International Union of Pure and Applied Chemistry
Division VIII
Chemical Nomenclature and Structure Representation

Approved Minutes for the Division Committee Meeting

Date: **Saturday, 19 September 2020**

Time: **08:00 EST (05:00 PDT, 13:00 BST, 14:00 CEST, 15:00 EEST, 24:00 NZST)**

Venue: **Online (GoToMeeting platform)**

1. Welcome, introductory remarks and housekeeping announcements

Alan Hutton (ATH) welcomed everybody to the meeting, extending a special welcome to those who were attending the Division Committee (DC) meeting for the first time. Because of the Covid-19 pandemic, the (online) format of the meeting was different from previous years, and he described the working protocols and arrangements for the meeting. He noted that the time was early in the morning in the USA, midnight in New Zealand, and that the meeting would have to move along swiftly to cover the agenda items that were dealt with during the normal one-and-a-half day physical meeting.

2. Attendance and apologies

Present for all or part of the meeting: Alan T. Hutton (President, ATH), Risto S. Laitinen (Secretary, RSL), Michael A. Beckett (MAB), Edwin C. Constable (ECC), Ture Damhus (TD), Safiye Erdem (SE), Adeyinka Fasakin (AF), Richard M. Hartshorn (RMH), Robin Macaluso (RM), Elisabeth Mansfield (EM), Leah R. McEwen (LRM), Gerard P. Moss (GPM), Warren Powell (WP), Amelia P. Rauter (APR), Molly A. Strausbaugh (MAS), Erik Szabó (ES), Augusto Tomé (AT), Clare A. Tovee (CAT), Jiří Vohlídal (JV), Andrey Yerin (AY)

(For the Division VIII membership in 2020-2021 and the group photo of the meeting attendees, see Appendix 1)

Apologies: Thomas Engel (TE), Steve Heller (SH), Karl-Heinz Hellwich (KHH), Ebbe Nordlander (EN), Michelle M. Rogers (MMR)

Absent: Neil Burford (NB), Rafał Kruszyński (RK), G. Jeffery Leigh (GJL), Alan McNaught (AM), Ladda Meesuk (LM), Jozsef Nagy (JN), Maria A. Petrova (MAP), Dušan Sladić (DS), Keith T. Taylor (KTT), Guoqiang Yang (GY)

3. Introduction of attendees

A short round of introductions was made. Augusto Tomé (AT), Adeyinka Fasakin (AF) and Safiye Erdem (SE) were attending the meeting of the Division Committee for the first time. ATH reported that there had been a slight improvement in the condition of our current TM Karl-Heinz Hellwich, but that he still needed full-time care and would not be able to participate in the affairs of the Division for some time. Quite apart from the personal tragedy, this was a serious loss for the Division, both in terms of the many projects in which he played a key role, as well as for the depth of his institutional knowledge. ATH also informed the DC that friend and supporter of the Division, Prof. Alexander Senning, had passed away due to Covid-19. He was a past member of the Advisory Subcommittee of the DC and had worked actively in chemical nomenclature within Denmark.

4. Approval of agenda

The draft agenda was approved (see Appendix 2).

5. Approval of minutes of meeting in Paris, 6–7 July 2019

While the substance of the minutes from the Paris DC meeting received approval, there were some last-minute technical changes. It was decided that the minutes should be distributed as soon as possible and that final comments should be sent to RSL by Sunday, September 27. If nothing of consequence was detected, the minutes would be considered approved and posted on the Division webpage.

> [*Secretary's note:* There were no comments by September 27, 2020, and the minutes were considered approved and posted to the IUPAC webpage on September 30, 2020]

6. Matters arising

It was agreed that most points arising from the Paris minutes would be discussed in connection with other items on the Agenda. The discussion about Division Rules, which was started in Paris (see Item 7 in Paris minutes), was briefly continued.

> ***Action:** A working group consisting of ATH, TD, GPM and RSL was formed with a brief to consider Division Rules and suggest modifications.

ATH reported that *Pure and Applied Chemistry* was receiving fewer papers for publication because of the reduced number of conferences resulting from Covid-19 cancellations. He noted that this did not directly affect Division VIII, since the outcomes of the work of the Division Task Groups would in any case be published in *PAC* as Recommendations or Technical Reports.

ATH noted that Covid-19 had actually improved the financial situation of IUPAC, because of the significantly reduced travel expenses resulting from the meetings of Task Groups being conducted online.

ATH reported that the Secretariat and the organizers in Montreal had started the planning for the (now to be virtual) General Assembly scheduled for August 2021, including Council and all associated meetings. The continuation of the Covid-19 pandemic would create a problem for all meetings next year. The IUPAC Statutes and Bylaws require that the Council meetings, and in particular the elections, must be carried out face-to-face, and electronic voting is currently not allowed. If the Council meeting cannot be arranged with physical in-location attendance, there might be a mechanism to circumvent this requirement. IUPAC could be placed under the Swiss law, which would be quite logical, since the Union was established in Switzerland. The Swiss legislation allows electronic remote voting. It was generally noted that the pandemic situation renders the meeting of committees in physically the same location challenging, but that virtual meetings across multiple time zones could be even more challenging.

7. Interactions between Division VIII and other bodies in relation to documents and projects involving chemical nomenclature

ATH invited RMH, in his capacity as IUPAC Secretary-General, to make some general observations. RMH continued and expanded the discussion on the effect of the Covid-19 pandemic on the functioning of IUPAC. The main change was that face-to-face meetings had been minimized, if not made impossible. He noted that electronic meetings generally ran smoothly, but that it was more difficult to do creative work in Task Groups remotely. There had been fewer project proposals submitted than normal this year, though a reduced number was quite common during the first part of the biennium. As Acting Chair of the *PAC* Editorial Advisory Board, RMH endorsed the comments made earlier by ATH regarding the paucity of manuscripts being received, and he encouraged members to consider submitting proposals for themed or special issues.

Division I. The Division VIII contact person is Risto Laitinen (RSL) (the Division I counterpart is Roberto Marquardt). RSL reported that there had been no contact since Paris.

Division II. The contact person is Robin Macaluso (RM), who is also a TM in Division II during the 2020-2021 biennium (the formal Division II counterpart is Daniel Rabinovich).

Division III. The contact person is Amélia Rauter (APR), who is also the Vice-President of Division III (and thus serves as the Division III counterpart).

Division IV. The contact persons are Jiří Vohlidal (JV) and Andrey Yerin (AY) (the Division IV counterpart is still unclear). JV and AY are members of the Subcommittee on Polymer Terminology (SPT), providing natural overlap. The overlap with SPT was considered so important that two contact persons were needed.

Division V. Risto Laitinen (RSL) is the contact person (with M. Clara Magalhães as the Division V counterpart). There is very little overlap between the work or activities of the two Divisions.

Division VI. Edwin Constable (ECC) is the contact person (the Division VI counterpart is yet to be established).

Division VII. Ture Damhus (TD) is the contact person for Division VII (with Helle Møller Johannessen as the Division VII counterpart). She has kindly sent us a summary of the highlights of the activities of Division VII (see Appendix 3).

> ****Action:*** The reactivation of contact relationships with Divisions IV and VI will be continued. RSL will interact with the Secretaries of these two Divisions.

ICTNS (Interdivisional Committee on Terminology, Nomenclature and Symbols). Ture Damhus is the representative of Division VIII on ICTNS and Gerry Moss is a Titular Member. TD reported that there had been plenty of discussion about developing the review process. He reported that the role of Division Representative in ICTNS was rather problematic, since it was sometimes difficult to know exactly how to behave. There were generally no problems for the Division Representative to review documents with no authors from the home Division. If there were authors from the home Division, and there had been a Division review, the ICTNS representative could have no further say. If he had objections to the paper while it was reviewed in the Division, but was overruled, there would not be much sense in repeating these objections as an "anonymous" ICTNS reviewer. Therefore, the Division Representative cannot work as an ordinary referee in the usual sense. It would be useful to act in cooperation with other members within the Division and in particular with people from other Divisions. This underlined the need for the Web Discussion Board (see Item 15.3). GPM noted that it was useful for several people look at the documents, in particular when dealing with nomenclature, and comments should be shared. For one person, a detailed assimilation of 300 pages of a book in great detail during a short timeframe was not possible.

> [*Secretary's note:* This topic was further discussed under Agenda Item 11, where an Action was invoked.]

InChI (International Chemical Identifier) Subcommittee. Andrey Yerin (AY) and Clare Tovee (CAT) continue as the Division VIII representatives on the InChI Subcommittee. The contact with the InChI Subcommittee needs to be expanded, since much of its work is directly related to Division VIII interests (see also Item 8).

SPT (Subcommittee on Polymer Terminology). Jiri Vohlidal and Andrey Yerin are the Division VIII Representatives.

CPCDS (Committee on Printed and Cheminformatics Data Standards). CPCDS is a Standing Committee of IUPAC. AY is a member of the Subcommittee on Cheminformatics Data Standards (SCDS), as well as a member of the InChI Subcommittee, which is a Division VIII subcommittee. Thomas Engel is Division VIII Representative on SCDS. Leah McEwen, who is the Chair of CPCDS is also *ex officio* a member of Division VIII.

CAS. The interaction with Chemical Abstracts Service (currently named only as CAS) is important for Division VIII. Molly Strausbaugh is our representative on CAS and is also NR for the USA on Division VIII.

ISO (International Organization for Standardization). Edwin Constable is the Division VIII representative to ISO. The joint project work in connection with nanoparticles and carbon nanotubes is in its initial stages.

CCDC (Cambridge Crystallographic Data Centre). Clare Tovee is an AM on the Division VIII Committee and is thus the natural contact between the two organizations.

JCBN (IUBMB–IUPAC Joint Commission on Biochemical Nomenclature). GPM is the Chair of JCBN, TD and APR are Associate Members, and ATH is a member *ex officio*. JCBN is a Joint Commission of IUBMB and IUPAC. GPM noted that while IUPAC dealt with chemical nomenclature, IUBMB dealt with enzyme nomenclature. JCBN had not met this year due to restrictions resulting from the Covid-19 pandemic.

REACH (Registration, Evaluation, Authorization and Restriction of Chemicals). Michelle Rogers is the Division contact to REACH, and together with TD and MAS forming a scoping group. Since MMR was unable to attend the DC meeting, there was no report. TD mentioned that there was a potential external contact from Denmark available to participate in these activities.

ACS (American Chemical Society) Nomenclature Committee. Molly Strausbaugh is the contact person for the ACS Nomenclature Committee. There is an extended report on its current activities in Item 9 of the minutes of the Division VIII Committee meeting in Paris. MAS reported that an update of the ACS Style Guide could now be found on the web.

RSC (Royal Society of Chemistry). Gerry Moss is the Chair of the RSC Committee on Standards in Nomenclature, Terminology, Units and Symbols and is therefore the contact person for Division VIII. He reported that RSC and national (UK) IUPAC activities have now been merged.

8. InChI report

ATH drew attention to two biannual reports from Steve Heller as InChI Trust Project Director, which are attached to these minutes as Appendices 4(a) (January 2020 report) and 4(b) (August 2020 report). It was noted that the name of Clare Tovee was missing in the reports as an InChI Subcommittee member, and her role as the Division VIII representative should be mentioned.

> [*Secretary's note:* The Chair and Secretary of the InChI Subcommittee were subsequently alerted to this omission and action has been taken.]

9. JCBN report

GPM reported that JCBN did not have a formal meeting this year due to travel restrictions resulting from the Covid-19 pandemic. The work on enzymes is, however, continuing.

10. CPCDS report

ATH reported that a CPCDS newsletter had been circulated to the Division. The Chair of CPCDS, Leah McEwen (LRM), indicated that Thomas Engel was another representative from Division VIII with interests in the work of CPCDS, and was Division VIII Representative on SCDS (see item 7). LRM outlined the current work on the updating of the Gold Book. She also mentioned that CPCDS had been collaborating with the Polymer Division's SPT to develop Wikipedia. The information is

presented on Wikipedia as images, which cannot be modified, and they carry IUPAC labels. LRM pointed out that much of the work of CPCDS is about chemical structure representation, and that the effects of the Covid-19 pandemic had slowed down progress. RMH had asked for a five-page discussion document about the possibilities of the structural representation of molecules. LRM indicated that there should be a "Colour Book" on cheminformatics, explaining how emerging digital standards can be used. The scoping project was complete, and the work could now start, with the first theme being chemical representation.

11. Division review of the outcomes of projects (i.e., Recommendations and Technical Reports) prior to submission to ICTNS

ATH initiated discussion on the process involved when reviewing the outcomes of projects prior to submission to ICTNS. There was a feeling that the Division Committee (DC) should supply a more consolidated review outcome, rather than relying on individual reports from members. It was proposed that when the document to be reviewed is received from the Task Group Leader (TGL), the Division President (DP) should send the document to the entire DC, requesting that responses be sent to all DC members in addition to the TGL. These responses should highlight matters that might benefit from discussion within the Division. If necessary, such discussion (initiated by the DP) could take place via e-mail or, if easier, via a virtual meeting of those interested. The DP would then, if needed, provide a consolidated response for onward transmission to the TGL and ICTNS. It was recognized that not all DC members will feel able to respond to all requests for reviews, depending on areas of expertise, etc.

ATH noted that the above *modus operandi* had worked well for two recently reviewed projects, namely for a Division VII project on anti-doping/prohibited substances (Abatte) and for a Division IV project on polymeric conjugates (Vert).

TD noted that when the Division supplies a consolidated opinion, the role of the Division Representative (DR) on ICTNS would be diminished, as he would not be able to add anything new. There was no need to duplicate the work. ATH remarked that the policy should be that if the review comes from the Division, the DR should not be asked to add anything else.

ATH also noted that a point to consider in the preparation of Recommendations and Technical Reports was whether the whole list of names of the DC roster needed to be included as a postscript to the article, as was currently the practice. There was a feeling that only the authors, who actually participated in the project process and writing of the article, should be listed on the front page.

> **Action:** ATH will write to Jürgen Stohner (Chair of ICTNS) to suggest that the Division Representative should not be asked to review a project outcomes document when the Division has already submitted a consolidated opinion. He also undertook to raise the issue of whether the listing of all DC members at the end of an article was really necessary.

12. Division review of project proposals submitted to the IUPAC Secretariat

ATH described the procedure for the evaluation of project proposal submissions. When the proposal is received by IUPAC, it is sent by Fabienne Meyers (FM) to all members of the relevant Division(s)

with the request to reply directly back to her (only). However, if there were issues of general interest to the Division, there might be a need for discussion. DC members are encouraged to raise issues for debate, and the DP should identify those that need to be debated and facilitate this by e-mail or online. In order to preserve anonymity and confidentiality, there was no need to distribute the individual reviews to the DC. At the end of the review period, FM sends the reviews submitted by the Division members, as well as those by the nominated external reviewers, to the DP, who makes a recommendation for acceptance and budget allocation in the context of any discussion that happened within the Division.

ATH noted that, as mentioned in Agenda Item 11 (above), it was recognized that not all DC members will feel able to respond to all requests for reviews, depending on areas of expertise, *etc*. Further, he observed that it was not necessary to respond to all of the several questions posed in the review form template – DC members should only engage with those questions that were relevant or within their competencies.

As an alternative to the protocol outlined above, ATH put forward the idea that the procedure be similar to that agreed above for the review of project outcomes (Agenda Item 11), but with the difference that the initiator be FM rather than the DP. This would mean that all DC members would receive copies of all the reviews. In the ensuing discussion ES brought up the importance of the anonymity of referees and the problem of potential conflicts of interest, *e.g.,* more positive reviews might be received for proposals from within the Division than for those coming from external sources. ATH conceded that it was important that only FM received the reports and then sent them together with those from the external referees to the DP, who would then decide whether there were issues to discuss with the whole DC. The meeting thus agreed to adopt the protocol outlined in the first paragraph of the minute of this Agenda Item (above).

RM enquired about the accepted timeframes for the assessment of projects, whether for a proposal or an outcome. ATH responded that the timeframe was normally given when the review was requested and this was usually one month. RM noted that the time of response was only a problem if the project involved more than one Division, when there could be serious delays. In such cases, the time-frames of the reviews have to be coordinated.

13. Reports on Division VIII projects

13.1. Nomenclature and associated terminology for inorganic nanoscale particles (2019-016-3-800, Scott Brown, Edwin C. Constable)

ECC reported that a preliminary meeting had been held. The next meeting was being set up. It was, however, difficult to make coherent progress.

13.2. Building Broader and Deeper Links Between OPCW and IUPAC (proposal 2018-022-2, Richard Hartshorn)

RMH reported that project was almost complete. Nomenclature was not relevant for this project – it was more important to consider education.

13.3. Alignment of principles for specifying ligands and substituent groups across various areas of nomenclature (2017-033-1-800, Ture Damhus)

TD reported that after some e-mail activity in the fall of 2019, there had been little joint work in the Task Group. The Paris minutes thus represent well the status of the project. The Chairman has made it clear that he needs the support of everyone to continue elaborating the several substantial pieces of text that already exist into an all-inclusive document. Timely communication within the group is key to further progress.

13.4. Graphical representation standards for chemical reaction diagrams (2003-045-3-800/2012-033-1-800/2017-036-2-800, Keith T. Taylor)

GPM reported that the work was progressing. The last document was from February 2020.

13.5. IUPAC International Chemical Identifier (InChI) projects

The reports on the individual InChI projects (below) are covered in the written reports submitted by SH [see Appendices 4(a)(January 2020) and 4(b)(August 2020)].

13.5.1. Enhanced recognition and encoding of stereoconfiguration by InChI tools (2019-017-2-800, Andrey Yerin)
13.5.2. InChI extension for mixture composition (2015-025-4-800, Leah McEwen)
13.5.3. Identifying International Chemical Identifier (InChI) Enhancements – QR codes and Industry Application (2015-019-2-800, Richard Hartshorn)
13.5.4. Implementation of InChI for chemically modified large biomolecules (2013-010-1-800, Keith Taylor)
13.5.5. Handling of Inorganic Compounds for InChI V2 (2012-046-2-800, Richard Hartshorn and Hinnerk Rey)
13.5.6. Redesign of Handling of Tautomerism for InChI V2 (2012-023-2-800, Marc Nicklaus)
13.5.7. InChI requirements for Representation of Organometallic and Coordination Compound Structures (2009-040-2-800, Colin Batchelor)
13.5.8. InChI Open Education Resource (OER) (proposal 2018-012-2, Robert Belford)

13.6. Corrections, Revisions and Extension for the Nomenclature of Organic Chemistry - IUPAC Recommendations and Preferred Names 2013 (the IUPAC Blue Book) (2015-052-1-800, Karl-Heinz Hellwich)

GPM reported that before the Paris meeting the review of the backlog of corrections to the Blue Book had been finally completed and was now on the web. The next stage of the project, which was to put an html version of the Blue Book on the web, had been started. So far up to the end of Chapter P-10 and Appendix 1 had been completed.

The statistics reveal that 92% of the book had been covered, with over 4700 images captured (structures and other graphics).

It was pointed out that there were frequent mentions of the term ‘chiral center’ in the book. The question was raised as to whether these should be changed to ‘chirality center’ (the center itself not

being chiral). It was agreed that although the book does use ‘chirality center’, the term ‘center of chirality’ was preferred.

WP had submitted a short discussion document on ‘The Future of Blue Book’ (see Appendix 5). The Blue Book had been translated into Japanese and during that work new corrections were pointed out. GPM acknowledged Renji Okazaki from Japan for his careful proof-reading of the html Blue Book chapters. GPM also queried whether Appendix 3 of the Blue Book should contain the CAS names, when CAS uses a systematic name.

TD pointed out the importance of good proof-reading, especially as the creation of the html version required redrawing of the formulae.

13.7. Nomenclature of carbon nanotubes and related substances (2013-056-1-800, Elisabeth Mansfield)

EM reported that there had been some progress made on the draft, including revisions to the text and new figures. Unfortunately, everyone had commented on a different draft and revisions were often conflicting. The reconciliation of these had taken significant time but a new draft had been established during Covid times and will be redistributed to project Task Group Members soon to make sure everything was captured. The new draft will be moved to an online editing platform to make sure further comments are collected on the same draft.

13.8. End-of-line hyphenation of systematic chemical names (2014-003-2-800, Albert Dijkstra)

The final Recommendation, after public review, had been sent for publication. Albert Dijkstra and Jan Reedijk had submitted written reports to the DC (see Appendix 6).

