

OPEN

DATA DESCRIPTOR

ReSpecTh: Reaction kinetics, spectroscopy, and thermochemical datasets

Tamás Turányi¹✉, István Gy. Zsély¹, Máté Papp¹, Tibor Nagy², Tibor Furtenbacher¹, Roland Tóbiás¹, Péter Árendás^{1,3} & Attila G. Császár¹✉

Description of a large number of datasets related to gas-phase reaction kinetics (Re), high-resolution molecular spectroscopy (Spec), and thermochemistry (Th), called ReSpecTh, is presented. The datasets contain accurate and validated experimental, empirical, and computed, machine-searchable data, and, whenever possible, the corresponding metadata. ReSpecTh data and the accompanying utility codes can be used in several engineering and scientific fields either separately or simultaneously, such as simulation of combustion reactions, atmospheres of planets and exoplanets, and stellar and interstellar environments.

Background & Summary

Knowledge of accurate data, and related metadata, has an eminent place in modern science and engineering, a statement true for many fields of chemistry. In the fourth age of quantum chemistry¹, the availability of high-level, complex algorithms and related sophisticated computer codes makes it feasible to (a) validate experimental data, (b) generate new, empirical data augmenting existing measured data, and (c) use experimental, empirical, and computed data on a common basis to help diverse scientific and engineering applications. This paper provides a description of the result of our efforts to collect published experimental kinetic, high-resolution spectroscopic, and thermochemical data and to include them in datasets under the common name ReSpecTh. The aim of the ReSpecTh project has been to help to (a) understand our wider natural environment (*e.g.*, star-forming regions, the interstellar medium, the atmospheres of planets and exoplanets), (b) answer important questions related to the origin of life on Earth and the special role of humans in the preservation of life's diverse forms, (c) improve our understanding of complex chemical reaction systems, (d) protect the quality of life on our planet, (e) validate and develop gas-phase kinetic reaction mechanisms to understand and model processes in combustion and atmospheric chemistry, (f) develop more efficient, low-emission combustion devices for both current and novel fuels, and (g) assist air-quality prediction and emission limit settings (*e.g.*, via an improved scientific understanding of the greenhouse effect and of combustion systems, as well as research and development related to climate change).

An improved understanding of the various processes and environments mentioned requires the availability of carefully validated data of traceable origin. Indeed, several databases and information systems exist in the fields of reaction kinetics²⁻¹⁰, high-resolution molecular spectroscopy¹¹⁻¹⁷, and thermochemistry¹⁸⁻²⁹. About a decade ago, we set up a dedicated website, <https://respecth.hu/>, freely accessible via a login system after registration, containing a large number of datasets in reaction kinetics, high-resolution molecular spectroscopy, and thermochemistry. At present, the number of registered users of ReSpecTh is close to 1000. After a decade of development, this paper describes the ReSpecTh datasets, publicly available at an OSF (Open Science Framework) repository, <https://osf.io/nbmzv/>³⁰.

The ReSpecTh datasets have several unique and noteworthy features. The kinetic branch contains searchable datasets of gas-phase chemical kinetics, focusing on elementary reactions relevant to combustion. It includes detailed experimental data and tools for assessing the uncertainty of reaction rate parameters. The spectroscopic datasets are built upon data mining, using methods of discrete mathematics novel to molecular physics, such

¹Institute of Chemistry, ELTE Eötvös Loránd University, H-1117, Budapest, Pázmány Péter sétány 1/A, Hungary.

²HUN-REN Research Centre for Natural Sciences, Magyar Tudósok körútja 2, 1117, Budapest, Hungary. ³Budapest University of Economics and Business, Budapest, Hungary. ✉e-mail: tamas.turanyi@ttk.elte.hu; attila.csaszar@ttk.elte.hu

as graph (network) theory. Many of the collected experimental and computed data influence the values and the uncertainties of the derived (empirical) data. It is also important to note that a considerable part of the thermochemical data are based on detailed spectroscopic line-by-line information.

Creation of the ReSpecTh datasets followed the FAIR (Findable, Accessible, Interoperable, and Reuseable) principle^{31,32}. The measured and theoretically determined data in our datasets are *findable* thanks to sophisticated, detailed search engines designed for user convenience. Each XML file has a unique identifier referenced throughout the database. In the case of chemical kinetics data, for example, the core of a measurement data file is formed by measured ignition delay times (IDT), flame velocities, laminar burning velocities (LBV), and temperatures or concentrations. Each data file is rich with metadata describing in detail the measuring device, includes a link to the original publication containing the measurement data (BibTeX fields and DOI link) along with the CAS, InChI, and SMILES identifiers of the chemical substances measured. Within the spectroscopy datasets each source has a standardized unique identifier, composed from the year of the publication and the first letters of the names of the authors. Each transition of each source also receives a unique identifier. The data are *accessible* at the OSF repository³⁰; furthermore, the ReSpecTh website provides easy downloads by data file, both individually and in groups after a free registration using the user's login credential. *Interoperability* of data is also ensured, because the files store the data in a well-described, standardized format. In the case of the kinetic data this is the ReSpecTh Kinetics Data (RKD) Format³³, while in the case of the spectroscopic data the format required by the MARVEL (Measured Active Rotational-Vibrational Energy Levels) code^{34,35}. The XML-based RKD data format is standalone, and can be used outside our environment. There are already existing tools^{36,37} that can convert these files into input files of popular chemical kinetic modelling program packages^{37–40}. Optima++³⁶ can also convert the RKD-format data files to plain text files. The data can be *reused* because the meaning of each tag used within ReSpecTh is described, allowing the users to understand, interpret, or even re-code the information. Correcting errors or adding new information in the formats applied is well handled and these changes do not invalidate the unmodified part of the files.

The rate of data accumulation in the fields of combustion research, as well as of reaction kinetics, has been steadily increasing: the current literature contains an extremely large number of experimental combustion data and detailed combustion mechanisms⁴¹. This leads to typical problems of 'big data': storage, search, and sharing. Sometimes it is argued that a combustion database based on direct and indirect experimental data is not of 'big data' type, since the amount of data for a given fuel and at a given set of conditions is limited. However, because of the hierarchical nature of the combustion chemistry of fuels⁴², the experimental data for hydrogen, syngas, methane, ethane *etc.* are also applicable for the validation of the high-temperature combustion mechanism of a higher hydrocarbon, like propane. Furthermore, there is a demand for mechanisms applicable in wide ranges of conditions (temperature, pressure, and equivalence ratio), which requires a comprehensive set of data.

Traditional scientific papers have been ideal for sharing new ideas and describing procedures and methods, but they are far from being ideal for storing data. In these papers, neither the figures and the tables, nor the typical supplementary files are in machine-readable form. The common search engines, designed to explore the scientific literature, can find information by keywords, but not by the experimental data themselves. These difficulties lead to poor and ineffective access to measurement data in general, and in reaction kinetics in particular.

Only a few databases contain information on the determination of direct experimental and theoretical reaction rate coefficients. The most famous and comprehensive one is the NIST Chemical Kinetics Database². This contains more than 38 000 separate reaction records for over 11 700 distinct reactant pairs, abstracted from over 12 000 papers. The website of an IUPAC Task Group on 'Atmospheric Chemical Kinetic Data Evaluation'³ provides evaluated kinetics and photochemical data, related to a wide range of atmospheric chemical processes. The EUROCHAMP-2020 database⁴ has a much narrower focus: it provides a compilation of recommended rate coefficients for reactions of volatile organic compounds with the OH and NO₃ radicals, ozone, and Cl atoms under atmospheric conditions. The JPL database⁴³ provides a series of evaluated sets of rate constants, photochemical cross sections, heterogeneous parameters, and thermochemical parameters compiled by the NASA Panel for Data Evaluation. The Master Chemical Mechanism⁴⁴ contains information on detailed gas-phase chemical processes involved in the tropospheric degradation of a series of primary emitted volatile organic compounds (VOCs). Currently, the degradation of methane and 142 non-methane VOCs is represented.

In the field of combustion kinetics, Frenklach⁴⁵ suggested collaborative data processing for the analysis and development of detailed reaction mechanisms. Frenklach emphasized⁴⁶ that the simultaneous application of large amounts of experimental data for testing a combustion mechanism may provide a new level of understanding. The related informatics infrastructure was called PrIME (Process Informatics Model). Unfortunately, neither the original PrIME website⁶, nor its local version at DLR Stuttgart are functioning. The PrIME collection contained interlinked experimental, bibliographic, and species information data. The PrIME experimental data format defined limited types of experimental data and new formats were not defined to accommodate other data types.

KAUST Cloudflame⁷ is a cloud-based cyberinfrastructure for managing combustion research and enabling collaboration. The infrastructure includes both software and hardware components and is freely offered to anyone with a valid professional or educational affiliation. This website provides a front-end for data-search tools, web-based numerical simulations, and discussion forums. It includes a collection of combustion data, which mainly focuses on liquid fuels and is limited to jet-stirred reactors (JSR) and ignition-delay-time (ST-IDT) measurements. There is no search engine for experimental conditions, and the data files use the csv format and an earlier version of the ReSpecTh Kinetics Data Format (RKD, see below). The last data uploads occurred in 2019.

The SciExpeM database^{47,48} contains more than 38 000 indirect experimental data points in more than 2000 RKD format data files. The main aim of this database is to develop simulation models for the combustion kinetics of green fuels related to the Industry 4.0 concept.

ChemKED^{8,49} is a GitHub repository limited to IDT measurements. The data files are handled by PyKED, a Python-based software applicable for validating and interacting with ChemKED files. There is no search engine connected to these data. A major update took place in 2018, with limited updates in 2021.

The UConn Combustion Database⁹ contains the data measured in the Combustion Diagnostics Laboratory of the University of Connecticut (UConn). This data collection contains IDTs measured in rapid compression machines (RCM) and laminar burning velocity measurements. The data are in Excel tables and the related data are provided in downloadable zip files. There is no search engine for these data.

The Fundamental Kinetics Database Utilizing Shock Tube Measurements¹⁰ summarizes the published shock tube experimental measurements performed under the supervision of Prof. Hanson. The database covers the years from 1974 to 2019 inclusively. The database is a collection of ignition delay time, speciation, and rate measurements in shock tubes, with data given in MS Word and PDF files. There is no search engine for these data, and computer reading of such files cannot be straightforwardly automated.

There are few comprehensive databases of detailed combustion mechanisms available on the internet. GITHUB hosts minor collections of combustion mechanisms in CHEMKIN format^{50–52}. The SciExpeM web site⁴⁷ contains 207 mechanisms for 30 fuels. There are websites containing mechanisms developed within a research group. These websites are maintained by the Combustion Chemistry Centre, University of Galway, Galway, Ireland⁵³, the CRECK Modeling Group, POLIMI, Milano, Italy⁵⁴, the Combustion Research Group, UCSD, San Diego, CA⁵⁵, the Combustion Kinetics Laboratory, USC, Los Angeles, CA⁵⁶, the Clean Combustion Research Center, KAUST, Kingdom of Saudi Arabia⁵⁷, and LLNL, Livermore, CA, USA⁵⁸.

