



HAL
open science

A machine-readable online database for rate coefficients in radical polymerization

Joren van Herck, Simon Harrisson, Robin Hutchinson, Gregory Russell, Tanja
Junkers

► **To cite this version:**

Joren van Herck, Simon Harrisson, Robin Hutchinson, Gregory Russell, Tanja Junkers. A machine-readable online database for rate coefficients in radical polymerization. *Polymer Chemistry*, 2021, 10.1039/D1PY00544H . hal-03269449

HAL Id: hal-03269449

<https://hal.science/hal-03269449>

Submitted on 26 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ARTICLE

A machine-readable online database for rate coefficients in radical polymerization

Received 00th January 20xx,
Accepted 00th January 20xx

Joren Van Herck,^a Simon Harrison,^b Robin A. Hutchinson,^c Gregory T. Russell,^d Tanja Junkers^{a,*}

DOI: 10.1039/x0xx00000x

An online database created and curated by an IUPAC subcommittee is introduced. It is designed to act as central access point for finding reliable kinetic data on radical polymerizations. The database can be accessed via a webinterface or via Python code available for download, and at the moment consists mostly of propagation rate coefficient data. Expansion to other coefficients, and eventually also other types of polymerization, is anticipated in time. Monomers can be searched by name, CAS number, SMILES notation or their InChI and InChIKey descriptors. The database returns values for Arrhenius parameters for the chosen reaction, key information for measurement conditions, and information as to whether the respective value has been subject to benchmarking by IUPAC. The aim of the database is to simplify the search for coefficients, and to unify modelling approaches in the community. Further, since the database is designed to be fully machine-readable, it allows for direct integration into software code, which enables advanced machine-learning and other computer-based research.

Introduction

The importance of precise knowledge of rate coefficients in radical polymerization has long been understood. Without detailed knowledge of the kinetics and mechanism of each step in the polymerization, reversible-deactivation radical polymerization could not have matured to the standard that synthetic polymer chemists are accustomed to today, and the ability to model and predict polymerizations would be severely limited. Consequently, the IUPAC Polymer Division features a Subcommittee on Modelling of Polymerization Kinetics and Processes that deals with elucidating kinetics and mechanisms of polymerization reactions, with a particular emphasis on radical polymerization. One of the most impactful outcomes of this Subcommittee to date has been a series of publications benchmarking kinetic rate coefficients to provide a solid and common basis for the research community.^{1,2,3,4,5,6,7,8} Most activities were focused in the past on propagation rate coefficients, k_p , for radical polymerization, based on the PLP-SEC technique (pulsed laser polymerization – size exclusion chromatography).⁹ Other coefficients for initiation (including efficiency), termination or transfer reactions have also been studied, but their summary and benchmarking are more

difficult compared to the propagation rate coefficients, where PLP-SEC opened a precise and very reproducible measurement method.¹

Yet, despite a broad range of coefficients being available, many researchers still use coefficients stemming from PLP-SEC sub-datasets or – even worse – from methods nowadays deemed unreliable. Especially in the field of modelling, some groups use sets of coefficients that are close but not identical to the benchmark values. This creates modelling “bubbles” in which each group uses their own set of data. In consequence, this means that newly determined coefficients increasingly are model dependent (and hence dependent on the specific set of underlying input parameters). A more unified approach is certainly desirable. The benchmarked kinetic data are summarized in a series of publications endorsed by the IUPAC subcommittee.^{1–8} However, not all researchers are necessarily aware of these papers, or may falsely assume that they should not rely on a series of publications that spans decades. Convenience of access would also promote more widespread use of the benchmarked data. Finding accurate and up-to-date data is not simple, and especially for researchers not in the field it easily becomes confusing. For all these reasons a central access point for the available data would thus be very useful.

