Constructing a Kinetics Database: A NSF-Sponsored Workshop

Workshop Report (7/20/07)

prepared by David M. Stanbury

Abstract: An NSF-sponsored workshop entitled "Constructing a Kinetics Database" was held at the NIST campus in Gaithersburg, MD, on April 19-21, 2004. Approximately 50 scientists attended the workshop to review the current status of databases on chemical kinetics, and to discuss the needs for maintaining the existing databases, the needs for extending the scope of the included data, and the processes through which these needs can be met. A significant consensus was achieved on many issues, and a concrete plan was established to follow up on them.

Table of contents

Introduction

Chemistry is widely regarded as "the central science", and chemical reactions are often regarded as the "heart of chemistry". Chemical kinetics is the science of quantitative description of the rates of chemical reactions, and it is these rates that determine the time frame of reactions and the chemical product distributions. The technical literature on chemical kinetics is vast and expanding. It constitutes an inportant component of fundamental science, and it is of great use to a wide range of applied subjects. Its value and utility are attested by the existence of several on-line kinetics databases that are available to the general public. Despite the existence of these databases, it should be noted that all of them are limited in scope, serving, for example, the fields of atmospheric or radiation chemistry, and all of them require continued support in order to remain current.

A number of general questions relate to the current and future status of chemical kinetics databases. Are the current databases adequate for the community needs? Should the scope of the databases be expanded? Is there a simple way to ensure the continued maintenance of these databases? What would be the value of a comprehensive set of kinetics databases?

To assess the current situation, address the above questions, and set a program for future action, a workshop was convened entitled "Constructing a Kinetics Database". The workshop had ~50 participants, was held at the NIST campus in Gaithersburg, MD on April 19-21, 2004, and was supported by NSF. Its co-organizers were David Stanbury (Auburn University) and Michael Frenkel (NIST-Boulder). The workshop consisted of 22 talks, each of which was followed by liberal time for discussion. A poster session (supported by Exxon/Mobil) and banquet were valuable additional components to the workshop. The final morning of the workshop was devoted to a general discussion and formulation of recommendations. One of the key outcomes was the establishment of a continuing committee to oversee the implementation of the recommendations.

Workshop Program Summary

A detailed workshop program can be found in Appendix A. The program was structured around four themes: 1) a sampling of the variety of kinetic data being produced, 2) a sampling of the various uses of kinetic data, 3) a nearly comprehensive review of the existing on-line kinetic databases, and 4) an investigation of the operation and value of other chemical databases. A detailed summary of the individual presentations can be accessed here.

Emergent Concepts and Issues

A number of important factors came to light during these presentations and the ensuing discussions as described below.

Within the first few hours of the workshop it became obvious that the scope and diversity of chemical kinetics is vast, that the various subfields use widely divergent language to describe their data, and often the work in one subfield can be virtually incomprehensible to workers in other subfields.

Review of the current kinetics databases revealed the following: While several current databases have been tremendously useful and exemplify both the utility and potential of a broadly based kinetics database, there are many glaring deficiencies. For example, other than that administered by the Notre Dame Center for Radiation Chemistry Data (CRDC) for radiation chemistry and free radicals, there is no general solution-chemistry kinetics database. On the other hand, databases, which have suffered lapses in funding, may only cover a particular period, exist in an obsolete computer format, or disappear altogether. In general, the data in existing databases is incomplete, for example the gas phase database extends only to n-butane, but there are seven independent heptane databases. Data also tends to be in different formats and cannot be transferred between databases. There are differences in how quantitave properties are defined and the units in which they are expressed. Often the data cannot be easily applied to specific problem without constructing an individual database precisely adapted to the application. There is no defined, generally agreed-upon overarching structure for the databases. There is no facile way to handle bulk quantities of kinetic data.