13.9. Nomenclature for metallacycles containing transition metals (2013-030-1-800, Alan Hutton)

ATH reported that progress since the Task Group meeting in Paris had been minimal. The main discussions in Paris were around the naming of charged metallacycles and it was decided that TD would continue the work on that topic. He has since provided the Task Group Chair with several further thoughts and proposals regarding these species. It now remains for this work to be incorporated into the draft document, along with improvements to the earlier sections, for consideration by the wider Task Group. To a certain extent, further progress will be limited until there is clarity on the developments around the kappa notation that is being considered by the Alignment Project.

13.10. Nomenclature for polyhedral boranes and related compounds (2012-045-1-800, Michael Beckett)

MAB reported that the manuscript was finally accepted whilst we were in our DC meeting in Paris in July 2019. Proofs were received in November 2019 and again in December 2019. The Recommendation was published on-line in December 2019 and in print in February 2020: *Pure and Applied Chemistry*, 2020, **92**, 355–381. He expressed his gratitude to all the task group members and to DC members who had contributed to the many overlapping discussions held during other Task Group meetings.

13.11. Revision and extension of IUPAC recommendations on carbohydrate nomenclature (2012-039-2-800, 2015-035-2-800, 2017-026-1-800 Johannes Vliegenhart)

Johannes Vliegenhart had provided a written report to the DC (see Appendix 7). While progress had been made, the work was ongoing and it appears that several issues remain to be resolved before a document can be finalized.

13.12. A comparison of assignment of hydro prefixes, added and indicated hydrogens in IUPAC, CAS and Beilstein nomenclature systems (2012-037-1-800, Andrey Yerin)

AY reported that the project was very close to completion and had gone through several rounds of review by the Task Group Members. The latest draft had been sent to the members on 17 September 2020 and contained just a few minor unresolved questions. It was hoped the document could be submitted for publication by the end of this year. It is a Technical Report and does not require wide public review, though an internal review may be formally necessary.

13.13. Terminology and nomenclature of inorganic and coordination polymers (2011-035-1-800, Richard Jones); for short TINCOPS

Due to the illness of Richard Jones, Lars Öhrström was acting as Task Group Chair. Recent progress involved updating the old document from 1984. TD reported that the structures had had to be redrawn and that he had provided corresponding names. One of the remaining problems involved the ranking of heteroatoms in the chain; however, there was a similar problem for the metallacycles.

13.14. Brief guides to the nomenclature of organic and inorganic chemistry ('Essentials' of organic and inorganic nomenclature) (2010-055-1-800, Richard Hartshorn and Karl-Heinz Hellwich)

This project had been completed.

Erik Szabó informed the meeting that Slovac translations of both the organic and the polymer Brief Guides had recently been published in the official magazine of the Slovac Chemical Society. The magazine is available online here: http://schems.sk/chemzi_pdf/ChemZi_1_2020.pdf, with the organic Brief Guide on pages 30-33 and the polymer Brief Guide on pages 34-35.

13.15. Glossary of small molecules of biological interest (2009-022-2-800, Gareth Owen)

With the retirement of Marcus Ennis this project was now run by Gareth Owen. There were no new developments to report.

13.16. Preferred names for inorganic compounds (2006-038-1-800, Ture Damhus)

This 'Inorganic PINs' project had been merged with the Alignment project (see Agenda Item 13.3 and further details in Paris minutes).

13.17. Nomenclature of phosphorus-containing compounds of biochemical importance (2006-019-1-800, Gerard Moss)

GPM reported that this project had been discussed in the JCBN meeting last year. It was decided to terminate it and submit a proposal for a new project.

***Action:** APR will prepare a project proposal.

13.18. Polymer projects (with Division IV)

There was no reported progress for these projects, with the exception of Agenda Item 13.18.6. During the discussion, a few comments were made for some individual projects.

13.18.1. Nomenclature of sequence-controlled polymers (2019-041-3-400, Patrick Theato)

13.18.2. Graphical representation of polymer structures (2017-039-2-800, Patrick Theato and Andrey Yerin)

13.18.3. Nomenclature for polymeric carriers bearing chemical entities with specific activities and names (2014-034-2-400, Michel Vert)

ATH reported that this was work in progress and that a document had been circulated for comments.

13.18.4. Structure-based nomenclature for regular linear, star, comb and brush polymers with different types of constitutional repeating units (CRU) (2013-031-3-800 and 2019-036-1-800, Jiazhong Chen)

This project was apparently nearing completion, since GPM reported having seen a document in ICTNS.

13.18.5. Definitions and notations relating to stereochemical aspects in polymer science (2009-047-1-400, Karl-Heinz Hellwich and Graeme Moad)

13.18.6. Revision of IUPAC Recommendations on macromolecular nomenclature – Guide for authors of papers and reports in polymer science and technology (2008-020-1-400, Philip Hodge) (Web-based IUPAC recommendations on polymer nomenclature)

Philip Hodge had reported that a concise guide to polymer nomenclature for authors of papers and reports in polymer science and technology was now published: P. Hodge, K.-H. Hellwich, R. C. Hiorns, R. G. Jones, J. Kahovec, C. K. Luscombe, M. D. Purbrick and E. S. Wilks, *Pure and Applied Chemistry*, 2020, **92**,797-813. The project was thus completed.

13.19. Survey of definitions and use of common solid-state chemistry terminology (2015-053-1-200, Robin Macaluso)

RM reported that a Recommendation document had been submitted to Division II. ATH noted that when the document is accepted by the Task Group, it should first go to Division II and then come to Division VIII. RM remarked that completion will still take a few more months.

13.20. Nomenclature of homodetic cyclic peptides produced from ribosomal precursors (2015-003-2-300, Martin Reaney)

GPM observed that the project completion date was 2 December 2019.

13.21. IUPAC colour book data management (2013-052-1-024, Kinnan)
[Note: This project had been closed and repurposed as *Backup, maintenance and redevelopment of the IUPAC Gold Book website (2016-046-1-024, Stuart Chalk)*]

LRM provided some background to this project. It had become clear that a large-scale review of terms needed to be performed to catch up from 2006. An analysis had been started of what currently appeared in the Gold Book and what had been published in *PAC* in the intervening years. Terms were being seen from different eras where disparate previous IUPAC bodies might have been involved. A more recent example was provided by the 2019 Recommendation on 'Nomenclature for boranes and related species', a collaboration between Divisions II and VIII. There were several recommendations in this document that commented on existing terms that would need to be reviewed with reference to what currently is included in the Gold Book.

LRM explained that the current project had now been closed and repurposed as *'Development of an IUPAC recommended term management system for expansion of the coverage of the IUPAC Compendium on Chemical Terminology' (2019-032-1-024, Stuart Chalk)*. ATH informed the DC that Clare Tovee had been nominated to represent the Division on this project Task Group.

ES enquired how the terms in the Gold Book would be updated: the effort could not be restricted to a single project, since it was a continuing long-term task, which extended beyond the project lifetime. When a new definition was published, it should immediately and automatically be included into the Gold Book. LRM replied that the Task Group agreed with this sentiment. There was an ICTNS liaison to discuss this issue and there could also be an editorial team. TD noted that as the ICTNS representative from Division VIII, he had not heard about this. He also remarked that authors of Recommendations do not generally indicate which terms were new and needed to be added to the Gold Book and which were already existing terms. The terms were sometimes used without actually defining them. The end result could be catastrophic. LRM noted that CPCDS could give data on the use of terms, but it was the responsibility of the Divisions to provide the actual definitions.

13.22. Rules for Abbreviating Protecting Groups (2011-044-1-300, Margaret Brimble)

APR reported that this project already had an outcome, but there had been comments received from GPM and KHH which necessitated a revision. The work was ongoing.

14. Future projects/activities

14.1. New edition of Nomenclature of Inorganic Chemistry, the 'Red Book'

There had not been any real progress since the Paris meeting. However, the Action Item from the Paris minutes (Agenda Item 11.2 therein) required ATH to canvass interest for participating in a subcommittee, which might form a Task Group/Editorial Board. ATH reported that the following members had responded and indicated their interest in participating in this project: MAB, ECC, TD, RMH, ATH, RK, RSL, RM, ES, CAT, JV and AY.

The Action Item from the Paris minutes was reiterated:

> ***Action:** RSL will prepare a first draft of the contents and distribute it to the whole Division. A Task Group will be formed.

14.2. New edition of 'Principles of Chemical Nomenclature'

In the 2019 meeting in Paris GJL proposed that a new edition of 'Principles of Chemical Nomenclature' should be produced. After some further discussion, it was decided that this would be considered later, when plans for a new 'Red Book' had been clarified. Possibly a revision of 'Principles' should run concurrently with that of the 'Red Book'.

14.3. A common language in anti-doping – prohibited substances (2020-017-1, Vincenzo Abatte)

This project was distributed to Division VIII DC members for comments earlier in summer of 2020. After receiving comprehensive feedback, the proposal was sent back for revision.

> [*Secretary's note:* This project has since been accepted and commenced in November 2020 as *'A database of chemical structures and identifiers used in the control of WADA prohibited substances' (2020-017-2-700, Abatte)*.]

14.4. UVCB nomenclature for industrial chemicals and the impact of ECHA on nomenclature for the registration of substances that are intentionally produced as complex mixtures of chemicals.

MMR was active on this project, but as she was not able to attend the DC meeting, there were no new developments to report.

14.5. Proliferating IUPAC terminology to denote that names are (maybe) acceptable (recommended, retained, preferred, alternatively used, sometimes used, widely used, etc.) or not acceptable (not recommended, (strongly) discouraged, not included in these recommendations, deprecated, etc.) or to characterise them otherwise (common, traditional, trivial, etc.).

ATH reported that TD had sent a document highlighting the problem to ICTNS. The main response from ICTNS was that there existed a standard use of the terminology. There did not appear to be the basis to make a project. ECC noted that an important question was whether the rules originated from *PAC* or IUPAC. TD responded that there was no easy way to know.

14.6. Other projects.

ATH noted that after the publication of the inorganic and organic 'Brief Guides', TD had proposed last year to JCBN that writing a 'Brief Guide to Biochemical Nomenclature' would be useful. However, since then there had been no progress. It was pointed out that GPM's chapter in 'Principles of Chemical Nomenclature' could be a good starting point.

ATH reported that there were two new project proposals from the Polymer Division's SPT, one on 'Nomenclature and terminology for supramolecular science' and another on 'Use of internet for polymer terminology dissemination'. These will come to Division VIII for evaluation in due course.

15. Membership matters

15.1. Status of Division VIII Committee membership

See the current membership roster in Appendix 1.

15.2. Division VIII representatives in other IUPAC bodies: CCE (Committee on Chemical Education), PAC (Pure and Applied Chemistry) Board, ICTNS, COCI (Committee on Chemistry and Industry), JCBN, CPCDS

Division VIII has representation on (or members of the Division Committee are members of) the following bodies:

CCE:	RM is the representative from Division VIII
PAC Editorial Advisory Board:	ATH is the representative from Division VIII
ICTNS:	TD is the representative from Division VIII
COCI:	MMR is a Member of COCI and thus also the contact person for Division VIII
JCBN:	GPM is the Chair, APR and TD are Associate Members, and ATH is *ex officio* a Member.
CPCDS:	KTT and AY are members of the Subcommittee on Cheminformatics Data Standards (SCDS); Leah McEwen (LRM) is the Chair of this subcommittee.

15.3. Division VIII Advisory Subcommittee

There was general agreement that it would be useful to reactivate the Division VIII Advisory Committee. TD and ATH noted that it would be more convenient to review proposals and documents if the previously existing Discussion Board could be brought back to operation.

> ***Actions:** RSL will write to the Advisory Subcommittee members, send them the approved minutes of the 2019 Paris DC meeting, and enquire how they would like to serve (whether to receive draft documents, participate in developing nomenclature, or in some other way). If needed, the membership of this subcommittee would be updated and reviewed.

ATH will write to the Secretary-General to discuss possibilities for the reactivation of the Web Discussion Board or similar facility.

15.4. Division VIII Emeritus Fellows Program

ATH confirmed that there were three Emeritus Fellows in Division VIII and their names were listed in the Division membership roster (see Appendix 1). A more comprehensive account of their biographies will appear on the IUPAC Division webpage in due course.

15.4. Nominating Committee for 2021 elections

ATH indicated that there would be a need for two DC members and three individuals from outside IUPAC to form a Nominating Committee for the Titular Member elections to be held the following April/May. It was likely that MMR would chair this committee, whose task will be to sift through the nominations and select ten or more persons for the election to fill the seven available TM positions.

16. Publicity

16.1. Division VIII (and related) publications since the 2019 Division Committee meeting

The list of publications is presented in Appendix 8.

16.2. IUPAC website

The general opinion was that the IUPAC website had improved and seemed to work reasonably well.

16.3. IUPAC contact form/nomenclature consultancy/naming service

ATH thanked everybody (and in particular GPM, AY and TD) who had responded to chemical nomenclature queries that had been received via the contact information form on the IUPAC website. This was one of the services to the chemistry community that ensured that IUPAC continued to be held in high esteem.

16.4. IUPAC/IUBMB nomenclature website

GPM reported that the website (https://www.qmul.ac.uk/sbcs/iupac/) was heavily used, but that he no longer has access to usage information or statistics.

17. Any other business

TD reported that he had received a book on mathematical stereochemistry (S. Fujita, *Mathematical Stereochemistry*, de Gruyter, Oldenburg 2015, 437 pp.) and presented it to the DC for information.

18. Dates and venue for next meeting

The 51st IUPAC General Assembly (9 – 15 August 2021) and 48th World Chemistry Congress (13 – 20 August 2021) are scheduled to be held in Montreal, Canada, but the format of the meetings (virtual *vs.* face-to-face, or a combination thereof) will depend on the world-wide Covid-19 pandemic situation. It is possible that the Division VIII DC meeting will be on 12/13 August, with Task Group meetings preceding those dates.

19. Adjournment

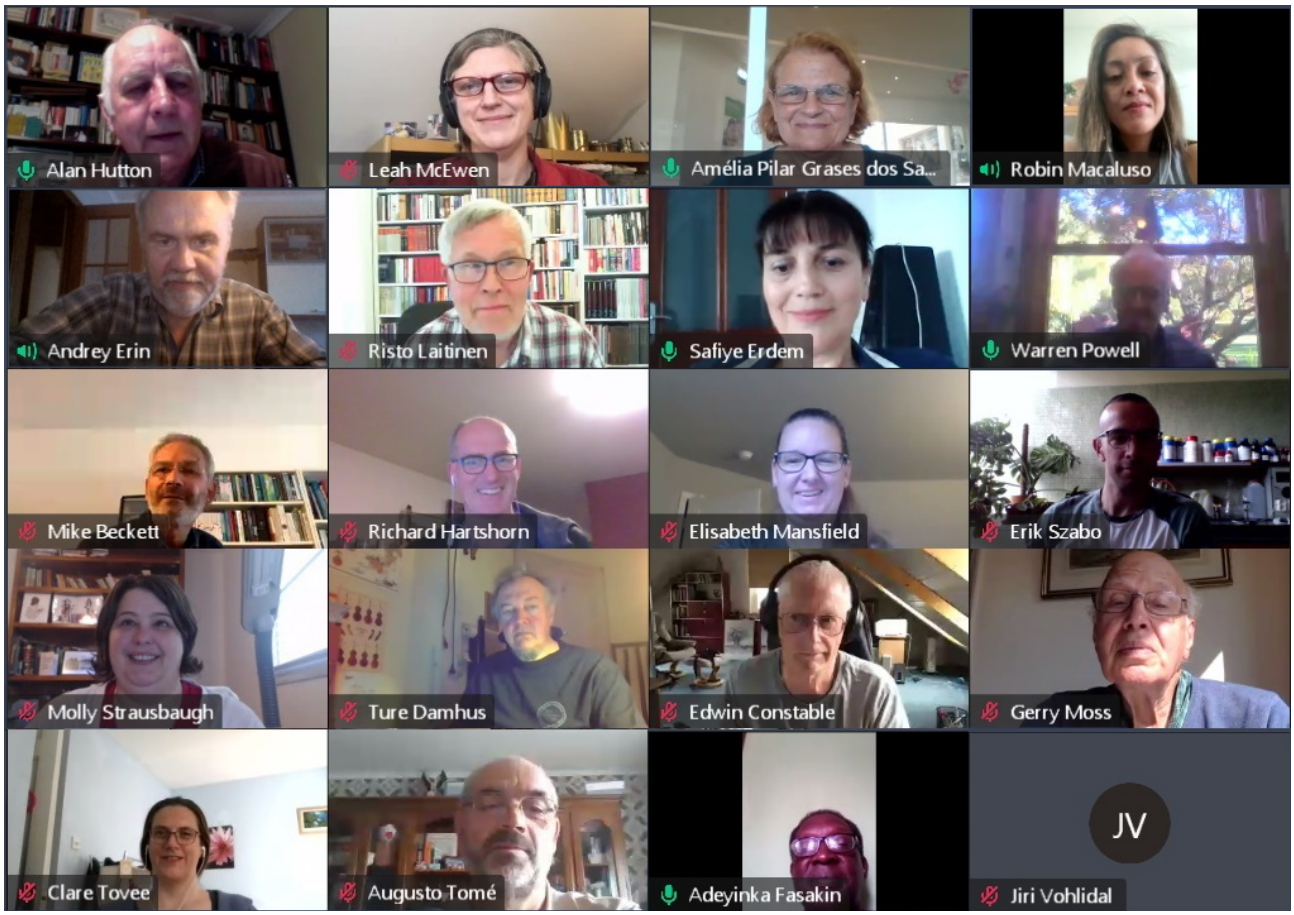
ATH thanked the participants for their attendance and contributions. He noted that the agenda had been satisfactorily worked through in 4½ hours and adjourned the meeting at 12:33 EST (09:33 PDT, 17:33 BST, 18:33 CEST, 19:33 EEST, 04:33 NZST).

