NFDI4Chem
ENHANCE
YOUR
DATA.

JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

NFDI4Chem
Chemistry Data Days 2023

Johannes Gutenberg-Universität Mainz
6/7 June 2023

Contents

Sponsors

This conference would not be possible without the generous support of these sponsors:

Deltablot

Schedule

Tuesday, 6 June 2023

11:00	*Registration & light Lunch*
12:00	**Welcome** NFDI4Chem
12:15	**Overview NFDI & NFDI4Chem + Services** NFDI4Chem
12:45	Guillermo Restrepo (MPI Leipzig) *"Chemical space: mathematical, computational and ethical questions arising form its historical unfolding"*
13:45	*Coffee Break*
14:00	**Session ELNs & Data Management Tools** ELNs in Chemistry - an introduction (NFDI4Chem) eLabFTW (Nicolas Carpi) Sciformation (Felix Rudolphi) Chemotion (Patrick Hodapp) LOGS-ELN (Jacob J Lopez)
16:00	*Coffee Break*
16:15	**Session GDCh / VAA / JCF** *Workshop "'Berufseinstieg in der Chemie"', zusammen mit JCF/VAA/GDCh-Karriereservice* or ***Barcamp***
18:00	**Poster Session, Buffet and Drinks Reception**

Wednesday, 7 June 2023

08:30	*Registration*
09:00	Sonja Herres-Pawlis (RWTH Aachen University) *How to make your data FAIR – a case study from a working group from synthetic chemistry* Paul Czodrowski (JGU Mainz) *No data, no AI party* Luc Patiny (Zakodium) *Using the browser for teaching and processing chemical data*
10:30	*Coffee Break*
10:45	**Workshops** Getting started with Chemotion Basics of Research Data Management (RDM): Crash course Scientific Publishing – A Peek into the Editorial Office and Beyond (with Wiley-VCH and JCF)
12:15	*Lunch*
12:45	Kevin Jablonka (EPFL) *Transforming chemistry with large language models* Ilka Paulus (DFG) *German Research Foundation: Funding Opportunities and more*
13:45–14:00	**Closing** NFDI4Chem

Abstracts

Tuesday, 6 June 2023

Chemical space: mathematical, computational and ethical questions arising form its historical unfolding

Guillermo Restrepo

Max Planck Insitute for Mathematics in the Sciences
Leipzig/Germany

E-Mail: guillermorestrepo@gmail.com

Chemical space, understood as all known chemicals and reactions, is central not only because it constitutes the raw material of the discipline, but because it encodes patterns that shed light on the future of chemistry. In this talk I will discuss the evolution of the chemical space from 1800 up to the result of an ongoing research programme on processing electronic repositories of information with mathematical and computational tools. Chemical space has been expanded at an exponential rate with doubling times of 16 is marked by three statistical regimes on the production of chemicals with sharp 1860 and 1980. Interestingly, despite its rapid growth, chemical space has been, with the passage of time, more and more concentrated on the deep exploration of some of its those where organic chemistry lies. The expansion space leads to several questions, some and computational nature and others and ethical tone, which are merged questions. Mathematical and computational questions include the maximum size space, the required memory to store it, the conditions to increase its expansion rate and the historical nature of the ontology of chemistry. Social questions touch upon the institutional structures of chemistry accounting for the historical evolution and whether it is possible to tune them. Ethical questions are related to the kind of space we should aim at, for the sake of knowledge and of our presence

[1] Llanos, E. J.; Leal, W.; Luu, D. H.; Jost, J.; Stadler, P. F.; Restrepo, G. *Exploration of the chemical space and its three historical regimes*. *P. Natl. Acad. Sci. USA* **2019**, *116*, 12660-12665.

[2] Lipkus, A. H.; Watkins, S. P.; Gengras, K.; McBride, M. J.; Wills, T. J. J. *Org. Chem.* **2019**, *84*, 13948-13956.

[3] Szymkuć, S.; Badowski, T.; Grzybowski, B. A. *Angew. Chem. Int. Ed.* **2021**, *60*, 26226–26232.

[4] Restrepo, G.; Jost, J. *The evolution of chemical knowledge: a formal setting for its analysis. Springer: Berlin* **2022**.

[5] Restrepo, G. *Chemical space: limits, evolution and modelling of an object bigger than our universal library* **Digital Discovery 2022**, *1*, 568–585.

Who's the fairest of them all? The place of the electronic lab notebook eLabFTW in data management

Nicholas Carpi

Deltablot, Paris/France

E-Mail: contact@deltablot.email

You've heard of the FAIR acronym. But what does it mean really and how can you apply its principles in your daily research activities?
We will discuss how Electronic Lab Notebooks (ELN) can help you achieve FAIRness without hassle. This talk will be articulated around eLabFTW, an open source, field agnostic ELN used by many institutions all over the world. We will also introduce the ELN file format, produced by The ELN Consortium, which allows seamless exchanges between ELN softwares, or even other types of software like data repositories (e.g. Dataverse).
After this talk, you will have a good understanding of the challenges and caveats of standardization of research data, exchange formats and the constant re-invention of the wheel.