Many essential areas of chemical kinetics, such as electrochemistry, homogeneous catalysis, solid state, etc., are not represented at all in existing databases. There is limited quality control of the literature data itself and, since data must be extracted by hand from articles, transcription errors may occur upon entry into the database. There is no consistency between databases in expressing errors, distinguishing between precision and systematic errors or even necessarily in including error estimates. Users are discouraged from utilizing multiple databases as their activation energy for doing so increases with the number of databases needed.

One debated topic was whether kinetic databases should contain just rate constants for elementary steps or whether they should also contain overall empirical rate laws (bulk kinetics). One one side, it was argued that elementary step rate constants are of great value because they can be assembled into complex mechanisms to model reaction systems that have not been previously studied. I.e., these rate constants are “transportable”. An argument in favor of compiling bulk kinetics data is that the elementary step rate constants are often unknown, the

concept of "elementary step" may not be readily applied to heterogeneous reactions, and there is considerable user demand for tabulations of bulk kinetics.

The 16 Questions

One of the specific objectives of the workshop wass to obtain answers to a list of 16 questions. These questions were posed to the participants at the opening of the workshop, and answers were taken by vote during the closing session. The questions and answers are as follows:

1) Are the current gas-phase kinetics databases adequate for their purposes, or should improvements be made? Real strengths, outstanding example, improvements should be and are being made. Is there reason to believe that they will continue to function well for the forseeable future? Major funding challenges and collaboration challenges.
2) Should the new database initially be limited to thermal liquid solution-phase reactions of simple compounds? NO!
3) Should it be limited initially to isothermal reactions? NO!
4) Should it include elementary steps, overall reactions, or both? **Both**
5) Should it be limited to spatially uniform reactions (no traveling waves, etc.)? Yes initially.
6) Should it be limited to reactions having strictly reproducible kinetics (no chaos or non-determinate behavior)? Bad question: beyond kinetics – system dynamics is the issue.
7) Should it be limited to reactions that are monophasic, or should more complex reactions be included? Include complex behavior.
8) Should it be limited to reactions having rate laws consisting of a single ordinary differential equation? NO.
9) If the database should include more than thermal liquid solution-phase reactions of simple compounds, what should it include? All of the below and more. **Priorities need to be set!!**
   a) Electrochemical kinetics?
   b) Heterogeneous catalysis?
   c) Gas-liquid reactions?
   d) Solid-state reactions?
   e) Colloid reactions?
   f) Photochemical and radiation-induced reactions?
   g) Precipitation and dissolution kinetics?
   h) Macromolecular reactions?
   i) Polymerization reactions?

j) Enzyme-catalyzed reactions?

10) What organization should maintain the database? Databases should be coordinated, federated, and sustained; they may be distributed—NIST is a logical choice for the overall coordination role

11) What steps can be taken to ensure the long-term success of the database? See report.

12) Should a subscriber fee be charged for users of the database? (Government support can be unpredictable) tbd.

13) Should the database simply compile published kinetic data, or should it assess the data critically? Both.

14) Should the database attempt to include the older literature or should it just attempt to keep up with the new literature as it becomes published? All data are valuable, some archival data are invaluable. Critical to begin to collect all new data!

15) Should it be limited to results published in peer-reviewed journals that are generally available? (No patent literature, commercial or government reports, or theses) Primary emphasis should be on peer reviewed publications, in some subfields there may be exceptions – ALL data should have clear provenance.

16) Should a committee be created to follow up on the recommendations? If yes, who should be on the committee, and what should be their charge? Yes, to be developed

General Recommendations

Generators of kinetic results in a particular subdiscipline need an easy way to enter their data in a manner that prompts them for the kinetic property, appropriate units, error estimates, rate law, and the method (spectroscopic, thermal, manometric, etc., relaxation or other technique, and conditions of measurement (temperature, pressure, concentration, media, etc.) familiar to their area. Data should be checked for internal consistency and against other, similar data before publication and entry into the database. Data entered into the database should be in a standard format. Where feasible, the data should also be critically evaluated by a panel of peers and appropriate annotations made in the database. Consideration should be given to eventually collecting the experimental data (spectra, etc) from which the kinetic data is derived.