Appendix 1

DIVISION VIII MEMBERSHIP 2020 – 2021

Name	Status	Term	NAO
Prof. Alan T. Hutton	President	2018-2021	South Africa
Dr. Michelle Rogers	Vice-President	2020-2021	USA
Prof. Risto S. Laitinen	Secretary	2020-2023	Finland
Prof. Michael A. Beckett	TM	2020-2021	United Kingdom
Prof. Edwin Constable	TM	2020-2021	Switzerland
Dr. Karl-Heinz Hellwich	TM	2020-2021	Germany
Dr. Elisabeth Mansfield	TM	2020-2021	USA
Prof. Ebbe Nordlander	TM	2020-2021	Sweden
Prof. Amélia Pilar Rauter	TM	2020-2021	Portugal
Prof. Jiří Vohlídal	TM	2020-2021	Czech Republic
Prof Neil Burford	AM	2020-2021	Canada
Dr. Thomas Engel	AM	2020-2021	Germany
Prof. Robin Macaluso	AM	2020-2021	USA
Dr. Erik Szabo	AM	2020-2021	Slovakia
Prof Augusto Tomé	AM	2020-2021	Portugal
Dr. Clare A. Tovee	AM, *CCDC rep.*	2020-2021	United Kingdom
Dr. Maria Atanassova Petrova	NR	2020-2021	Bulgaria
Dr. Ture Damhus	NR	2020-2021	Denmark
Prof. Safiye Erdem	NR	2020-2021	Turkey
Mr. Adeyinka Fasakin	NR	2020-2021	Nigeria
Prof. Rafał Kruszyński	NR	2020-2021	Poland
Dr. Ladda Meesuk	NR	2020-2021	Thailand
Prof. József Nagy	NR	2020-2021	Hungary
Prof. Dušan Sladić	NR	2020-2021	Serbia
Ms. Molly Strausbaugh	NR, *CAS rep.*	2020-2021	USA
Prof. Guoqiang Yang	NR	2020-2021	China
Dr. Andrey Yerin	Invited observer	2020-2021	Russia
Prof. Richard M. Hartshorn	*Ex officio (Sec Gen)*	2020-2021	New Zealand
Dr. Steve Heller	*Ex officio (InChI)*	2020-2021	USA
Leah R. McEwen	*Ex officio (CPCDS)*	2020-2021	USA
Dr. Gerard P. Moss	*Ex officio (JCBN)*	2020-2021	United Kingdom
Prof. G. Jeffery Leigh	*Emeritus Fellow*	2019-	United Kingdom
Dr. Alan McNaught	*Emeritus Fellow*	2019-	United Kingdom
Dr. Warren Powell	*Emeritus Fellow*	2019-	USA

Attendees of the Virtual Meeting of the Division VIII Committee
GoToMeeting platform on Saturday, September 19, 2020



Appendix 2

International Union of Pure and Applied Chemistry
Division VIII
Chemical Nomenclature and Structure Representation

Draft agenda for the Division Committee Meeting

Date: **Saturday, 19 September 2019**

Time: **08:00 EST (05:00 PDT, 13:00 BST, 14:00 CEST, 24:00 NZST)**

Venue: **Online (GoToMeeting platform)**
https://global.gotomeeting.com/join/979234541

Agenda

1. **Welcome, introductory remarks and housekeeping announcements**
2. **Attendance and apologies**
3. **Introduction of attendees**
4. **Approval of agenda**
5. **Approval of minutes of meeting in Paris, 6–7 July 2019**
6. **Matters arising**
7. **Interactions between Division VIII and other bodies in relation to documents and projects involving chemical nomenclature**
8. **InChI report (Subcommittee on the IUPAC International Chemical Identifier)**
9. **JCBN report (IUBMB–IUPAC Joint Commission on Biochemical Nomenclature)**
10. **CPCDS report (Committee on Publications and Cheminformatics Data Standards)**
11. **Division review of the outcomes of projects (i.e. Recommendations and Technical Reports prior to submission to ICTNS)**
12. **Division review of project proposals submitted to the IUPAC Secretariat**
13. **Reports on Division VIII projects**

13.1. Nomenclature and associated terminology for inorganic nanoscale particles (2019-016-3-800, Scott Brown, Edwin C. Constable)

13.2. Building Broader and Deeper Links Between OPCW and IUPAC (proposal 2018-022-2, Richard Hartshorn)

13.3. Alignment of principles for specifying ligands and substituent groups across various areas of nomenclature (2017-033-1-800, Ture Damhus)

13.4. Graphical representation standards for chemical reaction diagrams (2003-045-3-800/2012-033-1-800/2017-036-2-800, Keith T. Taylor)

13.5. IUPAC International Chemical Identifier (InChI) projects

- *13.5.1. Enhanced recognition and encoding of stereoconfiguration by InChI tools (2019-017-2-800, Andrey Yerin)*
- *13.5.2. InChI extension for mixture composition (2015-025-4-800, Leah McEwen)*
- *13.5.3. Identifying International Chemical Identifier (InChI) Enhancements – QR codes and Industry Application (2015-019-2-800, Richard Hartshorn)*
- *13.5.4. Implementation of InChI for chemically modified large biomolecules (2013-010-1-800, Keith Taylor)*
- *13.5.5. Handling of Inorganic Compounds for InChI V2 (2012-046-2-800, Richard Hartshorn and Hinnerk Rey)*
- *13.5.6. Redesign of Handling of Tautomerism for InChI V2 (2012-023-2-800, Marc Nicklaus)*
- *13.5.7. InChI requirements for Representation of Organometallic and Coordination Compound Structures (2009-040-2-800)*
- *13.5.8. InChI Open Education Resource (OER) (proposal 2018-012-2, Robert Belford)*

13.6. Corrections, Revisions and Extension for the Nomenclature of Organic Chemistry - IUPAC Recommendations and Preferred Names 2013 (the IUPAC Blue Book) (2015-052-1-800, Karl-Heinz Hellwich)

13.7. Nomenclature of carbon nanotubes and related substances (2013-056-1-800, Elisabeth Mansfield)

13.8. End-of-line hyphenation of systematic chemical names (2014-003-2-800, Albert Dijkstra)

13.9. Nomenclature for metallacycles containing transition metals (2013-030-1-800, Alan Hutton)

13.10. Nomenclature for polyhedral boranes and related compounds (2012-045-1-800, Michael Beckett)

13.11. Revision and extension of IUPAC recommendations on carbohydrate nomenclature (2012-039-2-800, 2015-035-2-800, 2017-026-1-800 Johannes Vliegenhart)

13.12. A comparison of assignment of hydro prefixes, added and indicated hydrogens in IUPAC, CAS and Beilstein nomenclature systems (2012-037-1-800, Andrey Yerin)

13.13. Terminology and nomenclature of inorganic and coordination polymers (2011-035-1-800, Richard Jones); for short TINCOPS

13.14. Brief guides to the nomenclature of organic and inorganic chemistry ('Essentials' of organic and inorganic nomenclature) (2010-055-1-800, Richard Hartshorn and Karl-Heinz Hellwich)

13.15. Glossary of small molecules of biological interest (2009-022-2-800, Marcus Ennis)

13.16. Preferred names for inorganic compounds (2006-038-1-800, Ture Damhus)

13.17. Nomenclature of phosphorus-containing compounds of biochemical importance (2006-019-1-800, Gerard Moss)

13.18. Polymer projects (with Division IV)
- *13.18.1. Nomenclature of sequence-controlled polymers (2019-041-3-400, Patrick Theato)*
- *13.18.2. Graphical representation of polymer structures (2017-039-2-800, Patrick Theato and Andrey Yerin)*
- *13.18.3. Nomenclature for polymeric carriers bearing chemical entities with specific activities and names (2014-034-2-400, Michel Vert)*
- *13.18.4. Structure-based nomenclature for regular linear, star, comb and brush polymers with different types of constitutional repeating units (CRU) (2013-031-3-800 and 2019-036-1-800, Jiazhong Chen)*
- *13.18.5. Definitions and notations relating to stereochemical aspects in polymer science (2009-047-1-400, Karl-Heinz Hellwich and Graeme Moad)*
- *13.18.6. Revision of IUPAC Recommendations on macromolecular nomenclature – Guide for authors of papers and reports in polymer science and technology (2008-020-1-400, Philip Hodge) (Web-based IUPAC recommendations on polymer nomenclature)*

13.19. Survey of Definitions and Use of Common Solid-State Chemistry terminology (2015-053-1-200, Robin Macaluso)

13.20. Nomenclature of Homodetic Cyclic Peptides Produced from Ribosomal Precursors (2015-003-2-300, Martin Reaney)

13.21. IUPAC color book data management (2013-052-1-024, Kinnan)
[Note: This project has been closed and repurposed as *Backup, maintenance and redevelopment of the IUPAC Gold Book website (2016-046-1-024, Stuart Chalk)*]

13.22. Rules for Abbreviating Protecting Group (2011-044-1-300, Margaret Brimble)

14 Future projects/activities

14.1. New edition of Nomenclature of Inorganic Chemistry, the 'Red Book'

14.2. New edition of Principles of Chemical Nomenclature

14.3. A common language in anti-doping – prohibited substances (2020-017-1, Vincenzo Abatte)

14.4. UVCB nomenclature for industrial chemicals and the impact of ECHA on nomenclature for the registration of substances that are intentionally produced as complex mixtures of chemicals.

14.5. Proliferating IUPAC terminology to denote that names are (maybe) acceptable (recommended, retained, preferred, alternatively used, sometimes used, widely used, etc.) or not acceptable (not recommended, (strongly) discouraged, not included in these recommendations, deprecated, etc.) or to characterise them otherwise (common, traditional, trivial, etc.).

14.6. Other projects.

15. Membership matters

15.1. Status of Division VIII Committee membership (see current membership roster in Appendix)

15.2. Division VIII representatives in other IUPAC bodies CCE, PAC Board, ICTNS, COCI, JCBN, CPCDS

15.3. Division VIII Advisory Subcommittee

15.4. Division VIII Emeritus Fellows Program

15.4. Nominating Committee for 2021 elections

16. Publicity

16.1. Division VIII (and related) publications since the 2019 Division Committee meeting

16.2. IUPAC website

16.3. IUPAC contact form/nomenclature consultancy/naming service

16.4. IUPAC/IUBMB nomenclature website

17. Any other business

18. Dates and venue for next meeting

19. Adjournment

Appendix 3

Highlights of the activities in Division VII

The three subcommittees (SC) of Division VII:

- Drug Discovery and Development (**DDD)**, Chair Gerd Schnorrenberg
- Toxicology and Risk Assessment (**TRA)** (Chair Vincenzo Abbate, succeeding John Duffus)
- Nomenclature for Properties and Units in Clinical Chemistry (**NPU**) Chair Ulla Magdal Petersen, succeeding Helle Johannessen

The three subcommittees works on projects and submit new ones in line with their specific objectives.
I present some highlights, albeit that this is by far not exhaustive:
DDD continues with their book series 'Successful Drug Discovery' (Editors : Janos Fischer, Wayne E. Childers and Christian Klein). Each issue has 15 chapters. Its structure follows that of the first volume consisting of three parts : I. General Aspects, II. Drug Class Studies and III. Case Studies. A new Project *Analysis of Phase III Failures chaired by* Michael Liebmann, received appoval after official IUPAC review. Currently the project has been started, the option to make it a joint project with the Subcommittee Toxicology and Risk Assessment and the Subcommittee NPU is ongoing.

TRA SubCommittee is particularly interested in working on nomenclature for psychoactive drugs of abuse. As I recall, there is a collaboration with one or more representatives from DIV.VIII in one of the projects led by Vincenzo Abbate.
Project 2014-019-1700: The emerging problem of Novel Psychoactive Substances It consists of two parts: i) synthetic cannabinoids – A first technical report has been published: The ongoing challenge of novel psychoactive drugs of abuse. Part I. Synthetic cannabinoids (IUPAC Technical Report), Pure and Applied Chemistry, 90(8), pp. 1255- 1282, 2018, https://doi.org/10.1515/pac-2017-0605. ii) Synthetic cathenones – in preparation. V. Abbate reported that the team has two new members: Claude Guillou and Azell Carter
- Project 2016-045-2-700: Safety of Engineered Nanomaterials and Project 2017-035-2- 600: Consideration of nano-enabled pesticides for industry and regulators. The article "Key challenges for evaluation of the safety of engineered nanomaterials" published by L.J. Johnston, N. Gonzalez-Rojano, K.J.Wilkinson and Baoshan Xing,
A CI article on WADA & anti-doping has been submitted by Vincenzo Abbate and David Cowan for publication in the Jan - March 2020 issue.

Project 2019-029-1-600 - PER AND POLYFLUROALKYL SUBSTANCES (PFASS) IN THE ENVIRONMENT: INFORMATION FOR EMERGING ECONOMIES ON PFASS ANALYSES IN ENVIRONMENTAL MEDIA AND THEIR IMPACTS ON HUMAN HEALTH. This is a newly approved interdivisional project where at least two participants of DivVII/TRA (Abbate and Hogstrand) are involved. Just started, first meeting in February 2020.

Nomenclature, Properties and Units (**NPU**) is a joint Committee in collaboration with International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), and provides a terminology (called NPU Terminology) for Properties and Units in the Clinical Laboratory Sciences.

The NPU terminology has evolved from 1966 as projects by joint Committee-SubCommittee on NPU under the aegis of IFCC and IUPAC. All the IFCC-IUPAC articles are to be find at www.npu-terminoly.org.
The NPU terminology is a coding system and terminology for identification and communication of examination results from clinical laboratories in the health area. It identifies types of result values, for use in reporting laboratory results. The definitions have a uniform structure and use a referenced vocabulary. Some examples:

NPU24866 Urine—Morphine; mass concentration = ? microgram per litre

NPU02187 Blood—Glucose; substance concentration = ? millimole per litre

The variety of examinations from clinical laboratories has increased over the last 50 years from a mere few hundred types to well over 30 000 NPU definitions and it seems that it will never end.
The NPU Terminology has to keep up with the development within the laboratory area, and develop new NPU definitions for new examinations.
The importance of stereochemistry in drugs is gaining greater attention in medical practice, and there is a demand for new NPU definitions in this area. Another issue is
The Committee on NPU need a basic knowledge of the subject, and hopefully a reliable source for identification and naming these substances.
Will there be an interest in DIV.VIII to enter into a collaboration with representatives from C-NPU to help on this matter?
Best regards, Helle

Helle Møller Johannessen
Health Informatics Specialist, MI
NPU Office
Department of Data Quality and Registries
The Danish Health Data Authority
Ørestads Boulevard 5, DK-2300 Copenhagen S

DANISH HEALTH DATA AUTHORITY

Appendix 4(a)

InChI Trust Project Director's Report
January 2020 (1/27/20)

Summary:

The project to develop the InChI algorithm for all defined chemical structures continues to move forward with two major working groups starting implementation of their standard: Organometallics/Coordination structures and extended tautomers. As mentioned in the last report, a contract was awarded to Alex Clark to start work on extending the algorithm to handle coordination compounds/organometallics. The plans for extending InChI to handle more tautomers is ready for programming and testing.

In June 2019 I was notified that the NIST lab that initially developed and programmed the InChI algorithm was no longer interested in the InChI standard and my 22 years as a Guest Researcher was terminated. In July 2019 I became a Guest Researcher at NIH/NLM/NCBI in the PubChem Project, where I continue my InChI activities and work.

Items covered in this report:

Membership/Support
InChI RFP/Contracts
InChI development and maintenance work
IUPAC InChI subcommittee and working groups
August 2019 – December 2019 Activities:
  Meetings attended & Talks/ Posters given
  Manuscripts
  InChI Trust Web Site
  PIDapalooza
  InChI Usage
  Technical Issues
  GitHub
Some History, Organizational Planning,
  and Project Sustainability
Plans for 2020
2020/2021 possible workshops/symposia

Membership/Support:

Memberships for the Chinese Chemical Society (CCS) and Google are in process.
As mentioned, numerous times in the past in most organizations, since InChI works and it is not high on their immediate priority lists, actual real progress is slow without a dedicated champion within an organization.
Mcule has joined the Trust as a supporter.
As of January 1, 2020:

Members (9):

IUPAC
ACS/Chemical Abstracts Service (CAS)

Chinese Chemical Society (CCS) (pending agreement)
Elsevier/Relx Group
Royal Society of Chemistry (RSC)
Springer Nature
John Wiley & Sons
Informa/Taylor & Francis
US National Institutes of Health (NIH)

Associates (12):

ACD Labs
Bio-Rad
CCDC
ChemAxon
Google (pending agreement)
Mcule
OntoChem
OpenEye
Sigma Millipore
University of California
US Food and Drug Administration
US National Institute of Standards and Technology

Certification Suite:

US Environmental Protection Agency

Active Supporters (41):

AKos Consulting and Solutions
American Chemical Society Division of Chemical Information (CINF)
Biochemfusion ApS
Caltech Library Services, Pasadena, CA, USA
Centre for Molecular Informatics, Cambridge University, UK
Chemistry Department, Clemson University, SC, USA
Chemistry Department, University of Arkansas at Little Rock
Chemistry Department, University of California, Riverside, CA, USA
Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, NC, USA Tropsha
Faculty of Science, University of Paderborn, Germany
Gesellschaft Deutscher Chemiker e.V. (GDCh), Germany
Guide to Pharmacology, UK
Imperial College London, UK
Institute for Chemoinformatics and Bioinformatics, University of Applied Sciences Gelsenkirchen, Recklinghausen Section, Germany
Institute of Chemical Technology, Prague, Czech Republic
Institute of Organic Chemistry, KIT Karlsruhe
International Union of Crystallography
Leadscope, Columbus, OH, USA
Leibniz-Institut für Analytische Wissenschaften – ISAS, Dortmund, Germany
Ludwig-Maximilians-Universität München, Munich, Germany
Molecular Materials Informatics, Inc
National Center for Biomedical Ontology, Stanford University, CA, USA

National Chemical Laboratory, Pune, India
NextMove Software, Cambridge, UK
Open Babel
RJB Computational Modeling
Royal Netherlands Chemical Society
School of Chemistry, University of Leeds, UK
SciencePoint, Redmond, WA, USA
StructurePendium Technologies GmbH
Technical University of Vienna, Austria
The Chemistry Development Kit, Eindhoven, The Netherlands
TW2Informatics Limited
University of California, Davis, Genome Center, CA, USA
University of Indiana, Bloomington, IN, USA
University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia
University of Southampton (Chemistry), UK
University of the West Indies, Mona Campus, Jamaica
Xemistry GmbH, Königstein, Germany
ZINC

InChI RFP/Contracts

After a long drought in initiating the programming of the standards that working groups have developed and greed upon there are two major projects underway - Organometallics and Extended tautomers. A second contract for work on programming InChI for coordination compounds, organometallics, and some inorganics is expected to be awarded to Alex Clark once further details and outcomes are clarified. The initial results of the work performed are given below in the section on Organometallics. An initial contract to examine the issues associate with extending the InChI capabilities in the area of tautomers was approved and work has been started by Igor Filippov.

InChI Development & Maintenance Work

Igor Pletnev continues to do a superb and a very responsive job as the InChI programmer. The release of version 1.06 is expected shortly. Version 1.06 with include the "any atom" feature originally suggested to Igor by the FDA/Yulia Borodina. It has been delayed due to many minor bugs that were fixed. There continues, as expected, to be useful feedback on minor issues and bugs as noted below. Gerd Blanke continues to do excellent work on the RInChI algorithm.

Igor has reported that many people have been in contact with him regarding bugs, errors, and issues with the algorithm. Without all this external help, primarily in the past from the InChI SourceForge list (established by Alan McNaught in May 2005), the algorithm would not be as good as it is. A copy of Igor's latest report is attached.

Besides the SourceForge inchi-discuss list, Igor had many valuable comments/issues, and advice reported in private correspondence by several experienced chemoinformaticians/developers which routinely use InChI.
There are a significant number of issues (and related fixes) Igor found in his own testing processes.
Some minor questions/reports also came from occasional InChI users, typically via Richard Kidd, and from other sources (like Google auto-fuzz).
In summary, we have many users and many people from all different areas of chemistry using and working to help improve the InChI algorithm.
SourceForge InChI url: https://sourceforge.net/projects/inchi/

IUPAC InChI Subcommittee & Working/task groups

IUPAC InChI Subcommittee members

The following are the current members of the IUPAC InChI subcommittee. With Leah McEwen becoming the chairperson of CPCDS in January 2020, she will rotate off the subcommittee and Jonathan Goodman will become the subcommittee secretary.

- Steve Heller, Chair
- Jonathan Goodman, Secretary

- Members:
- Steve Bachrach
- Bob Belford
- Gerd Blanke
- Evan Bolton
- Marc Nicklaus
- Carmen Nitsche
- Hinnerk Rey
- Wendy Warr
- Tony Williams
- Andrey Yerin
- Shuli You

- Igor Pletnev – Technical Advisor
- Dmitrii Tchekhovskoi – Technical Advisor

Working Groups

1. **Mixtures**
2. **Monomer Atoms**
3. **Variability**
4. **Isotopologues**
5. **Positional Isomers**
6. **Resolver**
7. **Polymers**
8. **Reactions**
9. **Organometallics**
10. **Inorganics**
11. **Large Molecules**
12. **Extended Tautomers**
13. **QR Codes**
14. **Education/Academic/Training**
15. **Extended Stereochemistry**
16. **GitHub**

Chemical mixture composition (MInChI)

MInChI: a chemical notation for mixtures
Update on IUPAC Project: 2015-025-4-800
Submitted January 2020, by *L. McEwen (Working group Chair)*

The goal of this project has been to articulate what can be said, definitively and in an actionable way, what is known about the chemical composition of a given mixed substance. Applying InChI notation in this context enables the development of an unambiguous machine-readable linear notation for mixed substances of uniform properties that can resolve to unique components, supporting the practical need to connect data and information on mixtures and individual components and enabling further downstream computation and analysis on properties, composition, etc. The initial phase has focused on components with molecular structures that produce a well-defined standard InChI identifier. Separate, discrete components allow for easy association of arbitrary concentrations. Other types of components may be considered for future phases of the project.

Mixture compositions are commonly described in terms of the source components, what is initially measured into the mix. The MInChI notation is tailored for this primary use case, to describe what is definitively reported at the time of mixing. Describing how the components interact once they are mixed is beyond the scope of MInChI (although this could suggest a potential application for RInChI in conjunction with MInChI if deemed of interest to explore in the community). While MInChI includes information on intended composition of mixed substances, it is not intended to function as a canonical identifier of mixtures due to process-related dependencies for combining substances. The component structure layer of the notation does provide a high level of consistency between descriptions of similar mixtures and comparisons of these ensembles can be done through trivial string manipulation, although literal comparison cannot be assumed.

The MInChI specification drafted by the project has been implemented as a proof of concept by Collaborative Drug Discovery (CDD) as described by Clark et al.[1] Several follow up presentations at InChI workshops and ACS meetings have highlighted the outcomes of this work, including a recent webinar. CDD has recently received a grant to continue development of this application of MInChI for drug discovery, including potential incorporation of outputs from other InChI working groups exploring organometallics, polymers and other types of molecules.