Sciformation ELN: Integrated Multidisciplinary Solution

Felix Rudolphi

Sciformation Consulting GmbH, Bad Vilbel/Germany

E-Mail: info@sciformation.com

The digital transformation revolutionized our daily lives. In contrast to that, a substantial number of scientists still use pen and paper to document experiments and findings. To be reusable according to the FAIR principles, data must be in digital form. To close the still-existing gap, digital tools must

1. offer instant benefits for users,
2. be sufficiently flexible for various scientific disciplines, and
3. store data in structured form, to avoid just dumping data.

Sciformation ELN is more than just an Electronic Lab Notebook – it is an integrated solution covering the complete research process. When planning and documenting experiments, it provides information on starting materials and lab equipment, then helps users to collect and process analytical data, and to compile pieces of information into tables and reports. The high level of integration saves substantial amounts of time and creates benefits beyond the reach of isolated solutions. At the same time, it offers flexible interfaces to integrate with 3rd-party software.
Scientists of various disciplines can document their findings on one platform, using an extensible template system to cover specific parts. Sciformation ELN stores parameters (temperature, voltage, duration,etc.) in structured form, so that experiments and measurements can easily be filtered.
Users can attach literature references and analytical data to experiments, to have all related data in one place. In many cases, analytical data can automatically be processed: extensible Python scripts generate preview imagery and extract graphs, allowing to search for peaks and similar spectra. Analytical data can be annotated on-screen, or interpreted using locally installed 3rd-party software (e.g. MestreNova) and re-uploaded with one click.
Information on starting materials and literature can automatically be extracted using web robots, saving users' time for creative work. Using barcodes, the software can make stock-keeping of chemicals and equipment much more time-efficient. Sciformation ELN can be used to book devices and keep track of necessary maintenance. We have created chemical shop solutions for clients, so that materials can be transferred and billed in-house.
Users can publish their experiments and analytical data at https://sciflection.com in a FAIR way, and they can import interesting experimental setups back into their own ELN with two clicks.

Chemotion ELN: An open-source electronic lab notebook for FAIR data

Patrick Hodapp 1, Felix Bach 2, Sonja Herres-Pawlis 3, Nicole Jung 1, Oliver Koepler 4, Johannes Liermann 5, Steffen Neumann 6, Matthias Razum 2, Christoph Steinbeck 7

1 Karlsruher Institut für Technologie/ Karlsruhe, Germany
2 Leibniz-Institut für Informationsinfrastruktur, Karlsruhe/ Germany
3 RWTH-Aachen University, Aachen/ Germany
4 Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek, Hannover/ Germany
5 Johannes Gutenberg-Universität, Mainz/ Germany
6 Leibniz-Institut für Pflanzenbiochemie, Halle an der Saale/ Germany
7 Friedrich-Schiller-Universität, Jena/ Germany

E-Mail: patrick.hodapp@kit.edu

Electronic Lab Notebooks (ELNs) are a key prerequisite to a comprehensive documentation of research processes, the digital storage of research data, and their reuse. ELNs can be used to plan, record, store and – in combination with repositories – disclose experiments or research data. In the long run, the benefit of ELNs is the option to store and manage data in a standardized way and to enrich the data with (automatically generated) information such as metadata, identifiers and descriptors. For scientists, ELNs offer advantages such as faster research processes and a faster access to information. Selected benefits of the ELN Chemotion – an ELN that was designed for the discipline Chemistry – will be presented in this talk to show exemplarily the use of research data management tools. The ELN offers special fea-tures for chemical work and includes diverse functions that allow the use of the ELN also in other disciplines. Both, the chemistry specific as well as the generic and adaptable modules will be presented in brief.
Chemotion ELN can be used in combination with the open access Chemotion repository. The disclosure of research data to the public is possible by a direct transfer of information from the ELN to the repository. The interoperable systems ELN and repository guarantee on the one hand an easy process for the disclosure of information and on the other hand the availa-bility of comprehensive data including primary data and descriptions.

LOGS-ELN: the electronic lab notebook that understands data

Carolin Frohnapfel, Martin Haehnke, Fabian Hiller, Sina Kazemi, Corrado Surmanowicz, Andreas Wolf and Jakob J. Lopez

SIGNALS, FiZ – Frankfurter Innovationszentrum Biotechnologie, Frankfurt am Main/Germany

E-Mail: jakob.lopez@signals.company

LOGS-ELN is a browser-based electronic lab notebook for scientists conducting experiments in biochemistry, chemistry, physical sciences, material sciences and related fields. LOGS-SDMS is a webserver-based scientific data management system that automatically gathers measurements from instruments in the lab, making these available via browser and API, for researchers and other platforms (e.g. ELN and LIMS).
Besides the conventional ELN features (high-end text editor, experimental notes, dated entries, versioning & audit trail, import of images & documents, etc.) this ELN understands dataformats.
Users can upload spectroscopic data from various vendors manually or automatically (e.g. NMR, EPR, UV/Vis, Raman, etc.) and LOGS-ELN will parse the format files, resulting in a searchable database of measurements, and the ability to display spectra within a built in spectrum viewer. Data can be linked and be displayed within the ELN seamlessly, via drag & drop functionality.
A Python API (LOGS-Python) allows researchers do interact with LOGS on a script basis, e.g. to query and download specific dataset in order to process and re-upload them for documentation, or for the exchange of data with third party platforms.