Ideally, users of kinetic data would have an apparently single, seamless database that can be searched through a single, user-friendly, subdiscipline-adapted search engine that can generate readable tables in a format familiar to the user.

The various kinetics communities should be involved in developing standards for data collection and use.

The database should have a sustained financial base, which should be provided, at least initially, by government and industry. A long-term support mode needs to be developed.

An effort should be made to induce journals to require submission of kinetic data for checking by the database prior to publication and automatic entry into the database following peer review. Government funding agencies may also wish to require this under their grant conditions so as to assure adequate dissemination of taxpayer funded results.

As much as possible, the database should make use of tools already under development, such as the ThermoML database architecture, data checking and data entry tools, the IUPAC/NIST chemical identifier (INChI) and structure drawing entry system, AnIML (Analytical and Instrument Markup Language), and subdatabase architectures developed by the Collaboratory for MultiScale Chemical Science (CMCS).

A continuing committee should be established that would work from existing examples of successful databases, such as the Gas Phase Kinetics Database, the NASA data panel, the IUPAC data panel and the Cambridge Crystallographic Database. The committee would assist in writing proposals and securing funding to develop the needed software and data storage methods. The committee would coordinate domain interactions with editors and publishers and the broader scientific community. The software would then be further developed in a phased, extensible (modular) approach to improve the scope, quality and accessibility of the data over time.

In one model, the committee would develop a set of questions for response by the various kinetic subdisciplines (domains), e.g. kinetic properties measured, symbol, units, uncertainty,

compound id., reaction definition, kinetic method and conditions. The committee would locate and designate "domain champions", who would secure a consensus from their domains in answering these questions. Each domain would be asked to develop a domain data utilization model, including a glossary defining kinetic properties, terms and units, based on the IUPAC Gold/Green Books, that would draw on a consensus of data generators, disseminators, and users. The results of the domain queries would be analyzed by the central committee and an initial overall structure for the database structure and input engines determined. This would be subject to a second review by the domains. Prototype data acquisition software would then be developed.

In another model, the committee would review the initial efforts of the NIST gas-phase kinetics database workers in developing a data-acquisition software package and attempt to use it as an initial framework for designing other such packages for each kinetics subfield.

Specific Recommendations

1) There is great value in chemical kinetic data, and there is great value in compiling this data in widely accessible form of on-line databases. The current kinetic databases should be maintained, and new datbases should be established to compile data from subfields that are not served by the existing databases.

2) Despite the ideal of a single unified database, the diversity of kinetic data is so great that it would be unreasonably difficult to design and implement a useful database that would be comprehensive and serve all subfields. Each subfield should design and implement its own database, but the individual databases should comply to certain standards so, in aggregate, they would constitute a linked family of databases.

3) The model of Thermo ML should be adopted as a means for data collection and input. To do so would require the construction of a standard computer language (dictionary) for data input, (call it KineticsML?).

4) As with ThermoML and the Cambridge Crystallographic Database, data collection should be performed at the time that papers are published in the journals. Journals will be solicited to encourage their authors to participate in this process.

5) In parallel with the development of KineticsML, "reader" software packages should be created and made publicly available so as to make the information in the databases accessible.

6) A continuing committee to oversee the development of these databases should be established. The committee members should be:

Chair: David Stanbury, Dept. of Chemistry, Auburn University
Tom Allison, NIST-Gaithersburg
Nick Delgass, Chemical Engineering, Purdue University
David Dixon, Dept. of Chemistry, U. of Alabama
Michael Frenkel, NIST-Boulder
Jeff Manion, NIST-Gaithersburg
Chuck Kolb, Aerodyne

Immediate tasks for the continuing committee include developing a plan for initial funding, determining the initial database format, enabling the development of data capture and data viewer software, and facilitating subfiled activities as described below.