The trajectory for MInChI on a technical level will look to incorporate the specification into the RInChI codebase for a combined executable. Similar approaches are used for developing the layers and include many shared or similar points of information. Planning is underway to fold this work into the next RInChI code release, scheduled to start development mid-year 2020. The CDD prototype code-base for MInChI is open source and can inform implementation of property information such as concentration.

While the technical implementations of RInChI and MInChI share many commonalities, the use cases for these notations likely span divergent chemistries and communities. The MInChI project will continue to explore approaches to notating more complex or specialized forms, such as formulations, buffers and hydrates. Establishing and expanding the user group for MInChI across sectors and into areas such as materials, agriculture, consumables and others will also be a high priority in the coming year.

[1] Clark, A. M.; McEwen, L.R.; Gedeck, P.; Bunin, B. A. (2019), Capturing mixture composition: an open machine-readable format for representing mixed substances. J. Cheminform. 11, 33, DOI: https://doi.org/10.1186/s13321-019-0357-4

Pseudo-atom biopolymer monomer representation atoms for InChI

Evan Bolton has prepared a draft proposal (below) for adding 'monomer' atom support to InChI (with some simple examples). Evan has commented that it seems that the 'monomer' atom would be both necessary and highly desirable to simplify the InChI and speed up its computation. By itself, 'monomer' atom support is very (tremendously?) useful to add. It is envisioned that this would only be part of the non-standard InChI.

Purpose:
Enable chemically modified biopolymers identity to be rapidly accessed (by comparing InChI/Key strings). Add formal support to IUPAC InChI for dealing with biologics using a monomer-based, pseudo-atom approach for non-standard InChI. Simplify the biopolymer InChI by reducing its complexity and length. Make possible for InChI to handle even larger biological molecules. Dramatically increase the speed to compute an InChI for large molecule biologics.

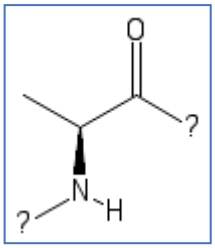
Comment:
For the purposes of this document, a 'defined atom' is any known element in the periodic table. A 'monomer atom' is a well-characterized pseudo-atom biopolymer monomer.
A 'monomer' atom represents all atoms and bonds within the biopolymer monomer it represents. Each 'monomer' atom is to be predefined in the InChI software code. Use of 'monomer' atom approach can dramatically sped up as tautomer and stereo processing in the InChI algorithm as the portion of the molecule covered by 'monomer' atoms can be ignored. [For example, simply covering the 20 natural amino acids would dramatically reduce the size and complexity of most biologics that primarily contain natural amino acids.]
Well-defined semantics for 'monomer' atom representation approaches already exist in many file formats, including chemical formats such as CTAB/MOL/SDF, SMILES, and HELM, providing a direct source of input structures for InChI processing.
Any part of a molecule not covered by 'monomer' atoms would be subject to usual InChI processing.
It is imagined that support for 'monomer' atom would only be accessible using non-standard InChI using command line switches. [If added to the standard InChI, one would need to determine what to do about existing InChI strings (and InChIKeys?) containing supported 'monomer' atom residues as the 'monomer' atom compacted InChI would not be identical (although the InChIKeys could be made to be identical?).]
The InChI software could provide a conversion facility to enable facile conversion to/from 'monomer' atom compacted InChI strings.

Limitations:
1. ***The use of three letters codes consisting of 26 possibilities each provides a maximum of 17,576 'monomer' atoms***, with the first being capitalized 'A-Z' and the next two being lower case 'a-z'. [If more possibilities are needed, this could be made more flexible. For example, the second and third characters could be allowed to be 'a-zA-Z', increasing the possible count of 'monomer' atoms to 70,304. With this said, it is imagined that only a small subset of the most frequently occurring monomers would be supported, a count on the order of 100 to 1000 'monomer' atoms with an aim to reduce the complexity of a biopolymer InChI.]
2. ***'Monomer' atoms are rigidly defined and cannot be chemically modified or substituted in any way***.
3. ***Bonding with a 'monomer' atom must follow special semantics to ensure correct connectivity***. For example, most proteins consist of a linear chain of connected amino acid monomers involving the bonding of an amine to a carboxylic acid to form an amide, making it important to understand the bonding direction such that there is an amine at one end of the chain and a carboxylic acid at the other end of the chain.
4. ***If a fixed set of 'monomer' atoms is used, adding additional ones in the future means it is possible to have to InChI that is different yet mean the same thing***. However, this should not be an issue

considering this is for non-standard purposes and one can decode previously created InChI and recode into the current InChI representation.

Approach:

1. **Representation:** A 'monomer' atom is to be represented using a three-letter code, where the first is upper case and the next two are lower case. [E.g., L-Alanine might be defined using the pseudo-atom biopolymer 'Ala'.]
2. **Treatment:** Each 'monomer' atom is to be treated like a defined atom.
    a. A single bond is always used to connect to a 'monomer' atom.
    b. Each 'monomer' atom would represent a specific set of atoms/bonds as well as any hydrogens, stereochemistry, and isotopic substitution within a given monomer. [E.g., for L-Alanine the internal representation might be: '[*]C(=O)[C@H](C)N[*]', see also the image to the right. *Please note that the 'OH' of the carboxylic acid is missing.*]



    c. Each 'monomer' atom can have two potential attachment points. Well-defined semantics for connecting other atoms and to a 'monomer' atom will require a bonding order convention to be used to indicate which terminus of a 'monomer' atom is being connected in a given bond. For example:
        i. Amino acid biopolymers have the N-terminus, where an atom may be attached or, if none specified, a hydrogen atom will be assumed, and the C-terminus, where an atom may be attached, if none specified an OH will be assumed. The amino acid bonding chain is assumed to always indicate the N-terminus first within the InChI. Specifying a hydrogen being attached to the 'monomer' atom will indicate that the N-terminus is not bonded, and the bond is to the C-terminus.
        ii. Nucleic acid biopolymers have the 5' end of deoxy-ribonucleotide, where an atom may be attached, if none specified, an OH will be assumed, and the 3' end of deoxy-ribonucleotide, if none specified a hydrogen atom will be assumed. Similar to amino acids, the nucleic acid bonding chain within the InChI is assumed to bond to the 3' end first, unless a hydrogen is specified to be on the 'monomer' atom and the bond it to be placed to the 5' end.
        iii. Glycan biopolymers have multiple potential bonding patterns. Multiple variants of 'monomer' atoms can be created in such cases. For example, the 1- and 6-positions, where an atom may be attached. Different 'monomer' atoms may then be defined for different glycan bonding patterns, such as [1-, 4-] or [4-, 6-].
3. **Numbering:** A 'monomer' atom should be otherwise treated in a similar fashion to defined atoms for indexing purposes. However, special care to bonding conventions (see **Treatment**) for a given monomer will need to be taken into account.
4. **Charges:** A 'monomer' atom cannot be directly involved (i.e., atom index) in the charge layers ('/q' or '/p').
5. **Stereochemistry:** A 'monomer' atom cannot be directly involved (i.e., atom index) in the stereo layers ('/b', '/t', '/m', '/s').
6. **Isotopes:** A 'monomer' atom cannot be directly involved (i.e., atom index) in the isotopic layers ('/i', '/h', '/b', '/t', '/m', '/s').
7. **Hydrogens:** A 'monomer' atom cannot be directly involved (i.e., atom index) in the hydrogen layers ('/f', '/h', '/q', '/b', '/t', '/m', '/s', '/o', '/i') except when required for bonding conventions (see **Treatment**).
8. **Normalization:** Any input structure is to be subjected to a 'monomer' atom re-perception, in that any specified 'monomer' atom will be translated to an all atom/bond description and then any 'monomer' atoms will be (re)perceived providing a single, unified input approach for structures with or without 'monomer' atoms. Very limited attempts (if any) will be made to perform any normalization of 'monomer' atoms within a structure. The input to InChI involving a 'monomer' atom connection to a defined atom is considered rigid (i.e., unchanging) due to the circumstances

involving the rigid aspects of the 'monomer' atom. In many ways, a 'monomer' atom can be ignored with any bond(s) connected from a defined atom to a 'monomer' atom excluded from normalization processing (where charges, protons, bond orders, radical cancellation, salt disconnection, tautomerism, and other modifying aspects occur that alter the structure drawing). [Please note that the defined atom attached to a 'monomer' atom can be involved in normalization processing provided it does not modify the 'monomer' atom or the bond that connects the 'monomer' atom to the defined atom.]

9. **Input:** A 'monomer' atom can be provided via SDF/MOL file input using an appropriate SGroup definition, the predefined three letter code for the given monomer, or other supported semantic. In terms of API, a 'monomer' atom can be defined in a similar fashion to SDF/MOL file input, with: the 'IXA_MOL_SetAtomElement' function for the element symbol; or the 'IXA_MOL_SetAtomAtomicNumber' function specifying the predefined element integer.

Examples:
L-Alanine 'InChI=1S/C3H7NO2/c1-2(4)3(5)6/h2H,4H2,1H3,(H,5,6)/t2-/m0/s1' would become 'InChI=1/Ala'
L-Alanyl-L-Alanine 'InChI=1S/C6H12N2O3/c1-3(7)5(9)8-4(2)6(10)11/h3-4H,7H2,1-2H3,(H,8,9)(H,10,11)/t3-,4-/m0/s1' would become 'InChI=1/Ala2/c1-2'

Variability

After a good number of years of discussions on how to proceed with handling variability of structures to produce useful InChI's initial work to test ideas for this has been started at Cambridge University. Anthony Baston, a master's student under Johnathan Goodman, is making good progress on variable InChIs in the proof-of-concept case of alkanes. Anthony has a program which, subject to further testing, can take a list of InChIs and generate a canonical variable InChI, and also decode a variable InChI to a list of InChIs. For the final three months of the project, we will continue to test this and to investigate extensions to more complex substrates.

Isotopologues

Chairperson: Hunter Moseley

In the past 6 months there has been work on the development of a draft SD file representation of isotopologues to facilitate generation of isotopologue representations in InChI.
There are two major things for the working group to do:
1) Write up a manuscript describing the accepted InChI isotopologue extensions.
2) Find an example of the fixed hydrogen layer issue in InChI.

Positional Isomers

Chairperson: Jonathan Goodman

While considerable technical interest in positional isomers has developed in the past, no one is willing to take the lead for this area. A number of people have offered to chair the group but then withdrew. Besides not having anyone to lead this effort at this point in time it would seem best to merge all these variable structures (Markush, positional isomers, and so on) into one working group.

Resolver

Chairperson: Markus Sitzmann

There was an InChI Resolver Meeting at GCC2019, on 11-4-2019.
Attendees: Steve Heller (NIST), Evan Bolton (NIH/NCBI), Marc Nicklaus (NIH/NCI), Gerd Blanke, Markus Sitzmann (FIZ Karlsruhe)
An InChI resolver describes a list of web resources (paid, free) an organization provides which can be accessed by InChI/InChIKey/RInChI. An InChI resolver has to adhere to an agreed protocol which allows for the creation of federated resolver systems or federated searches on the web, respectively.
Roadmap:

- Gerd and Markus put together a first version for a final draft of the InChI Resolver protocol (paper document), the group agrees for a final document during Q1/2020
- For demonstration purpose, a openly available reference implementation of the InChI Resolver protocol is made available online which links (prototype) InChI resolver instance at NCI/CADD (CIR), PubChem and e.g. ChEMBL to a federated resolver system.
- During Spring/Fall ACS Meeting we open up to more interested parties (Evan?)

The current state of the project is available at https://inchi-resolver.org, the source code of the reference implementation is available at https://github.com/inchiresolver/inchiresolver.

Polymers

Chairperson: Andrey Yerin

As mentioned previously, with the release of version 1.05 a limited area of polymer chemistry can now be handled by the InChI algorithm. A number of issues were found after release 1.05 and Igor continues to work on these matters.

As a result of feedback from the community Igor has added some extensions and has done a redesign. More regarding this can be found in Igor's 27 page report which was submitted to the Trust for the February 2019 Board meeting.

Reactions

Chairperson : Gerd Blanke

Participants:
David Nicolaides (Biovia, Cambridge) Gerd Blanke (StructurePendium), Günter Grethe, Hans Kraut (InfoChem), István Öri (ChemAxon), Jan Holst Jensen (Biochemfusion), Jonathan Goodman (University of Cambridge)

Status of RInChI:
The group meets biweekly per Skype conference to further develop RInChI and to discuss actual issues like the participation at conferences and user group meetings. Because the group is really keen to debate technical details, we have a lack of personal meetings and interactions that would fasten the discussion processes we have. With that the development times take longer than anticipated.
After Biovia and ChemAxon have implemented RInChI into their standard cheminformatics software packages, Knime nodes for RInChI have been developed by Lhasa Ltd. This is the first known

implementation by a third party that has not been invoked by members of the RInChI group and shows that RInChI makes its ways into the cheminformatic community.
Out of personal discussions during conferences it has become clear that there seem to be multiple groups looking into RInChI to represent chemical reactions (e.g. ChEBI). During a personal meeting in Hinxton I got a quite positive feedback from the ChEBI developer; the reported issues are known general limitations of the InChI as such and are discussed within the InChI community anyway. Unfortunately, we have not received any feedback from other groups.

Working towards version 2.0 of RInChI:

- Finalized issues
  - Atom mapping for reactions
  - Stereochemistry representation
- Currently under discussion
  - Representation of failing reactions
- Remaining issues to be discussed and prepared in Q1 2020
  - Handling of reaction conditions (ProcAuxInfo)
  - Transfer to Open Source development
    - Build the necessary development and test environment on GitHub
    - May become blue print for other technical developments based on InChI

RInChI Presentations and Publications

- Participation at the IUPAC committee meeting in Paris, representation of the RInChI group at the IUPAC generally assembly in Paris (July 2019)
- Participations at the San Diego InChI meeting, RInChI Break out session (August 2019)
- Talk in the CINF session "InChIing forward" at the ACS Fall meeting (August 2019)
- Poster at the GCC conference in Mainz, Germany (November 2019)

Upcoming RInChI events
The upcoming events are mainly focused to discuss and establish the "ProcAuxInfo"

- AI for Reaction Outcome and Synthetic Route Bristol, March 2020 (Poster requested)
- Under discussion: Support for Günter Grethe to organize a CINF symposium about prediction and optimization methods for reactions at the ACS fall meeting 2020.

Organometallics

Chairperson: Colin Batchelor

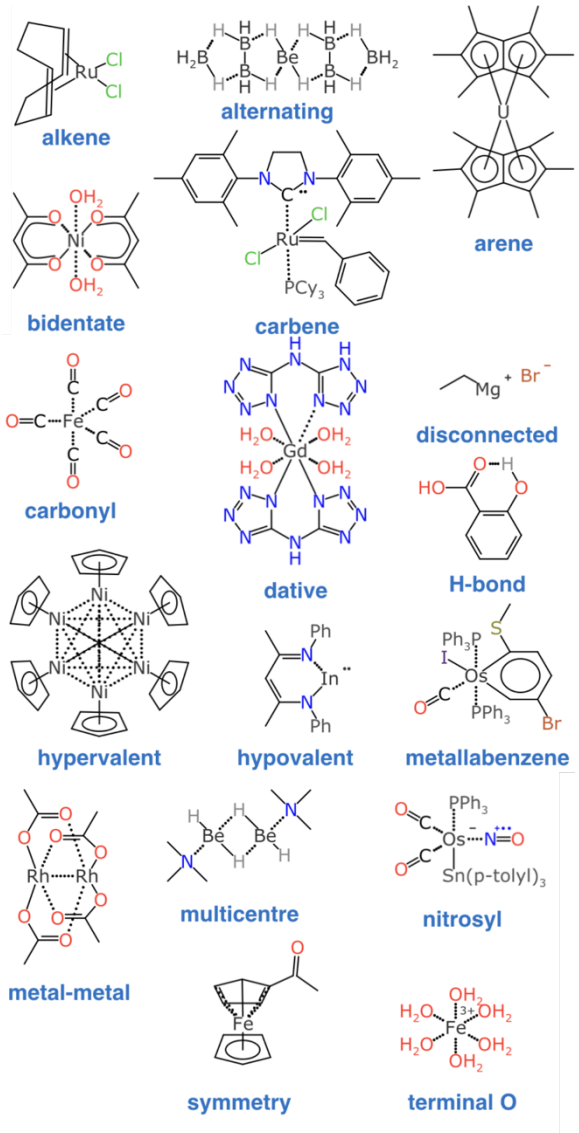
Extending the InChI identifier to include inorganic & organometallic compounds
Dr. Alex M. Clark, December 2019

Readers will be familiar with the InChI identifier, which has come to be thought of as the definitive standard way of turning a molecular representation into a canonical string. The general idea is that whenever two drawings of a compound represent the same thing, they produce the same identifier, even if they were drawn differently or as different tautomers. The value to information systems is immense: these identifiers can be used as an essentially instant way to disambiguate chemicals, or to index them in databases that are themselves not in possession of any chemical awareness.

One of the weaknesses of the InChI identifier is shared by most of the cheminformatics industry: it was designed and validated primarily for drug-like organic compounds that have bonding arrangements that fit the Lewis octet rule. Outside of this domain lies the rest of the periodic table with its vast diversity of exotic bond types, many of which seem to exist solely for the purpose of defying classification. The standard InChI generation algorithm partially solves this problem by simply disconnecting any bonds that involve metals.

While this entails a significant amount of destruction of information (i.e. loss of reversibility), in principle it could have only a small impact on the primary use case for InChI (i.e. indexing of chemicals). In practice, though, bond deletion is not compatible with multiple equally valid valence models, and interferes with downstream normalisation algorithms. The result is that generating standard InChI identifiers for inorganic/organometallic structures frequently gives rise to nonsensical descriptions which are mutually incompatible.

In Spring 2019, the InChI Trust commissioned an effort to gauge the scope of the problem and chart a course toward augmenting the InChI identifier so that it can play nice with all manner of different *coordination complexes* that are found throughout the inorganic/organometallic realm. The results of this initial step can be found in the public GitHub repository [https://github.com/aclarkxyz/data_coordinchi], which also includes a detailed writeup of the technology.



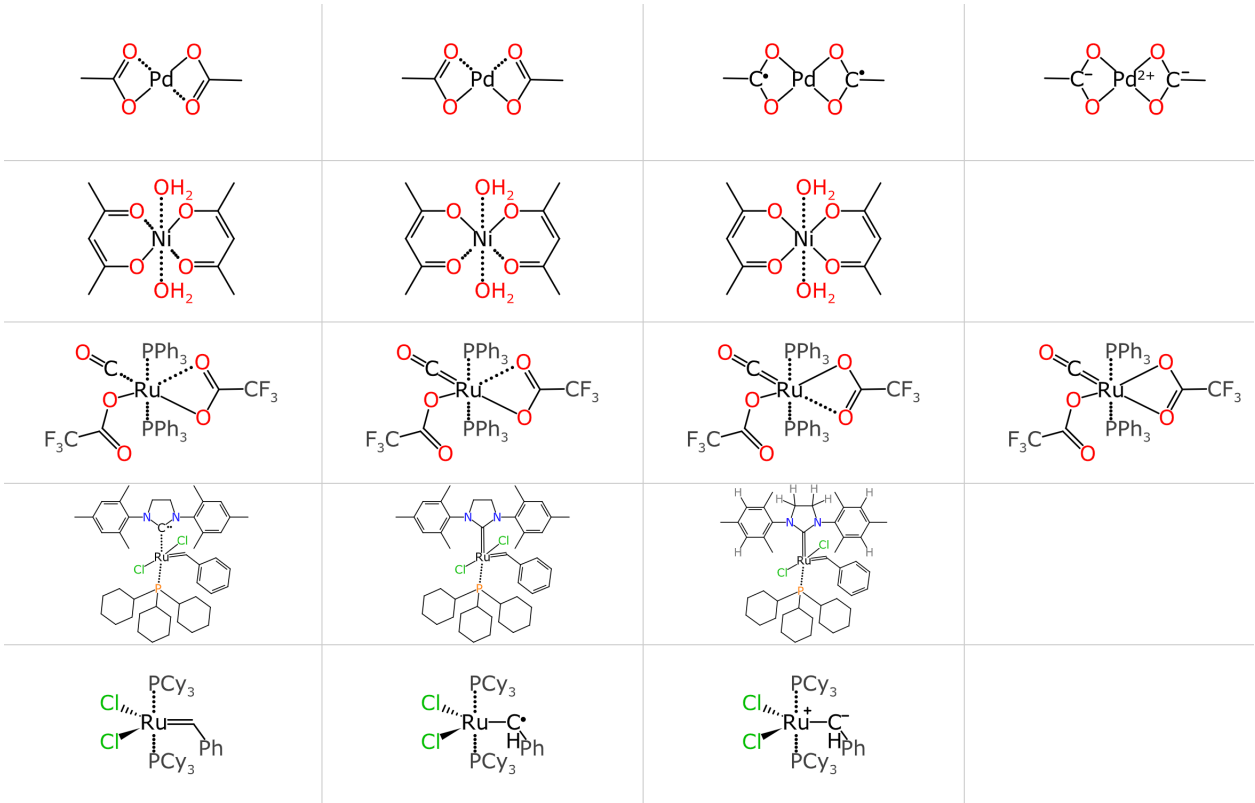
The first order of business was to gather data. As many are aware, there is very little software support for representing coordination compounds, and consequently even less useful data. One of the starting points was a set of 500 or so complexes that I had curated myself over the years [https://pubs.acs.org/doi/abs/10.1021/ci200488k]. Another source was PubChem [https://pubchem.ncbi.nlm.nih.gov], but unfortunately the vast majority of coordination complexes were

either drawn incorrectly at the time of submission or broken beyond recognition by the autocorrect algorithms. The most useful source turned out to be the Cambridge Structural Database [https://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd] which was made available for this project, along with some secret information that unlocked valuable cheminformatics properties, which will hopefully be unveiled at a later date. About half of the million structures is made of non-polymeric coordination complexes, which are fair game for this task.