There are a few websites which contain utility codes for the analysis and development of detailed combustion mechanisms. The LLNL Mech Checker website⁵⁹ provides tools⁶⁰ to improve the performance of chemical mechanisms. The site contains the Thermodynamic Checker and the Ignition Delay Time Diagnostic tool. The Thermodynamic Checker (a) refits the NASA polynomial coefficients to correct for the discontinuities in the thermodynamic functions between the lower and upper temperature ranges of their definition, and (b) plots thermodynamic data for isomers, so the user can determine if there are any outliers. A similar tool is provided by the authors of OpenSMOKE++³⁷. The Ignition Delay Time Diagnostic tool utilizes the Zero-RK solver to run single ignition delay calculations. The results are then analyzed to provide the user with information about the mechanism and flag expected problem areas. A similar web tool is available at the CloudFlame website⁷ for screening gas-phase chemical kinetic models based on collision limit compliance.

Clearly, the availability of a searchable database of experimental combustion data, direct rate coefficient determinations, and combustion mechanisms should significantly assist combustion research. One of the aims of creating the reaction kinetics branch of ReSpecTh was to provide a hub, an all-in-one place, for measurements, rate coefficient determinations, reaction mechanisms, and tools for combustion simulations.

Note that databases of different types are also of high value in combustion science. For example, there is an effort to create databases of DNS simulation results of turbulent flames (see, e.g., the BLASTNET⁶¹ network).

In the field of high-resolution rovibronic spectroscopy, several data repositories have been set up (for example, in alphabetical order, CDMS¹⁴, GEISA¹³, HITRAN¹⁶, and JPL¹⁷, as well as several others concerning individual molecules and their isotopologues, such as the carbon dioxide spectroscopic database CDSD-296¹⁵). These databases contain measured or empirically calculated, mostly rovibrational transitions for a number of molecules, typically in various formats and with even more different verification procedures. Thus, there is a need for standardized database files, which can be validated by well-established protocols, like MARVEL (Measured Active Rotational-Vibrational Energy Levels)^{34,35}.

Spectral lines are described by a number of parameters: position, intensity, width, shape, labels referring to the upper and lower quantum states⁶², as well as self and foreign shift/ broadening due to collisions with many different partners. Of all this information, at present we aim to store and validate only line positions, uncertainties, and assignments.

The spectral line parameters may undergo significant changes over time as more modern measurement techniques are made available, improving sensitivity, resolution, precision, and overall accuracy. Measurements of gradually increasing accuracy not only substantially alter the positions and the uncertainties of the individual lines, but often result in assignment changes, as well. Numerous conflicts have been detected among data reported by different experimental groups due to the use of inconsistent labels. Unfortunately, the assignment of spectral lines using first-principles computations remains a significant challenge even today⁶³, especially at higher excitation. Hence, there is a considerable need for developing algorithms capable of verifying and merging a large number of measured and assigned lines. In 2007, an IUPAC Task Group decided to attempt to refine our understanding about the measured rovibrational spectra of one of the (if not the) most important molecules, water. One of the outcomes of this effort was the MARVEL algorithm^{34,35,64–66}. Through an inversion process, MARVEL inverts the information contained in the measured lines, yielding empirical rovibronic energy levels. Then, the task is to ensure that all the experimental, empirical, and computed spectroscopic information available is self-consistent and the data put into a spectroscopic database are as complete as possible.

The MARVEL approach has been used to analyze a large number of spectra measured for a large number of molecules, including diatomics (¹²C₂^{67,68}, ¹²CH⁶⁹, ¹⁴NH⁷⁰, ¹⁶OH⁶⁹, ¹⁶O₂⁷¹, ⁴⁸Ti¹⁶O⁷², and ⁹⁰Zr¹⁶O⁷³), triatomics (several isotopologues of carbon dioxide^{74,75}, H₃⁺⁷⁶, H₂D⁺ and D₂H⁺⁷⁷, nine water isotopologues^{78–86}, two isotopologues of H¹⁶OCl^{87,88}, H₂³²S⁸⁹, ¹⁴N₂¹⁶O⁹⁰, and four sulphur isotopologues of S¹⁶O₂⁹¹), tetratomics (¹²C₂H₂⁹², H₂¹²C¹⁶O⁹³, and ¹⁴NH₃^{94,95}), and even a pentatomic molecule, H₂¹²C₂¹⁶O⁹⁶.

For stable molecular species, a number of thermochemical data compilations are available^{20–24,27,28}, utilized in a number of scientific and engineering applications. These databases are characterized by rather different levels of accuracy and occasionally contain results which do not agree within the stated uncertainties. The most important thermochemical data are the enthalpies of formation, the bond dissociation energies, the reaction

free energies, and the temperature-dependent ideal gas partition functions, $Q(T)$, and quantities which can be derived straightforwardly from $Q(T)$ ^{97–100}.

Enthalpies of formation and heat capacities of stable molecules have traditionally been determined by calorimetry. Derivation of some additional thermochemical quantities, such as entropies, has been based on results of detailed spectroscopic measurements. Short-lived intermediates are usually not amenable to direct calorimetric measurements, and their spectroscopic characterization also requires special efforts and sophisticated instrumentation. Thus, experimental thermochemical data for a large number of species of interest is either missing from the databases or have relatively large uncertainties (or their uncertainties are significantly underestimated)¹⁰¹.

A new era in the treatment and validation of thermochemical data came with the introduction of the ATcT (Active Thermochemical Tables) approach^{25,28,29,101}, yielding high-quality empirical thermochemical quantities, in particular enthalpies of formation, as a function of temperature. The ATcT approach has been used successfully in a number of cases, including the determination of the enthalpies of formation of radicals²⁶.

The recent blossom of research in the field of thermochemistry is due to several important developments. Besides new measurement techniques, the ability of first-principles quantum chemistry to produce accurate thermochemical quantities has improved tremendously, even for open-shell systems and molecules containing heavier elements. Modern first-principles thermochemistry relies on advanced techniques of electronic-structure theory, allowing the computation of accurate thermochemical quantities with well-defined uncertainties^{26,102}. For the determination of temperature-dependent, effective quantities, one should use variational-like techniques of nuclear-motion theory, readily available in the fourth age of quantum chemistry¹. One of the key questions in electronic-structure theory concerns the accuracy of the computed molecular quantities, including relative energies. The composite focal-point analysis (FPA) approach developed more than three decades ago by one of the authors^{103,104} provides a clean and straightforward way to determine the uncertainties of computed relative energies. It should also be noted that the accuracy of the computed first-principles enthalpies of formation can be improved significantly by the use of properly chosen reaction schemes¹⁰⁵, which also offer improved capability for computational thermochemistry of larger systems.

Traditionally, the computational approaches to thermochemistry relied either on simple assumptions, like the harmonic-oscillator¹⁰⁶ and rigid-rotor¹⁰⁷ models, or at best on effective Hamiltonians (EH)¹⁰⁸ fitted to experimental data. These modeling efforts have resulted in relatively poor predictions with unknown uncertainties, especially at higher temperatures. First-principles thermochemistry is able to determine highly accurate partition functions and related thermochemical data, with well-defined uncertainties: the direct summation technique, based on accurate rovibrational energy levels with known small uncertainties, offers a straightforward way to determine accurate thermochemical quantities over a wide temperature range (even up to temperatures of 6000 K).

According to old conventions of thermochemistry, the definition of the enthalpy of formation does not reference atomic information and the enthalpies of formation of all elements in their standard states are zero, independent of the temperature. Advantages of atom-based thermochemistry (AT) over the conventional element-based one have been pointed out several times^{109–114}. One problem with the element-based convention is that the elements in their standard states may be gases, liquids, or solids of different constitution. Another one is that for *ab initio* quantum chemical energy and enthalpy calculations, which are bound to have an ever increasing role in the determination of thermochemical quantities, the natural reference is the atom in its ground electronic state.

Methods

Reaction kinetics. Experimental chemical kinetics data are called *direct* if the aim of the experiment is to determine the rate coefficient of an elementary reaction at a set of given conditions (temperature, pressure, and bath gas). A typical experimental method to obtain direct kinetic data is laser photolysis – laser-induced fluorescence detection (LP-LIF). High-level first-principles computations can also be used to determine the rate coefficient of an elementary reaction at a given temperature, pressure, and bath gas, and these results, when accompanied by proper uncertainty estimates, can be treated similarly to those of direct measurements.

Experimental kinetics data are called *indirect* if rate coefficients cannot be determined directly from them, but the measured value can be reproduced *via* simulation using a detailed reaction mechanism. The indirect combustion data in ReSpecTh contain ignition delay times, measured in shock tubes and rapid compression machines (RCMs), as well as laminar flame velocities and concentration profiles, observed using facilities like laminar and turbulent flow reactors, micro-flow reactors, jet-stirred reactors (JSRs), shock tubes, and burner-stabilized flames.

Our kinetics dataset³⁰ contains more than 162 000 kinetics data points in more than 3600 XML data files of indirect experimental data. The order of adding the data roughly followed the hierarchy of fuels, as defined by Westbrook and Dryer¹¹⁵ (see also Fig. 1).

The ReSpecTh Kinetics Data Format. In our previous studies^{116–133}, the important elementary reaction steps under various conditions were identified using local sensitivity analysis and the results of several direct determinations of the corresponding rate coefficients were encoded in the ReSpecTh Kinetics Data (RKD) format. Both the direct and the indirect data files in the reaction kinetics branch of ReSpecTh were encoded in the RKD format, the latest version of which is ver. 2.5³³.

Most of the RKD specifications are directly derived from the XML elements and attributes defined in PrIme⁶, extended in various directions, making the two incompatible. The detailed RKD Format Specification³³ describes not only the possible tags and attributes, but also how they should be interpreted. The RKD format (a) contains specifications which provide unambiguous definitions for the storage of experimental combustion data

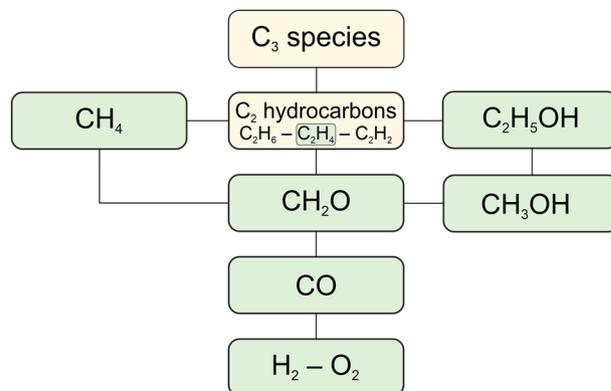


Fig. 1 The hierarchy of hydrocarbon fuels according to Westbrook and Dryer¹¹⁵. Green shadings indicate fuels present in ReSpecTh.

and rate coefficient determinations, and (b) describes an XML document type definition which allows flexible data representation and easy extension of the format specification.

Our aim with the provision of a data-format specification for combustion experiments, along with experimental and theoretical rate-coefficient determinations, was the support of long-term, laboratory-independent storage of related data and metadata. Storage of data in the RKD format allows an unambiguous interpretation of the data for simulations, as well as sharing and archiving the data in databases. Both might have very important long-term impacts on the combustion community. The unambiguous data format prevents misinterpretation and caters all the information needed to set up the related simulations. Sharing RKD format files in databases makes combustion research considerably more efficient.

According to the traditional workflow, researchers had to collect the experimental data from individual publications. Keyword-based search engines for scientific publications are not effective when someone is looking for a specific experimental condition. Even if relevant publications were found, the data had to be extracted from figures or files readable only by humans. This requires a lot of effort and is certainly error-prone. Unfortunately, even in some recent publications, the measured data are reported only in figures, or key information is missing or remains ambiguous. These potential problems are eliminated by having the experimental results reported in RKD format files.