7) Each subfield should decide on the type of data to be collected, and a sample data input form should be circulated for approval in each subfield. As a starting point, specific subfields and their database mavens should be the following:

solution-phase inorganic - David Stanbury
solution-phase organic - Claude Bernasconi
solution-phase free radicals - Steve Mezyk
gas-phase kinetics - Jeff Manion
polymers/TGA & solid state - Sergey Vyazovkin
heterogeneous catalysis - Nick Delgass and Fabio Ribiero
electrochemistry - Dennis Evans
enzymology - Tom Leyh
computational - David Dixon

Additional subfields will be added at a later date.

Followup Meeting

A small meeting was held on September 21, 2005 on the campus of the University of Alabama, Birmningham. The attendees were

Tom Allison	NIST-Gaithersburg
Dave Dixon	U. Alabama, Tuscaloosa

Michael Frenkel	NIST-Boulder
Michael Frenklach	UC, Berkeley
Karl Mueller	Penn. State U.
Larry Rahn	Sandia Nat. Labs
Fabio Ribiero	Purdue U.
David Stanbury	Auburn U.
Sergei Vyazovkin	U. Alabama, Birmingham

The purpose of this meeting was to convene a small group of the people most interested in the project and to recruit the participation of two recent recipients of NSF Cyberinfrastructure grants related to the project: the two Cyberinfrastructure grantees were Karl Mueller and Michael Franklach, neither of whom was a participant in the Workshop. During this meeting the highlights of the Workshop were recapped, and the suitability of the NIST ThermoML system as a model for a kinetics database was revisited. Michael Frenkel, director of the ThermoML project, reiterated his offer to provide the basic code used in the ThermoML project, so that it could be adapted for a kientics database. The main outcome of the meeting was that Mueller and Frenklach assumed the responsibilities of following up on the project.

The Cyberinfrastructure grantees have created Web sites for their kinetics-database-related projects. For the Mueller/Penn. State project the web site is $Chem_xSeer$: http://chemxseer.ist.psu.edu/

For the Frenklach/Berkeley project, PrIMe, the website is http://primekinetics.org/

Appendix 1: Workshop Participant List

Speakers

Tsang, Wing	301-975-2507 NIST, Gaithersburg	wing.tsang@nist.gov
Manion, Jeffrey	301-975-3188 NIST. Gaithersburg	jeffrey.manion@nist.gov
Huff, Hartz, Kara	412-268-6023 Carnegie Mellon U.	karah@andrew.cmu.edu
Bernasconi, Claude	831-459-2035 UC Santa Cruz	bernasconi@chemistry.ucsc.edu
Leyh, Thomas	718-430-2857 Albert Einstein Univ.	leyh@aecom.yu.edu
Dennis Evans	520-626-0318 U. of Arizona	dhevans@email.arizona.edu
Vyazovkin, Sergey	205-975-9410 U. of Alabama, Birmingham	vyazovkin@uab.edu
Darensbourg, Don	979-845-5417 Texas A&M U.	djdarens@mail.chem.tamu.edu
Smith, Greg	650-859-2000 SRI Int.	gregory.smith@sri.com
Kustin, Ken	858-450-2979 San Diego	kmkustin@ix.netcom.com
Barckholtz, Tim	908-730-3423 ExxonMobil	tim.barckholtz@exxonmobil.com

Clarke, Mike	703 292 4967	
	NSF, Boston College	mclarke@nsf.gov
Madden, Keith	547-631-7279	
	Notre Dame U.	madden.1@nd.edu
Huie, Bob	301-975-2559	
	NIST	robert.huie@nist.gov
Allison, Tom	301-975-2216	
	NIST	thomas.allison@nist.gov
Kolb, Charles	978-663-9500 ext. 290	
	Aerodyne	kolb@aerodyne.com
Atkinson, Roger	909-787-4191	
	UC Riverside	roger.atkinson@ucr.edu
Campbell, Carelyn	301-975-4920	
	NIST, Gaithersburg	carelyn.campbell@nist.gov
Rheingold, Arnie	858-822-3879	
	U Calif., SD	arheingold@ucsd.edu
Mallard, Gary	301-975-2564	
	NIST, Gaithersburg	gary.mallard@nist.gov
Chaka, Anne		
	NIST, Gaithersburg	anne.chaka@nist.gov
Stein, Steve	301-975-2505	
	NIST, Gaithersburg	stephen.stein@nist.gov
Chirico, Rob	303-497-4126	
	NIST, Boulder	chirico@boulder.nist.gov