From these collections, a diverse set of coordination complexes was assembled, which were determined to be correct - which for this purpose is defined as:

1. complete bond graph, i.e. no disconnections between bonded atoms
2. all hydrogens accounted for correctly, whether virtual or actual
3. charges correctly localised onto their connected components

Note that the above criteria are quite loose. Condition (1) says nothing about how the bond types are labelled; condition (2) can be satisfied by creating an atom for each hydrogen, or defining a formula for H-counts, or storing the number as an explicit property, or some combination thereof; and (3) allows charges to be placed anywhere, as long as they add up to the right value without jumping across space. This means that almost all coordination complexes have a huge number of valid ways to draw, and most of them have more than one variant which might be chosen by at least one chemist within the limitations of a particular drawing tool. For the dedicated training set, I went through 500 diverse structures and ensured that each one had several different sketches, each of them valid in its own way.



The primary goal of the coordination-enhanced InChI identifier is to be able to create the same output string for every structure that describes a particular molecular entity correctly. At this stage of exploration, we are ignoring stereochemistry (which is fortunately a separate problem), and are also not too worried about being able to reverse the process (i.e. get back a meaningful drawing). The question is: can we devise a canonical

labelling method that is invariant to the multitude of different ways that coordination complexes can be drawn?

The short answer is yes: there is at least one way - as demonstrated - which works well and has an acceptably small set of known edge cases, with supporting evidence of this claim drawn from real world structures (i.e. the Cambridge Structural Database). The approach that turned out to be productive involves using the bond type metadata to perceive electron delocalisation subgraphs, within which electrons (and charges) are considered to be mobile and thus smeared across that proportion of the graph. This includes the familiar concept of aromaticity and double/single bond resonance, as well as ligands coordinated to metal centres. Once this treatment has been applied to the input structure, it is possible to compose a method for labelling each of the atoms and bonds, and selecting a canonical ordering for the atoms and bonds. This can be used to generate a line notation string, much the same as is done for the standard InChI and various implementations of canonical SMILES. The technical details (and the known edge cases) are described in detail on the GitHub site.

For the future of InChI, this alternate method of labelling represents an integration challenge, since it cannot coexist with the normalization procedures that the standard InChI algorithm uses. The first official implementation may be done by grafting on an additional layer to the standard InChI, which can be used as an alternate option when coordination bonds are part of a particular use case. The proof of concept approach for coordination complexes also does not attempt to address tautomerization (which may not even be viable for inorganic compounds) and defers the issue of stereochemistry (which has several extra geometries that simple organic compounds do not).

In conclusion, we now have an openly available demonstration that an InChI-like canonical identifier is possible for coordination complexes, and it does not rely on introducing any specific drawing conventions besides within reason. And we also have a validation set of real-world coordination complex examples that can be used to verify correctness for any further refinements.

Lastly, mention should be made of CDD's continued involvement in the project,
cf. https://www.collaborativedrug.com/cdd-awarded-phase-2-sbir-grant-mixtures-formats-drug-discovery-formulation/

Inorganics

Chairperson: Hinnerk Rey

A decision on how to proceed with this awaits the outcome of the Organometallics work. Hinnerk is working with Ture Damhus (Division VIII) to develop what needs to be done.

Large molecules, biopolymers/proteins/biological polymers/macromolecules/biomolecules, etc.

Chairperson: To be determined (TBD)

As mentioned previously, the working group chair has bowed out. There is nothing further to report at this time. It would seem necessary to find someone willing to chair and be active in moving this area forward.

Extended Tautomers

Chairperson: Marc Nicklaus

The scientific part of the project has been completed. More than 80 tautomeric transforms have been identified from various sources of experimental literature. Two papers forming the scientific background of this project have been submitted for publication: "Toward a Comprehensive Treatment of Tautomerism in Chemoinformatics Including in InChI V2" and "Tautomer Database: A Comprehensive Resource for Tautomerism Analyses" (preprints available at https://doi.org/10.26434/chemrxiv.10794962.v1 and https://doi.org/10.26434/chemrxiv.10790369.v1). See also L. Guasch, M. Sitzmann, M. C. Nicklaus, "Enumeration of Ring–Chain Tautomers Based on SMIRKS Rules", *J. Chem. Inf. Model.* (2014), 54(9): 2423-2432; https://doi.org/10.1021/ci500363p. Their contents constitutes the scientific material that the Working Group will use in its decision what types of tautomerism should be recommended to include in InChI V2.

The vast majority of the SMIRKS rules match at least hundreds (and many match millions) of molecules in large small-molecule databases such as PubChem. Fewer than ten of these rules are currently covered to a large extent by current (V.1.05) Standard InChI. A somewhat higher number are covered in a non-standard InChI generated with additional tautomerism options KET and 15T turned on. In general, about three times as many molecules would be affected if these rules are implemented in their entirety in InChI V2 in comparison to what InChI V1's current algorithm does.

First coding tests are being initiated to investigate how (subsets of) these new tautomeric rules can best be added to InChI, be it as an extension to the existing InChI code or as a rewrite of the code. These two alternatives are very different. As Johnathan Goodman has pointed out:

Marc's work is extensive, and it would be good to implement it. There are two obvious routes:

"(i) Rewrite the whole InChI code to incorporate it

(ii) Write a program to add an extra layer to current InChI. It would take an InChI and generate a layer which either indicated that the InChI was the preferred tautomeric form, or else how the preferred form could be generated from it.

Option (i) is a large amount of work, which would generate new and incompatible InChIs. Option (ii) could be achieved much more quickly, and would retain backward compatibility."

The first step in being able to decide which choice to make has been taken with this initial contract. I hope that there will be sufficient information from this initial contract so that we may have a useful discussion about the options at the InChI meeting in San Francisco.

QR Codes

Chairperson: Richard Hartshorn,

A 15-page manuscript is being submitted to the IUPAC PAC journal on the recommendations for the use of InChI QR Codes on Labels for Chemicals: Linking labels to digital resources. The manuscript is attached to this report. Some of the specific recommendations that involve the Trust, extracted from the manuscript are: IUPAC and the InChI Trust should set up and run an InChI resolver. Those running the resolver should work with the major (open and public chemical database providers) to ensure that the resolver contains as wide a coverage of the compounds mentioned in the chemical literature and suppliers as possible; we acknowledge that this again is far from trivial to support in the long term.

That IUPAC and the InChI Trust should commission a QR InChI app (ideally for IOS and Android) that would read the InChI QR code and then enable the users to make the most of the InChI enabled services that exist worldwide. This would significantly increase the worldwide exposure for the InChI.

That there should be discussions with industry associations to understand in more detail how the QR codes we recommend can be used within their current labelling systems and requirements that need to be met by the chemical suppliers.

Education/Academic/Training

August 2, 2019 – January 15, 2020 Status Report

This report details efforts of the InChI Trust OER Working group, along with efforts of the IUPAC Task Group to develop an OER, project no. 2018-012-3-024 (https://iupac.org/projects/project-details/?project_nr=2018-012-3-024). This report covers the time span from August 2, 2019 to January 15, 2020.

The primary mission of the OER task group is to bring about a greater awareness of InChI in the education community, while also facilitating its use and adoption across other communities. There are three major facets to this endeavor; creating original educational material, and creating a resources that allows educators and scientists to share material with others, with the latter being the mission of the OER website (https://www.inchi-trust.org/oer/), and general outreach (running workshops, symposia, giving talks, posters, etc.).

(a) Original Material Developed: During this time, one tutorial on the applications of InChI to an Organic Chemistry class was uploaded by Walker, and three tutorials involving InChI with python scripts were uploaded by Cornell and Belford.

(b) Work on OER site: During this time 44 items were added to the site (mostly by Cornell), raising the content from 42 to 86 published items, 59 of which are tagged as OER. Cornell also cleaned some of the script as AJAX broke upon one of the updates.

(c) The following posters, oral presentations and other activities were performed by members of the task group.

1. VII Jornades sobre l'Ensenyament de la Física i la Química, Institut d'Estudis Catalans (Calle del Carme, 47, Barcelona)
    - "Resources from the InChI OER: Spreadsheets to Teach and Learn about the InChI", Poster manned by Jordi Cuadros, 10-24-2019.
2. 258th ACS National Meeting (San Diego),
    - "InChI Open Education Resource (OER)" talk in Chemical Nomenclature & representation: Past Present and Future symposium given by Belford
    - "InChI Open Education Resource (OER)" SciMix poster manned by Belford

Future Efforts:

1. 259th ACS National Meeting (Philadelphia), March 22-26
    - "Cheminformatics for Chemists" Symposium Organized by Belford & Qin
    - "InChI OER Integration with the LibreText", CINF & SciMix poster sessions to be manned by Belford, Qin and Scalfani
2. 26th IUPAC International Conference on Chemical Education (Capetown, SA), July 13-17.
    - Jordi Cuadros will attend and present

3. 26th Biennial Conference on Chemical Education (Corvallis, Oregon), July 18-23
    - Belford, Gupta & Bulchotz will attend
4. 260th ACS National Meeting (San Francisco)
    - "Uses and Applications of InChI: Past, Present and Future, symposium organized by Belford and Qin.

Extended Stereochemistry

Chairperson: Andrey Yerin

- Project is approved. Stated start date: 09 September 2019
https://iupac.org/projects/project-details/?project_nr=2019-017-2-800
- The first working document have been sent to the group and is related to support of MOL V3000 enhanced stereo by InChI tools.
- This year we expect the task group meeting. The dates and place are not clear yet.

This working group is just getting started. Andrey put in a project plan for some support in April 2019 but it was misplaced and only with the help of Fabienne Meyers last month was the proposal submitted for review by IUPAC. Preliminary report on enhanced recognition and encoding of stereoconfiguration by InChI tools.

An improvement in treatment of stereoconfigurations by InChI tools was rated as the most important task by the participants of the InChI Workshop in Bethesda in August 2017.

In accord with the communication with the interested parties and analysis of existing InChI tools the following tasks related to stereoconfigurations are listed as needing improvements:

1. Support of enhanced stereo designation in accord with the specifications of MOL file V3000;

2. Correct recognition of stereoconfigurations in traditional Haworth and chair representations of carbohydrate structures currently incompletely interpreted by InChI tools;

3. Recognition and encoding of stereoisomers with chirality axis related to atropisomerism;

4. Possibility to recognize and encode additional types of stereoisomers other than double bonds and tetrahedral chirality;

5. Several specific improvements to recognize more complex cases of tetrahedral chirality and longer stereoisomeric allenes.

This works assumes development of new additional procedures to recognize additional types of stereoisomers and specification of notation to encode them in InChI string. The most tasks are straightforward and should be relatively easy to implement.

All tasks except the second assume changes in InChI procedures and encoding principles. It seems that all or at least most of them can be implemented in InChI version 1.x leading to just additional recognized stereogenic units or assuming minimal changes in InChI format.

The second task can be done via external module that should recognize Haworth and chair representations and convert them into structures with strict definition of stereo centers for further treatment via the standard InChI procedures.

The corresponding project proposal was submitted to IUPAC Secretariat in April 2019.

The current task group is represented by five members:

Andrey Yerin (ACD/Labs, Russia) Gerd Blanke (StructurePendium Technologies GmbH, Germany) Igor Pletnev (Moscow State University, Russia) Burt Leland (OpenEye Scientific Software Inc., USA) Jürgen Kammerer (Sanofi-Aventis Deutschland GmbH, Germany)

While the work assumed by the project is not started yet waiting the project approval, the documents, presentations and communication at the InChI Workshops and during the preparation of the project proposal make a good starting point for the project having detailed specifications for most tasks with representative examples.

GitHub

There has no progress about any GitHub/InChI activities as we are waiting for InChI version 1.06 release and associated source code. This should occur by late spring of 2020.

August 2019 – December 2019 Activities/Status

Meetings Attended and Talks/Posters Presented

GDCh poster, November 2019
A number of conference call meetings with Ray Boucher, Richard Kidd and Ian Bruno were held over the past six months to deal with issues that needed to be addressed between Board meetings.
I met on a regular basis with members of NIH/NCBI, particularly Evan Bolton, to discuss InChI issues. On July1 I was appointed as a Guest Researcher at NIH/NLM/NCBI. Going forward, I will be working from there on the InChI project
I met with Guenter Grethe in San Diego on January 10 to discuss the RInChI project.

Manuscripts, Talks, and Posters

There were no new manuscripts published.

PIDapalooza 2020

Ian Bruno and Henry Rzepa will give an InChI presentation at the January 29/30, 2020 PIDapalooza (Persistent Identifier) meeting in Lisbon.

InChI Trust web site

The Trust web is up on the InChI Trust cloud server and working well. Updates, such as these reports, are behind schedule for updating.

InChI Usage

Numerous publications now use InChI as part of their efforts in merging and analyzing database structures. Clearly InChI is being used on a very regular basis in many organizations and research projects and publications.

InChI Trust Videos - Access numbers/Views as noted below continue to increase slowly every year:

InChI & the Islands – 1,379 (1/20), 1,327 (7/19), 1,269 (12/18) , 1,208 (7/18), 960 (1/17); 804 (1/16); 728 (7/15); 526 views (7/14)

The Googleable InChIKey – 2,423 (1/20), 2,301 (7/19), 2,115 (12/18) 1,985 (7/18), 1,379 (1/17), 1,037 (1/16); 915 views (7/15), 597 views (7/14)

The Birth of the InChI – 2,013 (1/20), 1,948 (7/19), 1,848 (12/18), 1,791 (7/18), 1,365 (1/17), 1,084 (1/16), 984 views (7/15), 687 views (7/14)

What on earth is InChI? - 7,884 (1/20), 7,348 (7/19), 6,750 (12/18), 6,102 (7/18), 4,188 (1/17), 3,331 (1/16), 2,956 (7/15), 2486 views (12/14); 1977 views (7/14)

IUPAC InChI (Google lecture - 2008) - 998 (1/2), 978 (12/18). 950 (1/17); 946 (7/16); 931 (1/16); 922 views (7/15)

IUPAC InChI (Google lecture - 2006) - 940 (1/20), 893 (7/19)

Some History, Organizational Planning, and Project Sustainability

I repeat what was in my last report as the Trust has not yet had their meeting to discuss these matters.

As mentioned in my previous reports and in discussions with a number of people about the long-term future, direction, and organization of the InChI project need to be addressed – in a technical, administrative, and financial sense.

The technical issue of how to maintain and expand the InChI algorithm appears to be easier to deal with. Having one programmer maintain and add to the algorithm, with additional pieces (such as RInChI and MInChI) coming from other programmers seems to be working well. The SourceForge group of programmers who test and provide feedback has been working well, but is not the ideal or accepted way to do oversight. The idea of using GitHub to have people around the world offer additions to the algorithm seems sensible in principle, but owing to the nature of InChI being an international standard, there are complications. Who decides at IUPAC and/or the Trust if more features are needed? Or does the community (whomever they are) decide? And where do the resources come from those who "decide" what is needed?

As for administrative and financial matters, things are a bit more problematical. IUPAC has never had a project like this which requires ongoing work and support and, after some 20 years, there does not seem to be another project like this in the works. That was the main reason the Trust was established some 10 years ago. At the recent IUPAC General Assembly in Paris in July 2019 members were informed that the IUPAC financial situation was, to be generous, not improving, and various (primarily travel) expenses would be reduced. While the Trust has seemed to be working well for the past decade the issue of ongoing support from the current Members and Associates is less clear. Much of the financial support really comes via individuals within these organizations who believe in the project. We have seen in the past few years when

some of the "founding" members change jobs their replacements do not have the same interest and enthusiasm for InChI. This has resulted in the loss of some members over the years.

This is not a unique issue. For many years scientists who developed databases had both funding problems to maintain and add to their databases as well as not receiving credit within their institutions for their efforts as this work was not considered publishable research. Well-known databases, such as Beilstein, which had been around for over a century, disappeared when the German government decided to terminate support. The world has changed since I first started working in the area in the 1970s. I and a number of my colleagues worked for various Government agencies and were able to move forward, while I doubt this would have been the case had we been in an academic setting.

In the past few years there have been considerable efforts to develop data and related standards in science from a number of groups (e.g., IUPAC, RDA, Pistoia, Allotrope, FAIR,). InChI could continue as a standalone activity if proper institutional, political, and financial support were available. InChI could be adopted by an established organization whose long-term goals and plans could include InChI. For example, the RSC, ACS, or EBI could be a possible long-term home. NIST, as the US standards agency and the organization that developed the InChI algorithm has shown no interest at all in taking the lead for this standard. The last initial suggestion would be the NIH/NLM/NCBI PubChem, an organization which has the expertise and is one of major users of InChI.

The issue of how IUPAC, the InChI subcommittee and its working groups, and the Trust work together needs to be updated. I believe the actual standards development work is now all done by the individual working groups. When the project first started almost 20 years ago, the area being developed was small (less than 1000 atoms) organic molecules. When the InChIKey standard was developed the subcommittee, along with other experts in the field, met at NIST to decide how to do it. The meeting in 2012, located near NIST, the last one until 2019, agreed that the working groups had moved from general organic chemistry to various detailed subcategories of chemistry (e.g., inorganics, polymers, reactions, organometallics ands related coordination complexes, and extended tautomers). As these areas were highly specialized there was little for the subcommittee to do other than "oversight", which did not work out as well as one might have hoped. The subcommittee members, for the most part, did not have the necessary detailed and intimate knowledge of these areas to be able to perform any useful working group oversight. In addition, while it was relatively easy to find expert and knowledgeable chemists in these subdisciplines who offered to volunteer to develop standards for creating InChIs, it was a lot harder to get these chemists to deliver a written standard for their area. In most cases the task was more difficult and complex that initially thought. In some cases, these volunteers had their day jobs that took priority. In other cases, people moved on from one job to another or resigned to other work within their organizations. It is easy to say what kind of person we need for a working group – a person interested in InChI, a person competent in the particular areas of chemistry of the working group, and a person who actually does put in the time to do what is needed in a timely manner. In practice, finding such people has proved very difficult to say the least.

In the past, a number of the working groups were established with a narrow remit (e.g., Inorganic Chemistry) but as their efforts progressed it has become clear just within the last year or so that these areas of chemistry have considerable overlap and much more interaction and coordination is needed. For example, some of the issues for a standard coordination compounds overlaps with inorganic chemistry. It would seem that as the project has evolved the role of the InChI subcommittee will not be able to provide oversight, arbitration, and approval of new extensions and standard, but rather it would be best for the subcommittee to help organize the working groups in more productive and functioning groups.

The issue of how to assemble the working groups for the current and future areas of InChI expansion would seem to be a critical mission of the subcommittee. Finding people who are interested in InChI and have the time and willing effort to expand InChI capabilities is clearly a difficult task as we have discovered over the last few years, as we are dealing with experts who are volunteering their time and often have necessary distractions that side track things. As a subcommittee member pointed out to me: "The responsibility of the

subcommittee is to oversee the working groups, facilitate the communication between the working groups, and provide whatever support possible to allow them to come to as rapid a solution for a standard in a particular area as possible."