Archiving data is also a very important issue. It occurs frequently that important details are missing from the published paper and that the research group does no longer exist. In other cases, the authors can be contacted, but the paper logs have been lost, while the electronic lab logs have been deleted, or use a format or storage media that cannot be read anymore. Therefore, it is very important to make all experimental data available in public databases, which allows the community to use them in the long run.

The XML data format allows (a) flexible data representation, (b) easy extension of the format specification, and (c) the introduction of new tags which do not invalidate the older ones. The RKD format was substantially revised once, indicated by a change from ver. 1.6 to 2.0, then it was extended several times during the last few years, indicated by changes in the first decimal of the version number. All these minor extensions enabled new features, preserving, at the same time, backward compatibility. Further extensions are planned by the addition of new types of experiments, like encoding extinction limits, soot measurements, and experimental data measured in stretched flames.

The PrIME data format included references to the bibliographic database of the PrIME server and the names of the species were linked to the species identification dataset of PrIME. Therefore, using a PrIME experimental data file required a functioning PrIME server, allowing these connections. The RKD format was designed to work standalone. Using an RKD format file does not require any link to the ReSpecTh website. All bibliographic data are stored in XML format BibTeX fields and all publications are identified with DOI links (if they exist). This means that all original publications can be immediately inspected by clicking on the provided DOI link. The species defined in the experimental data files are identified using CAS, InChI, and SMILES descriptors, which allows straightforward linking to internet databases describing properties of these substances.

Data files obeying the RKD format contain two main sections. The first one identifies the source of the data and provides information about the creation of the file. The data source is not only referenced to the literature using BibTeX keywords, but a detailed description shows the exact location of the information within the article. There is information about the creator of the file, the version of the RKD format, the date of the first creation, and the date of the last modification. The second section is divided into a few subsections. In the indirect experimental RKD files there are three main subsections. The first one identifies the experimental setup (type of experiment, apparatus, and measurement-related keywords, *e.g.*, the definition of the ignition delay time used in the measurement). The second one provides data about the experimental conditions that were kept constant during the experiment (*e.g.*, pressure, temperature, equivalence ratio, and initial concentration). The third subsection provides the varied experimental condition(s) and the measured values (*e.g.*, ignition delay time, laminar burning velocity, and species concentration). The proper definition of the ignition delay time should be emphasized, since the application of an inappropriate definition may lead to significant errors in simulated ignition delay times¹³⁴. The uncertainty of the measured data can also be provided. The rate determination RKD

Fuel	#files	#datasets	#data points	#models	References	
					mechanism comparison	optimization
hydrogen	177	208	2327	21	Olm <i>et al.</i> ¹¹⁶	Varga <i>et al.</i> ¹¹⁷
syngas	387	423	4852	17	Olm <i>et al.</i> ¹¹⁸	Varga <i>et al.</i> ¹¹⁹
methanol	287	584	14849	19	Olm <i>et al.</i> ¹²¹	Olm <i>et al.</i> ¹²¹
ethanol	192	504	16808	17	Olm <i>et al.</i> ¹²⁰	Olm <i>et al.</i> ¹²⁰
methane	1294	1294	11153	14	Zhang <i>et al.</i> ^{126,127}	
ethylene	219	397	4323	14	Su <i>et al.</i> ¹³⁰	Su <i>et al.</i> ¹³⁶
hydrogen/NO _x	207	332	4949	18	Kovács <i>et al.</i> ¹²²	Kovács <i>et al.</i> ¹²²
syngas/NO _x	61	145	1834		Kovács <i>et al.</i> ¹²⁵	Kovács <i>et al.</i> ¹³²
methanol/NO _x	74	243	2574	21	Kovács <i>et al.</i> ¹²⁵	Kovács <i>et al.</i> ¹³²
ammonia	351	527	5278	44	Kawka <i>et al.</i> ¹²³	
					Szanthoffer <i>et al.</i> ^{124,133}	
butanol	266	475	89347	29	Bolla <i>et al.</i> ¹²⁹	
pentane	72	226	3629			Wang <i>et al.</i> ¹³¹

Table 1. Indirect reaction kinetics datasets collected and made available within ReSpecTh.

format files (referring to direct measurements and theoretical calculations) contain two main subsections. The first one describes the elementary chemical reaction for which the rate coefficients were determined. The second subsection contains the values of the measured or calculated rate coefficients.

A valid indirect experimental RKD format file must contain all information needed to set up a simulation using the experimental data encoded. In the case of concentration measurements, one dataset may contain more than one concentration profile, when several concentrations are measured in the same experiment. These concentration profiles, called data series, are stored together in a single RKD format file. Some experiments need additional profiles to consider, *e.g.*, temperature-distance profile in burner-stabilized flame experiments, or pressure-time profiles in ignition-delay-time experiments. These profiles can be stored in the same XML file.

The utility code `Optima++`³⁶ can be used to create new RKD format data files and check the consistency of external RKD format files. `Optima++` can be applied for an automatic testing of reaction mechanisms using selected datasets.

Detailed combustion mechanisms. The indirect combustion experimental data have been used for testing published reaction mechanisms. For the combustion of each investigated fuel, detailed reaction mechanisms, in Chemkin format¹³⁵, were collected from the authors' websites, the supplementary files of publications, or obtained directly from the authors. These mechanisms were also converted to OpenSMOKE++³⁷ and Cantera³⁸ mechanism formats, providing the transport files and the thermodynamic files in NASA polynomial format. The aim has been to acquire, for each fuel, all mechanisms published during the last 20 years. These mechanisms are related to the combustion of hydrogen^{116,117}, syngas^{118,119}, methanol¹²¹, ethanol¹²⁰, methane (ignition delay time)¹²⁶, ethylene^{130,136}, H₂/O₂/NO_x mixtures¹²², syngas/NO_x mixtures¹²², ammonia^{123,124,133}, methanol/NO_x mixtures¹²⁵, butanol¹²⁹, and pentane¹³¹ (see Table 1). Furthermore, using the collection of indirect and direct data, new, significantly more accurate reaction kinetic models were created mainly via kinetic parameter optimization for the combustion of hydrogen¹¹⁷, syngas¹¹⁹, methanol¹²¹, ethanol¹²⁰, H₂/O₂/NO_x¹²² and methanol/NO_x¹²⁵ fuels. These six optimized mechanisms can be downloaded in Chemkin¹³⁵, OpenSMOKE++³⁷, and Cantera³⁸ formats. Until now, no thematic collection of combustion mechanisms existed, and the present project provides a unique collection of combustion mechanisms.

Utility codes. A series of computer codes are available in ReSpecTh which can be used for the analysis, reduction, and optimization of detailed reaction mechanisms, for sensitivity and uncertainty analyses, as well as for the quantification of the uncertainty of experimental data and model parameters. In most cases, the codes are available as source codes, accompanied by manuals and sample input/output files. We also host several similar computer codes that were developed by authors we have collaborated with. Most codes have both Windows and Linux versions.

`Optima++` is the central code in the reaction kinetics branch of ReSpecTh. It has a command-line version and a GUI-based version. `Optima++` reads the data provided in text files and prints out XML data files in proper RKD format. The code checks the data for completeness (whether all physical conditions are provided and the experiment is well-defined) and identifies possible contradictions (*e.g.*, whether the reaction temperature was defined multiple times). `Optima++` itself cannot be used to perform combustion simulations, but it reads RKD format data files, and sets up input files for simulation codes FlameMaster³⁹, OpenSMOKE++^{37,137}, Cantera¹³⁸, or Zero-RK⁴⁰. `Optima++` is able to run these simulation packages on the same PC, on an external PC, or on an external computer cluster. Then, `Optima++` reads the output files of the simulation codes and processes them.

`Optima++` is able to test a combustion mechanism against experimental data. It reads the names of the RKD data files to be used and carries out all simulations using a selected mechanism. Using the estimated uncertainty of the experimental data encoded in the RKD format indirect data files, an error function value is calculated that shows the squared difference between the experimental data and the simulation results compared to the variance

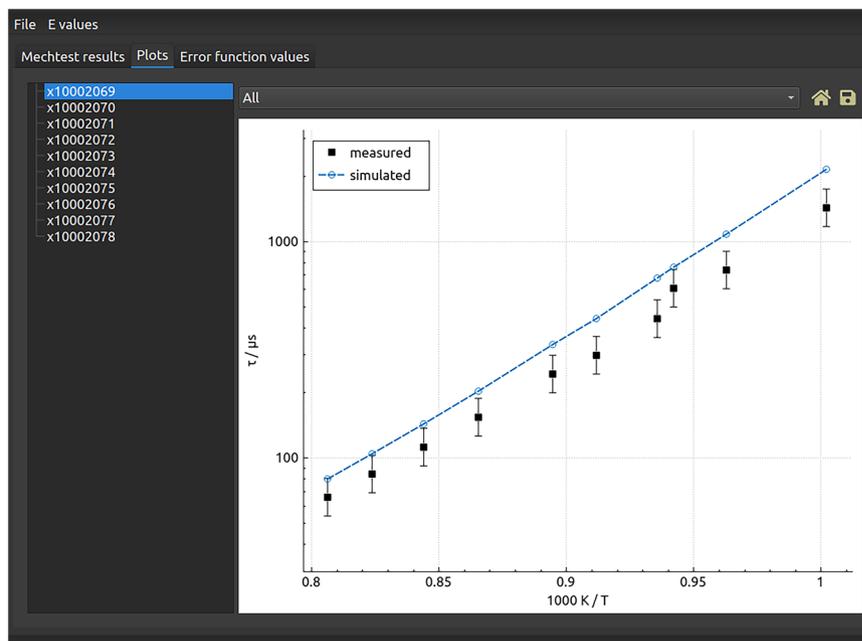


Fig. 2 Optima++-generated figure that compares the experimental data (solid squares), the experimental 2σ uncertainty (error bars), and the simulation results (dots interconnected by dashed lines). The list of data files used are given on the left hand side.

of the experimental data. Optima++ is also able to create a series of figures which makes a visual comparison of the experimental data and simulation results (see Fig. 2). Figure 3 illustrates the performance of several mechanisms on the same set of experimental data. Optima++ can be used to carry out a local sensitivity analysis; it (a) identifies the most important reaction steps and thermodynamic parameters at each experimental data point, and (b) provides a list of the important reactions and thermodynamic parameters at all data points.

Finally, Optima++ is able to carry out the optimization of detailed reaction mechanisms^{41,139,140}. The code fits the Arrhenius A , n , and E parameters of selected reaction steps and certain thermodynamic parameters to selected experimental data. These thermodynamic parameters are NASA polynomial coefficients¹⁴¹ a_1 , a_6 , and a_7 , which are the temperature-independent parameters of the heat capacity at constant pressure, enthalpy of formation, and molar entropy. Optima++ allows using the FOCTOPUS^{140,142}, Nelder–Mead simplex¹⁴³, and BOBYQA^{144,145} algorithms. In our detailed investigations¹⁴² we have found that FOCTOPUS should be the default method, but the Nelder–Mead simplex and the BOBYQA methods allow fast and accurate optimization in simpler cases, especially when only the Arrhenius A parameters are fitted. The optimized mechanism obtained is then to be tested with Optima++ on a wider range of experimental data.