Ribeiro, Fabio	765-494-7799	
	Purdue U., Chem. Eng.	fabio@purdue.edu
Bartels, David M.	574 631 5561	
	Notre Dame U.	bartels@hertz.rad.nd.edu
Frenkel, Michael	303-497-3952	
	NIST, Boulder	frenkel@mail.boulder.nist.gov
Cabelli, Diane	631-344-4361	
	Brookhaven Nat. Lab.	cabelli@bnl.gov
Johnson, Mike	505-646-3627	
	New Mexico State Univ.	johnson@nmsu.edu
Stanbury, David	334-844-6988	
	Auburn Univ.	stanbury@auburn.edu
Bill Koch	NIST	

Attendees

Raul Miranda	U.S. Dept. of Energy Basic Energy Sciences	Raul.Miranda@science.doe.gov
Prof. David A. Dixon	205-348-8441 Department of Chemistry The University of Alabama Lloyd Hall Box 870336 Tuscaloosa, AL 35487-0336	dadixon@bama.ua.edu
Malcolm W. Chase	802-382-9325 1115 East Munger Street Middlebury VT 05753	mwcpwc@sover.net

Reuben H. Simoyi	503-725-3895 Professor of Chemistry, Portland State University Portland, OR 97207	rsimoyi@comcast.net
Stuart Kirkham Cole	757-431-1002 McKim & Creed 448 Viking Drive, Suite 100 Virginia Beach, Virginia 23458	KCole@mckimcreed.com
Holly Ellis	334-844-6991 Chemistry Department, Auburn University	ellishr@auburn.edu
Michael J Clarke	703 292-4967 National Science Foundation Arlington, VA 22230	mclarke@nsf.gov
Hai Wang	University of Delaware	hwang@me.udel.edu
Louis J. Stief	Emeritus Scientist NASA/Goddard Space Flight Center, Greenbelt, MD	stief55@earthlink.net
Dwight Tardy	319-335-1355 Prof. of Chemistry, University of Iowa	dwight-tardy@uiowa.edu
Regina J. Cody	301/286-3782 NASA Goddard Space Flight Center Code 691 Greenbelt, MD 20771	Regina.Cody@nasa.gov
William J. Cooper	910-962-2387 Department of Chemistry and Center for Marine Science	cooperw@uncw.edu

	5600 Marvin K. Moss Lane UNC, Wilmington, NC	
Chad Sheng	NRC post doc at NIST - Gaithersburg	chad.sheng@nist.gov
Vadim D. Knyazev	202-319-6742 The Catholic University of America	knyazev@cua.edu
Paul Romani	301-286-1525 NASA - Goddard Space Flight Center	Paul.Romani@nasa.gov
Pedi Neta	NIST-Gaithersburg	pedi@nist.gov
Peter C. Preusch	301-594-5938 Program Director Pharmacology, Physiology, and Biological Chemistry Division National Institute of General Medical Sciences	PREUSCHP@nigms.nih.gov
Kent H. Casleton	304-285-4573 US Department of Energy, National Energy Technology Laboratory Morgantown WV	KENT.CASLETON@netl.doe.gov
James M. Caruthers	765-494-6625 School of Chemical Engineering Purdue University West Lafayette, IN 47907-1283	caruther@ecn.purdue.edu
W. Nicholas Delgass	765-494-4059 School of Chemical Engineering Purdue University	delgass@ecn.purdue.edu

Stephen Peter Mezyk 562-985-4649
Cal. State University Long Beach smezyk@csulb.edu
Long Beach, CA, 92660

Pauline Ho 505-237-8995
Reaction Design
pho@reactiondesign.com

Appendix 2: Workshop Program

Monday, April 19, 2004: NIST Gaithersburg

9:00 AM	Koch, Bill: NIST	Welcoming remarks
	David M. Stanbury: Opening Remarks: charge, objectives, & issues.	