Wednesday, 7 June 2023

How to make your data FAIR – a case study from a working group from synthetic chemistry

Sonja Herres-Pawlis

RWTH Aachen University, Aachen/Germany

More and more digital research data are generated in Chemistry. So, new concepts are essential: In which data formats can data be stored in the long term? How and where can data be stored? Which information of the experiment / the simulation should be noted in the metadata? How can these data be made accessible for group members and other researchers? How can these data be made findable for researchers and AI algorithms? Researchers need to be trained in these topics and concepts to apply them successfully in their daily research processes. NFDI4Chem[1-3] tackles these challenges by providing several tools and support. This talk highlights best practices in using the electronic lab notebook Chemotion as well as different repositories which are already now freely available and productively working.[4]
Moreover, NFDI4Chem offers a large array of teaching and training courses and materials – for all career stages as well as all RDM levels. For example, we offer regularly different workshops on research data management or electronic lab notebooks. Furthermore, we believe that young chemists and students are key to the cultural change. Therefore, we are increasingly paying attention on education, e.g., providing teaching courses, teaching materials, and knowledge bases. This presentation will highlight the first teaching programs implemented in theoretical and practical student courses and collected feedback of students thereof. We have collected several years of experience on the integration of RDM into an inorganic student´s lab course in the fifth semester, as a hands-on experience where students prepare special chemical substances and directly use the electronic lab notebook Chemotion to document this. Further, we deepen the RDM discussion in a Master lecture on sustainable polymerisation catalysis with case studies on good and bad RDM. Here, best practices from NFDI4Chem directly serve as case studies and are used for teaching.

[1] S. Herres-Pawlis, O. Koepler, C. Steinbeck, „NFDI4Chem: Shaping a Digital and Cultural Change in Chemistry", Angew. Chem. Int. Ed., 58, 32, 10766-10768, August 2019, *https://doi.org/10.1002/anie.201907260*

[2] K. C. Steinbeck, O. Koepler et al., "NFDI4Chem – Towards a National Research Data Infrastructure for Chemistry in Germany", Research Ideas and Outcomes, 6, e55852.2020, June 2020 *https://doi.org/10.3897/rio.6.e55852*

[3] J. Ortmeyer, J. D. Jolliffe, „Treatment of research data", Nachr. Chem., 70, 10, 16-17, Oktober 2022 *https://doi.org/10.1002/nadc.20224131398*

[4] F. Fink, H. Hüppe, N. Jung, A. Hoffmann, S. Herres-Pawlis, Sharing is Caring : Guidelines for Sharing in the Electronic Laboratory Notebook (ELN) Chemotion as applied by a Synthesis-oriented Working Group, Chem. Methods, 2022, 2, e202200026. *https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cmtd.202200026*

No data, no AI party

Paul Czodrowski

Johannes Gutenberg-Universität Mainz

E-Mail: czodpaul@uni-mainz.de

The current AI hype also has arrived in the field of chemistry. However, data quality (or even existence) is far from sufficent high quality to profoundly develop and use AI for chemistry-related approaches. We will show cases from our own AI research and how it is impacted by research data management.
One major topic will cover our contributions to the blind prediction competition EuroSAMPL. Here, we measure by UV/Vis spectroscopy pKa values of drug-like molecules and only make the structures available to the community but not the data. This is followed by prediction efforts of the community to reliably predict our measured values. After disclosure of our measured data, we can compare the different prediction methods and plan the next round of blind predictions with more fine-tuned rationales.

Using the browser for teaching and processing chemical data

Luc Patiny

Zakodium Sàrl, Lonay/Switzerland

For the last 20 years we are developing websites allowing to store, visualise and process chemical data that are also used for teaching. During this presentation we will show what can be done today in the browser using as example:

[1] Example 1: *https://www.chemcalc.org*

[2] Example 2: *https://www.c6h6.org*

[3] Example 3: *https://www.nmrium.org*

Transforming chemistry with large language models

Kevin Jablonka

École polytechnique fédérale de Lausanne, Lausanne/Switzerland

E-Mail: kevin.jablonka@epfl.ch

Data-driven techniques have made a lot of progress in the chemical sciences. They typically rely on large amounts of data structured in tabular form. The reality of (experimental) chemistry, however, is that data often cannot be easily collected in this form.
Recent advantages of the application of large language models (LLMs) to chemistry indicate that they might be used to address this challenge. My talk showcases how LLMs can autonomously use tools, leverage structured data as well as soft inductive biases, and in this way transform how we model chemistry.

DFG – German Research Foundation: Funding Opportunities and more

Ilka Paulus

Deutsche Forschungsgemeinschaft: DFG, Bonn/Germany

E-Mail: Ilka.paulus@dfg.de

The German Research Foundation (Deutsche Forschungsgemeinschaft) is the central self-governing research funding organisation in Germany. The DFG funds research projects, creates competitive opportunities and conducts procedures for the review, evaluation, selection, and decision of research proposals. Further, DFG helps to shape appropriate conditions and standards of academic research. This talk will give a glance at DFG as an organisation and introduce some of the funding programmes for early career and experienced researchers. The topics “Good Research Practice” and “Research Data Management” will also be introduced briefly.