Minimal Spline Fit is a utility code for the calculation of the random noise in an experimental dataset¹⁴⁶. It determines a minimal-parameter spline that captures the noise-free evolution of a noisy dataset and it estimates the standard deviation of the noise based on the fitted curve. The code fits Akima splines with an increasing number of control points (m), and fits polynomials with increasing order ($m = 0, 1, 2, \dots$). The optimal fitted curve can be deduced from the evolution of the fitting error as a function of the parameters of the fitting curve.

FluxViewer++ is a reaction path visualization tool, which reads the flux information file generated by Optima++. The code draws the species as boxes and the element fluxes as interconnecting arrows. The boxes and arrows can be moved by the mouse to create an aesthetic arrangement. The boxes, the arrows, and the other visualization settings can be saved into a session file for later reuse. The final figure can be saved in commonly used image formats (.png, .jpeg, and .bmp) in a chosen resolution. Development and change of the connections with reaction progress can be inspected as an animation. The frames of the animation can also be saved to create a video or an animated GIF file.

SEM is a mechanism reduction code based on the method of simulation error minimization (SEM)^{147,148}. Application of SEM requires significantly more CPU time than that of the widely used DRG or DRGEP methods, but the reduced mechanism found by SEM is much smaller, at the same level of agreement between the full and reduced mechanisms (“simulation error”). Furthermore, the allowed simulation error is an input parameter for SEM.

ReactionKinetics is a Mathematica-based package that investigates reaction mechanisms using a selected set of structural and graph-theoretical approaches, as well as techniques related to the time evolution of the concentration sets obtained from the considered mechanisms. The hydrogen and syngas reaction mechanisms present in the mechanism collection of ReSpecTh have been analyzed¹⁴⁹ with ReactionKinetics.

High-resolution spectroscopy. The MARVEL algorithm^{34,35,64–66} is applicable to molecules of arbitrary size and complexity and it was run over the entire set of input spectroscopic data present in ReSpecTh. This

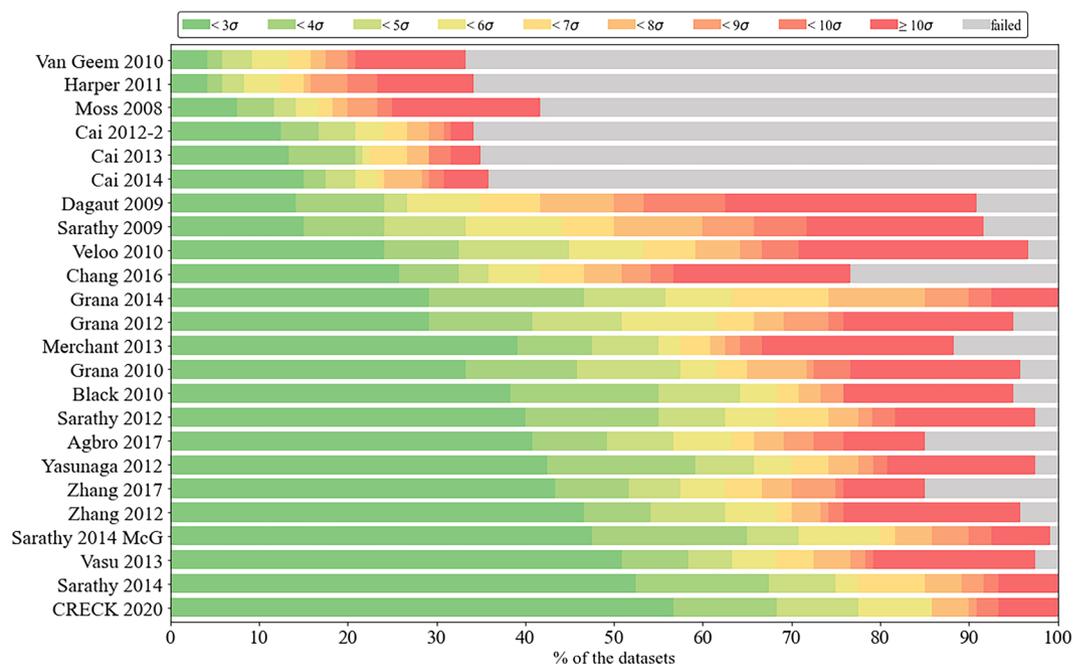


Fig. 3 Comparison of the performance of 24 detailed butanol combustion reaction mechanisms on *n*-butanol combustion experimental data (Optima++-generated figure). Various colored bars (olive, green, yellow, and orange) show the percentage of the data points that were reproduced within integer multiples of the experimental uncertainty σ . Red bars show the percentage of data that were reproduced higher than 10σ uncertainty, while the grey bars show the percentage of failed simulations.

involved the following steps. (1) For each molecule, all measured transitions reported in the literature were collected, converted into a predefined format, and compiled into a dataset, serving as the input for the MARVEL code. Although the supplementary information attached to articles nowadays is of considerable help during the addition of dataset entries, the transitions reported in older papers, unfortunately, had to be entered into the dataset manually or with the help of a scanner and a character recognition software. As a result, even a carefully prepared dataset may contain errors and typos which need to be checked and corrected before performing a calculation with the MARVEL code. (2) Using a least-squares method, the best possible set of empirical rovibronic energy values was determined with the help of MARVEL, followed by the validation of the measured lines with these energies. (3) When a discrepancy was found between an observed position and its MARVEL-predicted counterpart, the indicated assignment issue, line-position problem, or underestimated uncertainty was resolved. After resolving a conflict, MARVEL was rerun repeatedly to guarantee that all conflicts were eliminated. At the end of the iterative MARVEL process, self-consistent datasets emerged, that is, ones without internal conflicts. Nevertheless, these may not be correct datasets, as the input data may contain errors that cannot be revealed without additional information. The list of molecules for which spectroscopic datasets were created is given in Table 2.

As seen in Table 2, we collected and validated more than one million measured transitions. Using these transitions, almost 200 000 rovibronic energy levels were derived.

Utility codes. The browser-based custom code MARVELOnline, with a graphical user interface, was used during the collection and validation of the spectroscopic data. It is available at <http://kkrk.chem.elte.hu/marvelonline/>.

TransitionFileConverter is a tool which converts older versions of a MARVEL input file into the latest one. Over time, there have been several formats for the transition files. Therefore, the need arose for a tool which can be accessed online to convert these older formats to the format required by the current version of MARVEL. A line in the current version of the input transitions file contains the following data: (a) line position, (b) initial uncertainty of the line position, (c) an actual (MARVEL) uncertainty, (d) descriptors, usually rovibrational assignments and symmetry information, characterizing the upper and lower states involved in the transition, and (e) line tag. There is no limit to the number of descriptors used for identifying a state, but each state must have a unique label with the same number of descriptors, separated by a white space. Since from quantum theory MARVEL utilizes only the Ritz principle⁶², it is not required that all descriptors, or even just one, have a clear physical meaning. The line tag provides provenance, it is used for an easy identification of the transitions, and it must be unique for each line in the transition file.

CheckTransitionFile inspects whether the transition file generated by the user meets the requirements of the current MARVEL version. In addition, this tool also investigates the fulfillment of some basic requirements: (a) the initial/actual transition uncertainties need to have nonzero values, (b) the upper and lower states have to be different, and (c) the quantum states in the transition file must have the same number of labels.

Molecule	Isotopologue	#lines	#levels	Year	Reference
AlH	²⁷ AlH	682	259	2018	Yurchenko <i>et al.</i> ¹⁵⁸
BeH	⁹ BeH	2201	1264	2018	Darby-Lewis <i>et al.</i> ¹⁵⁹
	⁹ BeD	2605	1495	2018	Darby-Lewis <i>et al.</i> ¹⁵⁹
	⁹ BeT	538	215	2018	Darby-Lewis <i>et al.</i> ¹⁵⁹
C ₂	¹² C ₂	31 323	7047	2020	McKemmish <i>et al.</i> ⁶⁸
CH	¹² CH	6348	1521	2022	Furtenbacher <i>et al.</i> ¹⁶⁰
CN	¹² C ¹⁴ N	40 516	8083	2020	Syme <i>et al.</i> ¹⁶¹
CP	¹² C ³¹ P	3264	948	2021	Qin <i>et al.</i> ¹⁶²
NH	¹⁴ NH	3002	1058	2019	Darby-Lewis <i>et al.</i> ⁷⁰
NO	¹⁴ N ¹⁶ O	11 136	4106	2017	Wong <i>et al.</i> ¹⁶³
O ₂	¹⁶ O ₂	30 671	15 946	2019	Furtenbacher <i>et al.</i> ⁷¹
OH	¹⁶ OH	15 938	1624	2022	Furtenbacher <i>et al.</i> ¹⁶⁰
PN	³¹ P ¹⁴ N	1735	61	2021	Qin <i>et al.</i> ¹⁶⁴
TiO	⁴⁸ Ti ¹⁶ O	56 240	10 761	2019	McKemmish <i>et al.</i> ¹⁶⁵
ZrO	⁹⁰ Zr ¹⁶ O	22 549	8088	2018	McKemmish <i>et al.</i> ⁷³
CaOH	⁴⁰ Ca ¹⁶ OH	3 240	1 955	2020	Wang <i>et al.</i> ¹⁶⁶
H ₂ O	H ₂ ¹⁶ O	309 290	19 027	2024	Furtenbacher <i>et al.</i> ⁸³
	H ₂ ¹⁷ O	27 045	5278	2020	Furtenbacher <i>et al.</i> ⁸⁶
	H ₂ ¹⁸ O	66 166	6865	2020	Furtenbacher <i>et al.</i> ⁹⁵
	HD ¹⁶ O	54 740	8819	2010	Tennyson <i>et al.</i> ⁸⁰
	HD ¹⁷ O	485	162	2010	Tennyson <i>et al.</i> ⁸⁰
	HD ¹⁸ O	8728	1864	2010	Tennyson <i>et al.</i> ⁸⁰
	D ₂ ¹⁶ O	53 534	12 269	2014	Tennyson <i>et al.</i> ⁸²
	D ₂ ¹⁷ O	600	338	2014	Tennyson <i>et al.</i> ⁸²
D ₂ ¹⁸ O	12 146	3350	2014	Tennyson <i>et al.</i> ⁸²	
H ₂ S	H ₂ ³² S	44 325	7436	2018	Chubb <i>et al.</i> ⁸⁹
H ₃ ⁺	H ₃ ⁺	1610	652	2013	Furtenbacher <i>et al.</i> ⁷⁶
	H ₂ D ⁺	195	86	2013	Furtenbacher <i>et al.</i> ⁷⁶
	D ₂ H ⁺	154	72	2013	Furtenbacher <i>et al.</i> ⁷⁶
HOCl	H ¹⁶ O ³⁵ Cl	20 349	5760	2023	Ecseri <i>et al.</i> ⁸⁸
	H ¹⁶ O ³⁷ Cl	10 266	3933	2023	Ecseri <i>et al.</i> ⁸⁸
SO ₂	³² S ¹⁶ O ₂	40 325	15 130	2018	Tóbiás <i>et al.</i> ⁹¹
	³³ S ¹⁶ O ₂	15 647	5852	2018	Tóbiás <i>et al.</i> ⁹¹
	³⁴ S ¹⁶ O ₂	31 088	10 893	2018	Tóbiás <i>et al.</i> ⁹¹
	³⁶ S ¹⁶ O ₂	31	—	2018	Tóbiás <i>et al.</i> ⁹¹
NH ₃	¹⁴ NH ₃	29 450	4961	2015	Al Derzi <i>et al.</i> ⁹⁴
C ₂ H ₂	¹² C ₂ H ₂	37 813	11 213	2018	Chubb <i>et al.</i> ⁹²
H ₂ CO	H ₂ ¹² C ¹⁶ O	39 662	5029	2021	Al-Derzi <i>et al.</i> ⁹³
H ₂ C ₂ O	H ₂ ¹² C ₂ ¹⁶ O	3879	353	2011	Fábri <i>et al.</i> ⁹⁶

Table 2. Spectroscopic (MARVEL) datasets available within ReSpecTh.