Session I: types of kinetic data available (talks from data producers)

Session chair: Wing Tsang		NIST, Gaithersburg
gas-phase		
9:15	Manion, Jeffrey	NIST. Gaithersburg
inorganic solution-phase		
9:40	Kara Huff-Hartz	Carnegie-Mellon
organic solution-phase		
10.:05	Bernasconi, Claude	UC Santa Cruz
10:30	Coffee break	
heterogeneous catalysis		
10:45	Ribeiro, Fabio	Purdue, Chem. Eng.
polymers		
11:10	Vyazovkin, Sergei	U. Alabama, Birmingham
heterogeneous stoichiometric (electrochem, phase transfer, fuel cells etc)		
11:35	Evans, Dennis	Arizona
12:30	Lunch in NIST cafeteria	
1:30 PM	Session I, continued	
	Continuing Session Chair: Mike Johnson (New Mexico State)	
organometallic		
1:31	Don Darensbourg	TAMU

Session II: community needs for kinetic data (data users)

Session Chair: Mike Johnson (New Mexico State)

atmospheric/combustion modeler

2:00 Smith, Greg SRI Int.

solution-phase nonlinear dynamics

2:30 Kustin, Ken San Diego

3:00 coffee break

solution-phase radiation modeler

3:15 Bartels, David M. Notre Dame

scrubbing technology

3:45 Barckholtz, Tim ExxonMobil

6:00 PM Poster session/mixer

7:30 PM dinner

Tuesday, April 20

9:00 AM Session III: currently available kinetics databases

Session chair: Diane Cabelli (Brookhaven NL)

NIST/NDRL & NDRL

9:01 Madden, Keith Notre Dame

NIST gas-phase

9:30 Allison, Tom NIST

NASA/JPL atmospheric

10:00 Kolb, Charles Aerodyne

10:30 coffee break

IUPAC gas-phase

10:45	Atkinson, Roger	UC Riverside
Solid-state diffusion		
11:15	Carelyn Campbell	NIST, Gaithersburg
12:30	Lunch	
1:30 PM	Session IV: suitable data collection and dissemination technologies	
	Session chair: M. Frenkel	NIST-Boulder
Cambridge crystallographic database		
1:31	Rheingold, Arnie	UCSD
NIST Webbook		
2:00	Mallard, Gary	NIST, Gaithersburg
Computation of rate constants		
2:30	Chaka, Anne	NIST, Gaithersburg
3:00	coffee break	
Chemical identifier and XML dictionaries in chemistry		
3:15	Stein, Steve	NIST, Gaithersburg
NIST ThermoML		
3:45	Chirico, Rob	NIST, Boulder
Biochemical		
4:15	Leyh, Thomas	Einstein
6:00	cash bar	
7:00	Banquet	

Wednesday, April 21

9:00 AM Session V: breakout session, review, and recommendations

Session chair: David Stanbury

a) breakout sessions;

b) reconvene for overview and final recommendations.

11:30 AM Workshop adjourns

Appendix 3: Involvement of Underrepresented Groups at the Workshop

Considerable effort was made to recruit participants from underrepresented groups, and that effort was reasonably successful given the highly skewed composition of the pool of qualified kineticists. One helpful tool in identifying candidate participants turned out to be the on-line version of the ACS Graduate Directory of Research; this on-line database enabled a search for people with the keywords "kinetics" and "female".

Out of the ca. 50 participants, 7 appeared to be female and 2 had apparent black-african ancestry. All of these underrepresented-group participants were highly active and vocal during the workshop and added valuable expertise and insights.

Appendix 4: Workshop Report Dissemination

A copy of this report will be posted on the Workshop web site maintained by David Stanbury. A copy will also be submitted to the Chemistry division at NSF. All Workshop participants will receive emails directing them to these web site reports.