Poster

Open-Loop Recycling of Polylactide (PLA) with Guanidine Iron(II) Catalyst - From Laboratory to FAIR Data

Lisa Burkart 1, Alexander Eith 1, Alexander Hoffmann 1, and Sonja Herres-Pawlis 1

1 RWTH Aachen University, Aachen/Germany

E-Mail: lisa.burkart@ac.rwth-aachen.de
E-Mail: sonja.herres-pawlis@ac.rwth-aachen.de

Plastics are an integral part of today's world. Unfortunately, they remain one after their end-of-life (EoL). The pollution of the environment coursed by plastics and the fossil resources used to produce conventional plastics is a major problem for today's society. The implementation of a circular plastics economy is an important step towards a sustainable future. Bioplastics such as polylactide (PLA), which are biodegradable and sourced from renewable materials, are promising alternatives for fossil-based plastics. Effective waste management and recycling are also important tools for a circular economy. Closed-loop recycling provides monomers which can be re-polymerized to the corresponding polymer, which can reduce the production costs of the polymer. An open-loop recycling approach leads to a variety of new value-added products. The previously reported non-toxic guanidine iron catalyst [FeCl2(TMG5NMe2asme)] was evaluated for its activity in the alcoholysis and aminolysis of PLA under mild conditions.[1] Using 1H NMR spectroscopy kinetic and thermodynamic studies were conducted for the methanolysis of PLA. Besides methanol, ethanol, n-butanol, i-butanol, allyl alcohol, benzyl alcohol, and benzylamine were used as nucleophiles for the depolymerization of PLA. Further, the solvent free methanolysis and ethanolysis of PLA were performed successfully.[2]
Besides the work in the laboratory, conducting experiments and the analysis of the obtained results, the FAIR processing of the obtained data is an important part of the publication process. The data should be findable, accessible, interoperable, and re-usable.[3] This is obtained by providing the data in open-source formats with a unique identifier. Key point is to provide data which is readable and processable for everyone. This will help to understand the work of other researchers better. Therefore, efficient planning and the prevention of errors can be optimized whereas misinterpretation of data can be minimized.

[1] R. D. Rittinghaus, P. M. Schäfer, P. Albrecht, C. Conrads, A. Hoffmann, A. N. Ksiazkiewicz, O. Bienemann, A. Pich, S. Herres-Pawlis, ChemSusChem 2019, 12, 2161-2165, doi: 10.1002/cssc.201900481.

[2] L. Burkart, A. Eith, A. Hoffmann, S. Herres-Pawlis, Chem. Asian J. 2023, 18, e202201195, doi: 10.1002/asia.202201195.

[3] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 'T Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. Van Schaik, S.-A. Sansone, E.

Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, Sci. Data 2016, 3, 160018, doi: 10.1038/sdata.2016.18.

nmrXiv: A highly visible, and consensusdriven NMR data repository and computational platform

V. Nainala 1, N. Sharma 1, N. Rayya 1, S. Kuhn 2, O. Koepler 3, S. Neumann 4, C. Steinbeck 1

1 Friedrich Schiller University, Jena/Germany
2 University of Tartu, Tartu/Estonia
3 Leibniz Information Centre for Science and Technology (TIB), Hannover/Germany
4 Leibniz Institute for Plant Biochemistry, Halle/Germany

E-Mail: chandu.nainala@uni-jena.de

Nuclear magnetic resonance (NMR) allows researchers to obtain rich structural information from the vibrations of the molecules in their natural environment while they are still intact. It is the standard method of determining the absolute chemical structure (chirality and, more commonly to determine the relative configuration of diastereomers). NMR is also widely used in natural product dereplication, unknown identification and metabolomics [1].
Owing to the importance of NMR data, several databases have been developed to host literaturederived and experimental NMR data of synthetic molecules and natural products over the past few decades. However, the majority of these databases cannot accept NMR data depositions from the community members. They do not meet the needs of the natural products/metabolomics communities in several ways. That includes closed licenses, lack of experimental data, limited search capabilities and more importantly, they are not FAIR compliant.
nmrXiv is currently being developed with the ultimate goal of accelerating broader coordination and data sharing among researchers by creating a computational platform for the management, sharing, and analysis of NMR data. nmrXiv is a voluntarily FAIR, cloud-based infrastructure that uses opensource code and deployment methods (openness ensures platform sustainability). Curation and analysis standards on nmrXiv for NMR data and metadata are non-prescriptivist by nature. nmrXiv also provides its users and curators with an intuitive and easy-to-use interface to create, edit and annotate their NMR studies online. Standards are built with input and contributions from the community and made compatible with existing and potential newly developed formats and approaches. Engagement of the analytical chemists and adjacent communities through webinars, workshops, tutorials, and community calls will ensure that nmrXiv meets the researchers' needs and implements adequate incentive mechanisms for data contributions and curation.

1 Nuzillard, Jean-Marc. (2021). Taxonomy-focused Natural Product Databases for Carbon- 13 NMR-based Dereplication. 10.20944/preprints202105.0701.v1.