Subgraph can display a spectroscopic subnetwork built around a given state. The user can choose the depth of the display, *i.e.*, how many neighboring levels should appear on the figure (the maximum value is 10). This tool is particularly useful to understand which transitions and which states affect the energy value of a particular state.

EnergyTable allows the tabulation of the empirical energy values deduced at the end of a MARVEL analysis. It is possible to group these values, and to obtain important statistical information (*e.g.*, mean and standard deviations, maximum and minimum values), as well, for the individual groups.

Thermochemistry. Computation of quantum-chemical energy and enthalpy results, including reaction enthalpies, is based more naturally on the electronic ground states of the atoms and not on the historically preferred elemental states. Since these two possible protocols can be converted into each other straightforwardly, it was proposed, following earlier suggestions^{109–113}, that first-principles thermochemistry, like NEAT¹¹⁴, standing for a “network of computed reaction enthalpies leading to atom-based thermochemistry”, should employ the ground electronic states of atoms. In atom-based thermochemistry, the enthalpy of formation for a gaseous compound corresponds simply to the total atomization energy (TAE) of that species; it is always positive, and it reflects the bonding strength within the molecule.

The NEAT dataset collected and put in the OSF (Open Science Framework) repository³⁰ contains computed reaction enthalpies for 355 reactions. This dataset was used to determine 0 K enthalpies of formation for 188 species.

Data Records

All ReSpecTh datasets are made publicly available at the following OSF repository:³⁰ <https://osf.io/nbmzv/>. The datasets of each branch of ReSpecTh are put into separate folders, *i.e.*, the top level of the OSF repository contains three folders: *ReactionKinetics*, *Spectroscopy*, and *Thermochemistry*.

Reaction kinetics. The *ReactionKinetics* folder contains two sub-folders called *indirect* and *direct*. The *direct* folder contains the results of direct experimental or theoretical determination of reaction rate coefficients. These types of experiments determine the rate coefficient of an elementary reaction at a given set of conditions (temperature, pressure, and bath gas).

As to indirect experimental combustion data, collections were created according to fuel and measurement types. A search in the indirect experimental datasets is allowed according to the chemical composition of the initial mixture, experiment type, and the ranges of temperature, pressure, and equivalence ratio. This way, it is straightforward to locate all experimental data within ReSpecTh according to a given set of experimental conditions, which can significantly assist mechanism development. It is also possible to search according to bibliographic data of the experimental publication, such as author names, DOI number, and the year of the publication. All data files found by a search condition can be downloaded in a single step.

At the conditions of the indirect experimental data, the most important reaction steps have been identified for the combustion of the fuels hydrogen, syngas, methanol, ethanol, hydrogen/NO_x, syngas/NO_x, methanol/NO_x, methane, ethylene, and ammonia. The direct experimental and theoretical determinations of these rate coefficients were encoded in RKD format data files. ReSpecTh contains 8451 data points in 443 XML data files in this category. The direct data cover not only high-temperature, but also room-temperature and sub-room-temperature ranges; therefore, these data are also of interest in astronomy or atmospheric chemistry. This collection of experimentally or theoretically determined gas-phase rate coefficients is also used on the so-called k-evaluation website¹⁵⁰. This is an interactive website that allows the determination of the temperature-dependent uncertainty of the rate coefficients, using both the data stored there and the data added by the user.

Note that both the categories of indirect and direct methods are organized according to the type of information and not the experimental method. For example, the laminar burning velocities are measured by one of the five main experimental approaches and these approaches have several subcategories. Furthermore, the category of direct measurements covers several very different experimental methods. However, the measured quantity is the same and therefore the exact experimental method is indicated only as auxiliary information.

High-resolution spectroscopy. The folder *Spectroscopy* contains two files: *Spectroscopic_Datasets.zip* and *Spectroscopic_Datasets.pdf*. The spectroscopic datasets are organized into folders, they can be found in the file *Spectroscopic_Datasets.zip*. Each folder contains the experimentally measured transitions, collected for the given molecule, in a text file. The folder also contains, where it is available, the empirical rovibronic energy levels of the molecule. As the transitions measured for each molecule have been labeled in a molecule-specific way, each transition and energy file contains a different number of columns. Thus, before using a spectroscopic dataset, it is recommended to review the reference of the chosen molecule. These references can be found in the file *Spectroscopic_Datasets.pdf*.

The file *Spectroscopic_Datasets.zip* contains information for a total of 23 molecules and 39 isotopologues. Table 2 lists the molecules and the isotopologues treated, together with the number of measured transitions (#lines) collected, the number of empirical rovibrational energy levels (#levels) determined, and the reference to the article where the results were originally published. It is important to note that the OSF repository contains not only the current (latest) version of a given spectroscopic database, but all previous versions are available. For example, for the H₂¹⁶O molecule, three datasets are available in the OSF repository: the so-called PART III⁸¹, W2020⁸³, and the latest W2024⁸⁶ dataset.

Thermochemistry. The folder *Thermochemistry* contains three sub-folders: *AccurateThermochemistry*, *BURCAT*, and *NEAT*. The folder *AccurateThermochemistry* contains accurate thermochemical functions for the ¹²C₂, D₂O, H₂¹⁶O, and ¹⁶O₂ molecules, organized into subfolders. Each sub-folder contains a “README” file, which specifies the columns of the thermochemical files, given in text format. The folder *BURCAT* contains Burcat’s Thermodynamic Database, both in text and XML format, and a “README” file, which describes the format of these files. The folder *NEAT* contains two files. The file *NEAT_reactions.txt* contains the full list of reactions in NEAT format. The file header contains the description of the columns. The file *NEAT_enthalpies.txt* lists the computed NEAT enthalpies of the species. This file also has a header showing the column descriptions. For more details, we recommend to consult ref. ¹¹⁴.

Technical Validation

Reaction kinetics. When direct experimental data are published, the main text of the article, or the supplementary material accompanying the paper, usually includes a table containing the measured rate coefficients, while the text describes the temperature, the pressure, and the bath gas that is related to the rate coefficient. These articles usually contain figures that plot the rate coefficient values as a function of temperature (so-called Arrhenius plots) or pressure (so-called fall-off curves). The contents of the RKD format data files were converted to similar plots and these were compared with the published figures. A similar procedure was followed

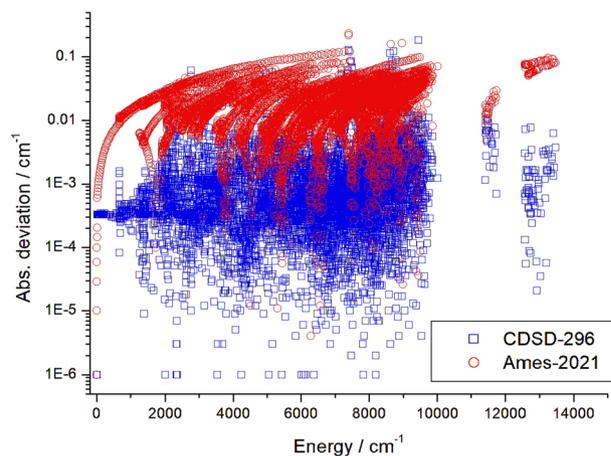


Fig. 4 Absolute deviations between the empirical rovibrational energies of the $^{16}\text{O}^{12}\text{C}^{18}\text{O}$ isotopologue of carbon dioxide obtained through the MARVEL approach⁷⁵ and those of CDS-296¹⁵ (blue squares) and Ames-2021¹⁵⁶ (red circles).

for the theoretically determined rate coefficients in case the calculated rate coefficients were provided. If only the Arrhenius parameters were published in a theoretical paper, the rate coefficients were calculated at several temperatures and stored in the RKD format data files.

Most of the indirect experimental data, like ignition delay times or laminar burning velocities, were not published in digital form, but provided as dots in the figures. In the case of a newer publication, the person creating the RKD datafile contacted the original authors by e-mail and requested the original values used to create the plots. For earlier publications, these values are typically not available. In this case, the dots of the figures were digitalized. The accuracy of the digitalization was checked in two ways. First, the data in the RKD file were plotted using the `Optima++` code and these plots were compared with the figures of the original publication. A complementary route was also followed, making simulations, with the help of the RKD files, with a new mechanism. Most of the data were well reproducible with these simulations.

Both the indirect and direct RKD data files were checked in such a way that the data points were compared to each other and the results of simulations. A small fraction, about 10 %, of the collected data was found to be potentially not reliable. These data were indicated by gray shading in the supplementary tables of our publications.

During the last decade, a part of the reaction kinetics data was collected by undergraduate students. The data files created by them were always systematically checked by PhD students or postdocs. In most cases, the data collected earlier were checked by a data evaluator, comparing them against the original publication.

High-resolution spectroscopy. A major reason behind the introduction of a network-based approach to high-resolution spectroscopy^{34,151} was the possibility to develop new tools for the validation of the diverse set of experimental high-resolution spectroscopic results present in line-by-line datasets. In fact, the MARVEL algorithm^{34,35,64} does provide several tools itself, based on graph theory^{65,66,152}, which help the validation of experimental line positions and the detection of incorrect entries in the list of transitions^{84,91,153–155}.

For each molecule reported in Table 2, an extensive validation process was carried out. It was found that the best way to identify incorrect or inaccurate literature entries is offered by the simulation of the measured spectrum, utilizing line positions deduced from the empirical energy levels, augmented with first-principles computed intensities. Given the uncertainties of the computed energy levels and transitions, a decision about the experimental line questioned could usually be made. When in doubt about which line causes the issue, the traditional techniques of experimental spectroscopists, based on combination differences, also proved to be helpful. Each publication, also indicated in Table 2, lists the transitions which had to be deleted during the MARVEL treatment of the data.

Effective Hamiltonian models are designed to act in a reduced space and describe part of the eigenvalue spectrum of a complete Hamiltonian. These models were also used to validate a number of measurements, especially those upon which they are based.

Figure 4 provides an example, for the case of the $^{16}\text{O}^{12}\text{C}^{18}\text{O}$ isotopologue of carbon dioxide⁷⁵ where rovibrational energies obtained by MARVEL are compared with those of an effective Hamiltonian model, contained in the CDS-296¹⁵ database. A comparison is also made there between MARVEL and a set of first-principles computed (Ames-2021¹⁵⁶) results. As expected and indeed seen in Fig. 4, the effective Hamiltonian model results in much more accurate energy levels than a first-principles computation. Nevertheless, Fig. 4 also shows that at about 7500 and 9000 cm^{-1} there are a number of MARVEL energy levels whose deviation from the EH results are much larger than the average (expected) deviation. In these cases, it is very difficult to decide whether the empirical MARVEL or the computed energies are correct, calling for new measurements, perhaps designed on the basis of network principles^{84,157}.

First-principles line lists help to detect problems with experimental datasets, and they were used indeed for the validation of MARVEL input data (for details, see the original publications listed in Table 2).