^a Polymer Reaction Design Group, School of Chemistry, Monash University, 19 Rainforest Walk, Building 23, Clayton, Vic 3800 (Australia)

^b Laboratoire de Chimie des Polymères Organiques, Université de Bordeaux/ENSCBP/CNRS UMR 5623, 16 avenue Pey Berland 33607 Pessac, France.

^c Department of Chemical Engineering, Dupuis Hall, Queen's University, Kingston, Ontario K7L 3N6, Canada

^d School of Physical and Chemical Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand.

Monomer			
Trivial name	n-butyl methacrylate	Abbreviation	BMA
CAS	97-88-1	SMILES	C=C(C)C(=O)OCCCC
InChI	InChI=1S/C8H14O2/c1-4-5-6-10-8(9)7(2)3/h2,4-6H2,1,3H3	InChIKey	SOGAXMICEFXMKE-UHFFFAOYSA-N
k_p data			
A	3801900 L · mol ⁻¹ s ⁻¹	E_a	22900 J · mol ⁻¹
Solution	bulk	Concentration	N/A
IUPAC	IUPAC Benchmarked	DOI	Reference
T_{min}	-20	T_{max}	90
Temperature	20 °C <input type="text"/> <input type="button" value="Change"/>	k_p (20°C)	316 (± 32) L · mol ⁻¹ s ⁻¹

Figure 1: Information available from the human interface of the database for propagation rate coefficients. Access is via <http://sql.polymatter.net/>.

At the same time, recent years have seen a dramatic increase in the use of artificial intelligence and machine learning algorithms in chemical sciences. Next to prediction of materials properties, chemometrics or their use in quantum chemistry calculations, their use in reaction optimizations has become more widespread.¹⁰ Machine learning algorithms depend on accessible data for their learning. Scientific publications are too complex for automatic data read-out, and a central database would be much more favourable for allowing such algorithms to connect with the knowledge that we possess of certain reactions.

Thus, we have set up within the framework of the IUPAC Subcommittee an online database that serves two purposes:

- Creation of a central storage point for kinetic rate coefficients, with special emphasis on benchmarked data
- Creation of the database in a machine-readable format

Database structure

In the longer term this database is intended to become a part of larger IUPAC online databases,¹¹ but for the moment it is implemented as a stand-alone database. On the back-end, an SQL database is employed. The above two criteria are met by accessing the database via one central point. For human users, this is the webpage via <http://sql.polymatter.net/>. Machine-codes can fetch the same data via the same page. The human interface of the database is programmed to be mostly self-explanatory and the type of information that can be accessed is exemplified in Figure 1.

To allow for automatic access via machine code, first the question of how to identify a monomer in an unambiguous and reliable way needs to be addressed. Polymer chemists and chemical engineers are used to certain nomenclature and abbreviations for monomers, even if these are often neither

fully consistent nor absolutely unique. For pragmatic reasons the database nevertheless allows for request of information via common names and abbreviations. Classical IUPAC naming might also be allowed, but such names can fall short as they are not always straightforward to derive or they may differ significantly from commonly used names. A work-around is the use of CAS numbers, however this excludes monomers that are not yet classified. A simpler way is the so-called Simplified Molecular-Input Line-Entry System (SMILES),¹² which directly translates a structure into a character string. This method is computer-friendly, and at the same time allows for direct interpretation of the code and translation into a structure. However, SMILES codes are dependent on the choice of atom in a structure selected to start the notation from. To overcome this, IUPAC has already in 2006 introduced the International Chemical Identifier (InChI) notation.¹³ Similar to IUPAC names, InChI codes are directly related to the chemical structure of a molecule. They are unique identifiers, and compared to CAS numbers they are not proprietary, and require no institution to assign codes to new structures, since they are derived strictly from chemical information. InChI codes can be quite long, yet with practice they can still be read by humans. A reduced form of the InChI is the so-called InChIKey, which represents a hashed form of the InChI that saves characters in the code and hence makes it easier to handle in searches online. InChIKeys are not straight forward to interpret yet are also practically unique for any given structure. Modern drawing software is typically able to interpret SMILES, InChI and InChIKey information, and can transform those directly into a chemical structure (and vice versa). Consequently, the database allows searching for any of the above information, while use of the InChIKey is recommended over all other formats for its relatively simple yet unique identifier. An example for the monomer methyl methacrylate is given in Table 1.