At the end of the day there will be progress, not by micromanagement of the working groups, but rather by offering whatever support the subcommittee can provide. The subcommittee receives no support from IUPAC/Division VIII and with the current (and future?) financial condition of IUPAC none is to be expected. This leaves either no support or the need to be supported by the Trust. As for what to do about slow progress, it is hard to predict which working groups will actually be able to work and produce. While it is relatively easy to create a working group and get chemists expert in their area of chemistry to offer to join, it is harder to guess in advance who will actually work and produce. When there is little progress, assuming it is not due to the difficulty of developing the standard for that area of chemistry, while it is easy to remove people from the working group, replacing them is not so simple. The new terms of reference for the subcommittee call, for example, "to actively participate in meetings", but when only perhaps half the subcommittee comes to Paris due (in the most part) no travel support, what should be done? Should a no-show in Paris result in being removed from the committee? Electronic only meetings have value, but face-to-face meetings cannot be easily replaced as many discussions are between 2-3 individuals, not the subcommittee as a whole. A decision needs to be made by the Trust as to what support it will provide to subcommittee members to meet. If the Trust decides to fund the subcommittee then one may consider asking the question, what is the role, if any, of IUPAC in this project going forward?

Plans for the remainder of 2020

For 2020 my current overall plans and goals are as follows:

1. Work to expand the current membership with two basic classes of members – Full and Associate as well as add to the number of Supporters. Continue to attend meetings and give talks on InChI where useful and appropriate.
2. Attend ACS meetings in Philadelphia and San Francisco.
3. Possibly give an InChI seminar for Steve Boyer/Google patent staff.
4. Attend and participate in sessions on InChI, InChI working groups, and related standards at the ACS meetings
5. Attend IUPAC and InChI and InChI Trust Board meetings in San Francisco.
6. Meet with groups to discuss adoption and usage of InChI.
7. Meeting Shuli You in Shanghai in mid-June to discuss Chinese participation in InChI
8. Attend the November 2019 GDCh meeting and have a poster or oral presentation on InChI.

2020/2021 possible workshops/symposia

NIH, where the largest InChI meeting was, has not been considered since there are no major chemistry meetings in Washington DC in the next 2 years.

2020

August 16 - 20, 2020 - ACS San Francisco, CA

October 29 - November 5 - Elsevier & Wiley visits and annual GDCh meeting, Germany

December 15 - 20, 2020 - Pacifichem Honolulu, HI. Perhaps participate a ½ day session at a hotel we if can find a session already going to be there. I could be there

as it is only 51 miles from where I am staying at that time of year. If I go, I would go just for the day, as the cost of Honolulu hotels is much higher than a round trip airfare from the next island.

2021

February 2021 – Workshop at University of Cambridge or EBI (if someone there is interested or RSC/London (need to reserve early for a room in London)

March 21-25, 2021 – ACS National Meeting, San Antonio TX

Spring 2021 – The 12th International Conference on Chemical Structures (ICCS), Noordwijkerhout, The Netherlands

July 13 - 17, 2021 Cape Town, ZA – 26th IUPAC International Conference on Chemistry Education (ICCE 2020)

August 22 - 26, 2021 – ACS National Meeting Atlanta, GA (same as San Diego/San Francisco assuming that all goes OK there)

Steve Heller

Appendix 4(b)

InChI Trust Project Director's Report
August 2020 (8/18/20 DRAFT)

Summary:

In spite of the world-wide COVID 19 pandemic the project to develop the InChI algorithm for all defined chemical structures continues to move forward with more progress to report here than in the past year. Progress included the pre-release of InChI version 1.06, initial implementing and testing a few new tautomer structures not previously handled by the original algorithm (released some 13 years ago developed by Dmitrii Tchekhovskoi) and work on extending the algorithm to handle coordination compounds/organometallics and inorganics. Lastly, as this report was being finalized, it was agreed that for purpose of the InChI working groups progress and more fruitful interactions and discussions, the Organometallics and Inorganics working groups will be merged into one.

While there is a world-wide COVID 19 pandemic as this report is being written I remain a Guest Researcher at NIH/NLM/NCBI in the PubChem Project led by Evan Bolton, where I am well supported and I continue my InChI activities and work.

Items covered in this report:

Membership/Support
InChI RFP/Contracts
InChI development and maintenance work
IUPAC InChI subcommittee
Working Group Reports
January 2020 – June 2020 Activities:
  Meetings attended & Talks/ Posters given
  Manuscripts
  PIDapalooza
  InChI Trust Web Site
  InChI Usage
  Technical Issues
  GitHub
Some History, Organizational Planning,
  and Project Sustainability (see attachment)
Plans for 2020
2020/2021 possible workshops/symposia

Membership/Support:

As mentioned, numerous times in the past in most organizations, since InChI works and it is not high on their immediate priority lists, actual real progress is slow without a dedicated champion within an organization. The Chinese Chemical Society (CCS) has joined the Trust as a member. With John Wiley having bought Bio-Rad in April 2020 I expect Bio-Rad (Greg Banik is a long time InChI supporter) will not renew its membership in 2021. The Google membership is still pending. Discussions with InfoChem joining the Trust have been going on for a number of years but have yet to materialize.

As of August 1, 2020:

Members (9):

IUPAC
ACS/Chemical Abstracts Service (CAS)
Chinese Chemical Society (CCS)
Elsevier/Relx Group
Royal Society of Chemistry (RSC)
Springer Nature
John Wiley & Sons
Informa/Taylor & Francis
US National Institutes of Health (NIH)

Associates (11):

ACD Labs
Bio-Rad
CCDC
ChemAxon
Google (pending agreement)
OntoChem
OpenEye
Sigma Millipore
University of California
US Food and Drug Administration
US National Institute of Standards and Technology

Certification Suite:

US Environmental Protection Agency

Active Supporters (42):

AKos Consulting and Solutions
American Chemical Society Division of Chemical Information (CINF)
Biochemfusion ApS
Caltech Library Services, Pasadena, CA, USA
Centre for Molecular Informatics, Cambridge University, UK
Chemistry Department, Clemson University, SC, USA
Chemistry Department, University of Arkansas at Little Rock
Chemistry Department, University of California, Riverside, CA, USA
Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, NC, USA Tropsha
Faculty of Science, University of Paderborn, Germany
Gesellschaft Deutscher Chemiker e.V. (GDCh), Germany
Guide to Pharmacology, UK
Imperial College London, UK
Institute for Chemoinformatics and Bioinformatics, University of Applied Sciences Gelsenkirchen, Recklinghausen Section, Germany
Institute of Chemical Technology, Prague, Czech Republic
Institute of Organic Chemistry, KIT Karlsruhe

International Union of Crystallography
Leadscope, Columbus, OH, USA
Leibniz-Institut für Analytische Wissenschaften – ISAS, Dortmund, Germany
Ludwig-Maximilians-Universität München, Munich, Germany
Mcule
Molecular Materials Informatics, Inc
National Center for Biomedical Ontology, Stanford University, CA, USA
National Chemical Laboratory, Pune, India
NextMove Software, Cambridge, UK
Open Babel
RJB Computational Modeling
Royal Netherlands Chemical Society
School of Chemistry, University of Leeds, UK
SciencePoint, Redmond, WA, USA
StructurePendium Technologies GmbH
Technical University of Vienna, Austria
The Chemistry Development Kit, Eindhoven, The Netherlands
TW2Informatics Limited
University of California, Davis, Genome Center, CA, USA
University of Indiana, Bloomington, IN, USA
University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia
University of Southampton (Chemistry), UK
University of the West Indies, Mona Campus, Jamaica
Xemistry GmbH, Königstein, Germany
ZINC

InChI RFP/Contracts

After a long drought in initiating the programming of the standards that working groups have developed and agreed upon there are two major projects underway which have produced initial results for further testing and examination - Organometallics/Coordination Compounds/ Inorganic and Extended tautomers. A contract for work on programming InChI for coordination compounds, organometallics, and some inorganics was awarded to Alex Clark and already has produced deliverable results. The initial results of the work performed are given below in the section on Organometallics, et al. An initial contract to examine the issues associate with extending the InChI capabilities in the area of tautomers was approved and work has been started by Igor Filippov and a few examples are being tested.

InChI Development & Maintenance Work

Igor Pletnev continues to do a superb and a very responsive job as the InChI programmer. The pre-release of InChI version 1.06 has finally arrived and is available for broad community testing in August - September. A copy of Igor's release notes is attached (v106-pre-RelNotes) Version 1.06 which includes the "any atom" feature originally suggested to Igor by the FDA/Yulia Borodina. It has been delayed due to numerous minor bugs that were fixed. There continues, as expected, to be useful feedback on minor issues and bugs as noted below. Gerd Blanke continues to do excellent work on the RInChI algorithm and is coordinating very well with the MInChI working group led by Leah McEwan.

Igor has reported that many people have been in contact with him regarding bugs, errors, and issues with the algorithm. Without all this external help, primarily in the past from the InChI SourceForge list (established by Alan McNaught in May 2005), the algorithm would not be as good as it is.

Besides the SourceForge inchi-discuss list, Igor had many valuable comments/issues, and advice reported in private correspondence by several experienced chemoinformaticians/developers which routinely use InChI. When the GitHub InChI web site is set up, we all expect to have even more input, assistance, and cooperation in furthering the development of InChI.

There are a significant number of issues (and related fixes) Igor found in his own testing processes.

Some minor questions/reports also came from occasional InChI users, typically via Richard Kidd, and from other sources (like the very helpful Google auto-fuzz).

In summary, we have many users and many people from all different areas of chemistry using and working to help improve the InChI algorithm.

SourceForge InChI url: https://sourceforge.net/projects/inchi/

IUPAC InChI Subcommittee

IUPAC InChI Subcommittee members

The following are the current members of the IUPAC InChI subcommittee.

- Stephen Heller, Chair
- Jonathan Goodman, Secretary

- Members:
- Steve Bachrach
- Bob Belford
- Gerd Blanke
- Evan Bolton
- Marc Nicklaus
- Carmen Nitsche
- Hinnerk Rey
- Wendy Warr
- Tony Williams
- Andrey Yerin
- Shuli You

- Igor Pletnev – Technical Advisor
- Dmitrii Tchekhovskoi – Technical Advisor

On August 12, 2020 we had a Zoom subcommittee meeting that lasted for over an hour. A number of issues were discussed as noted below in the Agenda that Jonathan Goodman prepared. We plan to have monthly Zoom meetings with the InChI subcommittee and invite one of the working groups to each meeting to discuss what they are doing and where they are headed, what resources, if any, are needed. The minutes from this meeting are attached (InChI-subcommittee-8-2020)

Agenda:

(i) Report from Steve Heller (attached - InChI-The-Future)

(ii) InChI 1.06

(iii) Future meetings:
Spring/Early summer 2021; in-person or hybrid in-person/virtual in Cambridge
Working groups to get together to discuss progress

(iv) Future virtual meetings
Would regular, short, Zoom meetings be helpful?

(v) Process for introducing changes
(a) Minor problem discovered
(b) New functionality requested
(c) Move to InChI V2

(vi) Working groups:

Working Group Reports

1. **Mixtures**
2. **Monomer Atoms**
3. **Variability**
4. **Isotopologues**
5. **Positional Isomers (to be merged into Variability)**
6. **Resolver**
7. **Polymers (to be removed in future reports)**
8. **Reactions**
9. **Organometallics**
10. **Inorganics (to be merged in Organometallics)**
11. **Large Molecules**
12. **Extended Tautomers**
13. **QR Codes**
14. **Education/Academic/Training (OER)**
15. **Extended Stereochemistry**
16. **GitHub**

Chemical mixture composition (MInChI)

Implementation of the MInChI specification is pending work on RInChI v2. Leah McEwan reported that the working groups understands from Gerd that this work is initiating and he should have everything he needs to get a start on incorporating MInChI. The MInChI working group is planning to schedule a joint meeting of key technical folks on both projects to review progress This meeting was originally considered for San Francisco, but now will, hopefully, take place in early fall.

Pseudo-atom biopolymer monomer representation atoms for InChI

In January 2020 a mini proposal (provided in the January 2020 Trust report) was disseminated by Evan Bolton for feedback. Feedback has been slow. The pseudo-atom approach is a way to extend/compact InChI representation with large molecules (many atoms represented as a single atom … so the 32K limit on atoms goes up substantially).

The 'Zz' atom approach enables variability for large molecules, which has similar but different concerns from small molecule variability being proposed by Jonathan Goodman, as was discussed at Cambridge in February 2020.

Variability (VInChI)

InChI and Variable structures: How can an InChI provide a canonical encoding of variable structures?

Working Group:
Jonathan Goodman
Gerd Blanke
Istvan Ori
Anthony Baston

A general solution to canonical variable structure identifiers is a challenging problem. We are moving towards this a step at a time. Anthony Baston, a Master's student at Cambridge, has written a python script which takes a list of InChI and generates a VInChI (variable structure InChI). It can also generate a list of InChI from a VInChI. This proof-of-concept program is restricted to isomeric acyclic alkanes. The program is available for testing.

There is a specific issue with canonicalization. The maximum common substructure found by RDKit is sensitive to the order of the inputs. Ways of addressing this issue are being investigated.

We need to build on this proof-of-concept in order to deliver a program which is more generally useful. Future extensions include broadening the scope of the molecules covered, including stereochemistry and making it possible to describe non-isomeric groups. Tools are also needed to add and subtract VInChI, ideally without enumeration as this will be time consuming for complex groups, to find common structures, and to generate large groups algorithmically rather than always relying on lists.

VInChI

input.txt
should contain either a list of InChI or a single VInChI

source activate my-rdkit-env
python vinchi_b.py input.txt

Isotopologues

Chairperson: Hunter Moseley

COVID 19 has delayed most of the efforts in this area.

We are preparing a draft the InChI isotopologue manuscript to be sent it to the team that developed that part of the standard.

As previously reported, there has been work on the development of a draft SD file representation of isotopologues to facilitate generation of isotopologue representations in InChI.

There are two major things for the working group to do:
1) Write up a manuscript describing the accepted InChI isotopologue extensions.
2) Find an example of the fixed hydrogen layer issue in InChI.

Positional Isomers

This area of extending the InChI algorithm into positional isomers and been subsumed by the working group on InChI variability led by Jonathan Goodman. This working group will be disbanded and merged into the Variability work group.

Resolver

Chairperson: Markus Sitzmann

The following is the short InChI Resolver status report (the longer version is at https://github.com/inchiresolver/inchiresolver/blob/master/README.rst)

1. I have implemented three prototype Resolver instances which are running here:

- (InChI Trust) Root InChI Resolver at https://root.inchi-resolver.org
- PubChem InChI Resolver Instance at https://pubchem.inchi-resolver.org
- InChI Resolver Instance of the NCI/CADD group, respectively the Chemical Structure Resolver at https://cactus.inchi-resolver.org

They all contain not much data yet but they demonstrate how InChI Resolvers supposedly should work together (I currently use them for testing and I used them for figuring out how the protocol should work). They all run on the same host but internally are separate instances only linked by URLs.

There is more information and a overview graphic here (well, the writing is still work in progress): https://github.com/inchiresolver/inchiresolver/blob/master/docs/prototype.rst

2. I have decided to use JSON:API (https://jsonapi.org/) as media type for the InChI Resolver protocol. The three resolvers above already fully implement it (beside some bugs I still find here and there). There is quite a bit of (client) software and support for many programming languages for this media type format (and on a very low level it is just json anyway). In my opinion it works well for the InChI Resolver project.

3. I currently prepare documentation/specification of the InChI Resolver protocol (it is still a rough version and I will improve it the next few days and weeks), the current state can be found here: https://github.com/inchiresolver/inchiresolver/blob/master/docs/inchi_resolver_protocol.rst

Besides that, the current version/code of the InChI Resolver is fully available and installable for everybody from GitHub (https://github.com/inchiresolver/inchiresolver), I think GitHub is generally also a good place for any feedback or collaborative work.

Polymers

Chairperson: Andrey Yerin

As Andrey states below, the first phase of InChI for polymers is finished and working and we await feedback, input, and user needs before looking into the matter of doing anymore. This working group report will be suspended and removed from these reports until the need for additional work arises.

There have been no activities with InChI for polymers. I treat this project as completed, at least the corresponding IUPAC project is marked as completed in 2015. The basic principles are already implemented, I do not know about any further requests from the community.

Looks like the use of polymer structures is limited and already implemented possibilities are generally enough.

Reactions

Chairperson: Gerd Blanke

Members
David Nicolaides (Biovia, Cambridge, up to May 2020) Gerd Blanke (StructurePendium), Günter Grethe, Hans Kraut (InfoChem), István Öri (ChemAxon), Jan Holst Jensen (Biochemfusion), Jonathan Goodman (University of Cambridge)

Status

The group meets biweekly per Skype conference to further develop RInChI and to discuss actual issues like the participation at conferences and user group meetings. Unfortunately, we lost David Nicolaides who had to step out because of increasing workload in his Biovia position.

RInChI is gaining more attraction as some feedback shows. It is used e.g. for reaction duplicate checks at Elsevier's new "Reaction workbench" and has been tested and implemented(?) by several companies for different purposes. The best feedback received came from Ontochem, they found a number of issues in the current documentation and made a few enhancement requests.

The RInChI group supports Günter Grethe organizing a full day symposium about reaction prediction and optimization at the ACS Fall meeting in 2020.

Citation of the RInChI publications as listed by Web of Science:

Working towards version 2.0

- Finalized issues
  - Atom mapping for reactions
  - Stereochemistry representation
  - Representation of failing reactions
- Currently under discussion
  - Handling of reaction conditions (ProcAuxInfo)

- Remaining issues to be discussed and prepared in Q1 2020
  - Transfer to Open Source development
    - Build the necessary development and test environment on GitHub
    - May become blue print for other technical developments based on InChI
    -

Presentations and Publications

- A poster describing the status of the current RInChI development has been presented at the AI for Reaction Outcome and Synthetic Route conference in Bristol, March 2020.

Upcoming RInChI events
Under discussion:

- Talk at the (digital) Biovia UGM in September 2020
- Poster equivalent at the (digital) GCC conference in November 2020

Organometallics

Chairperson: Colin Batchelor

This is a report from Alex Clark who is doing the implementation work. Alex is extending InChI to handle some inorganic and organometallic compounds.

The basic idea is to (1) extend the phase 1 algorithm so that it produces identifiers that successfully disambiguate various forms of stereochemistry commonly found in metal complexes; (2) refactor the output so that it can be proposed as an optional layer to add to the official InChI generator; (3) deliver a convincing validation set to show how well it works and what kinds of edge cases can be expected (and can also be used to test how well the core InChI code works for these materials, in case anyone wants to have a go at making that work).

Stereochemistry for inorganics is coming along nicely. I've been gathering examples of complexes with metal-centric stereochemistry, and adapted the "dot-hash" approach to incorporate this in a way that will work in the context of the InChI identifier. The main deliverables will be a training set that can be used as a hard-minimum truth, and an algorithm that demonstrates a solution. By the end of the project, the algorithm will be overhauled to deliver something that is more in line with InChI conventions. At present it looks like there will be time left over to document it quite thoroughly.

As noted in the next section report on inorganics it is clear that a merging of Organometallics, Coordination compounds, and Inorganics makes the most sense so that the overlapping areas are handled properly. I have contacted the chairs of both working groups regarding this matter.

Inorganics

Chairperson: Hinnerk Rey

The status mid 2019:

- InChI organometallics: there has been good progress from the contract with Alex Clark
- InChI inorganics: no progress, no group

I tried to revitalize the group / find new members (thanks to Ture as only one person, who replied and furthermore still wants to participate!).

My plan was to have a physical meeting as re-start. First, I thought of Cambridge UK (EBI meeting, which was not on the list for 2020), then on ACS fall meeting (where COVID came into play)..

I still see the urgent need of InChI for organometallics, which cannot be separated from InChI for Inorganics. Therefore, I think both groups need to be combined / need to collaborate. I think inorganics can be under the umbrella of organometallics. I will offer this in next meeting.

The high priority tasks are (at least from my perspective), to provide more samples (Elsevier had provided some a few years and Alex Clark has been proving some as well), which are more comprehensive (all kinds of organometallics, coordination chemistry, inorganics (alloys, ceramics, nano)) and cover a wider variability (from different sources like PubChem, RSC, ACS, Elsevier, …). I know, there have been some samples around, but as far as I know, they are not used / covered by the actual group of InChI organometallics. I think, some areas of the chemical space can be separated easily between both groups, others need more joint collaborations. And this needs to be sorted out as soon as possible.