Species	NEAT	ATcT
H ₂ O	918.034(150)	917.83(3)
C ₂ H ₂	1625.826(360)	1626.16(24)
CH	334.675(150)	334.66(23)
NH ₃	1157.341(290)	1157.25(4)
CO	1071.936(260)	1072.13(9)
CH ₃	1209.576(280)	1209.63(13)
CH ₂	752.409(200)	752.70(26)
OH	425.84(150)	425.62(3)
CO ₂	1597.927(370)	1598.27(9)

Table 3. Comparison of selected NEAT-based enthalpies of formation with those of the ATcT approach. All values correspond to 0 K and are in kJ mol⁻¹. The uncertainties quoted in parentheses are the expanded uncertainties using a coverage factor of 2 (*i.e.*, the quoted uncertainty is $\pm 2\sigma$ for both datasets).

Thermochemistry. To check the accuracy of the enthalpies of formation derived within the NEAT approach, they were compared to their counterparts within the Active Thermochemical Tables (ATcT) database²⁵. Table 3 shows the comparison between the NEAT data and those of ATcT for a few closed- and open-shell species. Clearly, the two sets of values agree within the uncertainty limits and, as expected, ATcT has significantly tighter bounds.

Code availability

No custom code was used to create the datasets available in the OSF repository³⁰.

Received: 17 November 2024; Accepted: 15 May 2025;

Published online: 17 June 2025

References

1. Császár, A. G. *et al.* The fourth age of quantum chemistry: molecules in motion. *Phys. Chem. Chem. Phys.* **14**, 1085–1106 (2012).
2. Manion, J. A. *et al.* NIST Chemical Kinetics Database, NIST Standard Reference Database 17, Version 7.0 (Web Version), Release 1.6.8, Data version 2024, National Institute of Standards and Technology, Gaithersburg, Maryland, 20899-8320., <https://kinetics.nist.gov>.
3. IUPAC Task Group on Atmospheric Chemical Kinetic Data Evaluation, <https://www.aeris-data.fr/en/projects/iupac-2/> (2023).
4. Database for the Kinetics of the Gas-Phase Atmospheric Reactions of Organic Compounds, <https://data.eurochamp.org/> (2023).
5. Database for the Kinetics of Gas-Phase Reactions of Organic Species, <https://www2.acom.ucar.edu/sections/laboratory-kinetics-database-kinetics-gas-phase-atmospheric-reactions-organic-species> (2023).
6. Frenklach, M. Prime website, <http://www.primekinetics.org> (2011).
7. Cloudflame, <https://ccrc.kaust.edu.sa/aramco/tools> (2023).
8. ChemKED-database, <https://github.com/pr-ometh-e-us/ChemKED-database> (2023).
9. UConn Combustion Database, <https://combdialab.engr.uconn.edu/database/> (2023).
10. Davidson, D. F. & Hanson, R. K. Fundamental Kinetics Database Utilizing Shock Tube Measurements, <https://exhibits.stanford.edu/data/catalog/kb621cw6967> (2019).
11. McClatchey, R. A. *et al.* AFCRL atmospheric absorption line parameters compilation. Tech. Rep., Air Force Cambridge Research Laboratories, Hanscom Field, Bedford (1973).
12. Rothman, L. S. *et al.* AFGL atmospheric absorption line parameters compilation: 1982 edition. *Appl. Opt.* **22**, 2247, <https://doi.org/10.1364/ao.22.002247> (1983).
13. Jacquinet-Husson, N. *et al.* The 2015 edition of the GEISA spectroscopic database. *J. Mol. Spectrosc.* **327**, 31–72 (2016).
14. Endres, C. P., Schlemmer, S., Schilke, P., Stutzki, J. & Müller, H. S. P. The Cologne database for molecular spectroscopy, CDMS, in the virtual atomic and molecular data centre, VAMDC. *J. Mol. Spectrosc.* **327**, 95–104 (2016).
15. Tashkun, S. A., Perevalov, V. I., Gamache, R. R. & Lamouroux, J. CDS-296, high-resolution carbon dioxide spectroscopic databank: An update. *J. Quant. Spectrosc. Radiat. Transf.* **228**, 124–131 (2019).
16. Gordon, I. E. *et al.* The HITRAN2020 molecular spectroscopic database. *J. Quant. Spectrosc. Rad. Transf.* **276**, 107949 (2022).
17. Homepage of ATcT: <https://atct.anl.gov/>.
18. NIST Chemistry WebBook, <https://webbook.nist.gov> (2023).
19. Thermochemical Database (TDB) Project, https://www.oecd-nea.org/jcms/pl_22166/thermochemical-database-tdb-project (2023).
20. Cox, J. D., Wagman, D. D. & Medvedev, V. A. *CODATA key values for thermodynamics* (Hemisphere Publishing Corporation, New York, 1989).
21. Gurvich, L. V., Veyts, I. V. & Alcock, C. B. *Thermodynamic properties of individual substances* (Hemisphere, New York, 1989).
22. Gurvich, L. & Veyts, I. *Thermodynamic properties of individual substances: elements and compounds* (CRC Press, Boca Raton, 1990).
23. Pedley, J. B. *Thermochemical data and structures of organic compounds* (Thermodynamic Research Center, College Station, 1994).
24. Burcat, A. Third millennium ideal gas and condensed phase thermochemical database for combustion. Tech. Rep., Qiryat Ha Technion, Haifa (2001).
25. Ruscic, B. *et al.* Introduction to active thermochemical tables: several “key” enthalpies of formation revisited. *J. Phys. Chem. A* **108**, 9979–9997 (2004).
26. Ruscic, B. *et al.* IUPAC critical evaluation of thermochemical properties of selected radicals. Part I. *J. Phys. Chem. Ref. Data* **34**, 573–656 (2005).
27. Goldsmith, C. F., Magoon, G. R. & Green, W. H. Database of small molecule thermochemistry for combustion. *J. Phys. Chem. A* **116**, 9033–9057 (2012).
28. Ruscic, B. Active thermochemical tables: water and water dimer. *J. Phys. Chem. A* **117**, 11940–11953 (2013).
29. JPL catalog search form. <http://spec.jpl.nasa.gov/ftp/pub/catalog/catform.html>.
30. Turányi, T. *et al.* ReSpecTh, <https://doi.org/10.17605/OSF.IO/NBMZV> (2024).
31. Wilkinson, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).

32. Jacobsen, A. *et al.* FAIR principles: interpretations and implementation considerations. *Data Intelligence* **2**, 10–29 (2020).
33. Varga, T. *et al.* ReSpecTh Kinetics Data Format Specification v2.5, https://respech.elte.hu/ReSpecTh_Kinetics_Data_Format_Specification_v2.5.pdf (2024).
34. Császár, A. G., Czákó, G., Furtenbacher, T. & Mátyus, E. An active database approach to complete spectra of small molecules. *Annu. Rep. Comput. Chem.* **3**, 155–176 (2007).
35. Furtenbacher, T., Császár, A. G. & Tennyson, J. MARVEL: measured active rotational-vibrational energy levels. *J. Mol. Spectrosc.* **245**, 115–125 (2007).
36. Papp, M. *et al.* Optima++ v2.3 A general C++ framework for performing combustion simulations and mechanism optimization, <https://respech.elte.hu/compProgs/optima> (2023).
37. OpenSMOKE++ home page, <https://www.opensmokepp.polimi.it> (2023).
38. Cantera YAML Input File Reference, <https://cantera.org/documentation/docs-3.0/sphinx/html/yaml/index.html> (2023).
39. Pitsch, H. FlameMaster v4.0 A C++ computer program for 0D combustion and 1D laminar flame calculations, <https://www.itw.rwth-aachen.de/en/downloads/flamemaster/> (2016).
40. Zero-rk: Zero order reaction kinetics, <https://github.com/LLNL/zero-rk> (2023).
41. Wang, H. & Sheen, D. A. Combustion kinetic model uncertainty quantification, propagation and minimization. *Progr. Ener. Combust. Sci.* **47**, 1–31, <https://doi.org/10.1016/j.pecs.2014.10.002> (2015).
42. Westbrook, C. K. & Dryer, F. L. Chemical kinetic modeling of hydrocarbon combustion. *Progr. Ener. Combust. Sci.* **10**, 1–57, [https://doi.org/10.1016/0360-1285\(84\)90118-7](https://doi.org/10.1016/0360-1285(84)90118-7) (1984).
43. Burkholder, J. B. *et al.* Chemical kinetics and photochemical data for use in atmospheric studies, evaluation no. 19, <https://jpldataeval.jpl.nasa.gov> (2019).
44. Master Chemical Mechanism, MCM v3.3.1, <https://www.mcm.york.ac.uk>.
45. Frenklach, M., Packard, A., Seiler, P. & Feeley, R. Collaborative data processing in developing predictive models of complex reaction systems. *Int. J. Chem. Kinet.* **36**, 57–66 (2004).
46. Frenklach, M. Transforming data into knowledge – process informatics for combustion chemistry. *Proc Combust Inst* **31**, 125–140 (2007).
47. SciExpeM database: <https://scixpem.polimi.it>.
48. Ramalli, E. *et al.* Automatic validation and analysis of predictive models by means of big data and data science. *Chem. Eng. J.* **454**, 140149, <https://doi.org/10.1016/j.cej.2022.140149> (2023).
49. Weber, B. W. & Niemeyer, K. E. ChemKED A human- and machine-readable data standard for chemical kinetics experiments. *Int. J. Chem. Kinet.* **50**, 135–148 (2018).
50. Stanford University. Combustion kinetics database (2024). https://github.com/IhmeGroup/Combustion_Kinetics_Database.
51. Oak Ridge National Laboratory. Combustion kinetics database (2024). <https://github.com/jiweiqi/CollectionOfMechanisms>.
52. University of Connecticut. Combustion kinetics database (2024). https://github.com/haowu80s/Combustion_Kinetics_Database.
53. Combustion Chemistry Centre, University of Galway, Ireland, <https://universityofgalway.ie/combustionchemistrycentre/mechanismdownloads/> (2023).
54. The CRECK Modeling Group, POLIMI, Milano, Italy, <https://creckmodeling.chem.polimi.it/menu-kinetics/menu-kinetics-detailed-mechanisms/> (2023).
55. Combustion Research Group, UCSD, San Diego, CA, <https://web.eng.ucsd.edu/mae/groups/combustion/mechanism.html> (2023).
56. Combustion Kinetics Laboratory, USC, Los Angeles, CA, <https://ignis.usc.edu:80/Mechanisms/Model>.
57. Clean Combustion Research Center, KAUST, KSA, <https://ccrc.kaust.edu.sa/> (2023).
58. Lawrence Livermore National Laboratory, Livermore, CA, <https://combustion.llnl.gov/mechanisms> (2023).
59. LLNL Mech Checker, <https://combustiontools.llnl.gov> (2023).
60. Killingsworth, N. J., McNenly, M. J., Whitesides, R. A. & Wagnon, S. W. Cloud based tool for analysis of chemical kinetic mechanisms. *Combust. Flame* **221**, 170–179 (2020).
61. Chung, W. T. *et al.* Turbulence in Focus: Benchmarking Scaling Behavior of 3D Volumetric Super-Resolution with BLASTNet 2.0 Data. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, vol. 36, 77430–77484 (2023).
62. Ritz, W. On a new law of series spectra. *Astrophys. J.* **28**, 237–243 (1908).
63. Mátyus, E. *et al.* Assigning quantum labels to variationally computed rotational-vibrational eigenstates of polyatomic molecules. *J. Chem. Phys.* **133**, 034113 (2010).
64. Furtenbacher, T. & Császár, A. G. MARVEL: measured active rotational-vibrational energy levels. II. Algorithmic improvements. *J. Quant. Spectrosc. Rad. Transf.* **113**, 929–935 (2012).
65. Furtenbacher, T., Árendás, P., Mellau, G. & Császár, A. G. Simple molecules as complex systems. *Sci. Rep.* **4**, 4654 (2014).
66. Császár, A. G., Furtenbacher, T. & Árendás, P. Small molecules – big data. *J. Phys. Chem. A* **120**, 8949–8969 (2016).
67. Furtenbacher, T. *et al.* Experimental energy levels and partition function of the ¹²C₂ molecule. *Astrophys. J. Suppl. S.* **224**, 44 (2016).
68. McKemmish, L. K. *et al.* An update to the MARVEL data set and ExoMol line list for ¹²C₂. *Mon. Not. R. Astron. Soc.* **497**, 1081–1097 (2020).
69. Furtenbacher, T., Hegedüs, S. T., Tennyson, J. & Császár, A. G. Analysis of measured high-resolution doublet rovibronic spectra and related line lists of ¹²CH and ¹⁶OH. *Phys. Chem. Chem. Phys.* **24**, 19287–19301 (2022).
70. Darby-Lewis, D. *et al.* MARVEL analysis of the measured high-resolution spectra of ¹⁴NH. *J. Mol. Spectrosc.* **362**, 69–76 (2019).
71. Furtenbacher, T. *et al.* MARVEL analysis of the measured high-resolution rovibronic spectra and definitive ideal-gas thermochemistry of the ¹⁶O₂ molecule. *J. Phys. Chem. Ref. Data* **48**, 023101 (2019).
72. McKemmish, L. K. *et al.* MARVEL analysis of the measured high-resolution rovibronic spectra of ⁴⁸Ti¹⁶O. *Astrophys. J. Suppl. S.* **228**, 15 (2017).
73. McKemmish, L. K. *et al.* MARVEL analysis of the measured high-resolution rovibronic spectra of ⁹⁰Zr¹⁶O. *Astrophys. J.* **867**, 33 (2018).
74. Ibrahim, M. T. I. *et al.* MARVEL analysis of high-resolution rovibrational spectra of ¹³C¹⁶O₂. *J. Comp. Chem.* **45**, 969–984 (2024).
75. Alatoon, D. *et al.* MARVEL analysis of high-resolution rovibrational spectra of ¹⁶O¹²C¹⁸O. *J. Comp. Chem.* **45**, 2558–2573 (2024).
76. Furtenbacher, T., Szidarovszky, T., Mátyus, E., Fábri, C. & Császár, A. G. Analysis of the rotational-vibrational states of the molecular ion H₃⁺. *J. Chem. Theor. Comput.* **9**, 5471–5478 (2013).
77. Furtenbacher, T., Szidarovszky, T., Fábri, C. & Császár, A. G. MARVEL analysis of the rotational-vibrational states of the molecular ions H₂D⁺ and D₂H⁺. *Phys. Chem. Chem. Phys.* **15**, 10181–10193 (2013).
78. Furtenbacher, T. & Császár, A. G. On employing H₂¹⁶O, H₂¹⁷O, H₂¹⁸O, and D₂¹⁶O lines as frequency standards in the 15–170 cm⁻¹ window. *J. Quant. Spectrosc. Rad. Transf.* **109**, 1234–1251 (2008).
79. Tennyson, J. *et al.* IUPAC critical evaluation of the rotational-vibrational spectra of water vapor. Part I. Energy levels and transition wavenumbers for H₂¹⁷O and H₂¹⁸O. *J. Quant. Spectrosc. Rad. Transf.* **110**, 573–596 (2009).
80. Tennyson, J. *et al.* IUPAC critical evaluation of the rotational-vibrational spectra of water vapor. Part II. Energy levels and transition wavenumbers for HD¹⁶O, HD¹⁷O, and HD¹⁸O. *J. Quant. Spectrosc. Rad. Transf.* **110**, 2160–2184 (2010).
81. Tennyson, J. *et al.* IUPAC critical evaluation of the rotational-vibrational spectra of water vapor. Part III. Energy levels and transition wavenumbers for H₂¹⁶O. *J. Quant. Spectrosc. Rad. Transf.* **117**, 29–80 (2013).