Data Mining Unveils Property-Activity Correlation and Predict De Novo Viral Transduction Enhancing Peptide Nanofibrils

Kübra Kaygisiz 1, Lena Rauch 2, Arghya Dutta 3, Tristan Bereau 3, Christopher V. Synatschke 1, Jan Münch 2, Tanja Weil 1

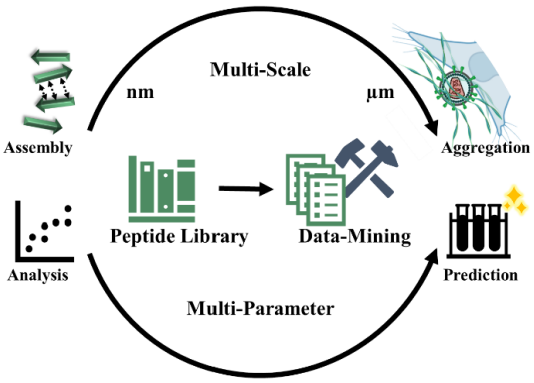
1 Max-Planck Institute for Polymer Research, Mainz/Germany
2 Institute of Molecular Virology, Ulm University Medical Center/Germany
3 Van 't Hoff Institute for Molecular Sciences and Informatics Institute, University of Amsterdam/Netherlands

E-Mail: kaygisizk@mpip-mainz.mpg.de

Self-assembling peptides can form hierarchical supramolecular fibrils that enable tuning of structural properties in a bottom-up approach. However, designing self-assembling sequences with desired bioactivity is challenging, as a small change in sequence can drastically alter physicochemical properties on multiple length scales.[1]

Here I present, how complementary physicochemical methods can be combined in the characterization of bioactive peptides to circumvent the individual scale limitations of the techniques and reveal fundamental design principles.

We chose viral infectivity enhancement as a biological readout because it offers a multi-parameter, multi-scale challenge for the optimization of biologically active supramolecular peptide assemblies. I show that we could identify a detailed sequence–structure–property relationship for our library of infectivity enhancing self-assembling peptides via a data mining approach. To this end, I created a peptide library consisting of 163 derivatives of an infectivity enhancing peptide sequence EF-C (QCKIKQIINMWQ) and characterized all peptides via experimental and bioinformatic descriptors for a multiparameter correlation. In contrast to previous reports, we find that not the self-assembled fibrils themselves, but rather µ-sized β-sheet-rich aggregates are responsible for bioactivity, i.e cell-interaction.[2] The dependencies we found apply to peptides we rationally designed based on sequence patterns from our peptide library, to short amyloidal peptides from pathogenic and functional origin not related to our library, and to inverse-engineered de novo minimalistic peptides generated by machine-learning.[3]



Insights into bioactivity and design motifs for de novo sequence creation: exploring the structure-property-activity relationship of self-assembling peptides through Data-Mining.

This work demonstrates that fundamental insights into self-assembly processes and promis-

ing design strategies of functional nanomaterials can be revealed by the powerful combination of physicochemical and computational methods.

1 S. Sieste, T. Mack, J. Münch, T. Weil et al., Adv. Funct. Mater., 2021, 31, 2009382.

2 K. Kaygisiz, L. Rauch-Wirth, A. Dutta, X. Yu, Y. Nagata, T. Bereau, J. Münch, C. V. Synatschke and T. Weil, ChemRxiv, 2023, DOI:10.26434/CHEMRXIV-2023-HFQXB.

3 K. Kaygisiz, A. Dutta, L. Rauch-Wirth, C. V Synatschke, J. Münch, T. Bereau and T. Weil, ChemRxiv, 2023, DOI:10.26434/chemrxiv-2023-1kqm3.

STRENDA – Standards for Enzymology Data

Carsten Kettner

Beilstein-Institut, Frankfurt am Main, Germany

E-Mail: ckettner@beilstein-institut.de

The STRENDA Commission (STandards for Reporting ENzymology Data, www.beilstein-strenda.org) made up of experts from the enzyme chemistry community and supported by the Beilstein-Institut, has developed the STRENDA Guidelines in tight consultation with the community. The aim is to improve the quality of enzyme function data in the literature. Today, more than 55 biochemical journals already recommend authors to refer to these guidelines when reporting enzyme kinetics data.
To enable scientists to easily prepare data for manuscripts, the STRENDA Commission has developed a web-based portal for the direct electronic submission of data by the authors prior to publication. This portal called STRENDA DB provides an assessment tool with which authors, journals' editors and reviewers can check whether the reporting of experimental data is compliant with the STRENDA guidelines and thus matches the instructions for authors from the journals. The data entered are stored in STRENDA DB and will be made publically accessible after they have been published in a journal. STRENDA DB is now recommended by more than 10 biochemical journals.
As data exchange standards and software support to aid research data management was considered lacking, the community developed the data exchange format EnzymeML. In addition, a first version of an application programming interface provides Python and Java libraries to be integrated in both applications and databases in order to drive a seamless data flow from the bench to publication platforms.