Table 1: Different identifiers for monomers on the example of methyl methacrylate

Trivial name	Methyl methacrylate
Abbreviation	MMA
CAS	80-62-6

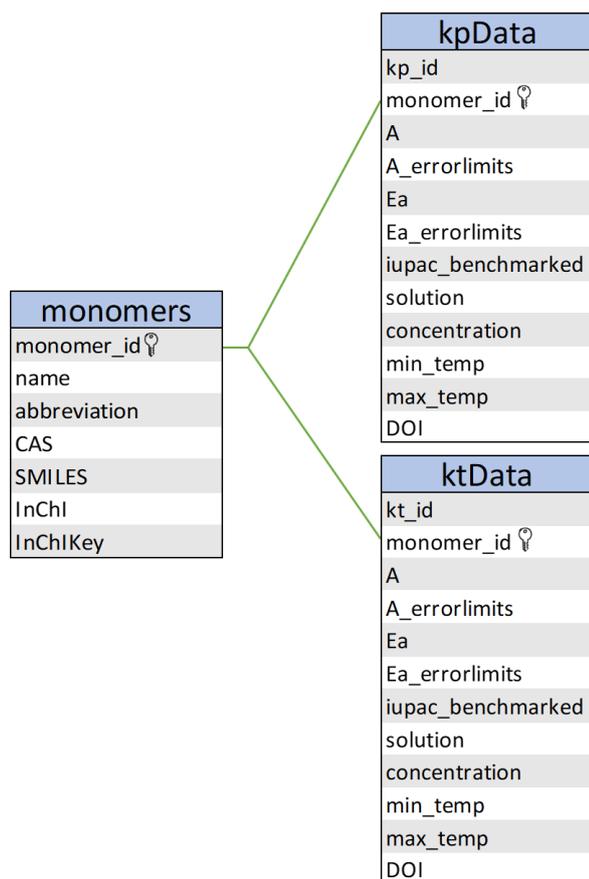


Figure 2: Internal structure of the kinetic database

SMILES	C=C(C)C(=O)OC
InChI	InChI=1S/C5H8O2/c1-4(2)5(6)7-3/h1H2,2-3H3
InChIKey	VVQNEPGJFQJSBK-UHFFFAOYSA-N

Different identifiers can be selected via the web interface. Once a monomer is selected, different information can be accessed. For now, the database focuses on propagation rate coefficients, and hence the frontend dropdown menu only has one option, i.e. k_p . Once additional data is included, the option will be extended to specify what kinetic parameter is available upon request for a monomer. Figure 2 gives an overview of the current database structure. For each monomer, subsections are defined; the example here shows propagation (k_p) and termination (k_t) rate coefficients. Data that is stored for the propagation rate coefficient are:

- Arrhenius parameters
- Error limits for Arrhenius parameters (currently not used)
- Temperature range in which data was measured
- Whether the given data is benchmarked by IUPAC
- Whether the data was obtained in bulk or solution
- solution concentration if data is not given for bulk conditions

- A link to the publication whence the data was retrieved

The webpage allows to specify a temperature and to display a rate coefficient which is calculated directly from the Arrhenius parameters for that specific condition. Individual coefficients for certain temperatures are not stored in the database. As this implies, the primary information in the database is Arrhenius parameters, not individual values of rate coefficients. While the database contains fields for error margins for these, no values are entered since a careful error analysis via joint confidence intervals is not easily expressed. A closer examination of errors will be subject of a forthcoming study. For the time being, a standard error of 10% is displayed on the web interface for IUPAC-benchmarked k_p , and 20% for non-benchmarked data. For more detailed discussions the reader and user of the database is referred to the individual publications.

