1999 International Conference on Web-Based Modeling and Simulation*. 1999, The Society for Computer Simulation International: San Diego.

[8] Kuljis, J. and Paul, R.J., *An appraisal of Web-Based Simulation: Whither We Wander*. Simulation Practice and Theory, 2001. **9**: p. 37-54.

[9] Page, E.H., Buss, A., Fishwick, P.A., Healy, K.J., Nance, R.E., and Paul, R.J., *Web-Based Simulation: Revolution or Evolution?* ACM Transactions on Modeling and Computer Simulation, 2000. **10**: p. 3-17.

[10] Foster, I. and Kesselman, C., *Globus: A metacomputing infrastructure toolkit*. Int. J. Supercomputing Applications, 1997. **11**: p. 115-128.