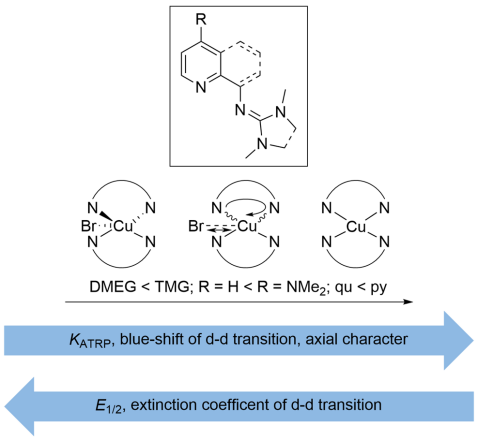
Structure-Activity Relationship of Highly Active ATRP Catalysts Based on Bidentate Ligands

Konstantin W. Kröckert, 1 Felix Garg 1, Michel V. Heinz 1, Justin Lange 1, Patricia P. Simões 1, Regina Schmidt 1, Olga Bienemann 2, Alexander Hoffmann 1, Sonja Herres-Pawlis* 1

1 RWTH Aachen University, Aachen, Germany
2 TU Dortmund University, Dortmund, Germany

E-Mail: konstantin.kroeckert@ac.rwth-aachen.de
E-Mail: sonja.herres-pawlis@ac.rwth-aachen.de

In atom transfer radical polymerization (ATRP), an equilibrium between a dormant and an active radical species mediated by a transition metal complex combines the advantages of radical polymerization with controlled polymerization characteristics. All components used, e.g. monomer, initiator, metal center, solvent and ligand, have an influence on the equilibrium. Catalyst systems based on copper complexes are by far best understood, and especially with N donor ligands they have excellent properties for ATRP. For these, several relationships between ligand structure and catalyst activity have been found so far. The present results show how structural changes in a series of bidentate ligands affect the coordination chemistry of the copper catalysts. We followed their behavior using a multi-method approach, tested the catalysts in ATRP, and found the most active copper complex for ATRP based on a bidentate ligand system.
+Major parts of the raw data in this publication are deposited in the Chemotion repository, allowing FAIR access. Therefore, the results of this publication are a best practice example for storing original scientific data, which is highlighted on the present poster



Conceptual structural change of copper(II) complexes based on bidentate ligands and influence on obtained properties

[1] K. W. Kröckert, F. Garg, M. V. Heinz, J. Lange, P. P. Simões, R. Schmidt, O. Bienemann, A. Hoffmann, S. Herres-Pawlis, Dalton Trans. 2022, 52, 13272-13287.

WWW.NMRIUM.ORG: Revolutionizing NMR Spectra Processing with a Free Web-Based Application

Hamed Musallam 1, Alejandro Bolanos 2, Luc Patiny 2

1 University of Jena, Department of Chemistry, Jena (Germany)
2 Zakodium, route d'Echandens, 1027 Lonay (Switzerland)

Nuclear Magnetic Resonance (NMR) spectroscopy is a widely used technique in chemistry and related fields for elucidating chemical structures and analyzing molecule properties. However, processing and analyzing NMR spectra can be challenging and often requires specialized software. To address this issue, NMRium (https://www.nmrium.org) is an innovative web-based application that provides a breakthrough solution, particularly in the context of making chemical knowledge collective and machine actionable[1].
In the poster, we will introduce NMRium and list its main features which will be valuable for the researchers.
NMRium supports various NMR formats (including Bruker, Varian, Jeol, and JCAMP-DX), molecule assignment, and superimposition, making it a valuable resource for both researchers and NMR facilities. Additionally, NMRium can be used for educational purposes, such as e-learning, as it enables the creation of exercise series for students without requiring the installation of any software.

1 Jablonka, K. M., Patiny, L., & Smit, B. (2022). Making the collective knowledge of chemistry open and machine actionable. Nature Chemistry, 14(4), 365-376.

2 Patiny, L., Musallam, H., Zasso, M., Bolaños, A., Kostro, D., Wenk, M., Ahkrin, S., Jeannerat, D., Ziegler, E., Liermann, J., Schloerer, N., & Wist, J. NMRium: ReactJS component to display and process NMR spectra [Computer software]. https://doi.org/10.5281/zenodo.5101777

NFDI4Chem: The current landscape of author guidelines in chemistry through the lens of research data sharing

N. A. Parks 1*(0000-0002-6243-2840), T. G. Fischer 2*(0000-0003-4480-8661), C. Blankenburg 2(0000-0002-7955-8378), V. F. Scalfani 3(0000-0002-7363-531X), L. R. McEwen 3(0000-0003-2968-1674), S. Herres-Pawlis 1(0000-0002-4354-4353), S. Neumann 2(0000-0002-7899-7192).