We believe that this database will make the available kinetic data much more accessible to the research community. Further, we hope that modelling software, be it Monte-Carlo codes used by certain research groups,¹⁴ or commercial software such as Predici,¹⁵ will automatically embed functions to access the database. Thus, instead of feeding coefficients and Arrhenius parameters for certain reactions into models, simply the InChIKey could be provided, and the modelling software would then automatically populate the coefficients that are known for the specific system. This will simplify modelling, and moreover unify it between different groups, hopefully leading to more consistent results and use of the latest recommended kinetic coefficients.

Data curation

With a database as described above, it is important that it be constantly updated, and in doing so to select data strictly on criteria such as reliability and reproducibility. As noted above, priority is given in the database to IUPAC-benchmarked kinetic coefficients, since these can be considered most accurate. However, to date this comprises only a limited number of monomers, and primarily propagation rate coefficients. Thus, to extend the use of the database, it is also of interest to enter non-benchmarked kinetic data. This creates in turn the question as to which data is deemed reliable. Benchmarking of kinetic data is a thorough process that requires stringent data treatment and confirmation of data by various research groups. However, in the vast majority of cases PLP-SEC delivers fairly reliable results, and the underlying problems are known. A large number of PLP-SEC data-sets exist in the literature that are not yet benchmarked, usually because a second independent confirmation is lacking. Some of this data could reasonably be added to the present databank. Therefore, the IUPAC Subcommittee has appointed a group of members who are responsible for the curation of the databank and to make reasonable choices as to which not-yet-benchmarked data can be included. Regardless of the choice, since the database

output contains direct information as to whether or not a certain coefficient is benchmarked, the user retains the ability to make changes if they believe they are in possession of a better data source. At the same time, data can be sent to the curators with a request for evaluation and inclusion into the database. In order to preserve the integrity of the data, only the curators have write access to the database. For the time being, a series of non-benchmarked k_p data that is deemed reliable has been added to the database (see Table 2), with more data to follow in time. For the time being, the authors of this article made the initial selection of monomers to be added.

Database development

The database itself is not static, and will evolve over time. While the general structure is not expected to change significantly, it is hoped that more sections will be added, covering more radical polymerization rate coefficients and potentially reactions from other types of polymerization too. Changes to the database will be published on the <http://sql.polymatter.net> webpage, and made transparent. We will further aim for broad communication via a variety of channels to keep interested researchers updated. To identify which monomers are covered in the database at a given point in time, a request can be made automatically, returning the full list of monomers for which data is available. The present list is provided in Table 2 for this paper, including whether it is IUPAC-benchmarked or included at the discretion of the curators.

Table 2: Monomers currently included in the database

Trivial name	$A / \text{L}\cdot\text{mol}^{-1}\cdot\text{s}^{-1}$	$E_A / \text{kJ}\cdot\text{mol}^{-1}$	benchmark marked
styrene	$4.27\cdot 10^7$	32.15	yes ¹
methyl methacrylate	$2.67\cdot 10^6$	22.36	yes ²
ethyl methacrylate	$4.07\cdot 10^6$	23.4	yes ³
butyl methacrylate	$3.80\cdot 10^6$	22.9	yes ³
dodecyl methacrylate	$2.51\cdot 10^6$	21	yes ³
cyclohexyl methacrylate	$6.31\cdot 10^6$	23	yes ⁴
glycidyl methacrylate	$6.17\cdot 10^6$	22.9	yes ⁴
benzyl methacrylate	$6.76\cdot 10^6$	22.9	yes ⁴
isobornyl methacrylate	$6.17\cdot 10^6$	23.1	yes ⁴
butyl acrylate	$2.24\cdot 10^7$	17.9	yes ⁵
methacrylic acid [#]	$1.54\cdot 10^6$	15	yes ⁶
methyl acrylate	$1.41\cdot 10^7$	17.3	yes ⁷
vinyl acetate	$1.35\cdot 10^7$	20.4	yes ⁸