Now I think it is worth to try an approach with starting a small group and the negotiate the results / proposals within a large auditorium /then vice versa, which I tried in the past).

I want to provide samples by end Q3 2020 (for Elsevier and also PubChem). This work should be coordinated with Alex Clark.

Large molecules, biopolymers/proteins/biological polymers/macromolecules/biomolecules, etc.

Acting Chairperson: Evan Bolton

Members (as of 2017):
Blanke, Gerd
Chalon, Didier
Drijver, Alex
Jensen, Jan
Yerin, Andrey
Berman, Helen

IUPAC Project: 2013-010-1-800

There is nothing to report at this time. However, there are other related activities taking place outside the InChI project, such as HELM, and it is expected that all of these will eventually be merged together into a coherent process of representing these structures.

InChI version 1.0.6, being released as this report is published, will allow new possibilities needed to better support large molecules, including Z atom support. Variability handling in InChI will help to inform how support for large molecules can be improved.

Extended Tautomers

Chairperson: Marc Nicklaus

Below is a summary from Marc Nicklaus of the initial results from the contract for implementing and testing a few extended tautomer capabilities for InChI.

1. Attempt to implement a limited number of new transforms - total of about 20 transforms out of 86 identified by the WG.
2. Attempt to extend the code to encompass handling of the new type(s) of tautomerism in as complete a manner as possible (complete as described by reference tautomer sets provided by the WG).
3. Compile detailed documentation about how the V1.06 code works if possible
4. Document in significant detail reasons for inability to extend the current code so that this can be used for further decisions about future development.
5. Implement as many of the high-priority rules as possible if code can be extended.

The outcome

- 6 rules added to InChI code (satisfies 1, 2 and 5)
- Comparison of these rules with the current tautomerism-related behaviour of InChI being undertaken by Marc Niklaus
- Current code not suited for the other rules (80 or so)
- Comments on current code have been provided (satisfies 3)

While this is about the specific extended tautomers Working Group, it is obviously part of a larger story that concerns InChI as a whole. Hence this is my personal opinion.

I am not sure we can go much further with implementing additional tautomeric rules in the current InChI 1.0x code. Igor Filippov tried several additional sets of (prototropic) rules but couldn't get them to work. Regarding ring-chain and valence tautomerism rules, even Dmitrii Tchekhovskoi stated at our meeting in December that InChI's chemistry model is most likely not amenable to those.

In Igor Filippov's report (attached - Implementing new tautomerism rules in InChI), it was stated that he was "coding in the dark," not really knowing how the current code operates. I doubt he wants to try it again it. No one really knows how this code works at its core - and this includes Dmitrii after all these years.

In general, I see only two possibilities for moving the core code forward in significant ways:
(a) Get Dmitrii Tchekhovskoi to pick the work up fully again
(b) Rewrite it from scratch
Option (a) is very unlikely to happen for a bunch of reasons.

What option (b) could look like depends on whether V2 InChIs should be a continuation of V1 InChIs, or there should be a clean break. I.e., do we want InChI[Key]s of non-capricious molecules such as water and benzene (and billion others) to stay the same; or is it OK if all V2 InChIs may be different from the V1 identifiers?

I don't know who can make this decision. Ultimately, it will be the users.

In any event, even if the all-new option is chosen, recoding InChI will be significant work. Dmitrii said at our December 2019 meeting that it took him 5-6 years of half of his time at NIST. I think that this is a realistic estimate, especially given all the new stuff people want to have in InChI. So, 2-3 years of a highly-skilled full-time developer. You do the math how much this will cost.

Otherwise, if no one puts half a million to a million dollars on the table, I think we are stuck with InChI V1.

Coming back to our project, let's do the comparison of our experimental tautomer-enhanced InChI with the current version, see what we have gained (and maybe lost), and then discuss further moves. I wish this comparison was already done but it got delayed for a whole bunch of different reasons including COVID-19. Hopefully, I'll have some numbers this fall.

QR Codes

Chairperson: Richard Hartshorn

As mentioned in the last report, a 15 page manuscript was submitted to the IUPAC PAC journal on the recommendations for the use of InChI QR Codes on Labels for Chemicals: Linking labels to digital resources. It appears responses and feedback from the numerous IUPAC groups who are asked to examine these recommendations have been slow.

Education/Academic/Training

With schools throughout the world affected by COVID 19 over the past months, there is not much to report on the InChI OER. Ehren and Martin both have created/are creating content involving InChI in the organic chemistry courses.

Extended Stereochemistry

Chairperson: Andrey Yerin

The project 2019-017-2-800 "Enhanced recognition and encoding of stereoconfiguration by InChI tools" proceeds quite in accord with the original plan, while obviously missing a possibility to meet.

Three tasks have been planned for 2020: Support of enhanced stereo in MOL V3000, atropisomers and additional tetrahedral stereo cases.

For the first two areas the documents are already created and shared with the group. The document for the third task is expected later this year.

We are planning to have online meeting in August/September for discussion of the documents and most of all an estimation from InChI development side of a possibility to implement. According to the preliminary considerations a support of atropisomers and additional tetrahedral stereo looks more promising for quick implementation.

GitHub

There has no recent progress about any GitHub/InChI activities as this matter is waiting for InChI version 1.06 release and associated source code, which is now here. Progress on this issue of GitHub should occur in the coming months.

January – June 2020 Activities/Status

Meetings Attended and Talks/Posters Presented

Attended the InChI Trust Board meeting in London in February.

Manuscripts, Talks, and Posters

Bruno, Ian; Rzepa, Henry; Blanke, Gerd. (2020, February). Connecting Chemistry Through PIDs. Zenodo. http://doi.org/10.5281/zenodo.3648508

Wendy Warr provided this mention of MInChI in a CDD press release:
https://www.collaborativedrug.com/cdd-awarded-phase-2-sbir-grant-mixtures-formats-drug-discovery-formulation/

Wendy, as a member of the InChI subcommittee, also submitted the following:
Jonathan mentions his talk given at the AI for Reaction Outcome and Synthetic Route Prediction meeting. It is published (by me *et al.* J) at https://eprints.soton.ac.uk/441628/.

Also in that publication https://eprints.soton.ac.uk/441628/ is:

UDM: a community-driven data format for the exchange of comprehensive reaction information
Dr. Jarek Tomczak, Pistoia Alliance

I note over 42 mentions of InChI in my latest report (*Chemical Information and Computation 2020, Number One. Cancelled 259th ACS National Meeting and Exposition, Philadelphia, PA, March 22-26, 2020, and Miscellaneous Other Meetings*, https://www.warr.com/morepubs.html). One "Philadelphia" talk in particular discusses InChI:

Curating ChemSpider: challenges in chemical data management
Mark Archibald, archibaldm@rsc.org. Royal Society of Chemistry, Cambridge, United Kingdom
You could cite the SciFinder abstract.

Others such as one by Coles *et al.* mention InChI but do not focus on it:

U.K. physical science data-science service: FAIR resource for chemistry in the United Kingdom
Simon J. Coles, s.j.coles@soton.ac.uk , Nicola Knight, N.Knight@soton.ac.uk. University of Southampton, Hampshire, United Kingdom.

Wendy Warr is not sure if it is indexed in SciFinder. It was in the CINF program but it was not in SciMeetings. Simon and I reconstructed it for my report.

Wendy Warr's report reference list includes Winter, R.; Montanari, F.; Noe, F.; Clevert, D.-A. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem. Sci.* **2019,** *10* (6), 1692-1701. It mentions use of InChI as a descriptor in AI.

Using a list of core InChI papers (which are listed below), Jonathan Goodman did a very impressive job finding 296 different citing papers, with increasing citations each year. The complete 353-page list is attached to this report (InChI-Citations-July-2020).

Toward a Comprehensive Treatment of Tautomerism in Chemoinformatics Including in InChI V2
Dhaked, DK; Ihlenfeldt, WD; Patel, H; Delannee, V; Nicklaus, MC
J. Chem. Inf. Model. 60, 1253-1275
DOI: 10.1021/acs.jcim.9b01080

International chemical identifier for reactions (RInChI)
G. Grethe, G. Blanke, H. Kraut and J. M. Goodman
J. Cheminformatics 2018, 10, 22.
DOI: 10.1186/s13321-018-0277-8

International chemical identifier for reactions (RInChI)
G. Grethe, J. M. Goodman and C. H. G. Allen
Journal of Cheminformatics 2013, 5, 45.
DOI: 10.1186/1758-2946-5-45

Many InChIs and quite some feat
By: Warr, WA (Warr, Wendy A.)
JOURNAL OF COMPUTER-AIDED MOLECULAR DESIGN
Volume: 29 Issue: 8 Pages: 681-694
DOI: 10.1007/s10822-015-9854-3
Published: AUG 2015

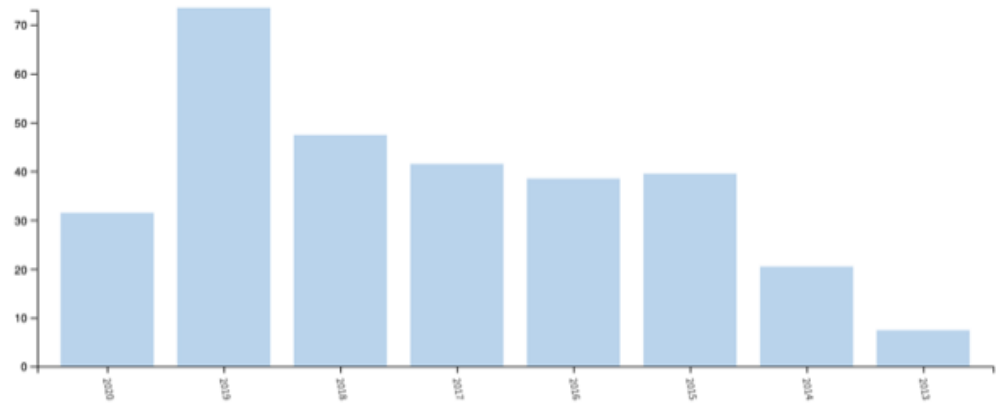
InChI, the IUPAC International Chemical Identifier
By: Heller, SR (Heller, Stephen R.)[1] ; McNaught, A (McNaught, Alan)[2] ; Pletnev, I (Pletnev, Igor) [3] ; Stein, S (Stein, Stephen)[1] ; Tchekhovskoi, D (Tchekhovskoi, Dmitrii)[1]
View Web of Science ResearcherID and ORCID
JOURNAL OF CHEMINFORMATICS 2015
Volume: 7
Article Number: 23
DOI: 10.1186/s13321-015-0068-4

InChI - the worldwide chemical structure identifier standard
By: Heller, S (Heller, Stephen)[1] ; McNaught, A (McNaught, Alan)[1] ; Stein, S (Stein, Stephen)[1] ; Tchekhovskoi, D (Tchekhovskoi, Dmitrii)[1] ; Pletnev, I (Pletnev, Igor)[1]
View Web of Science ResearcherID and ORCID
JOURNAL OF CHEMINFORMATICS
Volume: 5
Article Number: 7
DOI: 10.1186/1758-2946-5-7
Published: JAN 24 2013

Chemistry International | Volume 35: Issue 6
Current Status and Future Development in Relation to IUPAC Activities
Andrey Yerin, Alan McNaught and Stephen Heller
DOI: 10.1515/ci.2013.35.6.12 | Published online: 01 Nov 2013

InChI: connecting and navigating chemistry
By: Williams, AJ (Williams, Antony J.)
View Web of Science ResearcherID and ORCID
JOURNAL OF CHEMINFORMATICS
Volume: 4
Article Number: 33
DOI: 10.1186/1758-2946-4-33
Published: DEC 13 2012

The IUPAC International Chemical Identifier: InChI—A New Standard for Molecular Informatics
by Alan McNaught
CHEMISTRY International November-December 2006, p12-15



PIDapalooza 2020

Ian Bruno and Henry Rzepa gave an InChI presentation at the January 29/30, 2020 PIDapalooza (Persistent Identifier) meeting in Lisbon. Their presentation is available at:
Bruno, Ian; Rzepa, Henry; Blanke, Gerd. (2020, February). Connecting Chemistry Through PIDs. Zenodo. http://doi.org/10.5281/zenodo.3648508

Background:

PIDapalooza is a "festival" for Persistent Identifiers (PIDs) organised by CrossRef, DataCite, ORCID and the California Digital Library. It offers an energetic two days exploring various aspects of PIDs, including infrastructures, adoption, governance and policies. PIDapalooza 2020 took place in Lisbon, Portugal from 29-30 January, and this year featured a session on InChI.

Connecting Chemistry through PIDs:

This PIDapalooza session was co-authored by Gerd Blanke, Henry Rzepa and Ian Bruno. Gerd couldn't be there but Henry and Ian both were. The session was 30 minutes long and opened with a spot-the-difference quiz intended to convey to the non-chemists attending the session the challenges associated with reliable representation of chemical structures. Judging by the blank but engaged expressions on the faces of the 20 or so session attendees, I think we succeeded in this aim.

This provided a springboard for explaining why InChI is necessary, what it is and how it has been adopted and extended. Current project areas along with the governance and sustainability structures that support maintenance and development of InChI were described. We also compared InChI to principles laid out in a recent draft PID Policy for the European Open Science Cloud; despite one or two inherent differences it generally compares favorably except where the ability to resolve an InChI is concerned. Henry talked about how InChIs have been used at Imperial College to represent structures in theses and included in DataCite metadata. We then concluded with some general questions and discussion. Points to arise from the discussion included the following:

- Do we have a Java implementation? This was asked by Jinseop Shin of the Korean Institute of Science and Technology Information who has used InChIs to facilitate chemical structure search in patents. He indicated that they found it difficult interfacing to the InChI Generator in its current form.
- Is the algorithm used to generate the InChI transparent enough? We noted that it was captured in code that is Open Source but this seemed too opaque for Brian Matthews of STFC (one of the authors of the EOSC PID Policy); Brian also picked up on the small number of people who have detailed knowledge of the algorithm and considered this to be an area of risk.
- Ted Habermann of Metadata Game Changers raised the question of whether InChI could be considered a <related identifier> in the DataCite schema. Henry indicated that he had experimented with this and with putting it in <subject> metadata items and found the latter worked better, a proposal also discussed with Ted and for which he expressed support. Ultimately, both could be populated. Ted is someone we might want to draw on to help establish recommendations for including InChI in metadata schemas such as these.
- Noted in our slides was the question of whether InChI can/should be used to populate CrossRef-Schema based metadata produced by publishers. Brian Vickery who may be known to some from his career in the publishing industry has recently become Director of Product at CrossRef. He mentioned that he had come across a reference to InChI in some historic CrossRef Labs project and said he would go back and investigate this some more.
- Kerstin Lehnert of Columbia University and President of IGSN eV, the implementation body of the International Geo Sample Number (IGSN), was interested in understanding how InChI might relate to IGSN and other sample IDs such as RRID. Probably if the sample can be described by a chemical structure then InChI may have a place in the kernel metadata for these identifiers.

- Slides used for the session can be found at https://doi.org/10.5281/zenodo.3648507.

Summary:

It was heartening to get such engagement in InChI during this session and at PIDapalooza generally. There was a quiz on the Wednesday night run by Ed Pentz of CrossRef and the very last question was "What is InChI?" I'll take this as a sign that the wider PID community might be now expected to know the answer to this is rather than being some arcane trivia that no one could possibly know!

InChI Trust web site

The Trust web is up on the InChI Trust cloud server and working well. Updates, such as these reports are posted on the web site.

InChI Usage

Numerous publications now use InChI as part of their efforts in merging and analyzing database structures. Clearly InChI is being used on a very regular basis in many organizations and research projects and publications.

InChI Trust Videos - Access numbers/Views as noted below continue to increase slowly every year with *What on earth is InChI?* by far the most popular video.

What on earth is InChI? - 8, 713 (8/20) 7,884 (1/20), 7,348 (7/19), 6,750 (12/18), 6,102 (7/18), 4,188 (1/17), 3,331 (1/16), 2,956 (7/15), 2486 views (12/14); 1977 views (7/14)

InChI & the Islands – 1,455 (8/20) 1,379 (1/20), 1,327 (7/19), 1,269 (12/18) , 1,208 (7/18), 960 (1/17); 804 (1/16); 728 (7/15); 526 views (7/14)

The Googleable InChIKey – 2,625 (8/20) 2,423 (1/20), 2,301 (7/19), 2,115 (12/18) 1,985 (7/18), 1,379 (1/17), 1,037 (1/16); 915 views (7/15), 597 views (7/14)

The Birth of the InChI – 2,106 (8/20) 2,013 (1/20), 1,948 (7/19), 1,848 (12/18), 1,791 (7/18), 1,365 (1/17), 1,084 (1/16), 984 views (7/15), 687 views (7/14)

IUPAC InChI (Google lecture - 2008) 1,009 (8/20) 998 (1/2), 978 (12/18). 950 (1/17); 946 (7/16); 931 (1/16); 922 views (7/15)

IUPAC InChI (Google lecture - 2006) - 972 (8/20) 940 (1/20), 893 (7/19)

Some History, Organizational Planning, and Project Sustainability

Sorry, but I repeat what was in my last reports as the Trust has not yet had their meeting to discuss these matters (see attached – InChI-The-Future). There has yet to be any feedback on these thoughts and comments.

Plans for the remainder of 2020

For 2020 my current overall plans and goals are as follows:

1. Work to expand the current membership with two basic classes of members – Full and Associate as well as add to the number of Supporters. Continue to attend meetings and give talks on InChI where useful and appropriate.
2. Attend virtual ACS meeting in San Francisco.
3. Attend and participate in virtual sessions on InChI, InChI working groups, and related standards at the ACS meeting.
4. Attend virtual IUPAC and InChI and InChI Trust Board meetings.
5. Attend the November 2020 Virtual GDCh meeting and have a poster or oral presentation on InChI.

2020/2021 possible meetings/ workshops/symposia

2020

ACS San Francisco, GDCh meeting, Germany, and Pacifichem Honolulu have all been canceled, postponed, or will be virtual.

We plan to have monthly Zoom meetings with the InChI subcommittee and invite one of the working groups to each meeting to discuss what they are doing and where they are headed and what resources, if any, are needed.

2021

Steve Heller March 21-25, 2021 – ACS National Meeting, San Antonio TX
(Virtual or in person ??)

Spring 2021 – The 12th International Conference on Chemical Structures (ICCS), Noordwijkerhout, The Netherlands (Virtual or in person or postponed until 2021??)

July 13 - 17, 2021 Cape Town, ZA – 26th IUPAC International Conference on Chemistry Education (ICCE 2020) (Virtual or in person??)

August 13- 20, 2021– 51st IUPAC General Assembly/World Congress, Montreal, Canada

August 22 - 26, 2021 – ACS National Meeting Atlanta, GA (InChI working group meetings and Trust Board meetings)

December 2021 – PacifiChem. Ray & I have submitted abstracts for a session at this meeting which was postponed from December 2020 to December 2021 due to COVID 19.

Steve Heller

Appendix 5

The Future of the Blue Book

(contribution from Warren Powell)

It is disturbing to me to see the number of corrections to the 2913 edition of the Blue Book. I have 202 pages of Corrections (major ones that supposedly only affect nomenclature) and 128 pages of 'minor corrections that supposedly do not affect nomenclature'. And new corrections appear now rather regularly

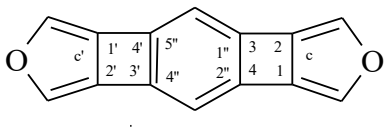
It has been 7 years since publications. Perhaps it is time to think about an update. It could take several years to do so. It took seventeen years to produce the 2013 edition; the one before was published in 1979, with a limited update in 1993. I know that Gerry is producing a html file that has the corrections, clarifications, and modifications that are in the major and minor corrections lists. I have a Word file that also has all of the corrections, clarifications, and modifications that are in the major and minor corrections lists and includes Supplements I, II, and III (in part). One of these files should be able to be the starting point for the creation of a new edition...

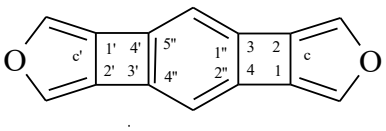
During this coronavirus, I have had lots of time so I have reviewed all of the corrections, both major and minor. As a result, I have a corrected and modified manuscript in Word file and that also includes most of the material in Supplements I, II, and III (in part). And I have compiled the following observations from the first 230 pages of the published book

(1) A lot of corrections to structures were, in fact, nor corrections, but different opinions on where locants should be placed, for example inside the ring or outside the ring or how many locants should be used, for example, only those that appeared in the name or all locants or only that defined structural features or which direction that the numbering should go.