82. Tennyson, J. *et al.* IUPAC critical evaluation of the rotational-vibrational spectra of water vapor. Part IV. Energy levels and transition wavenumbers for H₂¹⁶O, D₂¹⁷O, and D₂¹⁸O. *J. Quant. Spectrosc. Rad. Transf.* **117**, 93–108 (2014).
83. Furtenbacher, T., Tóbiás, R., Tennyson, J., Polyansky, O. L. & Császár, A. G. W2020: A database of validated rovibrational experimental transitions and empirical energy levels of H₂¹⁶O. *J. Phys. Chem. Ref. Data* **49**, 033101 (2020).
84. Tóbiás, R. *et al.* Spectroscopic-network-assisted precision spectroscopy and its application to water. *Nat. Commun.* **11**, 1708 (2020).
85. Furtenbacher, T. *et al.* The W2020 database of validated rovibrational experimental transitions and empirical energy levels of water isotopologues. Part II. H₂¹⁷O and H₂¹⁸O with an update to H₂¹⁶O. *J. Phys. Chem. Ref. Data* **49**, 043103 (2020).
86. Furtenbacher, T., Tóbiás, R., Tennyson, J., Gamache, R. R. & Császár, A. G. The W2024 database of the water isotopologue H₂¹⁶O. *Sci. Data* **11**, 1058 (2024).
87. Rácsai, B., Furtenbacher, T., Fusina, L., Di Lonardo, G. & Császár, A. G. MARVEL analysis of the high-resolution rovibrational spectra of H¹⁶O³⁵Cl. *J. Mol. Spectrosc.* **384**, 111561 (2022).
88. Ecseri, G. *et al.* Joint survey of the experimental high-resolution spectra of H¹⁶O³⁷Cl and H¹⁶O³⁵Cl with a reanalysis of the 2ν₂ band. *J. Mol. Spectrosc.* **397**, 111834 (2023).
89. Chubb, K. L. *et al.* MARVEL analysis of the measured high-resolution rovibrational spectra of H₂S. *J. Quant. Spectrosc. Rad. Transf.* **218**, 178–186 (2018).
90. Tennyson, J., Furtenbacher, T., Yurchenko, S. N. & Császár, A. G. Empirical rovibrational energy levels for nitrous oxide. *J. Quant. Spectrosc. Rad. Transf.* **316**, 108902 (2024).
91. Tóbiás, R. *et al.* Critical evaluation of measured rotational-vibrational transitions of four sulphur isotopologues of S¹⁶O₂. *J. Quant. Spectrosc. Rad. Transf.* **208**, 152–163 (2018).
92. Chubb, K. L. *et al.* MARVEL analysis of the measured high-resolution spectra of C₂H₂. *J. Quant. Spectrosc. Rad. Transf.* **204**, 42–55 (2018).
93. Al-Derzi, A. R. *et al.* An improved rovibrational linelist of formaldehyde, H₂¹²C¹⁶O. *J. Quant. Spectrosc. Rad. Transf.* **266**, 107563 (2021).
94. Al Derzi, A. R., Furtenbacher, T., Yurchenko, S. N., Tennyson, J. & Császár, A. G. MARVEL analysis of the measured high-resolution spectra of ¹⁴NH₃. *J. Quant. Spectrosc. Rad. Transf.* **161**, 117–130 (2015).
95. Furtenbacher, T. *et al.* Empirical rovibrational energy levels of ammonia up to 7500 cm⁻¹. *J. Quant. Spectrosc. Rad. Transf.* **251**, 107027 (2020).
96. Fábri, C. *et al.* Variational quantum mechanical and active database approaches to the rotational-vibrational spectroscopy of ketene, H₂CCO. *J. Chem. Phys.* **135**, 094307 (2011).
97. McQuarrie, D. A. *Statistical mechanics* (Harper and Row, New York, 2000).
98. Furtenbacher, T. *et al.* Definitive ideal-gas thermochemical functions of the H₂¹⁶O molecule. *J. Phys. Chem. Ref. Data* **45**, 043104 (2016).
99. Simkó, I. *et al.* Recommended ideal-gas thermochemical functions for heavy water and its substituent isotopologues. *J. Phys. Chem. Ref. Data* **46**, 023104 (2017).
100. Ruscic, B. & Bross, D. H. Thermochemistry. In *Mathematical Modelling of Gas-Phase Complex Reaction Systems: Pyrolysis and Combustion*, 3–114 (Elsevier, 2019).
101. Ruscic, B. Uncertainty quantification in thermochemistry, benchmarking electronic structure computations, and Active Thermochemical Tables. *Int. J. Quant. Chem.* **114**, 1097–1101 (2014).
102. Császár, A. G., Leininger, M. & Szalay, V. The standard enthalpy of formation of CH₂. *J. Chem. Phys.* **118**, 10631–10642 (2003).
103. Allen, W. D., East, A. L. L. & Császár, A. G. *Ab initio* anharmonic vibrational analyses of non-rigid molecules. In Laane, J., Dakkouri, M., van der Veken, B. & Oberhammer, H. (eds.) *Structures and conformations of nonrigid molecules*, 343–373 (Kluwer, Dordrecht, 1993).
104. Császár, A. G., Allen, W. D. & Schaefer III, H. F. In pursuit of the ab initio limit for conformational energy prototypes. *J. Chem. Phys.* **108**, 9751–9764 (1998).
105. Wheeler, S. E., Houk, K. N., Schleyer, P. R. & Allen, W. D. A hierarchy of homodesmotic reactions for thermochemistry. *J. Am. Chem. Soc.* **131**, 2547–2560 (2009).
106. Wilson Jr., E. B., Decius, J. C. & Cross, P. C. *Molecular vibrations: the theory of infrared and Raman vibrational spectra* (McGraw Hill, New York, 1955).
107. Kroto, H. W. *Molecular rotation spectra* (Dover, New York, 1992).
108. Watson, J. K. G. Simplification of the molecular vibration-rotation Hamiltonian. *Mol. Phys.* **15**, 479–490 (1968).
109. Sanderson, R. T. *Chemical Bonds and Bond Energy* (Academic Press, New York, 1976).
110. Sanderson, R. T. *Polar Covalence* (Academic Press, New York, 1983).
111. Gillespie, R. J., Spencer, J. N. & Moog, R. S. An approach to reaction thermodynamics through enthalpies, entropies, and free energies of atomization. *J. Chem. Educ.* **73**, 631 (1996).
112. von Szentpály, L. Atom-based thermochemistry: Crystal atomization and sublimation enthalpies in linear relationships to molecular atomization enthalpy. *J. Am. Chem. Soc.* **130**, 5962–5973 (2008).
113. von Szentpály, L. Atom-based thermochemistry: Predictions of the sublimation enthalpies of group 12 chalcogenides and the formation enthalpies of their polonides. *J. Phys. Chem. A* **112**, 12695–12701 (2008).
114. Császár, A. G. & Furtenbacher, T. From a network of computed reaction enthalpies to atom-based thermochemistry (NEAT). *Chem. Eur. J.* **16**, 4826–4835 (2010).
115. Westbrook, C. K. & Dryer, F. L. Chemical kinetic modeling of hydrocarbon combustion. *Prog. Energy Combust. Sci.* **10**, 1–57 (1984).
116. Olm, C. *et al.* Comparison of the performance of several recent hydrogen combustion mechanisms. *Combust. Flame* **161**, 2219–2234 (2014).
117. Varga, T. *et al.* Optimization of a hydrogen combustion mechanism using both direct and indirect measurements. *Proc. Combust. Inst.* **35**, 589–596 (2015).
118. Olm, C., Zsély, I. G., Varga, T., Curran, H. J. & Turányi, T. Comparison of the performance of several recent syngas combustion mechanisms. *Combust. Flame* **162**, 1793–1812 (2015).
119. Varga, T. *et al.* Development of a joint hydrogen and syngas combustion mechanism based on an optimization approach. *Int. J. Chem. Kinet.* **48**, 407–422 (2016).
120. Olm, C. *et al.* Development of an ethanol combustion mechanism based on a hierarchical optimization approach. *Int. J. Chem. Kinet.* **48**, 423–441 (2016).
121. Olm, C., Varga, T., Valkó, E., Curran, H. J. & Turányi, T. Uncertainty quantification of a newly optimized methanol and formaldehyde combustion mechanism. *Combust. Flame* **186**, 45–64 (2017).
122. Kovács, M., Papp, M., Zsély, I. G. & Turányi, T. Determination of rate parameters of key N/H/O elementary reactions based on H₂/O₂/NO_x combustion experiments. *Fuel* **264**, 1–11 (2020).
123. Kawka, L. *et al.* Comparison of detailed reaction mechanisms for homogeneous ammonia combustion. *Z. Phys. Chem.* **234**, 1329–1357 (2020).
124. Szanthoffer, A. G., Zsély, I. G., Kawka, L., Papp, M. & Turányi, T. Testing of NH₃/H₂ and NH₃/syngas combustion mechanisms using a large amount of experimental data. *Appl. Ener. Combust. Sci.* **14**, 100127 (2023).