tert-butyl acrylate	$1.90\cdot 10^7$	17.5	no ¹⁶
2-ethyl hexyl acrylate	$9.10\cdot 10^6$	15.8	no ¹⁷
stearyl acrylate	$1.86\cdot 10^7$	16.93	no ¹⁸
acrylonitrile	$1.79\cdot 10^6$	15.4	no ¹⁹
2-ethylhexyl methacrylate	$1.87\cdot 10^6$	20.4	no ²⁰
tert-butyl methacrylate	$2.4\cdot 10^6$	22.1	no ²¹
stearyl methacrylate	$3.45\cdot 10^6$	21.49	no ²²

[#] measured in 15% aqueous solution

Python code

On the webpage, a web scraping python tool is made available that can be used to automatically extract info from the database. The script relies on the 'requests' and 'Beautiful Soup' python packages,²³ both popular and powerful libraries for web scraping applications. Web scraping rather than direct SQL database access was chosen to avoid issues with database access user rights management. First, the python code sends a HTTP GET request to fetch the content of the webpage. The content is parsed to a more machine-readable HTML format via the 'Beautiful Soup' package, and the desired data is extracted and returned in the appropriate data structure. Three functions are defined. The first function returns a list of all available monomers in the database based on the given identifier (see Table 1 for valid identifiers). As on the webpage, the second function returns the complete information table. The third function extracts the Arrhenius parameters from the database and calculates the k_p value at a given temperature. Since the database is currently only populated with k_p data, the described functions apply only to this coefficient. In the future, similar functions for other coefficients will be added that will request the page content of the respective database table. For a more detailed and technical explanation of the code, the reader is directed to the documentation that is provided on the webpage. We provide Python code as a starting point for interested researchers, but note that the same method can in principle be applied via any programming language.

Conclusions

A machine-readable, central kinetic database for rate coefficients in radical polymerization is presented. For the moment, the database focuses on propagation rate coefficients, but in future will hopefully also contain copolymerization reactivity ratios, as well as termination and transfer rate coefficients, and kinetic data for chain initiation. The database is accessible via a human web-interface, but also allows for direct requests from Python code or other software. This will allow direct implementation of the data into modelling software, or machine learning algorithms. The data

will be centrally collected and curated by the IUPAC Subcommittee on Modelling of Polymerization Kinetics and Processes. This will set a strong standard for the polymer chemistry community and relieve researchers of the tedious work of finding suitable coefficients in the literature.

Author Contributions

Joren van Herck: Software, Data curation, Visualization, Writing – review & editing; **Simon Harrison:** Conceptualization, Writing – review & editing; **Robin Hutchinson:** Conceptualization, Writing – review & editing; **Greg Russell:** Conceptualization, Writing – review & editing; **Tanja Junkers:** Conceptualization, Data curation, Project administration, Writing – original draft, Writing – review & editing

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge supporting funding by IUPAC via project 2019-045-1-400 - Development of a Machine Accessible Kinetic Databank for Radical Polymerizations.