1 RWTH Aachen University, Aachen/Germany (04xfq0f34)
2 Leibniz Institute of Plant Biochemistry, Halle (Saale)/Germany (01mzk5576)
3 Rodgers Library for Science and Engineering, Tuscaloosa/USA (03xrrjk67)
4 Cornell University, Ithaca/USA (05bnh6r87)

E-Mail: parks@itc.rwth-aachen.de
E-Mail: tfischer@ipb-halle.de

NFDI4Chem's vision is the digitalisation of all key steps in chemical research to support scientists in their efforts to collect, store, process, analyze, publish, and re-use research data [1]. For that goal, we aim at making chemical research data FAIR (Findable, Accessible, Interoperable, Reusable) [2].
As the primary method of communicating research results, journals and their author guidelines have a tremendous impact on community behavior. To work with scientific journal editors to enhance recommendations on data publication, we organised the Editors4Chem workshop in 2021 [3]. The 2nd Editors4Chem workshop will be held in November of 2023. In addition to these workshops, we surveyed author guidelines from several publishers and journals active in chemistry research. The results not only impart important information for journals and publishers, but also for authors in terms of adjusting to new requirements for research data underlying a journal publication.
With this poster, we will present the results of a large-scale analysis [3] of author guidelines: To which extent is the publishing landscape supporting FAIR Data and Open Science practices? In which areas is this support lacking and what might be underlying reasons? How should authors navigate these changes?

1 C. Steinbeck, O. Koepler, F. Bach, S. Herres-Pawlis, N. Jung, J. Liermann, S. Neumann, et al., Research Ideas and Outcomes, 2020, 6: e55852. DOI: 10.3897/rio.6.e55852. Website: www.nfdi4chem.de.

2 M. Wilkinson, M. Dumontier, I. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al. Sci Data 2016, 3, 160018, DOI: 10.1038/sdata.2016.18.

3 T. G. Fischer, 1st Editors4Chem Workshop, www.nfdi4chem.de, 2021, URL: www.nfdi4chem.de/index.php/2021/11/23/1st-editors4chem-workshop/

4 N. A. Parks, T. G. Fischer, C. Blankenburg, V. F. Scalfani, L. R. McEwen, S. Herres-Pawlis, S. Neumann, Pure Appl. Chem. 2023, ASAP. DOI: 10.1515/pac-2022-1001.

www.nmrium.org: Ultimate online platform to teach NMR spectroscopy

Luc Patiny 1, Johannes Liermann 2

1 Zakodium, route d'Echandens, 1027 Lonay (Switzerland)
2 Johannes Gutenberg University Mainz, Department of Chemistry, 55099 Mainz (Germany)

E-Mail: luc.patiny@zakodium.com

Over the past decade, we are doing research on online teaching of NMR spectroscopy [1]. In this regard, we have developed an innovative web-based NMR processing and analysis software called NMRium [2] [3]. NMRium is a cutting-edge web application designed specifically for visualizing and processing NMR data. With the aid of NMRium, we can now offer students challenging exercises that require them to determine the structure of unknown compounds using NMR spectra and molecular formula. These exercises cover a wide range of topics, including integration, multiplicity analysis, Fourier transform (FT), phase correction, baseline correction, as well as 1D and 2D spectra, among others.

A selection of exercises is already accessible on the website: https://www.nmrium.org/teaching. We encourage everyone to utilize these series free of charge. If you wish to develop your own exercises, please reach out to us, and we will provide assistance throughout the process.

1 Patiny, L., Bolaños, A., Castillo, A. M., Bernal, A., & Wist, J. (2018). Teaching NMR spectra analysis with nmr. cheminfo. org. Magnetic Resonance in Chemistry, 56(6), 529-534.

2 Davies, A. N., & Patiny, L. (2021). NMRium browser-based nuclear magnetic resonance data processing. Spectroscopy Europe, 33(4).

3 Patiny, L., Musallam, H., Zasso, M., Bolaños, A., Kostro, D., Wenk, M., Ahkrin, S., Jeannerat, D., Ziegler, E., Liermann, J., Schloerer, N., & Wist, J. NMRium: ReactJS component to display and process NMR spectra [Computer software]. https://doi.org/10.5281/zenodo.5101777

Recent advancements in DECIMER.ai, an open platform for automated optical chemical structure identification, segmentation, and recognition in scientific publications

Kohulan Rajan 1, Henning-Otto Brinkaus 1, M. Isabel Agea 2, Achim Zielesny 3 and Christoph Steinbeck 1

1 Institute for Inorganic and Analytical Chemistry, Friedrich-Schiller-University Jena, Lessingstr. 8, 07743, Jena, Germany
2 Department of Informatics and Chemistry, Faculty of Chemical Technology, University of Chemistry and Technology Prague, Technicka 5, 166 28, Prague, Czech Republic.
3 Institute for Bioinformatics and Chemoinformatics, Westphalian University of Applied Sciences, August- Schmidt-Ring 10, D-45665 Recklinghausen, Germany

E-Mail: kohulan.rajan@uni-jena.de

The majority of chemical information published in literature is not available in machine-readable form in public databases. The process of extracting this information manually is a time-consuming and error-prone task [1]. The information published in the printed literature is often not structured in a manner that can easily be converted into a machine-readable format and stored in public-domain databases [2]. As an open-source platform, DECIMER.ai (Deep Learning for Chemical Image Recognition) automates the segmentation, classification, and translation of chemical structure depictions from printed literature into machine-readable representations by utilizing deep learning, computer vision, and natural language processing. DECIMER.ai [3] consists of three deep learning models: DECIMER Segmentation [4], which can detect and segment chemical structure depictions, DECIMER Image Classifier, which can identify whether an image contains a chemical structure or not; and DECIMER Image Transformer [5], an encoder-decoder model that converts chemical structure depictions into computer-readable files. These combined techniques can achieve high levels of accuracy without reliance on hand-picked rules. The optical chemical structure recognition (OCSR) engine of DECIMER.ai has outstanding performance on all benchmark datasets, and the segmentation and classification tools are the only openly available packages of their kind. This work is openly available on GitHub, and the web application is available at https://decimer.ai. The source code, trained models, and datasets developed in this work have been published under permissive licenses, allowing users to modify and redistribute the code easily.
DECIMER.ai aims to significantly reduce the workload and produce high-quality data for the research community and those developing and curating chemical databases.