125. Kovács, M., Papp, M., Zsély, I. G. & Turányi, T. Main sources of uncertainty in recent methanol/NO_x combustion models. *Int. J. Chem. Kinet.* **53**, 884–900 (2021).
126. Zhang, P., Zsély, I. G., Samu, V., Nagy, T. & Turányi, T. Comparison of methane combustion mechanisms using shock tube and rapid compression machine ignition delay time measurements. *Energy Fuels* **35**, 12329–12351 (2021).
127. Zhang, P., Zsély, I. G., Papp, M., Nagy, T. & Turányi, T. Comparison of methane combustion mechanisms using laminar burning velocity measurements. *Combust. Flame* **238**, 1–16 (2022).
128. Eckart, S., Zsély, I., Krause, H. & Turányi, T. Effect of the variation of oxygen concentration on the laminar burning velocities of hydrogen-enriched methane flames. *Int. J. Hydrog. Energy* **533–546** (2023).
129. Bolla, M. *et al.* Comparison and analysis of butanol combustion mechanisms. *Energy Fuels* **36**, 11154–11176 (2022).
130. Su, B. *et al.* Comparison of the performance of ethylene combustion mechanisms. *Combust. Flame* **260**, 113201 (2024).
131. Wang, P. *et al.* Mechanism development for larger alkanes by auto-generation and rate rule optimization: A case study of the pentane isomers. *Proc. Combust. Inst.* **40**, 105408 (2024).
132. Kovács, M. *et al.* Optimization of a methanol/NO_x combustion mechanism based on a large amount of experimental data. *Fuel* **375**, 132544 (2024).
133. Szanthoffer, A. G., Papp, M. & Turányi, T. Identification of well-parameterised reaction steps in detailed combustion mechanisms – a case study of ammonia/air flames. *Fuel* **380**, 132938, <https://doi.org/10.1016/j.fuel.2024.132938> (2025).
134. Su, B., Papp, M., Zhang, P. & Turányi, T. Dependence of ignition delay time on its definition – a case study on methane ignition. *Combust. Flame* **262**, 113364 (2024).
135. Reaction Design San Diego. *CHEMKIN 10101 and CHEMKIN-PRO 15101 Input manual* (2010).
136. Su, B., Nagy, T., Papp, M. & Turányi, T. Reduction-assisted parameter optimization of the ethylene chemistry in the aramcomech 2.0 combustion mechanism. *Combustion and Flame* **273**, 113976, <https://doi.org/10.1016/j.combustflame.2025.113976> (2025).
137. Cuoci, A., Frassoldati, A., Faravelli, T. & Ranzi, E. OpenSMOKE++: an object-oriented framework for the numerical modeling of reactive systems with detailed kinetic mechanisms. *Comput. Phys. Commun.* **192**, 237–264 (2015).
138. Goodwin, D. G., Moffat, H. K., Schoegl, I., Speth, R. L. & Weber, B. W. Cantera: an object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes, <https://www.cantera.org> (2023).
139. Frenklach, M. Systematic optimization of a detailed kinetic model using a methane ignition example. *Combust. Flame* **58**, 69–72 (1984).
140. Turányi, T. *et al.* Determination of rate parameters based on both direct and indirect measurements. *Int. J. Chem. Kinet.* **44**, 284–302 (2012).
141. McBride, B. J. & Gordon, S. Computer program for calculating and fitting thermodynamic functions. *NASA Reference Publication 1271* <https://ntrs.nasa.gov/api/citations/19930003779/downloads/19930003779.pdf> (1992).
142. Goitom, S. K. *et al.* Efficient numerical methods for the optimization of large kinetic reaction mechanisms. *Comb. Theor. Model.* **26**, 1071–1097 (2022).
143. Nelder, J. & Mead, R. A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965).
144. Powell, M. The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives. *Technical Report, Department of Applied Mathematics and Theoretical Physics* https://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf (2009).
145. King, D. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.* **10**, 1755–1758 (2009).
146. Nagy, T. & Turányi, T. Minimal spline fit a model-free method for determining statistical noise of experimental data series. *Proceedings of the European Combustion Meeting Paper* 336 (2021).
147. Nagy, T. & Turányi, T. Reduction of very large reaction mechanisms using methods based on simulation error minimization. *Combust. Flame* **156**, 417–428 (2009).
148. Zsély, I. G., Nagy, T., Simmie, J. M. & Curran, H. J. Reduction of a detailed kinetic model for the ignition of methane/propane mixtures at gas turbine conditions using simulation error minimization methods. *Combust. Flame* **1469–1479** (2011).
149. Tóth, J., Nagy, A. L. & Zsély, I. G. Structural analysis of combustion mechanisms. *J. Math. Chem.* **53**, 86–110 (2015).
150. Papp, M., Valkó, E., Nagy, T. & Turányi, T. Evaluation of high-temperature gas phase rate coefficients of elementary reactions and determination of their prior uncertainty limits <https://k-evaluation.elte.hu/>.
151. Császár, A. G. & Furtenbacher, T. Spectroscopic networks. *J. Mol. Spectrosc.* **266**, 99–103 (2011).
152. Newman, M. E. J. *Networks* (Oxford University Press, Oxford, 2010).
153. Tóbiás, R., Furtenbacher, T. & Császár, A. G. Cycle bases to the rescue. *J. Quant. Spectrosc. Rad. Transf.* **203**, 557–564 (2017).
154. Tóbiás, R., Furtenbacher, T., Tennyson, J. & Császár, A. G. Accurate empirical rovibrational energies and transitions of H₂¹⁶O. *Phys. Chem. Chem. Phys.* **21**, 3473–3495 (2019).
155. Árendás, P., Furtenbacher, T. & Császár, A. G. From bridges to cycles in spectroscopic networks. *Sci. Rep.* **10** (2020).
156. Huang, X., Schwenke, D. W., Freedman, R. S. & Lee, T. J. Ames-2021 CO₂ dipole moment surface and IR line lists: Toward 0.1 % uncertainty for CO₂ IR intensities. *J. Phys. Chem. A* **126**, 5940–5964 (2022).
157. Castrillo, A. *et al.* On the ¹²C₂H₂ near-infrared spectrum: absolute transition frequencies and an improved spectroscopic network at the kHz accuracy level. *Phys. Chem. Chem. Phys.* **25**, 23614–23625 (2023).
158. Yurchenko, S. N., Williams, H., Leyland, P. C., Lodi, L. & Tennyson, J. ExoMol line lists XXVIII: the rovibronic spectrum of AlH. *Mon. Not. R. Astron. Soc.* **479**, 1401–1411 (2018).
159. Darby-Lewis, D. *et al.* Synthetic spectra of BeH, BeD and BeT for emission modeling in JET plasmas. *J. Phys. B: At. Mol. Opt.* **51**, 185701 (2018).
160. Furtenbacher, T., Hegedüs, S. T., Tennyson, J. & Császár, A. G. Analysis of measured high-resolution doublet rovibronic spectra and related line lists of ¹²CH and ¹⁶OH. *Phys. Chem. Chem. Phys.* **24**, 19287–19301 (2022).
161. Syme, A.-M. & McKemmish, L. K. Experimental energy levels of ¹²C¹⁴N through MARVEL analysis. *Mon. Not. R. Astron. Soc.* **499**, 25–39 (2020).
162. Qin, Z., Bai, T. & Liu, L. Line lists for the X²Σ⁺–X²Σ⁺, A²Π–A²Π and A²Π–X²Σ⁺ transitions of CP. *J. Quant. Spectrosc. Rad. Transf.* **258**, 107352 (2021).
163. Wong, A. *et al.* ExoMol line list - XXI. Nitric oxide (NO). *Mon. Not. R. Astron. Soc.* **470**, 882–897 (2017).
164. Qin, Z., Bai, T. & Liu, L. Spectral transition properties of the A¹Π–X¹Σ⁺ system for PN. *Chem. Phys. Lett.* **782**, 139028 (2021).
165. McKemmish, L. K. *et al.* ExoMol molecular line lists - XXXIII. The spectrum of titanium oxide. *Mon. Not. R. Astron. Soc.* **488**, 2836–2854 (2019).
166. Wang, Y., Owens, A., Tennyson, J. & Yurchenko, S. N. MARVEL analysis of the measured high-resolution rovibronic spectra of the calcium monohydroxide radical (CaOH). *Astrophys. J. Suppl. Ser.* **248**, 9 (2020).

Acknowledgements

The work related to reaction kinetics received support from a series of National Research, Development, and Innovation Office (NKFIH) grants; the latest ones are K147024 and FK134332. The collection of direct experimental kinetic data was related to the activity of an IUPAC Task Group (No. 2014-028-2-100). The spectroscopy and thermochemistry data collection received support from NKFIH most recently through grants K138233 and PD145972.

Author contributions

Tamás Turányi: Conceptualization, Funding acquisition, Project administration, Supervision, Writing - original draft, Writing - review & editing. István Gy. Zsély: Data curation, Methodology, Writing - original draft, Writing - review & editing. Máté Papp: Software, Visualization, Writing - review & editing. Tibor Nagy: Formal Analysis, Funding acquisition, Methodology, Software, Writing - review & editing. Tibor Furtenbacher: Software, Visualization, Data curation, Writing - review & editing. Roland Tóbiás: Software, Visualization, Data curation, Writing - review & editing. Péter Árendás: Software, Visualization, Data curation, Writing - review & editing. Attila G. Császár: Conceptualization, Funding acquisition, Project administration, Supervision, Writing - original draft, Writing - review & editing.

Funding

Open access funding provided by Eötvös Loránd University.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.T. or A.G.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025