Notes and references

- [1] M. Buback, R. G. Gilbert, R. A. Hutchinson, B. Klumperman, F.-D. Kuchta, B. G. Manders, K. F. O'Driscoll, G. T. Russell and J. Schweer, *Macromol. Chem. Phys.*, 1995, **196**, 3267–3280.
- [2] S. Beuermann, M. Buback, T. P. Davis, R. G. Gilbert, R. A. Hutchinson, O. F. Olaj, G. T. Russell, J. Schweer and A. M. van Herk, *Macromol. Chem. Phys.*, 1997, **198**, 1545–1560.
- [3] S. Beuermann, M. Buback, T.P. Davis, R.G. Gilbert, R.A. Hutchinson, A. Kajiwarra, B. Klumperman and G.T. Russell, *Macromol. Chem. Phys.*, 2000, **201**, 1355–1364.
- [4] S. Beuermann, M. Buback, T. P. Davis, N. García, R. G. Gilbert, R. A. Hutchinson, A. Kajiwarra, M. Kamachi, I. Lacík and G. T. Russell, *Macromol. Chem. Phys.*, 2003, **204**, 1338–1350.
- [5] J. M. Asua, S. Beuermann, M. Buback, P. Castignolles, B. Charleux, R. G. Gilbert, R. A. Hutchinson, J. R. Leiza, A. N. Nikitin, J.-P. Vairon and A. M. van Herk, *Macromol. Chem. Phys.*, 2004, **205**, 2151–2160.
- [6] S. Beuermann, M. Buback, P. Hesse, F.-D. Kuchta, I. Lacík and A. M. van Herk, *Pure Appl. Chem.*, 2007, **79**, 1463–1469.
- [7] C. Barner-Kowollik, S. Beuermann, M. Buback, P. Castignolles, B. Charleux, M.L. Coote, R.A. Hutchinson, T. Junkers, I. Lacík, G.T. Russell, M. Stach and A.M. van Herk, *Polym. Chem.* 2014, **5**, 204–212.
- [8] C. Barner-Kowollik, S. Beuermann, M. Buback, R.A. Hutchinson, T. Junkers, H. Kattner, B. Manders, A.N. Nikitin, G.T. Russell and A.M. van Herk, *Macromol. Chem. Phys.* 2017, **218**, 1600357.
- [9] O. F. Olaj, I. Bitai and F. Hinkelmann, *Makromol. Chem.*, 1987, **188**, 1689–1702.
- [10] S. Lazzari, A. Lischewski, Y. Orlov, P. Deglmann, A. Daiss, E. Schreiner and H. Vale, *Adv. Chem. Eng.* 2020, **56**, 187–227.
- [11] <https://goldbook.iupac.org>
- [12] D. Weiniger, *J. Chem. Inf. Comput. Sci.* 1988, **28**, 1, 31–36
- [13] S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi, I. Pletnev, *J. Cheminf.* 2013, **5**, 7.
- [14] See for example: a) D. R. D'Hooge, P. H. M. Van Steenberge, P. Derboven, M. -F. Reyniers and G. B. Marin, *Polym. Chem.*, 2015, **6**, 7081–7096; b) M. Drache and G. Drache, *Polymers* 2012, **4**, 1416–1442.
- [15] M. Wulkow, *Macromol. React. Eng.* 2008, **2**, 6, 461–494.
- [16] B. Dervaux, T. Junkers, M. Schneider-Baumann F. E. Du Prez and C. Barner-Kowollik, *J. Polym. Sci. A* 2009, **47**, 6641–6654.
- [17] T. Junkers, M. Schneider-Baumann, S. S. P. Koo, P. Castignolles and C. Barner-Kowollik, *Macromolecules* 2010, **43**, 10427–10434.
- [18] A. P. Haehnel, M. Schneider-Baumann, L. Arens, A. M. Misske, F. Fleischhaker and C. Barner-Kowollik, *Macromolecules* 2014, **47**, 10, 3483–3496.
- [19] T. Junkers, S. P. S. Koo and C. Barner-Kowollik, *Polym. Chem.* 2010, **1**, 438–441.
- [20] R. A. Hutchinson, S. Beuermann, D. A. Paquet and J. H. McMinn, *Macromolecules* 1997, **30**, 12, 3490–3493.
- [21] G. E. Roberts, T. P. Davis, J. P. A. Heuts and G. E. Ball, *Macromolecules* 2002, **35**, 27, 9954–9963.

- [22] A. P. Haehnel, M. Schneider-Baumann, K. U. Hildebrandt, A. M. Misske and C. Barner-Kowollik, *Macromolecules* 2013, **46**, 1, 15–28.
- [23] <https://requests.readthedocs.io/en/master/> and <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

ARTICLE

TOC Figure