Examples.

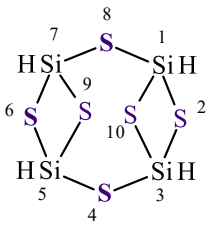
(a) Structure on page 580 Replace the structure with:





I don't see the correction to justify the replacement.

(b) Structure on page 167 Replace the structure with:



(1)(b)(cont'd)

What is a valid reason for replacing the structure as published? There is no error in numbering this example. The only difference is the direction of numbering which should not be an error.

(2) A lot of correction in the minor list were also listed in the major list. Lots of duplication. I easily found the seven examples listed below in the first 250 pages of the publication. As I went through the lists, the number of duplications seemed to increase particularly in Appendix 2. Certainly, the rest of the lists should be reviewed to remove these the duplication.
Examples:

(a) From Corrections….

Page 32, P-14.3.4.3, example 1 on this page. [corrected 19 September 2018]
For choropropanedioic acid (PIN)
read chloropropanedioic acid (PIN)

From Minor Corrections….

Page 32, P-14.3.4.3, example on this page. [corrected 21 November 2018]
For choropropanedioic acid (PIN)
read chloropropanedioic acid (PIN)

(b) From Corrections….

Page 57. P-51.1.4 title
For Position of the endings "ane", 'ene', and 'yne'
Read Position of the endings 'ane', 'ene', and 'yne'

From Minor Corrections……

Page 57. P-51.1.4 title
For Position of the endings "ane", 'ene', and 'yne'
read Position of the endings 'ane', 'ene', and 'yne'

(c) From Corrections….

Page 74, P-15.3.1.2.2.3, example 2.
For ethane-1,2- diylbis[azanylylidene(chloromethanylylidene)] (preferred prefix)
read ethane-1,2-diylbis[azanylylidene(chloromethanylylidene)] (preferred prefix)
[removed space from name]

From Minor Corrections…..

Page 74, P-15.3.1.2.2.3, example 2.
For ethane-1,2- diylbis[azanylylidene(chloromethanylylidene)] (preferred prefix)

(2)(c) (cont'd)

read ethane-1,2-diylbis[azanylylidene(chloromethanylylidene)] (preferred prefix) [removed space from name]

(d) From Corrections…. ….

Page 146, P-22.2.2.1.5.1, lines 1/2. [corrected 9 January 2019]
For ...rings containing only nitrogen; otherwise...
read ...rings only containing nitrogen heterotoms; otherwise..

From Minor Corrections….

Page 146, P-22.2.2.1.5.1, line 2.
For ...containing only nitrogen; otherwise...
read ...containing only nitrogen heteroatoms; otherwise.

[This example also illustrates another point: 'Which correction is correct?' There are many other examples that better illustrate this point.]

(e) From Corrections….

Page 192, P-24.7.2, example (3), Explanation line 4.
For ...where 4′ is lower than 5′)
read ...where 4′ is lower than 5′.

From Minor Corrections….

Page 192, P-24.7.2, example (3) explanation line 4.
For …where 4′ is lower than 5′)
read …where 4′ is lower than 5′.

(f) From Corrections….

Page 194, P-24.7.4.1, example 1. [modified 27 May 2020]
Delete the correction to the first name

From Minor Corrections….

Page 194, P-24.7.4.1, example 1.
For trispiro{bis(cyclohexane)-1,4′:1″,6′-furo[3,4-*d*][1,3]oxathiole-2′,14‴-[7]oxadispiro[5.1.5^{8}.2^{6}]pentadecane} (PIN)
read trispiro[bis(cyclohexane)-1,4′:1″,6′-furo[3,4-*d*][1,3]oxathiole-2′,14‴-[7]oxadispiro[5.1.5^{8}.2^{6}]pentadecane] (PIN)

For pentaspiro[tetracyclohexane-1,2′(5′*H*):1‴,5′:1⁗,4″(6″*H*):1⁗⁗6″-furan-3′(4′*H*),2″-furo[3,4-*d*][1,3]oxathiole] (the CAS index name)

(2)(f)(cont'd.)

read pentaspiro[tetracyclohexane-1,2′(5′*H*):1‴,5′:1⁗,4″(6″*H*):1⁗′,6″-furan-3′(4′*H*),2″-furo[3,4-*d*][1,3]oxathiole] (the CAS index name; note that multiple primes are not divided into groups of three)

[Do both corrections apply?]

(g) From Corrections….

Page 194, P-24.7.4.1, example 2. [modified 27 May 2020]
For 1λ^5,3λ^5,5λ^5,7λ^5-tetraspiro[tetraspiro[2,4,6,8,9,10-hexathia-1,3,5,7-tetraphosphaadamantane-1,2′:3,2″:5,2‴:7,2⁗ ′-tetrakis([1,3,2]oxathiaphosphetane)]-4′,7⁗ ″, 4″,7⁗ ‴,:4‴,7⁗ ⁗ ′,4⁗ ′,7⁗ ⁗ ″-tetrakis(pyrano[2,3-*c*]acridine)} [PIN, see SP-7.4(b)]
read 1λ^5,3λ^5,5λ^5,7λ^5-tetraspiro{tetraspiro[2,4,6,8,9,10-hexathia-1,3,5,7-tetraphosphaadamantane-1,2′:3,2″:5,2‴:7,2⁗ ′-tetrakis([1,3,2]oxathiaphosphetane)]-4′,7⁗ ″:4″,7⁗ ‴: 4‴,7⁗ ⁗ ′:4⁗ ′,7⁗ ⁗ ″-tetrakis(pyrano[2,3-*c*]acridine)} [PIN, see SP-6.4(b), in ref 8]

for octaspiro[2,4,6,8,9,10-hexathia-1,3,5,7-tetraphosphatricyclo[3.3.1.1^{3,7}]decane-1,2′λ^5:3,2″λ^5:5,2‴λ^5:7,2⁗λ^5-tetrakis[1,3,2]oxathiaphosphetane- 4′,7⁗ ″:4″,7⁗ ‴:=4‴,7⁗ ⁗ ′:4‴,7⁗ ⁗ ″-tetrakis[7*H*]pyrano[2,3-*c*]acridine] (the CAS index name)
read octaspiro[2,4,6,8,9,10-hexathia-1,3,5,7-tetraphosphatricyclo[3.3.1.1^{3,7}]decane-1,2′λ^5:3,2″λ^5:5,2‴λ^5:7,2⁗λ^5-tetrakis[1,3,2]oxathiaphosphetane-4′,7⁗′:4″,7⁗′′:4‴,7⁗′′′:4⁗′,7⁗′′′′-tetrakis[7*H*]pyrano[2,3-*c*]acridine] (the CAS index name; note that multiple primes are not divided into groups of three) [delete spaces from name]

From Minor Corrections….

For 1λ^5,3λ^5,5λ^5,7λ^5-tetraspiro[tetraspiro[2,4,6,8,9,10-hexathia-1,3,5,7-tetraphosphaadamantane-1,2′:3,2″:5,2‴:7,2⁗ ′-tetrakis([1,3,2]oxathiaphosphetane)]-4′,7⁗ ″, 4″,7⁗ ‴,:4‴,7⁗ ⁗ ′,4⁗ ′,7⁗ ⁗ ″-tetrakis(pyrano[2,3-*c*]acridine)} [PIN, see SP-7.4(b)]
read 1λ^5,3λ^5,5λ^5,7λ^5-tetraspiro[tetraspiro[2,4,6,8,9,10-hexathia-1,3,5,7-tetraphosphaadamantane-1,2′:3,2″:5,2‴:7,2⁗ ′-tetrakis([1,3,2]oxathiaphosphetane)-4′,7⁗ ″:4″,7⁗ ‴: 4‴,7⁗ ⁗ ′:4⁗ ′,7⁗ ⁗ ″-tetrakis(pyrano[2,3-*c*]acridine)] [PIN, see SP-7.4(b)]

forr octaspiro[2,4,6,8,9,10-hexathia-1,3,5,7-tetraphosphatricyclo[3.3.1.1^{3,7}]decane-1,2′λ^5:3,2″λ^5:5,2‴λ^5:7,2⁗λ^5-tetrakis[1,3,2]oxathiaphosphetane- 4′,7⁗ ″:4″,7⁗ ‴:=4‴,7⁗ ⁗ ′:4‴,7⁗ ⁗ ″-tetrakis[7*H*]pyrano[2,3-*c*]acridine] (the CAS index name)
read octaspiro[2,4,6,8,9,10-hexathia-1,3,5,7-tetraphosphatricyclo[3.3.1.1^{3,7}]decane-1,2′λ^5:3,2″λ^5:5,2‴λ^5:7,2⁗λ^5-

(2)(g)(cont'd.)

tetrakis[1,3,2]oxathiaphosphetane-4′,7‴″:4″,7⁗:4‴,7⁗″:4⁗,7⁗‴-tetrakis[7*H*]pyrano[2,3-*c*]acridine] (the CAS index name; note that multiple primes are not divided into groups of three) [delete spaces from name]

(3) Major changes were introduced, for example hydroxylamine. This cannot be called a correction. Therefore, to publish these changes as simple corrections is **wrong**. They are **clarifications** and **modifications** as well as corrections. Supplement V was an attempt to clarify the treatment of treatment of hydroxylamine nomenclature without a major change.

(4) Some structure are needlessly introduced that do not correct an error

(a) Page 192, P-24.7.2, example (3). [modified 27 May 2020] should be simply corrected as follows:

for trispiro[bis(cyclopentane)-1,1′:1,4′-cycloheptane-6′,2‴-[1,4]-dioxolane] (**I**) (PIN)
read trispiro[bis(cyclopentane)-1,1′:1,4′-cycloheptane-6′,2‴-[1,4]dioxolane](**I**) (PIN)

and

for (not trispiro[bis(cyclopentane)-1,1′:1,5′-cycloheptane-3′,2‴-[1,4]-dioxolane] (**II**)
read (not trispiro[bis(cyclopentane)-1,1′:1,5′-cycloheptane-3′,2‴-[1,4]dioxolane] (**II**)

Instead a benzene ring was fused and thus this became a major correction with new names.

(b) Page 192, P-24.7.2, example (3). [modified 27 May 2020]

I don't find anything wrong with the example as printed. So, what was the purpose in replacing four structures and four names as corrections. If there was a reason then this "correction" should be considered as a modification or an extension and not an explanation.

(c) Page 233, P-25.3.2.4 (j), example. (modified 23 October 2019)

Agsin, there is no error in this example. indeno[1,7-*kl*]aceanthrylene is a correct name for the structure. Whoever suggested this change was clearly was influenced by the explanation. Any change here would have to be considered a modification or clarification not a correction. The new name acephenanthryleno[5,4-*k*]aceanthrylene and structure is a good addition but is not a correction. Perhaps it could replace the two explanatory names and structures.

My understand of the status of the five Supplements written in the years following the publications of the book. I wrote five supplements:

(1) Supplement I: SEPARATION OF STRINGS OF PRIMES and the ORDERING OF NAME COMPONENTS DEPENDING ON NONALPHANUMERIC CHARACTERS

This supplement was approved and was awaiting submission for publication by Karl-Heinz. I think the separation of primes was introduced into Gerry's html version. I haven't seen the ordering of name components as a section, but I think there are application of it particularly for λ names and isotopic labeled names. These proposals have indeed been included in my Word File.

(2) Supplement II on the selection of the preferred IUPAC name PS-45.2.3 and P-45.5.1.

For most of the examples given in P-45.5.1 in the 2013 edition consideration was not given to an earlier hierarchical Rule P-45.2.3 where preference is given to lower locants in the order that they appear in the name. Accordingly, Rules 45.2.3 (page 526-ff) and P-45.5 (page 531-ff) as corrected were summarized in this document. Hence, as just a reorganization of these two subsections it should have been readily approved for publication. This reorganization has been included in my Word file.

(3) Supplement III on stereochemical criteria the strerseniority order for parent structures (PS3-45.2.3) and preferred IUPAC names (PS3-45.5).

This Supplement is concerned with stereochemical criteria for in the selection of both parent structures and preferred IUPAC names). The Sequence Rules, particularly Rules 4 and 5 as used in the 2013 publication, have proved to be much to complicated for these selections. Accordingly, the following seniority principles used for other structural criteria in selecting the senior parent structures and the preferred IUPAC name. The stereochemical criteria are worked into the fabric of P-44, Seniority Order for Parent Structures, and P-45, Selection of the Preferred IUPAC Name.

These recommendations have been worked out with Andrey Yerin and were read for review by the full Division Committee. I don't know if any of these recommendations have been worked into Gerry's html version. I think that this material has been included in my Word file, but I would have to check that out.

(4) Supplement IV. Ester Nomenclature, A complex subject. A reorganization of ester nomenclature particularly of polyesters. I don't think it has been reviewed by any other Committee Members.

(5) Hydroylamine nomenclature. This was not written as a Sypplement but an attempt to pull together in one place all of the nomenclature for hydroxylamine nomenclature This major change in the nomenclature of hydroxylamine was introduced Gerry's html version and my Word file but as corrections.

Appendix 6

End-of-line hyphenation of systematic chemical names

(Submitted by Albert Dijkstra and Jan Reedijk)

13.8. End-of-line hyphenation of systematic chemical names (2014-003-2-800)

Albert Dijkstra: I chair project 2014-003-2. I am glad to tell you that the Task Force looking after the Hyphenation Project has published its Recommendations, received comments, accommodated these comments in the recommendations and is now waiting for the proofs. Once these have been read and corrected, the Recommendations will at long last be published and as far as I am aware that is the end of the project.

Jan Reedijk: After the public review of our provisional recommendation paper on project 2014-003-2-800 in PAC, earlier this year, we have revised/updated the manuscript, so that it is now ready for publication as the final recommendation in PAC, after agreement/endorsement by the editor and the two DPs.

We also copied Fabienne for the technicalities and instructions for preparation of the electronic submission.

For easy reference and handling we attach already now:

1) The ms in pdf (24 pages) with line numbers kept; this is easiest to check.
2) The ms in Word, with line numbers **removed** (this is relevant when the publishers prepare a pdf with automatically added line numbers, generated by their pdf-making process).
3) A memo (pdf) in which we show how the public review comments have been addressed and the changes have been made.

The character of the text will make that small typos are not easy to discover, and as a team of authors we scrutinized several drafts of the manuscript before we arrived at the final draft.
As in the pdf of the preliminary recommendations, produced by deGruyter in December, it is crucial that we shall have enough time to scrutinize the new dG-produced pdf, to make sure that our own rules (of dividing words and names) are correctly followed.

Following an earlier request from Fabienne, we are also trying to prepare a short memo for publication in CI, to encourage readers to take note of the PAC recommendation document. This memo will be mailed to you for information in the near future.

Appendix 7

Revision and extension of IUPAC recommendations on carbohydrate nomenclature

(Submitted by Johannes Vliegenhart)

13.11. Revision and extension of IUPAC recommendations on carbohydrate nomenclature (2012-039-2-800, 2015-035-2-800, 2017-026-1-800)

Johannes Vliegenhart: The field of glycoscience is rapidly moving forward. Where in the past carbohydrate chemistry was the area wherein fundamental progress was made mainly by 'true' chemists, it now also the field of glycobiologists. Groundbreaking results in glycobiology, often concerning highly complex carbohydrate structures, have inevitably led to naming that does not always follow the strict IUPAC recommendations as formulated in the document 2-carb.

To give just two examples:
1. Sugar, in particular in composite names like in aminosugars.
2. Glycan means in carbohydrate chemistry a generic term for homopolysaccharides. In glycobiology the term comprises the side-chain of glycoproteins.

Such items are creating continuous discussion. It is practically impossible to follow strictly the 2-carb recommendations. Some adaptation to common usage of names is needed, to keep good understanding between the different groups of scientists. In fact, the journals in the field of glycoscience do so, by accepting names proposed by authors.

The task group tries to give in 3-carb recommendations that can widely be used and are clear.

As to the actual 3-carb significant progress has been made:
1. The text of the first part of 3-Carb, which mainly concerns carbohydrates as such, is now finished except for the discussion points as mentioned before, finetuning of the text and the redrawing of figures in an actual ChemDraw format.
2. The second part is dealing with glycoconjugates, the text of the glycoprotein document was finished. The figures have to re-drawn in another format. The glycolipids division needs further finetuning and the figures need to be extended with more examples. For glycoconjugates consisting of a carbohydrate moiety and dendrimers, fullerenes or comparable entities, nomenclature is still under discussion.
3. As to the part III on glyco-informatics, it was decided to adopt the proposal of Martin Frank. Needed is only a rewriting of the paper in a format that is in-line with the remainder of 3-Carb. It is concluded that this proposal fits the approach of International Protein database.

It is relevant that this part three gets soon International confirmation to avoid the confusion that could arise by circulation of different systems.

Appendix 8

Publications since Paris (July 2019)

Recommendations and Technical Reports

A. Fradet, J. Chen, K.-H. Hellwich, K. Horie, J. Kahovec, W. Mormann, R. F. T. Stepto, J. Vohlídal, E. S. Wilks, Nomenclature and terminology for dendrimers with regular dendrons and for hyperbranched polymers (IUPAC Recommendations 2017), *Pure Appl. Chem.* **91**(3), 523 – 561 (2019), https://doi.org/10.1515/pac-2016-1217.

M. Vert, J. Chen, K.-H. Hellwich, P. Hodge, T. Nakano, C. Scholz, S. Slomkowski, J. Vohlidal, Nomenclature and terminology for linear lactic acid-based polymers (IUPAC Recommendations 2019), *Pure Appl. Chem.* **92**(1), 193 – 211 (2020), https://doi.org/10.1515/pac-2017-1007.

M. A. Beckett, B. Brellochs, I. T. Chizhevsky, T. Damhus, K.-H. Hellwich, J. D. Kennedy, R. Laitinen, W. H. Powell, D. Rabinovich, C. Viñas, A.Yerin, Nomenclature for boranes and related species (IUPAC Recommendations 2019), *Pure Appl. Chem.* **92**(2), 355 – 381 (2020), https://doi.org/10.1515/pac-2018-0205.

K.-H. Hellwich, R. M. Hartshorn, A. Yerin, T. Damhus, A. T. Hutton, Brief guide to the nomenclature of organic chemistry (IUPAC Technical Report), *Pure Appl. Chem.* **92**(3), 527 – 539 (2020), https://doi.org/10.1515/pac-2019-0104.

P. Hodge, K.-H. Hellwich, R. C. Hiorns, R. G. Jones, J. Kahovec, C. K. Luscombe, M. D. Purbrick, E. S. Wilks, A concise guide to polymer nomenclature for authors of papers and reports in polymer science and technology (IUPAC Technical Report), *Pure Appl. Chem.* **92**(5), 797 – 813 (2020), https://doi.org/10.1515/pac-2018-0602.

C. M. Fellows, K.-H. Hellwich, S. V. Meille, G. Moad, T. Nakano, M. Vert, Definitions and notations relating to tactic polymers (IUPAC Recommendations 2020), *Pure Appl. Chem.* **92**(11), 1769 – 1779 (2020), https://doi.org/10.1515/pac-2019-0409.

Other relevant publications

J. Capitolis, S. Delacroix, X. Frogneux, É. Medina, N. Rey, L. Tinat, S. Carenco, Précis de nomenclature en chimie inorganique, *Actual. Chim*. No. 437, 12 – 17 (2019), [French translation of the Brief Guide to the Nomenclature of Inorganic Chemistry in: *Pure Appl. Chem*. **87**(9-10), 1039 – 1049 (2015)]; http://www.lactualitechimique.org/Precis-de-nomenclature-en-chimie-inorganique.

G. J. Leigh, IUPAC and the Periodic Table, *Chem. Int.* **41**(1), 6 – 9 (2019), https://doi.org/10.1515/ci-2019-0102.

E. Scerri, Looking backwards and forwards at the development of the Periodic Table, *Chem. Int.* **41**(1), 16 – 20 (2019), https://doi.org/10.1515/ci-2019-0104.

G. J. Leigh, A history of CNIC, *Chem. Int.* **41**(3), 39 – 43 (2019), https://doi.org/10.1515/ci-2019-0313.

E. Hepler-Smith and L. McEwen, A century of nomenclature for chemists and machines, *Chem. Int.* **41**(3), 46 – 49 (2019), https://doi.org/10.1515/ci-2019-0315.

L. R. McEwen, Towards a digital IUPAC, *Chem. Int.* **42**(2), 15 – 17 (2020), https://doi.org/10.1515/ci-2020-0203.