[1] Swain MC, Cole JM (2016) ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature. J Chem Inf Model 56:1894–1904

[2] Rajan K, Brinkhaus HO, Zielesny A, Steinbeck C (2020) A review of optical chemical structure recognition tools. J Cheminform 12:60

[3] Rajan K, Brinkhaus HO, Isabel Agea M, Zielesny A, Steinbeck C (2023) DECIMER.ai – An open platform for automated optical chemical structure identification, segmentation

and recognition in scientific publications. ChemRxiv. https://doi.org/10.26434/chemrxiv-2023- xhcx9

[4] Rajan K, Brinkhaus HO, Sorokina M, Zielesny A, Steinbeck C (2021) DECIMERSegmentation: Automated extraction of chemical structure depictions from scientific literature. J Cheminform 13:20

[5] Rajan K, Zielesny A, Steinbeck C (2021) DECIMER 1.0: deep learning for chemical image recognition using transformers. J Cheminform 13:61

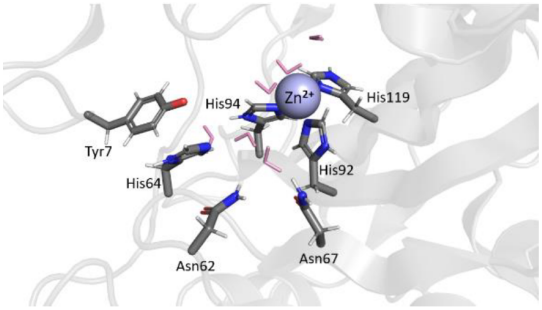
An old target revisited: A structure-based case study on human carbonic anhydrase II

Anna Riede 1, Lisa Johannknecht 1, Paul Czodrowski 1

1 Johannes Gutenberg University Mainz, Institute of Physical Chemistry, Germany
E-Mail: anriede@uni-mainz.de

Knowledge of the pKa value and thus the protonation state of proteins (macro- and microscopically with respect to individual amino acid residues) is a key issue in drug discovery given its significant impact on molecular recognition properties.[1] We use the Poisson-Boltzmann-Equation solver ZAP by OpenEye[2] embedded in a python-based workflow allowing for high-throughput calculations with minimal human intervention to predict individual amino acid pKa values of proteins and to elucidate possible pKa value shifts and protonation effects that occur during the binding process between proteins and small molecule drugs.

A NMR study of the collaboration group of R. Linser showed pH-dependent chemical shift alterations for the metalloenzyme human carbonic anhydrase II (hCAII) in general and the amino acid residues Asn62 and Asn67 in particular, which might be explained by protonation state changes. The spatially closest titratable amino acids are His64 and Tyr7 located within the cone-shaped active site cavity (cf. Fig.).[3]



First, the robustness of the protein_pKa calculations with respect to the predicted value for His64 (and all other 11 His of hCAII) was examined. The sensitivity analysis included all (ligand-deleted and true) holo wild-type structures available in the Protein Data Bank (PDB), the largest open access collection of 3D-structural data of biological macromolecules. Thereby, common chemical data science methods were employed for retrieving, filtering, preprocessing, and visualization of the data. Next, calculations were performed on two individual holo hCAII neutron diffraction structures 3KKX and 4Y0J as models for the limiting pH states (6.0 and 9.0) in which the enzyme is physiologically active.[3,4] Furthermore, a structural analysis in PyMOL[5] and protonation state evaluation was performed to match the computationally predicted pKa values with the crystallographically observed protonation states. The first results look promising and will be presented on the poster.

We are currently investigating a dorzolamide-bound hCAII structure (pdb: 6BC9)[6] and are planning to enrich our computational approach with isothermal titration calorimetry measurements in the future. Within the framework of the NFDI4Chem Chemistry Data Days, insights into data resources, (pre-) processing and visualization in context of the research project described above will be showcased on the poster.

[1] P. Czodrowski, C. Sotriffer, G. Klebe, Journal of molecular biology, 2007, 1347-1356.

[2] OpenEye Toolkits, version 2021.2.0, OpenEye Scientific (https://www.eyesopen.com)

[3] S. Fisher, A. Kovalevsky, J.F. Domsic, R. McKenna, D. Silverman, P. Langan, Biochemistry, 2010, 49: 415-421.

[4] R. Michalczyk, A. Kovalevsky, R. McKenna, S. Fisher, et al. Proc Natl Acad Sci USA, 2015, 112, 5673-5678

[5] The PyMOL Molecular Graphics System, version 2.5.4, Schrödinger LLC (https://pymol.org).

[6] A. Kovalevsky, S. Fisher, R. McKenna, et al. Structure, 2018, 26, 383-390.